

Structural, functional and genetic analyses of novel candidate genes
for tuber quality traits in potato (*Solanum tuberosum* L.)

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Lena Katharina Schreiber

aus Gütersloh

Köln, 2011



Max Planck Institute for
Plant Breeding Research

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

Berichtersteller/in: PD. Dr. Christiane Gebhardt

Prof. Dr. Martin Hülskamp

Tag der mündlichen Prüfung: 13.12.2011

FÜR MEINE GROBTANTEN
IRMGARD UND MARIA

TABLE OF CONTENTS

Abbreviations.....	v
Figures.....	vii
Tables.....	ix
1 Introduction.....	1
1.1 The potato – basic information.....	1
1.2 Quantitative traits and linkage maps.....	2
1.3 Improved diagnostic markers derived from candidate gene approaches.....	2
1.4 Cold-sweetening as a result of starch degradation.....	3
1.5 Candidate genes for structural, functional and genetic analyses.....	8
1.5.1 Alpha-glucan water dikinase - GWD.....	8
1.5.2 Phosphoglucan water dikinase - PWD.....	10
1.5.3 Invertases.....	11
1.5.4 Leucine aminopeptidase - LAP.....	13
1.5.5 The putative invertase inhibitor – a member of the Kunitz type protease inhibitors ..	15
1.6 Objectives.....	16
2 Materials and Methods.....	17
2.1 Materials.....	17
2.1.1 Chemicals.....	17
2.1.2 Enzymes.....	17
2.1.3 Bacterial and yeast strains.....	17
2.1.4 Vectors.....	18
2.1.5 Plant material.....	18
2.1.6 Media and solutions.....	19
2.1.6.1 Media for cultivation of bacteria.....	19
2.1.6.2 Antibiotics.....	19
2.1.6.3 Media for cultivation of yeast.....	20
2.1.6.4 General solutions.....	20
2.1.7 Oligonucleotides.....	21
2.2 Methods.....	24
2.2.1 Cold-storage experiment.....	24
2.2.2 Preparation of genomic DNA from potato tubers.....	24
2.2.3 Determination of DNA concentration.....	25
2.2.4 Preparation of RNA from potato tubers.....	25
2.2.5 Synthesis of first strand cDNA.....	25

2.2.6	Plasmid preparation	26
2.2.7	Polymerase Chain Reaction (PCR)	26
2.2.7.1	Standard PCR protocol	26
2.2.7.2	Polymerases and modified PCR protocols.....	26
2.2.8	Gel-electrophoresis	27
2.2.9	Sequencing	27
2.2.10	Cloning, heterologous expression and purification.....	27
2.2.10.1	Cloning of invertase and putative inhibitor into a secretory vector	27
2.2.10.2	Storage of clones – glycerol stocks for bacterial and yeast cells.....	28
2.2.10.3	Heterologous expression of invertase and putative inhibitor in <i>Pichia pastoris</i> ..	28
2.2.10.4	His-tag purification of heterologous expressed proteins.....	29
2.2.10.5	Protein gels and Coomassie/silver staining.....	30
2.2.10.6	Western blots and immunodetection	30
2.2.10.7	Ammonium sulfate precipitation	31
2.2.11	Purification of native invertase from potato tubers	31
2.2.12	Activity assays.....	32
2.2.12.1	Invertase activity assay.....	32
2.2.12.2	Invertase inhibition assay.....	33
2.2.12.3	Trypsin activity assay.....	33
2.2.13	BAC-library screen	33
2.2.13.1	Construction of radioactive-labelled probe via nick-translation.....	33
2.2.13.2	Pre-hybridization and hybridization of BAC filter.....	34
2.2.13.3	PFGE- pulsed field gel electrophoresis	35
2.2.13.4	BAC sequencing	35
2.2.14	Expression analysis - Pyrosequencing	36
2.2.15	Association analysis.....	37
2.2.16	Haplotype calculation.....	37
2.2.17	Programs and software	38
3	Results	39
3.1	Glucan water dikinase - genomic organization	39
3.1.1	Functional domains of <i>GWD</i>	42
3.1.2	Structural comparison of <i>S. tuberosum</i> and <i>A. thaliana GWD</i>	43
3.1.3	Comparison of the <i>GWD</i> BAC sequence and the genomic sequence of <i>S. phureja</i>	44
3.1.4	Conservation of <i>GWD</i> among different species.....	44
3.1.5	Association analysis of <i>GWD</i>	48
3.1.6	<i>GWD</i> haplotype modeling	51

3.2	Phosphoglucan water dikinase - genomic organization.....	53
3.2.1	Structural comparison of <i>PWD</i> from <i>S. tuberosum</i> and <i>A. thaliana</i>	55
3.2.2	Functional domains of <i>PWD</i>	55
3.2.3	Comparison of the <i>PWD</i> BAC sequence and the genomic sequence of <i>S. phureja</i>	56
3.2.4	Conservation of <i>PWD</i> among different species	57
3.2.5	Association analysis of <i>PWD</i>	59
3.2.6	<i>PWD</i> haplotype modeling.....	62
3.2.7	Structural comparison of <i>GWD</i> and <i>PWD</i>	63
3.3	<i>Invertase 6</i> - genomic organization	64
3.3.1	Structural comparison of <i>Inv6</i> and <i>Pain-1</i> of <i>S. tuberosum</i>	65
3.3.2	Conservation of <i>Inv6</i> among different species.....	67
3.3.3	Association analysis of <i>Inv6</i> in the GABICHIPS population.....	68
3.3.4	Analysis of the distribution of associated <i>Inv6</i> alleles in the BIOSOL population.....	71
3.3.5	<i>Inv6</i> haplotype modeling and association.....	73
3.3.6	Expression analysis of <i>Inv6</i>	75
3.4	Leucine aminopeptidase - <i>LAP</i>	76
3.4.1	<i>LAP</i> - genomic organization.....	77
3.4.2	Conservation of <i>LAP</i> among different species.....	80
3.4.3	Association analysis of <i>LAP</i>	82
3.4.3.1	Association analysis of <i>LAP</i> in the GABICHIPS population.....	83
3.4.3.2	Analysis of the distribution of associated <i>LAP</i> alleles in the BIOSOL population.....	86
3.5	Analysis of the putative invertase inhibitor	88
3.5.1	Genomic organization of the putative invertase inhibitor	88
3.5.2	Conservation of the putative invertase inhibitor among different species	91
3.5.3	Expression analysis of the putative invertase inhibitor	91
3.5.3.1	Semi-quantitative expression in different potato tissues	91
3.5.3.2	Cultivar-dependent expression of the putative invertase inhibitor.....	92
3.5.3.3	Expression analysis on transcript level.....	93
3.5.4	Association analysis of the putative invertase inhibitor	96
3.5.5	Haplotype modeling and association	98
3.5.6	Functional characterization of the putative invertase inhibitor	100
3.5.7	Activity assays.....	101
3.5.7.1	Activity assay for heterologously expressed Invertase	101
3.5.7.2	Activity assay for heterologously expressed putative invertase inhibitor	102
3.5.7.3	Invertase inhibition assay.....	102
3.5.7.4	Inhibition assay of native invertase.....	103
4	Discussion	104

4.1	Candidate genes for SNP based association analysis: forward and reverse approaches ...	104
4.2	Forward candidate gene approach	104
4.2.1	Glucan water dikinase - a functional candidate in starch breakdown	105
4.2.1.1	<i>GWD</i> is a large and highly conserved gene	105
4.2.1.2	<i>GWD</i> is predominantly associated with starch quality traits	106
4.2.2	Phosphoglucan water dikinase – number 2 in starch degradation.....	108
4.2.2.1	<i>PWD</i> is located close to a chip quality and starch QTL on chromosome 9	108
4.2.2.2	<i>PWD</i> is highly conserved	109
4.2.2.3	<i>PWD</i> shows high association with starch and chip quality traits	109
4.3	Candidate recruitment based on <i>in silico</i> analyses - discovery of invertase 6.....	110
4.3.1	<i>Inv6</i> shows strong homology to the vacuolar invertase Pain-1	111
4.3.2	<i>Inv6</i> is highly associated with starch and chip quality traits	112
4.3.3	<i>Inv6</i> shows only low expression in tubers	113
4.3.4	Haplotype modeling is a powerful tool, but remains artificial.....	113
4.4	Reverse candidate gene approach	114
4.4.1	<i>LAP</i> - Comparative Proteomics revealed a novel candidate gene.....	115
4.4.2	<i>LAP</i> is associated with starch and chip quality traits.....	115
4.4.3	Structural analysis of <i>LAP</i> helps to speculate about the functional role.....	116
4.5	Putative invertase Inhibitor – To be or not to be an invertase inhibitor	117
4.5.1	The putative invertase inhibitor belong to a cluster of Kunitz-type protease inhibitors on chromosome3	118
4.5.2	The cDNA sequences of the putative invertase inhibitor are allelic variants	118
4.5.3	The putative invertase inhibitor shows an unique expression profile in tubers.....	119
4.5.4	The putative invertase inhibitor is associated with starch quality traits	121
4.5.5	In vitro inhibition assays do not prove invertase inhibition.....	124
4.6	Candidate gene approach and association analyses are powerful tools	126
	Literature	127
	Abstract	139
	Zusammenfassung	140
	Appendix.....	142
	Acknowledgements.....	160
	Erklärung.....	161
	Lebenslauf.....	162

ABBREVIATIONS

A	Adenine
aa	amino acid
ATP	adenosine triphosphate
BAC	bacterial artificial chromosome
BD	binding domain
bp	base pair (s)
BMGY/BMMY	buffered complex medium containing glycerol/methanol
Btn	Biotin
C	Cytosine
CBM	carbohydrate binding module
cDNA	complementary deoxyribonucleic acid
cM	centimorgan
CTAB	Cetyl trimethyl ammonium bromide
cTP	chloroplast target peptide
COA	chips quality in autumn (after harvest)
CQS4	chips quality after storage at 4°C
CQS8	chips quality after storage at 8°C
d	duplex
Da	Dalton
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
<i>E. coli</i>	<i>Escherichia coli</i>
e. g.	<i>exempli gratia</i> (Lat.) for example
EST	expressed sequence tag
<i>et al.</i>	<i>Et alii/et aliae</i> (Lat.) and others
EtBr	Ethidium bromide
Fru	fructose
G	Guanine
GWD	Glucan water dikinase
GLM	general linear model
Glu	glucose
H	histidine
InDel	insertion/deletion
Inh	Inhibitor
Inv6	Invertase 6
kb	kilo base pair(s)

kDa	kilo Dalton
KTI	Kunitz-type protease inhibitor
L	liter
LAP_A	Acidic leucine aminopeptidase (chloroplastic)
LAP_N	Neutral leucine aminopeptidase
LB	Lysogeny Broth
LD	linkage disequilibrium
M	Molar
MAS	marker-assisted selection
mL	milliliter
n	nulliplex
nM	nanomolar
ns	not significant
OD	optical density
P	phosphate
PCR	polymerase chain reaction
PFGE	pulse field gel electrophoresis
pl	isoelectric point
PVP	Polyvinylpyrrolidone
PWD	Phosphoglucan water dikinase
q	quadruplex
QTL	quantitative trait locus
RNA	ribonucleic acid
RT	room temperature
s	simplex
SNP	single nucleotide polymorphism
<i>S. tuberosum</i>	<i>Solanum tuberosum</i>
t	triplex
T	Tymidine
TP	time point
TSY	tuber starch yield
TSC	tuber starch content
TY	tuber yield
V	Volt
v/v	volume per volume
wocs	weeks of cold-storage
w/v	weight per volume
YPD/YPAD	Yeast extract peptone dextrose (adenine sulphate)

FIGURES

Figure 1: Synthesis of transient starch in chloroplasts.....	5
Figure 2: Synthesis of storage starch in heterotrophic tissues	6
Figure 3: Starch degradation pathway	7
Figure 4: Phenotypes of <i>GWD</i> mutants.....	9
Figure 5: Phenotypes of <i>PWD</i> mutants.	10
Figure 6: Localization and proposed functions of plant invertases.....	12
Figure 7: Experimental set up of the cold-storage experiment of the BIOSOL population.	24
Figure 8: Insert size of BAC clone BA202-H17 estimated by PFGE.....	40
Figure 9: Length and structure of BAC clone BA202-H17 (contig 13).	41
Figure 10: Genomic structure of <i>GWD</i> in <i>S. tuberosum</i>	42
Figure 11: Structural comparisons of <i>StGWD</i> and <i>AtGWD</i>	43
Figure 12: Phylogenetic tree of <i>GWD</i> deduced protein sequences	45
Figure 13: Sequence alignment of 11 <i>GWD</i> sequences from different species.....	48
Figure 14: Means plots of SNP 1 (A ₄₄₄ T) of LD1 of <i>GWD</i>	50
Figure 15: Means plots of SNP 7 (A ₄₈₂₆ G) of <i>GWD</i>	51
Figure 16: PFGE of <i>PWD</i> BACs digested with NotI.....	53
Figure 17: Genomic organization of <i>PWD</i> from <i>S. tuberosum</i>	54
Figure 18: Structural comparison of <i>StPWD</i> and <i>AtPWD</i>	55
Figure 19: Functional domains of <i>StPWD</i>	56
Figure 20: Phylogenetic tree of <i>PWD</i> deduced protein sequences.....	57
Figure 21: Alignment of eight amino acid sequences of different <i>PWD</i> orthologues.....	59
Figure 22: Means plots of SNP 8 (G ₁₁₁₄₀ A) of LD3 of <i>PWD</i>	61
Figure 23: Genomic organization of <i>Inv6</i> and <i>Pain-1</i> of <i>S. tuberosum</i>	65
Figure 24: Amino acid sequence alignment of <i>Inv6</i> and <i>Pain-1</i> from <i>S. tuberosum</i>	66
Figure 25: Alignment of the deduced amino acid sequences of different <i>Inv6</i> orthologues.....	67
Figure 26: Phylogenetic tree of <i>Inv6</i> protein sequences retrieved from diverse blast analyses.	68
Figure 27: Means plots of SNP 2 (C ₂₀₇₆ T) of LD1 of <i>Inv6</i>	70
Figure 28: Presence of three associated <i>Inv6</i> SNPs in potato tubers without cold-storage.....	71
Figure 29: Presence of three associated <i>Inv6</i> SNPs in potato tubers after 4 weeks of cold-storage....	72
Figure 30: Effect of the allelic dosage on low and high amount of reducing sugars.....	73
Figure 31: Semi-quantitative expression analysis of <i>Inv6</i>	75
Figure 32: Comparison of 2D protein gels of good and bad chipping genotypes.	76
Figure 33: Genomic organization of the identified tandem repeat of <i>LAP</i> in <i>S. tuberosum</i>	78

Figure 34: Amino acid sequence alignment of the LAP sequences of potato and tomato.	79
Figure 35: Phylogenetic tree of LAP sequences from different organisms.	80
Figure 36: Alignment of leucine aminopeptidase peptide sequences of different species.	82
Figure 37: Means plots of SNP 1 (G ₂₇₄₆ A) of the <i>LAP</i> locus in <i>S. tuberosum</i>	85
Figure 38: Presence of the associated <i>LAP</i> SNP 1 (G ₂₇₄₆ A) without cold-storage.	86
Figure 39: Presence of the associated <i>LAP</i> SNP 1 (G ₂₇₄₆ A) after 12 weeks of cold-storage.	86
Figure 40: Relationship between sugar content and the presence of the superior <i>LAP</i> SNP1 (G ₂₇₄₆ A).	87
Figure 41: Phylogenetic tree of 30 cDNA sequences of group C Kunitz-type inhibitors.	89
Figure 42: Protease inhibitor cluster in superscaffold 159.	90
Figure 43: Semi-quantitative expression analysis of subgroups 1421, 1422 and 1423.	92
Figure 44: Presence of the three inhibitor subgroups 1421, 1422 and 1423 at genomic DNA level.	93
Figure 45: Analysis of the transcript level of subgroup 1421 during cold-storage.	94
Figure 46: Analysis of the transcript level of subgroup 1423 during cold-storage.	95
Figure 47: Calculated expression level of subgroup 1422 during cold-storage.	96
Figure 48: Means plots of SNP 3 (A ₂₆₁ G) of the putative invertase inhibitor.	98
Figure 49: Western Blot of purified invertase and inhibitor alleles.	101
Figure 50: Analysis of invertase activity.	101
Figure 51: Inhibitor activity of the heterologously expressed Kunitz-type protease inhibitors.	102
Figure 52: Inhibition assay of invertase alleles.	102
Figure 53: Analysis of purified putative invertase inhibitor on partially purified native invertase.	103

TABLES

Table 1: Bacterial Strains.....	17
Table 2: Yeast strains.....	18
Table 3: List of used vectors for cloning, transformation and propagation.....	18
Table 4: List of Antibiotics.	19
Table 5: List of primer combinations.....	21
Table 6: List of used antibodies for immunodetection of invertase and putative invertase inhibitor.	31
Table 7: List of frequently used online tools, databases and software.....	38
Table 8: Obtained shotgun library data from GATC Biotech.....	41
Table 9: Comparison of StGWD and AtGWD sequences.....	43
Table 10: Comparison of the genomic sequence of <i>GWD</i> obtained from BAC sequencing with the released <i>S. phureja</i> sequence.....	44
Table 11: Overview of associated SNPs of <i>GWD</i>	49
Table 12: List of <i>StGWD</i> haplotype models based on 12 SNPs scored in 218 genotypes.	52
Table 13: Comparison of <i>StPWD</i> and <i>AtPWD</i> at sequence length level.	55
Table 14: Overview of associated SNPs of <i>PWD</i>	60
Table 15: List of <i>StPWD</i> haplotype models based on 12 SNPs scored in 218 genotypes.....	62
Table 16: Comparison of <i>StGWD</i> and <i>StPWD</i>	63
Table 17: Best hits of <i>Inv6</i> blasted against the <i>S. phureja</i> genome (PGSC, 2011).....	64
Table 18: Comparison of <i>Inv6</i> and <i>Pain-1</i> from <i>S. tuberosum</i> at different sequence level.	65
Table 19: Overview of associated SNPs of <i>Inv6</i>	69
Table 20: List of <i>Inv6</i> haplotype models.....	73
Table 21: List of haplotype models for LD1 of <i>Inv6</i>	74
Table 22: List of haplotype models for two SNPs of LD1 of <i>Inv6</i>	75
Table 23: Best hits of <i>LAP</i> blasted against the <i>S. phureja</i> genome (PGSC, 2011).....	77
Table 24: Comparison of the tomato and the potato <i>LAP</i> cDNA sequences.	78
Table 25: Overview of associated SNPs of <i>LAP</i>	84
Table 26: Kunitz-type protease inhibitor cDNA sequences blasted against different datasets of the <i>S. phureja</i> genome.	90
Table 27: Overview of associated SNPs of 1421.	97
Table 28: List of haplotype models for the putative invertase inhibitor.....	99
Table 29: Invertase and inhibitor alleles used for cloning and heterologous expression.....	100

1 INTRODUCTION

1.1 THE POTATO – BASIC INFORMATION

The word potato can be used synonymously for the plant (*Solanum tuberosum*) itself and for the underground storage stem which forms an edible tuber. It originates from the word “papas” which was used by the people of the Andean highlands and means root tuber. The German word “Kartoffel” got influenced most likely by the Italian word “tartufo”, due to its similarity to truffles. Potato has been domesticated 7000-10000 years ago in a single domestication event of *Solanum bukasovii* in the southern Peruvian highlands (Spooner et al., 2005). In the middle of the 16th century, potatoes were brought from Peru to Europe by the Spanish conquerors and became an important staple food in the 19th century. Nowadays, potato is the third most important crop worldwide after wheat and rice (FAO, <http://faostat.fao.org>). Although the production of potatoes declined in Europe within the last years, the worldwide production grew further, due to an extremely expanded production in developing countries like China and India. In the last 20 years, the participation in the production of these countries increased from 20 % to 52 % and surpassed the share of the developed countries. Potatoes are a good source of carbohydrate, protein, vitamins, riboflavin, folic acid and minerals like calcium and phosphorus (Milbourne et al., 2007). Besides direct consumption of potatoes, diverse non-food applications are emerging and potato serves as a source of starch for the production of textiles, adhesives and pharmaceuticals (Ellis et al., 1998). Although more than 4000 cultivated varieties are available (www.europotato.org), the demand for optimized potatoes increases due to a rising consumption of fresh and processed potatoes.

For cultivated and wild potatoes, it was shown that they possess different ploidy levels with the base of 12 chromosomes. Most of the potato species are diploid or tetraploid, but tri-, penta- and hexaploids have been identified as well (Hawkes, 1990; Hijmans and Spooner, 2001). Besides 98 cultivated species (landraces), 261 wild species have been found (Spooner et al., 2005). The latter ones are considered as valuable resource for breeders due to their broader genetic base (Milbourne et al., 2007). The cultivated potato *Solanum tuberosum ssp. tuberosum* is tetraploid ($2n = 4x = 48$) and displays autotetrasomic inheritance which means random pairing, recombination and segregation of the four sets of homologous chromosomes during meiosis (Milbourne et al., 2007). Besides a high level of heterozygosity, inbreeding depression and sterility of the F₁ population additionally challenge classical genetic analyses and therefore recessive alleles are often masked and cannot be identified.

1.2 QUANTITATIVE TRAITS AND LINKAGE MAPS

Due to technical development, it was possible to artificially haploidise tetraploid cultivated potatoes to generate fertile diploids which can be used for the construction of genetic maps (Milbourne et al., 2007). The first maps in potato were constructed by using RFLP markers and used as reference molecular maps (Bonierbale et al., 1988; Gebhardt et al., 1989). Diploid potato plants are self-incompatible and highly heterozygous, but the offspring segregates according to the F1 segregation pattern (Ritter et al., 1991; Gebhardt, 2005). Besides these reference molecular maps, diverse functional maps for agronomically important traits like pathogen resistance and tuber traits including starch content, chip colour, cold-sweetening and carbohydrate metabolism in general have been developed (Gebhardt and Valkonen, 2001; Schäfer-Pregl et al., 1998; Douches and Freyre, 1994; Menéndez et al., 2002; Chen et al., 2001). These molecular function maps are based on the usage of various types of DNA-based markers which originate from the natural DNA variation present in a population of different genotypes (Gebhardt, 2005).

Natural variation can be defined as the phenotypic variation within one species which is caused by spontaneous mutations. These mutations were maintained during evolution because they were resistant against the pressure of artificial and natural selection (Alonso-Blanco et al., 2009). Most of the phenotypic differences are due to allelic variants on different loci within the genome and are considered as quantitative trait loci (QTL). Natural variation in different types of plants has been exploited during the last thousand years of cultivation to improve crops according to desired traits (Alonso-Blanco et al., 2009).

1.3 IMPROVED DIAGNOSTIC MARKERS DERIVED FROM CANDIDATE GENE APPROACHES

Marker-assisted selection (MAS) makes use of diagnostic DNA markers that have an enormous potential to improve the efficiency and precision of conventional breeding (Collard and Mackill, 2008). Those markers are mainly DNA markers like simple sequence repeats (SSR), single strand conformation polymorphism (SSCP), restriction fragment length polymorphism (RFLP) and single nucleotide polymorphism (SNP; Vignal et al., 2002). They show variation at the DNA level which can be linked to a quantitative trait and are therefore in tight linkage disequilibrium (LD). In contrast to the linkage maps constructed from single-pair crosses, these markers are analyzed in large populations of unrelated individuals (Milbourne, 2007). In general, markers should be linked closely to target loci, easy detectable and cost-effective (Tanksley, 1983; Collard and Mackill, 2008). If such a reliable marker is obtained, first of all it can be used to screen the breeding material directly and unsuitable parents are excluded. Additionally, the offspring can be examined in early stages (i.e.

seedlings) and laborious and time-consuming phenotypic screens are avoided or at least reduced. Especially for tuber quality traits this is of great advantage because the phenotypically examination starts for the first time 5 years after the beginning of the breeding program, when sufficient tubers are available (Milbourne, 2007).

Very promising markers are the SNP markers which can be easily detected by PCR or amplicon sequencing with a specific primer at the genomic DNA of either the crossing population or the first generations of the offspring. Making use of the candidate gene approach, these diagnostic SNP markers are more and more directly identified in functional and positional candidate genes. In other words, these candidate genes are selected upon their known functional role in a pathway and/or upon their genetic position within or close to a QTL and subsequently analyzed for association with a phenotype of interest (Pflieger et al., 2001). Dealing with potato tubers, yield, amount of starch and processing performance are of general interest. Several association studies using strong functional candidates of the carbohydrate metabolism have already been done and resulted in high associations of vacuolar and cell-wall invertases and starch phosphorylases with starch and chip quality traits (Li et al., 2005; Li et al., 2008). These and further markers of other candidate genes can be combined and used as a powerful tool to optimize the selection of parents in breeding programs.

1.4 COLD-SWEETENING AS A RESULT OF STARCH DEGRADATION

Starch is the most abundant carbohydrate reserve in plants and the main component of the dry matter in potato tubers (Leszczynski and Lisinska, 1988). Although the biosynthesis and degradation of starch are complex processes, this metabolism is one of the best studied pathways in plants (Smith et al., 2003). Starch and sucrose serve as long- and short-term storage deposits of products of the photosynthetic carbon fixation which are not directly reused in the reductive pentose phosphate way (Fettke et al., 2009). During the photoperiod, starch is synthesized and stored as transitory starch in chloroplasts or as storage starch in non-photosynthetic organs as the amyloplasts in tubers or seeds. In the dark, the transitory starch in leaves is degraded to provide substrates for energy. In potato tubers, the long-term storage starch serves predominantly as an energy source for sprouting of the potato tubers. A phenomenon named cold-sweetening has been described early for potatoes and several other plants (Müller-Thurgau, 1882). Potatoes are mainly stored in the cold to avoid sprouting and diseases, but upon prevention of freezing, reducing sugars are mobilized as osmoregulators and transported into the vacuole. A high amount of reducing sugars like glucose and fructose prevents the formation of ice crystals in the vacuole which would damage the tonoplast and the whole cell. The major sources for these osmoregulatory sugars are starch and sucrose. Besides an

unwanted sweet taste of potato tubers for normal cooking, these sugars interfere in further processing steps of the food industry. One of the problems caused, is the Maillard reaction. This is a non-enzymatic reaction (Shallenberger et al., 1959) between reducing sugars and free amino acids which produces brown- and black pigmented products. Besides the unwanted discoloration and the bitter taste (Roe et al., 1990), the Maillard reaction is accused to generate acryl amide, a neurotoxin and a potential carcinogen (Mottram et al., 2002; Stadler et al., 2002). A major goal of potato breeders and the processing industry is to find a balance between the induction of sprouting and the expansion of diseases in too warm and optimal climates and the occurrence of cold-sweetening during cold-storage and a resulting impaired further processing capacity.

Starch exists as a water-insoluble granule having a defined and evolutionary conserved internal structure. The granule shape can be characteristic of a genus or a species, ranging from oval/round (genus *Solanum*) to irregular starch types in rice (Ellis et al., 1998). Potato starch has a complex quaternary structure and consists of polymers of glucose molecules (α -glucans) named amylose and amylopectin. Amylopectin, the major compound, is composed of α -1,4 linked glucans that are clustered together and hooked to longer spacer glucans by α -1,6 linkages. Amylose is often referred to as a smaller, essentially linear molecule with very few α -1,6 branches (Ball and Morell, 2003). Despite the simple composition of only two polyglucans, starch forms semi-crystalline structures which accumulate in plastids (Kötting et al., 2010). The semi-crystalline matrix allows the starch granule to grow to considerable size (Mouille et al., 1996). Starch in potato tubers possess a relatively high content of phosphate (0.5 %) compared to starch of potato and *Arabidopsis thaliana* leaves which contains around 0.1% of phosphorylated α -glucans (Mikkelsen and Blennow, 2005; Yu et al., 2001).

Transient starch is synthesized in chloroplasts of photosynthetic tissues by the conversion of triose-phosphate from recently fixed carbon of the Calvin cycle to Fructose-6-phosphate (Fru6P; Figure 1; Geiger and Servaites, 2001; Zeeman et al., 2007). Fru6P is then converted into Glucose-6-Phosphate and Glucose-1-Phosphate. The next step is the pyrophosphorylation into ADP-Glucose by AGPase which is then transferred to the non-reducing ends of amylose or amylopectin chains by branching, debranching and starch synthase. The starch biosynthetic pathway in leaves is assumed to be controlled by modulation of AGPase activity. It has been established that AGPase activity is allosterically regulated by levels of P_i as an inhibitor and by 3-PGA (3-phosphoglycerate from Calvin cycle) as an activator (Preiss, 1989). High rates of 3-PGA result in an increased level of starch synthesis. Additionally, it is known that AGPase is redox-regulated resulting in activation during the day and an inactivation during the night (Hendriks et al., 2003). Therefore, it is directly linked to the

diurnal rhythm. A surplus of triose-phosphates from the Calvin cycle is converted into sucrose and transported into sink tissues like seeds and tubers.

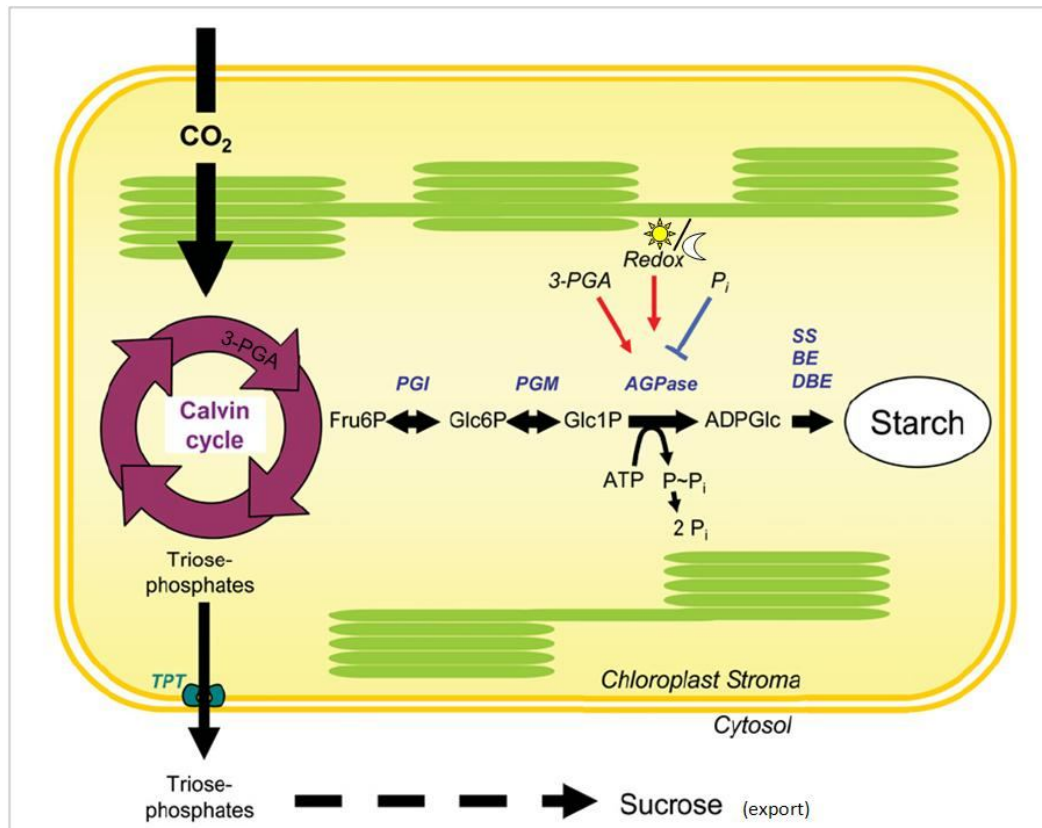


Figure 1: Synthesis of transient starch in chloroplasts

(Modified from Zeeman et al., 2007). Carbon is assimilated via the Calvin cycle and kept in the chloroplast for the synthesis of transient starch. A surplus of triosephosphates is converted to sucrose and exported to sink tissues. AGPase is allosterically and redox regulated. 3-PGA: 3-phosphoglycerate, Fru6P: fructose-6-phosphat, Glc6P: glucose-6-phosphate, Glc1P: glucose-1-phosphate, ADPGlc: ADP-glucose, TRT: triose-phosphate translocator, PGI: phosphoglucose isomerase, PGM: phosphoglucose mutase, AGPase: ADPGlc phosphorylase, SS: soluble or granule-bound starch synthase, BE: branching enzyme, DBE: debranching enzyme.

Storage starch is synthesized in heterotrophic tissues from imported sugars, mainly in form of sucrose which is transported via the phloem from the autotrophic tissues (source) to the sink tissues (Geiger and Servaites, 2001; Geigenberger, 2003). The sucrose is converted into hexoses like fructose and UDP-glucose via the sucrose synthase and subsequently imported as different hexose-phosphates into the amyloplast (Figure 2). Glucose-1-phosphate and glucose-6-phosphate are exchanged against P_i via the hexose phosphate translocator, whereas the ADP-glucose transport is mediated by the ADPGlc translocator. The latter one does not exist in *S. tuberosum*. As described for the transient starch synthesis, the hexoses need to be bound to ATP and are subsequently incorporated into the starch granule. Although the enzymes that synthesize amylopectin are known, the way in which they arrange the branching points to end up in the semi-crystalline structure is not fully understood (Zeeman et al., 2007). A certain part of the imported sucrose from source tissue is

converted into hexoses and used in the glycolysis and respiratory pathway for the synthesis of ATP which might be used amongst other things again for the starch synthesis.

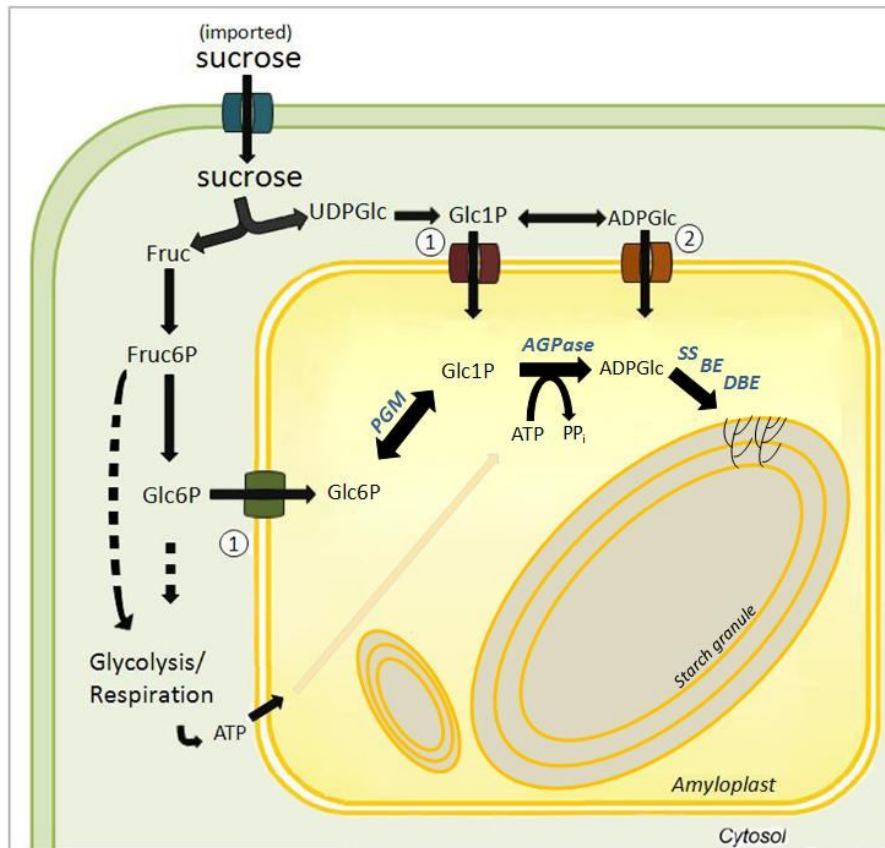


Figure 2: Synthesis of storage starch in heterotrophic tissues

(Modified from Geiger and Servaites, 2001; Zeeman et al., 2007). Sucrose is imported from source tissues into a cell of source tissue and either used for starch synthesis in chloroplasts and amyloplasts or used for the production of energy via glycolysis. Hexose-phosphates can be transported via hexose-phosphate translocators (1) or ADPG translocator (2). Fruc: fructose, Fruc6P: fructose-6-phosphate, UDPGlc: UDP-glucose.

The pathway of starch degradation has been extensively studied in the last years and the major steps in the pathway and the key genes/proteins have been identified (Figure 3; Zeeman et al., 2004; Zeeman et al., 2007). For the initiation of starch breakdown, the enzyme α -glucan water dikinase (GWD) is required (Smith et al., 2005). In a dikinase reaction, it phosphorylates the C⁶ position of α -glucans (Ritte et al., 2002). GWD is accompanied by the phosphoglucan water dikinase (PWD) which exclusively phosphorylates the C³ positions (Baunsgaard et al., 2005; Kötting et al., 2005; Ritte et al., 2006). It is suggested that PWD acts downstream of GWD and needs pre-phosphorylated starch as substrate (Fettke et al., 2009). The added phosphates are supposed to loosen the tight packing of the amylose and amylopectin chains to make them more accessible for degrading enzymes (Smith et al., 2005). Recently, a third enzyme involved in the initial attack was revealed, the so called laforin-like phosphoglucan phosphatase (SEX4) which is supposed to remove some of the beforehand introduced phosphates and their negative charge to additionally facilitate the accessibility for further enzymes (Kötting et al., 2009). The subsequent cleavage, degradation and linearization steps are

performed by exo- and endoamylases, debranching enzymes and disproportionating enzymes (Smith et al., 2003). These enzymes produce predominantly glucoses, maltoses and linear malto-oligosaccharides. Simultaneously, some of the oligosaccharides are further degraded into products like glucose-1-phosphates (performed by starch-phosphorylase; Duwenig et al., 1997), trioses and more maltoses which are thereupon transported out of the chloroplast into the cytosol. The best characterized transporters are the glucose transporter and the maltose transporter (MEX1; Weber et al., 2000; Niittylä et al., 2004). The fate of maltose and hexose-phosphates in the cytosol is still unclear. Smith et al. (2005) suggested several possible ways via heteroglycans or glucose intermediates which end up in the formation of sucrose. Sucrose can either be exported to other tissues (Turgeon, 1989) or can be transiently stored in the vacuole and be further degraded by acid invertase to glucose and fructose (Zeeman et al., 2004; Deschamps et al., 2008). In most plants, sucrose is the major form how carbon is transported because of the non-reducing nature of the disaccharide in which glucose and fructose are linked (Arai et al., 1992). Sucrose and its cleavage products are the central molecules for carbohydrate translocation, signaling and sensing in higher plants (Roitsch and González, 2004). As putative regulators of invertase, two invertase inhibitors have been characterized. One of the inhibitors showed high similarity to a Kunitz-type protease inhibitor, whereas the second inhibitor was similar to an apoplasmic invertase inhibitor from tobacco (Greiner et al., 1999; Glaczinski et al. 2002; Liu et al., 2010; Brummel et al., 2011).

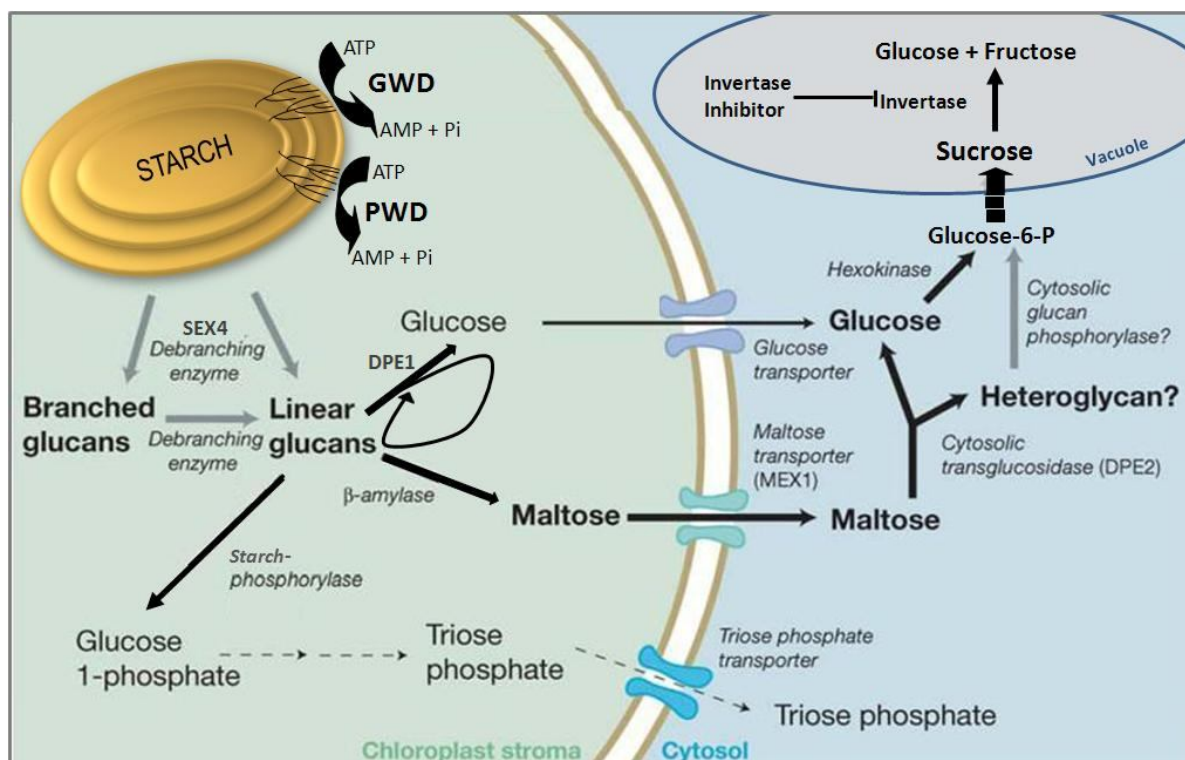


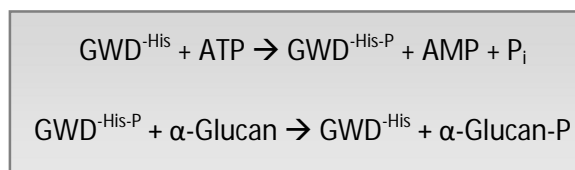
Figure 3: Starch degradation pathway

(Modified after Zeeman et al., 2004; Zeeman et al., 2007). The major net products of the starch degradation in the chloroplast are glucoses and maltoses which are exported into the cytosol and either used as energy or stored in the vacuole as sucrose. SEX4: laforin-like phosphatase, DPE1/2: disproportionating enzyme 1 and 2.

1.5 CANDIDATE GENES FOR STRUCTURAL, FUNCTIONAL AND GENETIC ANALYSES

1.5.1 ALPHA-GLUCAN WATER DIKINASE - GWD

The α -glucan water dikinase (GWD, EC 2.7.9.4) was initially identified as Ritte-1 (R1) in *Solanum tuberosum* by Lorberth et al. in 1998. In a dikinase reaction, GWD preferentially phosphorylates the C⁶ positions of the α -glucans in the amylopectin chains (Ritte et al., 2000; Ritte et al., 2006). In the first part of the dikinase reaction, the catalytic histidine of GWD becomes auto-phosphorylated and forms a stable phosphohistidine intermediate (Mikkelsen and Blennow, 2005). In the second step, this phosphate is transferred to the α -glucan molecule.



The phosphorylation ratio of starch in potato leaves is about 0.05 % (Lorberth et al., 1998), whereas in potato tubers it is up to 0.5 % (1:200 α -glucans; Mikkelsen and Blennow, 2005). The phosphate content of starches determines different physico-chemical properties, such as pasting properties, gel strength and stickiness (Mikkelsen et al., 2004). So far, the introduction of a negatively charged phosphate is assumed to disturb the tight and condensed packing of the amylopectin chains in the semi-crystalline structure of the starch granule to render it more accessible to further degrading enzymes (Edner et al., 2007).

Antisense repression in potato plants resulted in a reduction in the phosphate content of starch (Lorberth et al., 1998). Additionally, a starch excess phenotype in potato leaves was observed which was due to the incapability to degrade and mobilize starch. The starch content in the potato leaves increased in dry matter between 50 and 90 % (Lorberth et al., 1998). This phenotype only occurred in those parts of the leaf that had already undergone sink-source transition and was therefore not detectable in very young leaves (Figure 3A and 3B). Although no effects were observed on the level of soluble sugar contents in leaves, cold-stored potato tubers showed a reduction in cold-sweetening. The phenotype of a GWD-deficient *starch-excess 1 (sex1)* mutant of *Arabidopsis thaliana* supported the findings of reduced starch degradation and resulted in higher leaf starch content (Yu et al., 2001). The phenotype was reversed when the mutant was complemented with the wild-type gene (Figure 3C). Additionally, the starch granules are up to 90 % less phosphorylated and bigger in size compared to wild-type starch granules.

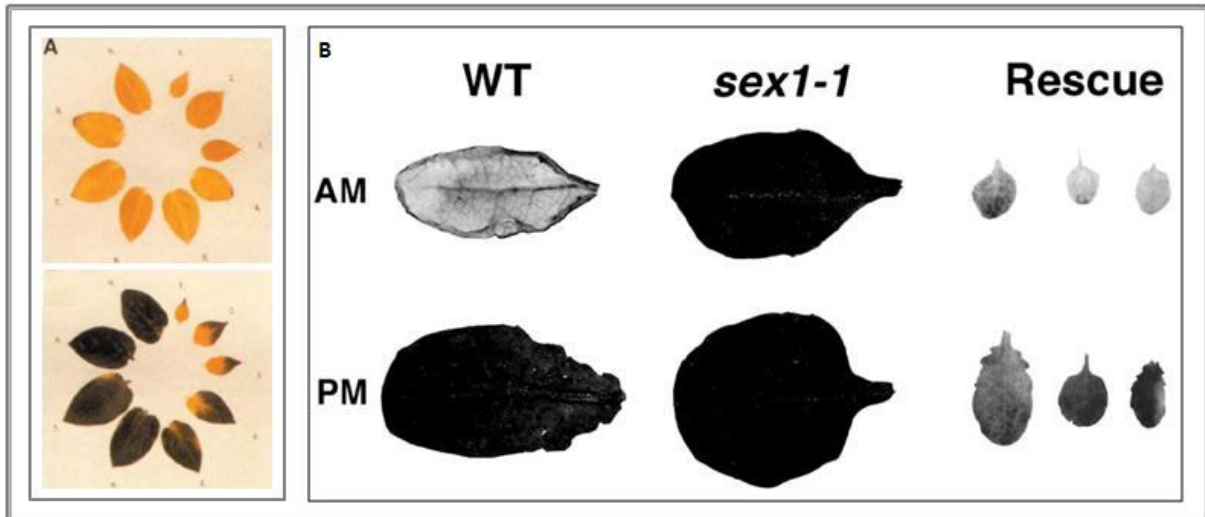


Figure 4: Phenotypes of GWD mutants.

A: Potato wild-type leaves and leaves containing a *GWD* antisense-construct, after storage in the dark for 72 hours (derived from Lorberth et al., 1998). The leaves were stained with Lugol's solution to visualize the starch (Amylose). **B:** Phenotype of *A. thaliana* mutant *sex1-1* and complementation (rescue) of the mutant with the wild-type gene (derived from Yu et al., 2011). WT: wild-type gene of *GWD*, AM: Harvest of leaves at the beginning of the photoperiod, PM: Harvest of leaves after 12 h of photoperiod

Including its chloroplastic target peptide (cTP), the GWD in potato forms a 163 kDa protein with a coding sequence of about 4.4 kb and corresponding 1464 amino acids (Lorberth et al., 1998). After the transport into the chloroplast, the cTP is cleaved and leads to a mature protein of 153 kDa. Besides the cTP, several other functional domains have been identified for GWD in *A. thaliana* (Yu et al., 2001). One of the most important regions is the catalytic histidine which is auto-phosphorylated in the intermediate step of the dikinase reaction. Mikkelsen et al. (2004) showed that the sequence and the mechanism are similar to a pyruvate, phosphate dikinase (PPDK) from *E. coli* and *Zea mays*. The histidine is located upstream of the C-terminal nucleotide binding domain. In the mature protein of GWD, a starch binding domain is located at the N-terminus. It belongs to the carbohydrate binding module family 45 (CBM 45) and consists of two domains (Glaring et al., 2011). Besides GWD, so far only α -amylase (AMY, EC 3.2.1.1) was identified to contain a starch binding domain of the same CBM family (Mikkelsen et al., 2006; Glaring et al., 2011). These domains are non-catalytic and mediate the binding of the protein to the substrate which are predominantly polysaccharides (Guillén et al., 2010).

In *A. thaliana*, a homolog to GWD has been identified as GWD2. Contrasting to GWD, it is not located in the plastids and does not show any differences in growth or starch and sugar levels (Glaring et al., 2007). No GWD2 homolog was identified in potato and therefore it is hypothesized that GWD is a single-copy gene and was mapped on chromosome 5 in potato (Chen et al., 2001). Several GWD homologues have already been identified in a great number of plants throughout the plant kingdom (Ritte et al., 2000).

1.5.2 PHOSPHOGLUCAN WATER DIKINASE - PWD

The phosphoglucan water dikinase (PWD) was initially identified in *A. thaliana* and declared mistakenly as a GWD homologue (AtGWD3). Later, PWD was described in more detail independently by two groups and was annotated as AtPWD which forms a separate subgroup from AtGWD (Kötting et al., 2005; Baunsgaard et al., 2005). On amino acid level, GWD and PWD share only 14 % identity and 24 % similarity in *A. thaliana* (Kötting et al., 2005). As a most characteristic fact, PWD was identified to act only on pre-phosphorylated starch granules. Kötting et al. (2005) showed in an *in vitro* binding assay that the pre-phosphorylation has to be done by GWD. However, the mechanism of the dikinase reaction is identical including an intermediate auto-phosphorylation step of a catalytic histidine. Besides the prerequisite of pre-phosphorylated starch, it was found that PWD specifically phosphorylates the C³ hydroxyl group of the glucose residue (Baunsgaard et al., 2005; Kötting et al., 2005). Whether the same or random α -glucans are phosphorylated by GWD and PWD remains still unclear.

Arabidopsis T-DNA insertion lines (*Atgwd3*) and RNAi silenced lines (*pwd*) displayed elevated levels of transitory starch because of a gradual starch accumulation during the diurnal rhythm (Baunsgaard et al., 2005; Kötting et al., 2005). Compared to the starch excess phenotype of *gwd* (*sex1*), it can be concluded that they resemble each other, but *AtGwd3* shows a general less severe starch accumulation and only a slight effect on plant development and growth (Kötting et al., 2005). Figure 4A shows clearly the reduced starch degradation in the night in the *Atgwd3* mutant compared to the wild-type plant (Baunsgaard et al., 2005). In general, starch turnover occurs in *pwd* plants, but the rate of starch degradation is lower than of the wild-type plants (Figure 4B; Kötting et al., 2005).

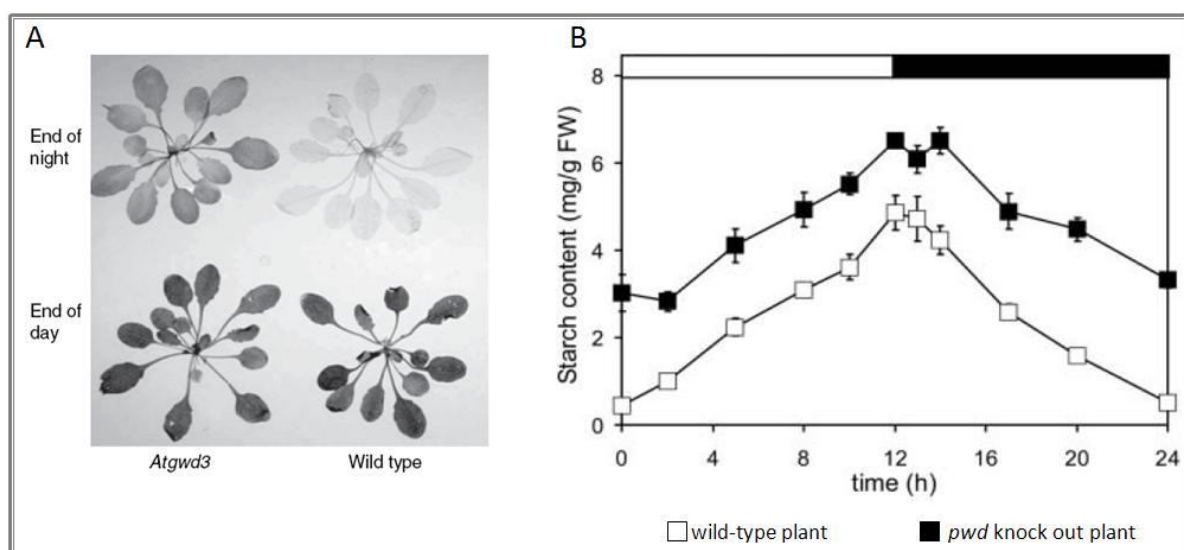


Figure 5: Phenotypes of PWD mutants.

A: Higher starch content can be observed after the dark period in *PWD* mutants (*Atgwd3*) compared to wild-type plants (derived from Baunsgaard et al., 2005). **B:** Starch turnover in the diurnal rhythm occurs in *PWD* knock-out mutants (*pwd*) less efficient than in wild-type plants (derived from Kötting et al., 2005)

Although the amino acid sequence similarity between GWD and PWD in *A. thaliana* is rather low, a plastid target peptide (cTP) and a conserved catalytic histidine as an active site have been identified for PWD (Kötting et al., 2005). The surrounding region of the catalytic histidine showed a high degree of conservation between PWD, GWD and other dikinases (Kötting et al., 2005). Similar to GWD, PWD contains a nucleotide binding domain at the C-terminal part and uses ATP as a phosphate donor (Kötting et al., 2005). The starch binding domain was characterized as a member of the carbohydrate binding module family 20 (CBM20; Christiansen et al., 2009).

To date, homologues have been identified in several higher plants like *Oryza sativa*, *Medicago truncatula* and *Zea mays* (Baunsgaard et al. 2005). No gene or mutant phenotype has been described for PWD in *S. tuberosum*.

1.5.3 INVERTASES

Invertases (EC 3.2.1.26) are ubiquitous enzymes that irreversibly hydrolyze sucrose into the reducing sugars glucose and fructose. Besides an important role in the partitioning of carbon between autotrophic source tissues (e.g. plant leaves) and heterotrophic source tissues (seeds, fruits and tubers), they also function in plant development and in the response to biotic and abiotic stress (Draffehn et al., 2010). Based on their solubility, subcellular localization and pH optimum, three different types of invertase isoenzymes can be distinguished: vacuolar, cell wall bound and neutral invertases (reviewed in Roitsch and González, 2004). Although vacuolar and cell wall invertases share some enzymatic and biochemical properties and show a high degree of sequence similarity, they are located in different compartments of the plant cell. Whereas vacuolar invertase contains a vacuolar target peptide and is mainly localized in the vacuole and to some extent in the cytoplasm and the ER, cell wall bound invertases are referred to as extracellular, apoplasmic or free-space invertases (Bhaskar et al., 2009; Roitsch and González, 2004). In both groups, characteristic and conserved sequence motifs for β -fructosidases (NDPNG; Sturm and Chrispeels, 1990) and catalytic sides (MWECP; Ji et al., 2005) can be identified.

Anti-sense mutants of cell wall bound and vacuolar invertase in carrot showed elevated levels of starch and sucrose and phenotypic alterations in early developmental steps (Tang et al., 1999). In *A. thaliana*, single and double mutants of vacuolar invertase resulted in changed expression levels for cell wall bound invertase and sucrose synthase (Huang, 2006). In potato, one vacuolar invertase is located on chromosome 3 and is named *Pain-1*. The activity of vacuolar invertase is correlated with the hexose/sucrose ratio in potato tubers (Pressey and Shaw, 1966). Zrenner et al. (1996) showed that *Pain-1* is differentially expressed during cold storage. Besides *Pain-1*, a second putative vacuolar

invertase was identified (G. Orjeda, Lima, Universidad Peruana; Liu et al., 2010). In potato, 4 cell wall invertases are encoded in two tandem repeats on chromosome 9 and chromosome 10, respectively. Liu et al. (2010) showed unique expression patterns for all 4 cell wall bound invertases. They are predominantly expressed in the aerial tissues, like leaves and flowers. Only one member possesses measurable expression levels in tubers, stolons and roots. No change of the expression level in the tuber was observed for cell wall bound invertases upon cold treatment.

The third group of invertases are neutral invertase which have not been identified and characterized in *S. tuberosum* so far. They are encoded by a small gene family of 9 members in the *A. thaliana* genome (Nonis et al., 2008). Neutral invertases were assumed to be localized in the cytoplasm upon their optimal pH between 6.8 and 8.0, but the identification of putative target peptides indicated localization in mitochondria and chloroplasts (Ji et al., 2005). Murayama and Handa (2007) demonstrated first proofs for the presence of neutral invertase in these organelles. Besides low transcript levels, only a faint enzyme activity was measured in different tissues, suggesting that neutral invertases are only expressed at basal level (Nonis et al., 2008). A role in the development and ripening of fruits is assumed (Nonis et al., 2007).

Due to the various roles of sucrose, many different functions can be assumed for invertases. Figure 5 shows an overview of proposed functions of plant invertases (Sturm, 1999).

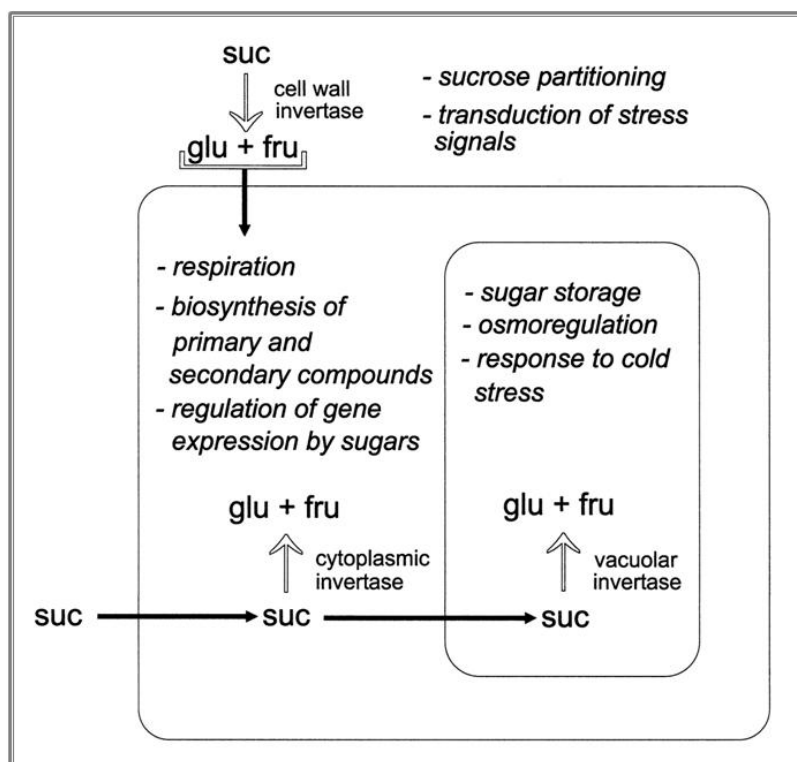


Figure 6: Localization and proposed functions of plant invertases

(Derived from Sturm, 1999). Most plants possess vacuolar, cell wall bound and cytoplasmic (neutral) invertases. Besides different subcellular locations, the generated hexoses (glu + fru) are assumed to have different fates and functions.

Based on the assumed presence of invertases in many different compartments of the cell, it can be concluded that invertases maintain the ratio of hexoses to sucrose and play an important part in signaling. A second important function is of course the mobilization of the reducing and active sugars glucose and fructose.

For the cell wall invertase a small polypeptide of 17 kDa was identified that exhibited the capacity to inhibit this invertase in tobacco (Weil et al., 1994). As reviewed in Rausch and Greiner (2004), cell wall and vacuolar invertases are supposed to be both post-transcriptionally regulated by these small inhibitors. The first evidence for an invertase inhibitor was presented by Schwimmer et al. (1961) and later confirmed by Pressey (1966, 1967). Invertase and inhibitor are assumed to form tight protein complexes. In the recent years, many inhibitor sequences have been identified in various plants, although it was revealed that this protein family is not well conserved. For tobacco, the two inhibitors for cell wall and vacuolar invertase share only 47 % of sequence similarity (Rausch and Greiner, 2004). Interestingly, the inhibitors displayed a certain similarity to pectin methylesterase inhibitor (PMEI; Carmadella et al., 2000) implying different substrate specificities for a large family of small inhibitors. A further putative invertase inhibitor was obtained from protein extracts of potato tubers and was identified to belong to the Kunitz-type protease family (Glaczinski et al., 2002). This inhibitor will be introduced in the following section.

In potato, one vacuolar (*Pain-1*) and four cell wall bound invertases were tested in candidate gene approaches for their contribution to starch and chip quality traits (Li et al., 2005; Li et al., 2008). *Pain-1* showed the highest association with tuber starch content and with chip quality and up to 12 % of the phenotypic variance of the tested traits can be explained. The two cell wall bound invertases on chromosome 9 displayed a moderate association with chip quality, whereas for the two remaining invertase either no or a negative association was shown (Li et al., 2008). The 6th invertase which was independently identified by *in silico* analyses by Liu et al. (2010) and G. Orjeda (personal communication), remains to be tested for association.

1.5.4 LEUCINE AMINOPEPTIDASE - LAP

A comparative 2D proteomics approach of different potato genotypes which were selected upon their processing qualities, revealed an interesting protein expression pattern for a leucine aminopeptidase (LAP; M. Fischer, unpublished data). The protein was solely detected in a pool of genotypes with good processing characteristics and was absent in the comparison group of pooled genotypes with bad processing characteristics. Upon this result, the leucine aminopeptidase was selected as a novel candidate for the candidate gene approach.

Leucine aminopeptidases have been characterized to be cell maintenance enzymes with a critical role in the turnover of peptides (Matsui et al., 2006). They are supposed to be activated upon mechanical wounding and as an inducible component of the defence response and subsequent higher levels of abscisic and jasmonic acids (Hildmann et al., 1992; Pautot et al., 1993; Gu et al., 1996; Chao et al., 1999). No clear indications of a contribution to the carbohydrate metabolism and sugar signaling pathways were obtained and described so far.

LAPs nomenclature is complex, but two major groups of either M17 or M1 enzymes were described. Both groups are metallopeptidases and cleave N-terminal residues from proteins and peptides, but they have distinctive structures, enzymatic mechanisms and biological roles. The LAP obtained from the proteomics approach was characterized as a member of the M17 group (EC 3.4.11.1) which are assumed to form hexamers and bind two cations. These proteins are highly conserved at the amino acid level and have been identified in bacteria, animals and plants (Matsui et al., 2006). LAPs contain a catalytic domain which is split in a C- and N-terminal part. Although they are named leucine aminopeptidases, they have broader substrate specificities. Besides the amino acid leucine, further peptides like arginine and methionine serve as N-terminal cleavage sites (Herbers et al., 1994). As a monomer, LAP possess a size of around 55 kDa, in alkaline pH (8.5) conditions, LAPs predominantly form hexamers (Sträter et al., 1999).

In potato, the first leucine aminopeptidase was identified by Hildmann et al. (1992) after the application of methyl jasmonate and abscisic acid and mechanical wounding. Interestingly, this leucine aminopeptidase belongs to the group of acidic or chloroplastic LAPs which is assumed to be unique in solanaceous plants (Chao et al., 2000). They can be detected in highly elevated transcript and protein levels after wounding. Besides acidic LAPs, neutral LAPs have been identified in almost all plants throughout the plant kingdom. The neutral LAPs are constitutively present, but only at a basal level which is supposed to be required for the cell maintenance in vegetative and reproductive tissues (Tu et al., 2003). It is hypothesized that the N-terminus may influence the half-life of a protein (Callis and Vierstra, 2000). Therefore, a cleavage by LAP_N at the amino terminal part of a protein is accompanied by a modification of the N-terminus and may alter the half-life or processing. This is assumed to be a normal and basal part of the ubiquitin-dependent protein degradation (Tu et al., 2003).

LAPs are best characterized in *Solanum lycopersicum*. For acidic leucine aminopeptidases, it is hypothesized that they are located in a tandem repeat and are >99% identical (Chao et al., 2000), whereas the neutral leucine aminopeptidase is assumed to be a single-copy gene (Tu et al., 2003). However, both LAPs have not been assigned to explicit loci on the tomato chromosomes.

1.5.5 THE PUTATIVE INVERTASE INHIBITOR – A MEMBER OF THE KUNITZ TYPE PROTEASE INHIBITORS

In potato tubers, Kunitz-type protease inhibitors occur numerously and are highly diverse. They possess a molecular weight of around 21 kDa. Most members of this family are organized in a single cluster on chromosome 3 (Heibges et al., 2003a). Besides patatin which represents around 40% of the storage proteins of tubers, they form the second major component of tuber proteins (Rosahl et al., 1986). They were initially isolated from soybean seeds by Kunitz et al. in 1945 and named by him. Up to now, many different inhibitors have been isolated and characterized as protease inhibitors (Stiekema et al., 1988). They generally inhibit serine, cysteine and aspartic peptidases and were classified into the groups A, B and C of Kunitz-type protease inhibitors. Group A and B display a higher degree of similarity among each other compared to group C (Heibges et al., 2003a). Some of the protease inhibitors were identified to inhibit non-proteolytic enzymes like amylase and invertase (Shewry, 1999; Glaczinski et al., 2002). In barley, a bi-functional inhibitor was characterized which showed inhibition of amylase and substilisin (Svendsen et al., 1986).

The invertase inhibitor protein has been purified from potato tubers. Thereupon, it was sequenced and several corresponding cDNA clones have been obtained (Glaczinski et al., 2002; Heibges et al., 2003a and 2003b). The sequences were identified as members of the group C of the Kunitz-type Inhibitors. Within group C, the sequences show at least 90% similarity (Heibges et al., 2003a), so it is not clear if the clones are allelic variants or isoforms. Most of the obtained clones were not full-length, but the estimated length for all of them amounts to 660 bp corresponding to 220 amino acids (Glaczinski et al., 2002). For the N-terminal part, a signal peptide of around 42 amino acids was identified which was assumed to mediate the transport in to the vacuole. Further characteristic differences for the putative invertase inhibitor were recognized by comparisons to known trypsin inhibitors of the same group C of Kunitz-type protease inhibitors (Glaczinski et al., 2002). The invertase inhibitors possess a characteristic methionine instead of an arginine at the active site for trypsin inhibition, hence this function is assumed to be reduced or impaired. A second prominent methionine was found close to the C-terminal part. Interestingly, both regions surrounding the two methionines displayed the highest level of variation and divergence.

Although invertase was clearly inhibited after adding the invertase inhibitor (Glaczinski et al., 2002), the obtained result was not reproducible in a heterologous system (Heibges, 2001). Additionally, it remains to be elucidated which of the numerous sequences encodes the invertase inhibitor.

1.6 OBJECTIVES

A candidate gene approach is a fast and powerful tool to analyze the contribution of candidate genes to agronomically important but quantitative traits. In *Solanum tuberosum*, starch content and chip quality of tubers represent some of these relevant quantitative traits for breeders and industry. The main objective of this thesis is to analyze functional and/or positional candidates according to their contribution to the starch and chip quality by using amplicon sequencing based association analysis. The candidate genes will be selected upon one or more different criteria and shall either:

- (I) have a known functional role in one of the pathways involved in starch and chip quality (functional candidate),
- (II) be a positional candidate upon close proximity to a QTL,
- (III) display a hypothetical but possible role due to *in silico* derived sequence homologies to known and analyzed candidate genes,
- (IV) have a regulatory function on a known and analyzed candidate gene and/or
- (V) be obtained from a reverse “-omics” approach based on a case and control population for processing quality traits.

As a prerequisite, the genomic organisations and sequences have to be identified to facilitate the analysis of important and functional parts of the candidate genes. Additionally, first functional evidences for the contribution shall be obtained for one of the selected candidates.

2 MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 CHEMICALS

All chemicals were purchased from the companies listed below. The company's names and the corresponding registered offices are stated:

Biozym Scientific GmbH (Hessisch Oldendorf), Carl Roth GmbH (Karlsruhe), Difco Laboratories (Detroit, Michigan, USA), GE Healthcare Europe GmbH (Freiburg), Invitrogen GmbH & Co. KG (Karlsruhe), Merck KGaA (Darmstadt), Peqlab Biotechnologie GmbH (Erlangen), Roche Deutschland GmbH (Grenzach-Wyhlen) and Sigma-Aldrich Chemie GmbH (Taufkirchen).

2.1.2 ENZYMES

Restriction enzymes and corresponding buffers were obtained from New England Biolabs GmbH (Frankfurt/Main), Roche Deutschland GmbH and MBI Fermentas GmbH (St. Leon-Rot).

2.1.3 BACTERIAL AND YEAST STRAINS

All bacteria and yeast strains are commercially available (Invitrogen, Karlsruhe) and were used or modified in accordance to the instructions supplied by the manufacturer, if not stated otherwise.

Table 1: Bacterial Strains.

Bacterial strain	Genotype	Reference
TOP10F' <i>E. coli</i>	F ⁻ {proAB, lacI ^q , lacZΔM15, Tn10 (Tet ^R)} <i>mcrA</i> , Δ(<i>mrr-hsdRMS-mcrBC</i>), φ80 <i>lacZ</i> ΔM15, Δ <i>lacX74</i> , <i>deoR</i> , <i>recA1</i> , <i>endA1</i> , λ ⁻ , <i>araD139</i> , Δ(<i>ara, leu</i>)7697, <i>galU</i> , <i>galK</i> , <i>rpsL</i> (Str ^R), <i>nupG</i>	EasySelect TM <i>Pichia</i> Expression Kit (Invitrogen)
ElectroMAXTM DH10BTM	F ⁻ , <i>mcrA</i> , Δ(<i>mrr-hsdRMS-mcrBC</i>), φ80 <i>lacZ</i> ΔM15, Δ <i>lacX74</i> , <i>recA1</i> , <i>endA1</i> , <i>araD139</i> , Δ(<i>ara, leu</i>)7697, <i>galU</i> , <i>galK</i> , λ ⁻ , <i>rpsL</i> , <i>nupG</i>	Calvin and Hanawalt, 1988
One Shot[®] <i>ccdB</i> SurvivalTM 2 T1	F ⁻ , <i>mcrA</i> , Δ(<i>mrr-hsdRMS-mcrBC</i>), φ80 <i>lacZ</i> ΔM15, Δ <i>lacX74</i> , <i>recA1</i> , <i>araD139</i> , Δ(<i>ara-leu</i>)7697, <i>galU</i> , <i>galK</i> , <i>rpsL</i> (Str ^R), <i>endA1</i> , <i>nupG</i> , <i>fhuA::IS2</i>	Bernard and Couturier, 1992

Table 2: Yeast strains.

Yeast strain	Genotype	Reference
MaV203	MAT α , <i>leu2-3, 112, trp1-901, his3Δ200, ade2-101, gal4Δ, gal80Δ, SPAL10::<i>URA3, GAL1::<i>lacZ, HIS3_{UAS GAL1}::HIS3@LYS2, can1^R, cyh2^R</i></i></i>	Vidal, 1997
X-33	<i>Pichia pastoris</i> wild-type strain	EasySelect™ <i>Pichia</i> Expression Kit
GS115	<i>Pichia pastoris</i> strain: mutation in histidinol dehydrogenase gene (<i>his4</i>)	EasySelect™ <i>Pichia</i> Expression Kit
KM71H	<i>Pichia pastoris</i> parent strain: mutation in argininosuccinate lyase gene (<i>arg4</i>)	EasySelect™ <i>Pichia</i> Expression Kit
GS115/His⁺ Mut^S Albumin	Expression and secretion of albumin; mutation in histidinol dehydrogenase gene (<i>his4</i>), positive control for secretion	EasySelect™ <i>Pichia</i> Expression Kit
GS115/pPICZ/lacZ Mut+ β-galactosidase	Expression of β -galactosidase, inducible with methanol, positive control for resistance against Zeocin, expression, purification and detection	EasySelect™ <i>Pichia</i> Expression Kit

2.1.4 VECTORS

The following vectors were used for heterologous expression approaches and as maintenance vectors for cloned alleles.

Table 3: List of used vectors for cloning, transformation and propagation.

Vector	Source	Resistance
pPICZα A	Invitrogen	Zeocin™
pESP-2	Stratagene (Waldbronn)	Ampicillin
pGEM - T	Promega	Ampicillin

2.1.5 PLANT MATERIAL

Genomic DNA of tetraploid *S. tuberosum* plants was obtained from the GABICHIPS population which was described in Li et al. (2008). It was provided by the breeding companies BIOPLANT GmbH (Ebster), SaKa Pflanzenzucht GbR (Windeby) and NORIKA GmbH (Groß Lüsewitz). The potato tuber material of the BIOSOL project was provided by BIOPLANT GmbH (Ebster) and consists of 40 potato cultivars. The set was selected upon good and bad chipping quality, whereas each group comprises 20 classified genotypes respectively. The 40 cultivars were phenotyped for the amount of reducing sugars (glucose and fructose) after different weeks of cold-storage (section 2.2.1)

2.1.6 MEDIA AND SOLUTIONS

All solutions and media were set up with double distilled Milli-Q ultrapure water (Millipore, Schwalbach) and either autoclaved or sterilized by filtration afterwards (Millipore; Corning, Corning NY).

2.1.6.1 MEDIA FOR CULTIVATION OF BACTERIA

LB (Lysogeny Broth): 1 % tryptone
 0.5 % yeast extract
 1 % NaCl
 pH 7.0
 (for plates: add 1.5 % agar)

For the salt-sensitive antibiotic Zeocin, the low-salt LB version was used.

Low Salt LB: 1 % tryptone
 (LB-Lennox) 0.5 % yeast extract
 0.5 % NaCl
 pH 7.5 (with 1 M NaOH)
 (for plates: add 1.5 % agar)

2.1.6.2 ANTIBIOTICS

All antibiotics were set up with double distilled Milli-Q ultrapure water (Millipore) and sterilized by filtration. Tetracycline was dissolved in 70 % EtOH.

Table 4: List of Antibiotics.

Antibiotic	Stock concentration	Working concentration
Ampicillin (Amp)	50 mg/ml	100 µg/ml
Gentamicin (Gent)	10 mg/ml	10 µg/ml
Kanamycin (Kan)	10 mg/ml	50 µg/ml
Tetracycline (Tet)	5 mg/ml	10 µg/ml
ZeocinTM (Zeo)	100 mg/ml	100 µg/ml

2.1.6.3 MEDIA FOR CULTIVATION OF YEAST

YPD/YPAD: 1 % Bacto yeast extract
2 % Bacto peptone
2 % dextrose (glucose)
(0.01 % adenine sulfate)
(for plates: add 2 % agar)

2.1.6.4 GENERAL SOLUTIONS

50X TAE: 2 M Tris
2 M acetic acid
50 mM EDTA (pH 8.0)

10X TBS: Tris buffered saline: 200 mM Tris
1.5 M NaCl
pH 7.5

10X PBS: Phosphate buffered saline: 200 mM Na₂HPO₄
1.5 M NaCl
pH 7.4

Ethidium bromide: 10 mg/ml stock solution

6X Loading-Dye: 0.25 % (v/v) Orange G
0.25 % (v/v) Xylene cyanol FF
40 % (w/v) sucrose in H₂O
in 1 X TAE

2.1.7 OLIGONUCLEOTIDES

Primers were custom synthesized by Metabion International AG (Martinsried) and Sigma-Aldrich Chemie GmbH (Taufkirchen). All used primers are listed in table 5.

Table 5: List of primer combinations.

	Primer name	5' – 3' sequence	T _a in °C
BAC library screens and genomic organization of GWD and PWD			
GWD probes	LS06_probe_2f	GTTGTCGTTGTGGATGAGTTGCT	57
	LS08_probe_4r	CTCCCTGATTTATGTCGTCTGAA	
	LS21_GWD_8f	ATAAAAGTCAAAGCAAAGAAGAGCCT	57
	LS20_GWD_6r	GGATAATGCCAGTGAAGAGTAAT	
BAC orientation			
	T3	AATTAACCCTCACTAAAGGG	54
	LS25_GWD3'fl_f	GCTTATCCAGGACGTGCTTTGAGTTTTATCTGC	
GWD BAC full-length			
	LS27_GWD5'fl_f	GGAGGATATGAGTAATTCCTTAGG	57
	LS24_GWD5'fl_r	GCAGTGTCTCGTATCTCCAGTCTCAG	
	LS25_GWD3'fl_f	GCTTATCCAGGACGTGCTTTGAGTTTTATCTGC	57
	LS26_GWD3'fl_r	ATCGTCGACCTCGAGTCACATCTGTGGTCTTGT	
PWD probe			
	LS15_PWD_4f	GCACGAGGGCTTGTTATACATTG	57
	LS16_PWD_4r	GAAGTTATGATTGTTGTCCGTTGG	
PWD BAC full-length			
	LS45_LeBAC_3f	ATGGAGGGTCCTTCTTTACTACAC	55
	LS46_LeBAC_1r	CTTCCAGGACCCTAATTCTTTG	
	LS47_PWD_9f	GATGGATTGGACAGAAAATGGATGG	59
	LS51_PWD_13r	CGTGAGAGCCTCCAAGCGATGTGAAC	
	LS09_PWD_f	TTCACATCGCTTGGAGGCT	52
	LS50_PWD_12r	TCACGAAGAGCATTGAGAGAG	
	LS39_PWD_2f	ATAGAGTGTCCCTCAGATTGT	55
	LS49_PWD_2r	TCTTACCATCTTCGCTGCCT	
	LS33_PWD_5f	ACGCAGCGAAGATGGTAAGA	53
	LS34_PWD_5r	CGAAACCTGAAAAACCACTAC	
	LS33_PWD_5f	ACGCAGCGAAGATGGTAAGA	53
	LS61_PWD_3r	GCAAGCATTGAGATAAACCCAAGTG	
	LS35_PWD_6f	TACAAAGCACTGGGTTTATCTGA	55
	LS36_PWD_6r	GGAGGTAGAAAGTTGTTTCTTAAG	
	LS35_PWD_6f	TACAAAGCACTGGGTTTATCTGA	55
	LS52_PWD_13f	GCATCAACAAGAGGAATAACACC	
	LS56_PWD_7f	CTCCACGAGGTAGGGTAAGGTCTG	56
	LS52_PWD_13f	GCATCAACAAGAGGAATAACACC	

Table 5: continued

	LS59_PWD_18f	GTGTTATTCCTCTTGTTGATGC	57
	LS16_PWD_4r	GAAGTTATGATTGTTGTCGTTGG	
	LS29_PWD_11f	CCCAACGGACAACAATCATAACTT	57
	LS30_PWD_11r	ATAAATTACTTGTGGAAACAGCCTCT	
<i>PWD</i> cDNA full-length			
	LS45_BAC_3f	ATGGAGGGTCCTTCTTTACTACAC	59
	LS51_PWD_13r	CGTGAGAGCCTCCAAGCGATGTGAAC	
	LS43_PWD_1f	ATGTGCTTTAATATGGATTCTATGC	54
	LS51_PWD_13r	CGTGAGAGCCTCCAAGCGATGTGAAC	
	LS09_PWD_f	TTCACATCGCTTGGAGGCT	52
	LS50_PWD_12r	TCACGAAGAGCATTAGAGAG	
	LS47_PWD_9f	GATGGATTGGACAGAAAATGGATGG	54
	LS50_PWD_12r	TCACGAAGAGCATTAGAGAG	
	LS48_PWD_10f	GAGAATCAGGGACATCGCTCATAG	53
	LS49_PWD_2r	TCTTACCATCTTCGCTGCCT	
	LS48_PWD_10f	GAGAATCAGGGACATCGCTCATAG	56
	LS34_PWD_5r	CGAAACCTGAAAAACCACTACTAC	
	LS33_PWD_5f	ACGCAGCGAAGATGGTAAGA	53
	LS52_PWD_13f	GCATCAACAAGAGGAATAACACC	
	LS57_PWD_9r	CAAGTAATGCGTCTTCTACAGGTGC	57
	LS31_PWD_1r	GACATCTTGAGGAGAACCAAATT	
	LS29_PWD_11f	CCCAACGGACAACAATCATAACTT	57
	LS30_PWD_11r	ATAAATTACTTGTGGAAACAGCCTCT	
Expression analysis of Kunitz-type protease inhibitors			
Group specific primer			
Group A	Turrà et al., 2009	ACCCAGTGAATCTCCTCTAC	60
		ACAGAAGTCATCATCAGAC	
Group B	Turrà et al., 2009	TACCTAGTGATGCTACTCCA	62
		AACAGAATTGATCATCAGAGG	
Group C	LS95_grpC	TACTCCCTGAAGTTTATGACCA	56
	LS98_grpC	CITIACCACCAGTTACCACC	
Allele/subgroup specific primers			
Group 1421	L121_21Spec_2f	CCCACAACCTGTCATGATAATC	56
	L105_1421_1r	TTGATATAAACAGCAGTCATGAG	
Group 1422	LS81_22Spec_f	ACAACCTTGTCATGCTGATGATA	54
	L102_1422_r	TTCGTC AACACACA ACTGTGA	
Group 1423	LS89_23Spec_f	TTTATGACCAAGACGGCGAA	53
	L103_1423_r	CGACATGTGCTGCAACACA	TD
Group 1421+ 1422	LS78_Mwb_f	AAAAGGCACGCCCGCCA	54
	LS76_M_r	TGCAGTAACTTGTATACATTTTCAT	

Table 5: continued

Pyrosequencing for relative expression analysis			
Amplicon	LS95_Py1_f	TACTCCCTGAAGTTTATGACCA	56
	LS98_Py2_r	CITIACCACCAGTTACCACC	
	L111_Py_Biotin_r	(Btn)CITIACCACCAGTTACCACC	
Sequencing	L106_PyroSeq1	GGAAAAGGCACGCCCGT	-
	L118_PyroSeq4	GGAAACCTTCAATGCCCAA	-
Association analyses			
<i>GWD</i>	LS06_probe_2f	GTTGTCGTTGTGGATGAGTTGCT	57
	LS08_probe_4r	CTCCCTGATTTATGTCGTCTGAA	
	LS21_GWD_8f	ATAAAAGTCAAAGCAAAGAAGAGCCT	57
	LS20_GWD_6r	GGATAATGCCCAGTGAAGAGTAAT	
<i>PWD</i>	LS60_PWD_19f	GGTCTGATGATCTATCTGATTGC	57
	LS31_PWD_1r	GACATCTTGAGGAGAACCAAACCTT	
Putative Invertase Inhibitor (Group C)	L170_g1421_f	TCAATAGAATTCCTTACCCG	55
	L171_g1421_r	CAATAACGATATCTAGACCTGATTAAC	
	LS65_kun_fl_r	CTACGCCTTGATGAACACAAATG	-
<i>LAP</i>	L164_StLAP_14f	GCTTCCTGGTCTTGGCTCA	60
	L165_StLAP_14r	GCAGCCAGGTCAGAAATCAA	
Invertase 6 (<i>Inv6</i>)	L183_inv6_7f	GTTCTCATCCCACCACCCG	56
	L173_inv6_1r	CGTTCACCAGATCCACCACTC	
Cloning for heterologous expression in <i>P. pastoris</i>			
Specific primer	L125_Inhpichia_f	ATCGAATTC CTTGTA CT CCCTGAAGTTTATGAC	55
	L126_Inhpichia_r	TGTATGCGGC CGCCTTGATGAACACAAATG	
	L127_Invpichia_f	GAGGAATTC ATGGCCACGCAGTACC	55
	L128_Invpichia_r	TGTATGCGGCCGC CAAGTCTTGCAAGGGGA	
General primer	5'AOX	GACTGGTTCCAATTGACAAGC	-
	3'AOX	GCAAATGGCATTCTGACATCC	-
	α -factor	TACTATTGCCAGCATTGCTGC	-
Confirmation of in-frame cloning			
For invertase alleles	Ex1_rev PAIN	GTTGAAAATGGTAAGCAGTTC	-
For inhibitor alleles	LS98_Py2_r	CITIACCACCAGTTACCACC	-
Semi-quantitative expression analysis			
Invertase 6 (<i>Inv6</i>)	L179	GCAAATATCAGCAGCTTCACTATCA	59
	L183	GTTCTCATCCCACCACCCG	
General primers			
Elongation factor	LS99_EF1a_f	GGAGGCACTCCCCGGTGACA	56
	L100_EF1a_r	GGCAGCCTTGGTCACTTTGGCA	

2.2 METHODS

2.2.1 COLD-STORAGE EXPERIMENT

All 40 genotypes of the BIOSOL population were stored for several weeks at 4°C in the dark. Time points for sampling were set to 0 weeks of cold-storage (wocs), 1 wocs, 2 wocs, 4 wocs and 12 wocs. At these 5 time points (Figure 7), 3 biological replicates of each cultivar were collected. The replicates were selected according to a similar tuber size and shape. The tubers were peeled, cut into small squares and frozen in liquid nitrogen for protein and RNA extractions or directly freeze dried for DNA extractions. The detailed extraction protocols are described in the following paragraphs.

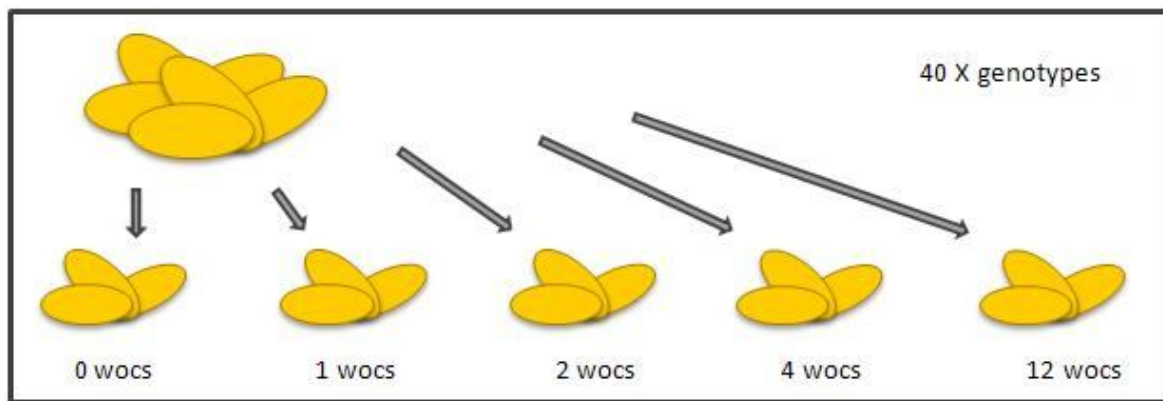


Figure 7: Experimental set up of the cold-storage experiment of the BIOSOL population.

Samples were taken at five different time points (after 0, 1, 2, 4 and 12 weeks) of cold-storage.

2.2.2 PREPARATION OF GENOMIC DNA FROM POTATO TUBERS

Total genomic DNA was extracted from tubers of all genotypes of the BIOSOL population. The tubers were peeled, cut into 1cm x 1cm squares and freeze dried subsequently. The squares were grinded into smooth powder using three grinding marbles (\varnothing 0.8 - 2 cm). One gram of tissue powder was mixed with 20 ml of modified CTAB-extraction-buffer (Wulff et al., 2002). 1% of β -mercaptoethanol and 100 mg PVP were added and incubated at 74°C for 30 min with slight and occasional agitation. The suspension was chilled to RT, mixed with 1 volume Chloroform/Isoamylalcohol (24:1 (v/v)) and vortexed. After a centrifugation step at 5000 x g at 4°C for 30 min, the supernatant was treated with RNase A (10 μ g/ml) at 37°C for 60 min. An equal volume of distilled Milli-Q ultrapure water was added and the pH was adjusted to pH 7.0 with HCl. Further purification was performed with Qiagen-tip 100 columns (Qiagen, Hilden) following the corresponding manual. The obtained DNA pellets were carefully dried and dissolved in 50 μ L HPLC-water (LiChrosolv®, Merck) at RT with constant shaking overnight.

CTAB-extraction-buffer: 100 mM Tris pH 8
 20 mM EDTA pH 8
 2 % CTAB
 2.5 M NaCl

2.2.3 DETERMINATION OF DNA CONCENTRATION

Genomic DNA quality and quantity was checked on a 0.8 % agarose gel (in 1 X TAE) by comparing band intensities with a molecular mass lambda ladder (MBI-Fermentas). The concentration of the DNA was determined by using the micro-volume spectrophotometer NanoDrop™ 1000 (PepLab Biotechnologie GmbH) measuring the absorption at a wavelength of 260 nm. The purity of the DNA sample was estimated by the ratios of 260/280 nm and 260/230 nm.

2.2.4 PREPARATION OF RNA FROM POTATO TUBERS

Isolation of total RNA from potato tubers was performed using the large scale RNA extraction kit of the PureLink™ Plant Reagent Kit (Invitrogen). Freshly peeled tubers were cut into small squares (1cm x 1cm), immediately deep frozen in liquid nitrogen and stored at -80°C. The pieces were grinded to a homogenous and fine powder using the Retsch® Schwingmühle MM400 (Retsch GmbH, Haan) with careful avoidance of thawing. Referring to the manual of the Plant Reagent Kit, 0.5 g of tuber powder was applied per extraction procedure. The obtained RNA pellet was resuspended in DEPC-treated H₂O and checked on a 1 % agarose gel to estimate the quality and purity of the RNA. The determination of the concentration was done by using the NanoDrop™ as previously described for genomic DNA with modified settings for RNA according to the supplier's manual. Afterwards, the RNA samples were diluted to stocks containing 500 ng/μL, treated with the DNA-free™ Kit (Ambion, Cambridgeshire, UK) to remove contaminations of genomic DNA and stored at -80°C. The same procedure was applied to the extraction of RNA from aerial potato tissues (leaves, stem, flowers).

2.2.5 SYNTHESIS OF FIRST STRAND cDNA

Following the instructions of the RevertAid™ H Minus First Strand cDNA Synthesis Kit (MBI Fermentas), the first strand of cDNA was synthesized by using oligo (dT)₁₈ primers provided with the kit. Around 3 μg of RNA were used in a 20 μl reaction volume. The synthesized cDNA was diluted 1:30 and used as template for the transcriptional expression analysis.

2.2.6 PLASMID PREPARATION

For small scale isolation of plasmid DNA from *E. coli* cells, the Qiaprep® Spin Miniprep Kit was used. Instructions were followed according to the Qiaprep® Miniprep handbook (Qiagen). Alternatively, the plasmids were isolated automatically using the Qiaprep® Spin Miniprep Kit in combination with the QIAcube® (Qiagen). Large scale plasmid preparations for yeast transformations and isolations of BAC plasmids were performed using QIAGEN-tip 100 according to Qiagen® Plasmid Purification handbook.

2.2.7 POLYMERASE CHAIN REACTION (PCR)

2.2.7.1 STANDARD PCR PROTOCOL

PCR was carried out in a standard PCR reaction volume of 25 µL, containing 1x Ammonium PCR buffer (Tris-HCl pH8.5, (NH₄)₂SO₄, 15 mM MgCl₂, 1 % Tween® 20), 2.5 M MgCl₂, 0.2 mM of each dNTP, 5 µM of each forward and reverse primer and 1 U Taq-Polymerase. As a template, genomic DNA or cDNA was used at concentrations of 50ng/µL. Standard cycling conditions for genomic DNA as a template were set up as follows: Initial denaturation for 4 min at 94°C; amplification by 34 cycles consisting of a denaturation step of 45 sec at 94°C, an annealing step of 30 sec at 55°C and an elongation phase of 75 sec at 72°C; final elongation was carried out at 72°C for 6 min, the PCR was stored at 16°C. If required, the annealing temperature was changed according to the given melting temperature ($T_a = T_m - 5^\circ\text{C}$) of the used primer combination and the elongation time was adjusted to the expected fragment size (1 kb is equivalent to 1 min). For amplification with cDNA as a template, the amount of cycles was increased to 40. PCR was performed in general either in a T3 Thermocycler (Biometra, Göttingen) or in a labcycler (Sensoquest, Göttingen).

2.2.7.2 POLYMERASES AND MODIFIED PCR PROTOCOLS

Most standard PCRs and colony PCR were carried out with self made *Thermus aquaticus* DNA polymerase (Taq-Polymerase; Pluthero, 1993). All PCRs for association analysis purposes were performed with Amplicon Taq III Polymerase (Amplicon, Scovlunde DK). For the amplification of PCR products which needed high accuracy and quality e.g. for later cloning, the proofreading (3'-5' exonuclease activity) KOD Hot Start Polymerase (Novagen, Darmstadt) was used according to the supplier's manual (section 2.2.10). Special treatment for the template was necessary in case of a colony-PCR. The picked *E. coli* colonies were resuspended in double distilled Milli-Q water and heated to 95°C for 10 min to render the DNA more accessible for the PCR. Yeast colonies were resuspended in 20mM NaOH (Wang et al., 1996) and boiled as described for *E. coli* cells. One microliter of this mixture was used as a template in the PCR. If an optimization of the cycling

conditions for several primer pairs was required, either a gradient PCR was performed, including a $\pm \Delta 5^\circ\text{C}$ range of different annealing temperatures or the PCR program contained *touchdown* steps to gradually decrease the annealing temperature for increased specificity (Don et al., 1991).

2.2.8 GEL-ELECTROPHORESIS

PCR fragments and products of restriction reactions were separated according to their size by gel electrophoresis. Depending on the size of the analyzed fragments, gels contained between 1-2 % agarose. The procedure was performed in 1X TAE. The loaded samples were mixed with 6X loading dye (section 2.1.6.4). The running conditions were set in general to 95 Volts for approximately 30 min. For visualization purposes, the gels were stained with 0.5 μg ethidium bromide (Serva Electrophoresis GmbH, Heidelberg) per mL of gel volume and analyzed on a transilluminator with UV-light of 302 nm (Gel Doc 2000, Bio-Rad Laboratories, München). The detected bands were compared to the size markers GeneRuler™ 100bp Plus DNA Ladder and 1kb DNA Ladder Plus (MBI Fermentas GmbH).

2.2.9 SEQUENCING

DNA sequences were determined by the MPIZ DNA core facility on Applied Biosystems (Weiterstadt, Germany) Abi Prism 377, 3100 and 3730 sequencers using BigDye-terminator v3.1 chemistry. Premixed reagents were supplied by Applied Biosystems. Oligonucleotides were purchased from Sigma-Aldrich Chemie GmbH (Taufkirchen) and Metabion International AG (Martinsried). Before sequencing, all samples were treated with ExoSAP-IT® (USB Affymetrix, Santa Clara California, USA) following the companies instructions and diluted according to the requirements of the MPIZ DNA core facility.

2.2.10 CLONING, HETEROLOGOUS EXPRESSION AND PURIFICATION

2.2.10.1 CLONING OF INVERTASE AND PUTATIVE INHIBITOR INTO A SECRETORY VECTOR

Cloning procedures were performed for a heterologous expression experiment. In order to clone selected invertase inhibitor and invertase alleles in frame with an α -factor signal (N-terminal) and the two tags c-myc and 6xHis (C-terminal) provided with the vector pPICZ α , suitable primers with restriction sites were designed (L125-L128; Table 5). The 5'-primer contains an EcoRI restriction site, whereas the 3'-primer possesses a NotI restriction site, for invertase and inhibitor alleles respectively. In the PCR reaction, the proofreading polymerase KOD Hot Start DNA Polymerase (Novagen,

Darmstadt) was used to obtain PCR products of high accuracy and quality. The amplified PCR products were cleaned up (Nucleospin® Extract II, Macherey Nagel, Düren) and restricted according to the manufacturer's instructions for 4 hours at 37°C (Fermentas). Linearization of the vector was performed in a restriction assay as described above. All insert restriction samples and the vector restriction reaction were separated on a 1% ethidium bromide-Agarose-Gel and a distinct band of each product of expected size was cut out. The clean-up procedure from the agarose gel was performed according to the supplier's manual (Silica Bead DNA Gel Extraction Kit, Fermentas). The ligation was carried out at 15°C overnight in a 10 µl reaction mix, containing the following reagents: 20-100ng vector, the insert in a ratio of 1:2 compared to the vector, 10x T4 Ligase Buffer, 1U T4 Ligase and nuclease-free water (T4 DNA Ligase, Promega). For the transformation reaction, 1µl of each ligation mixture was brought via electroporation into ElectroMAX™ DH10B™ or TOP10F' Electrocomp™ cells (Invitrogen) following the manual. Modifications of the protocol were carried out in the recovery phase after electroporation, LB-LS (low-salt) medium was used instead of SOC medium and cells were plated afterwards on LB-LS plates containing 25 µg/ml Zeocin™ (Maniatis et al., 1989, pages 5.10 to 5.13). Positive clones were analyzed by colony-PCR (primer 5'AOX and 3'AOX, Table 5) and plasmid DNA was extracted in a large scale format (section 2.2.6).

2.2.10.2 STORAGE OF CLONES – GLYCEROL STOCKS FOR BACTERIAL AND YEAST CELLS

To store positive clones for few weeks, the cells were grown on fresh plates containing the appropriate media and selective antibiotic. The plates were renewed every 3-4 weeks by selecting one single colony and streaking it onto a new plate. For long-term storage, a single cell was picked and grown in an overnight culture, containing the correct medium without a selection reagent. In the stationary phase, the cells were harvested and mixed with 15 % sterile glycerol (Maniatis et al., 1989). Cells were directly frozen in liquid nitrogen and stored at -80°C.

2.2.10.3 HETEROLOGOUS EXPRESSION OF INVERTASE AND PUTATIVE INHIBITOR IN *PICHIA PASTORIS*

Plasmid DNA containing a correct insert of either invertase or the putative invertase inhibitor in frame with the α -signal and the two tags was transformed via electroporation into the two *Pichia pastoris* strains GS115 and KM71H according to the manual provided by the EasySelect™ *Pichia pastoris* Expression Kit (Invitrogen). The obtained clones were checked by yeast colony PCR. For each construct, one positive clone was chosen for a heterologous expression in *P. pastoris*. Expression was optimized for buffered complex medium (BMGY/BMMY).

BMGY/BMMY: 1 % yeast extract
2 % peptone
100 mM potassium phosphate pH 6.0
1.34 % YNB
4x10⁻⁵ % biotin
1% glycerol or 0.5 % methanol

2.2.10.4 HIS-TAG PURIFICATION OF HETEROLOGOUS EXPRESSED PROTEINS

For the purification of a heterologous expressed protein, a 12 ml BMGY pre-culture was inoculated with a single colony of *P. pastoris* GS115 strain or KM71H strain containing the desired insert. The culture was incubated overnight at 30°C with shaking at 180 rpm. For the main culture, the pre-culture was centrifuged and the cell pellet was inoculated in 100 ml of BMMY medium to an approximate O.D of 3 and grown for 3 days with vigorous shaking. For the induction of protein expression, 0.5 % methanol was added every 24 hours. Afterwards, the medium containing the expressed and secreted protein was harvested by a centrifugation step of 5 min at 4°C at 4000 x g and the cell pellet was discarded. Ammonium sulfate was added gradually to a final concentration of 50 % at 4°C and stirred for at least 1 h at 4°C. The mixture was split into 4 x 50 ml centrifugation tubes (2.5 x 10 cm; Sorvall Instruments, Thermo Scientific) centrifuged at 10000 x g for 15 min at 4°C, the supernatants were discarded and the pellets were resuspended in an overall final volume of 5 ml binding buffer and pooled. The sample was passed through a 45 µm filter to get rid of large protein accumulations which may block later steps. For the purification, the liquid chromatography system ÄKTAprime plus (GE Healthcare, Freiburg) was used. The sample was applied to the system by a syringe and purification was carried out either with 5ml HisTRAP™ HP (GE Healthcare, Freiburg) or 5ml Protino® Ni-NTA columns (Macherey-Nagel, Düren). The systems settings are based on the program for Affinity Purification any HiTrap (described in ÄKTAprime plus manual), the modifications are listed below.

Binding buffer: 10 mM NaCl
20 mM Imidazole
50 mM phosphate puffer (KH₂PO₄)
pH 7.9 with KOH

Elution buffer: 10 mM NaCl
500 mM Imidazole
50 mM phosphate puffer (KH₂PO₄)
pH 7.9 with H₃PO₄

ÄKTAprime plus settings:	pressure limit: 0.5 MPa
	flow: 2.5 ml/min
	gradient: 0.1 ml
	fraction size: 1.0 ml
ÄKTAprime plus protocol:	equilibration/priming wash vol.: 25 ml
	auto zero
	sample application: 5 ml
	wash 1 vol.: 180 ml
	elution vol.: 25 ml and collection of fractions
	re-equilibration vol.: 25 ml

In average, six to nine fractions were pooled and concentrated in Amicon® Ultra-15 centrifugal filter devices (Millipore) to a volume of approximately 1.5 ml. In order to remove ammonium sulfate, PD-10 desalting Columns (GE Healthcare, Freiburg) were used according to the gravity manual. The elution volume of 3.5 ml was concentrated in Amicon® Ultra-4 centrifugal filter devices (Millipore) to 150 - 200 µl. An equal volume of 100 % glycerol was added to the sample. The protein concentration was determined by using the Quant-iT™ Protein Assay Kit (Invitrogen) and storage was carried out at -20°C.

2.2.10.5 PROTEIN GELS AND COOMASSIE/SILVER STAINING

Purified protein was checked on a 4-12 % NuPAGE® Novex® Bis-Tris Mini Gels running in NuPAGE MES SDS running buffer (Invitrogen). The samples were mixed with 4x LDS and 10 % DTT and heated to 95°C for 5 min before loading. The running conditions were set to 175-200V for approximately 45 min. Afterwards, the gels were stained either with PageBlue™ Protein Staining Solution (Coomassie-Staining) or PageSilver™ Silver Staining (both Fermentas) according to the manuals. The Spectra™ Multicolor Broad Range Protein Ladder (10-260 kDa; Fermentas) was loaded as size marker. For gel imaging, the Chemidox XRS System was used in combination with the Image Lab software (BIO-RAD).

2.2.10.6 WESTERN BLOTS AND IMMUNODETECTION

For Western Blot detection, samples were treated and gels were run as previously described (section 2.2.10.5). MagicMark™ XP Western Protein Standard (Invitrogen) was loaded as an additional size ladder. Protein samples were blotted on Amersham Hybond™-P PVDF Membrane (GE Healthcare)

using the XCell II™ Blot Module (Invitrogen). The procedure was carried out as described in the related manual. After blotting, the membrane was briefly checked by staining with Ponceau S solution (Sigma-Aldrich). Subsequent de-staining and washing was performed with 1 X TBS (section 2.1.6.4). The membrane was blocked for at least 1 h with 1.5 % milk powder in 1X TBS while shaking (blocking solution). The first antibody was applied to the membrane in fresh blocking solution and incubated at 4°C overnight.

Table 6: List of used antibodies for immunodetection of invertase and putative invertase inhibitor.

1. Antibody	concentration	reference
Inhibitor	1 : 1000	Draffehn, A., 2010
Invertase	1 : 1000	Glaczinski et al., 2002
c-myc	1 : 500	ECL™ Anti-Rabbit IgG, Horseradish peroxidase-linked whole antibody (from donkey; GE Healthcare)
2. Antibody		
Anti-rabbit	1 : 10000	Anti-c-Myc antibody (from rabbit; Sigma-Aldrich)

The membrane was washed three times for 10 min with 1 X TBS. The second antibody (Table 6) was added in fresh blocking solution and incubated for 2 h at RT. For the detection, the membrane was wrapped into foil and ECL Plus Western Blotting Detection Reagent (GE Healthcare) was applied according to the manufacturer's instructions. Imaging was performed in the Chemidox XRS System in combination with the Image Lab software (BIO-RAD). If necessary, blots were stored in 1 X TBS and stripped with Restore™ Western Blot Stripping Buffer (Thermo Scientific) following the supplied instructions.

2.2.10.7 AMMONIUM SULFATE PRECIPITATION

For a fast extraction of total protein, heterologously expressed protein was precipitated directly from the medium (BMMY) by adding 50 % ammonium sulfate. After centrifugation for 30 min at 5000g and 4°C, the obtained pellet was dissolved in 1X PBS.

2.2.11 PURIFICATION OF NATIVE INVERTASE FROM POTATO TUBERS

Native invertase was purified from potato tubers according to the procedure described in Glaczinski et al. (2002) with few modifications: The extraction was exclusively performed at 4°C, if not stated otherwise. A genotype BIOSOL population with high sugar accumulation (M. Fischer, unpublished

data) after cold-storage was chosen and 2 kg of potato tubers from this cultivar were peeled. Juice was extracted from potato slices with the help of the juice extractor (Braun Multipress). 0.8M Na₂SO₃ and 1 % (w/v) PVP were added per 100 ml of freshly extracted potato juice. The juice was filtered with Miracloth (Merck) and centrifuged at 13000 rpm for 30 min. The supernatant was adjusted to a concentration of 0.2 M Na acetate, pH 5.0. After centrifugation at 13000 rpm for 20 min, 30 % of ammonium sulfate was gradually added to the supernatant in the first protein precipitation step. This step was repeated by adding 50 % of ammonium sulfate. This obtained protein pellet was dissolved in 0.2 M NaCl. In total, 10 mg of the protein extract was applied in a gel filtration procedure. The column Superdex™ 200 10/300 GL (GE Healthcare) was equilibrated with 0.2 M NaCl and operated with the ÄKTAprime plus gel filtration standard protocol. All fractions were analyzed on a SDS PAGE gel and stained with Coomassie (sections 2.2.10.5 and 2.2.10.6). Fractions containing proteins of correct size were additionally checked on a western blot with a specific invertase (*Pain-1*) antibody and by activity assays (section 2.2.12.1). The four positive fractions were pooled, adjusted to 10 mM Na acetate (pH 5.0) and concentrated in an Amicon® Ultra-4 centrifugal filter device (Millipore) to a total volume of 150 - 200 µl. For storage at -20°C, 50% glycerol was added to the sample.

2.2.12 ACTIVITY ASSAYS

2.2.12.1 INVERTASE ACTIVITY ASSAY

The invertase activity test procedure was in accordance to Zrenner et al. (1996) with few modifications. Either unpurified invertase (in medium) or highly purified invertase (via His-Tag) was used in the assay. The invertase (10-30 µl) was incubated with 100 mM sucrose and 20 mM sodium acetate (pH 4.7) for 15 min at 30°C. The reaction was neutralized by adding 10 µl 1 M sodium phosphate (pH 7.2) and stopped afterwards by heating the mixture to 95°C for 10 min. This template was used in the Glucose (GO) Assay Kit (Sigma-Aldrich) for the determination of the amount of produced glucose according to the manual. The assay was scaled down from 5 ml to 1.5 ml end volume and the absorbance was measured at 540 nm in 1 mL cuvettes in a photometer. A second assay for more accurate absorptions was performed with the Glucose HK Assay Kit (Sigma-Aldrich) following the manufacturer's instructions. The used solutions were scaled down to a 96-well format (140 µl of reagent) and the samples were analyzed by measuring the absorbance at 340 nm in a plate reader.

2.2.12.2 INVERTASE INHIBITION ASSAY

In order to test for inhibition of invertase, the His-tag purified putative inhibitor alleles were added to the purified invertase. The putative inhibitor was added in the equimolar concentration and in a 5 fold higher concentration compared to invertase. Both enzymes were pre-incubated for 30 min at 37°C (Link et al., 2004). This mixture of invertase and putative inhibitor was used in the previously described invertase activity. The amount of produced glucose was calculated upon the measured absorbance at 340 nm.

2.2.12.3 TRYPsin ACTIVITY ASSAY

For the putative invertase inhibitor a weak but measurable trypsin inhibition was shown (Heibges et al., 2003b). Therefore, the heterologously expressed putative invertase inhibitor was tested for activity by using a trypsin inhibition assay. The trypsin inhibition assay was performed according to the protocol of Heibges et al. (2003b) with few modifications. In general, the assay was scaled down to a concentration of trypsin of 2.5 µg in a final reaction volume of 600 µL. Inhibitor protein was used in the following molar ratios compared to trypsin: (Inh) 1 : 1.75 (Tryp) and (Inh) 1 : 3.5 (Tryp). Due to similar molecular masses of trypsin and the putative inhibitor, the protein amount of the putative inhibitor corresponded to approximately 1.5 µg and 0.75 µg. The absorption was measured at 440 nm against water in a plate reader.

2.2.13 BAC-LIBRARY SCREEN

In order to characterize the genomic structure of several candidate genes, the two BAC (bacterial artificial chromosome) libraries "BA" and "BC" were screened. Both libraries were constructed from high-molecular weight genomic DNA fragments of the diploid *S. tuberosum* genotype P6/210 (Ballvora et al., 2002; Ballvora et al., 2007). The inserted genomic fragments in both libraries consist in average of 70-80kb and therefore lead to the calculated value of 6-8 fold coverage of the potato genome per library. Each library contains approximately 100.000 clones, represented on 4 filters per library, one filter consisting of 6 plates with 384 spots, whereas one spot corresponds to 12 clones.

2.2.13.1 CONSTRUCTION OF RADIOACTIVE-LABELLED PROBE VIA NICK-TRANSLATION

Via normal PCR amplification procedure (section 2.2.7.2), a probe was constructed, including a part of the candidate gene. After ExoSAP-IT® treatment, the PCR product was diluted to a final concentration of 200 ng and used in a nick-translation reaction (Kelly et al., 1970). In addition, the nick-translation reaction contained 1X Buffer A1 (0.2 mM dCTP, 0.2 mM dGTP, 0.2 mM dTTP,

Invitrogen), 1.25 U DNaseI/ 1U DNA-Polymerase (Invitrogen) and 3 μl α ^{32}P dATP in a final volume of 25 μl . The mixture was incubated at 16°C for 1 hour and purified afterwards using a Sephadex-G50 column (Amersham Biosciences) and 1X TE buffer. The collected radioactive probe was heated to 95°C for 5 minutes.

2.2.13.2 PRE-HYBRIDIZATION AND HYBRIDIZATION OF BAC FILTER

The BAC filters were pre-hybridized in hybridization buffer at 65°C for 1 hour in a hybridization glass-tube (4cm x 30cm). 500 μl of salmon sperm (10 mg/ml) were added. New BAC- filters were pre-treated for 30 min with 2 X SSPE at 65°C with continuous rotation, afterwards the pre-hybridization step was prolonged to four or more hours. For the continuous rotation, a hybridization oven was used (Bachofer, Reutlingen). After adding the radioactive probe, the hybridization was carried out overnight.

Hybridization buffer: 20 X SSPE
100 X Denhardtts
10 % SDS

20X SSPE: 3 M NaCl
0.2 M NaH_2PO_4
20mM EDTA pH 7.0

Several washing steps of the filters were performed on the next day, according to the following procedure:

- Washing 1: 2 X SSPE
0.1 % SDS
10 -15 min at 65°C

- Washing 2: 1 X SSPE
0.1 % SDS
10 - 15 min at 6°C

- Washing 3: 0.2 X SSPE
0.1 % SDS
20 min at 65°C

Afterwards, the filters were wrapped separately in Saran foil and exposed to a 35 x 43 cm film (KODAK BioMax XAR Film, Rochester) in an X-ray cassette (REGO X-Ray GmbH, Augsburg) for 1-3 days at -80°C. Positive clones were represented as spots consisting of two points. With the help of spotting-pattern moulds, one single positive BAC was identified. The corresponding clone was plated on a LB-plate containing tetracycline, analyzed by colony-PCR and Plasmid-DNA was prepared afterwards. Used BAC filters were washed with a washing solution (100mM NaOH, 10 mM EDTA, 0.1 % SDS) and stored at 4°C.

2.2.13.3 PFGE- PULSED FIELD GEL ELECTROPHORESIS

The BAC plasmids were purified as described above (section 2.2.6) and digested with NotI in a standard restriction assay at 37°C for 4h. Due to DNA fragments expected to be larger than 20 kb, pulsed field gel electrophoresis was performed instead of standard separation on agarose gels. Because of alternating directions of the electric field, large DNA fragments have to re-orientate and can pass the pores of the agarose gel more easily (Schwartz and Cantor, 1984) and are separated according to their correct size. The gel contained 1 % of Gold Agarose (SeaKem®, Lonza, Basel) and was run in 0.5 X TAE cooled to 14°C. The restriction reaction and the two size markers Lambda Ladder PFG marker and MidRange Marker II (both New England Biolabs GmbH) were loaded into the pockets and sealed with agarose afterwards. Running conditions were set as follows:

- initial time: 5 sec
- final time: 15 sec
- running time: 14 h
- voltage: 5 V/cm
- incl. angle: 120°
- pump: 80

The gel was stained afterwards for 20 min in a 2 % EtBr solution and visualized on a transilluminator (Typhoon, GE Healthcare).

2.2.13.4 BAC SEQUENCING

BAC sequencing of the GWD BAC was performed by the company GATC Biotech (Konstanz) with the help of high-throughput sequencing of a shotgun library of the BAC. Coverage of at least 8-10 folds was expected. The obtained shotgun-sequences were assembled to one single contig.

2.2.14 EXPRESSION ANALYSIS - PYROSEQUENCING

Pyrosequencing bases on a sequencing-by-synthesis method which relies on the detection of light produced by the release of pyrophosphate (PP_i) upon nucleotide incorporation (Ronaghi et al., 1996). The quantity of emitted light is proportional to the amount of used nucleotides in the sequencing reaction. SNP ratios can be easily quantified.

A 350 bp PCR template for the pyrosequencing approach was amplified with a biotinylated primer pair (Table 5) and contained all SNPs of interest. The different sequencing primers for the Kunitz type inhibitors are listed in table 5. The product was amplified from cDNA in a standard PCR reaction with the following modifications: The reverse primer was labelled with biotin and the amount of cycles was set to 50 to remove as much as possible of the components to avoid later interference in the pyrosequencing reaction. As a control, the cDNA templates were checked for genomic contamination despite of a preceding DNaseI treatment of the RNA. A special sequencing forward primer was designed for the SNP of interest which ended 3 bp before the SNP.

After template amplification, the PCR product was immobilized to biotin binding Streptavidin SepharoseTM beads (GE Healthcare). The mixture contained 20 µl of the PCR, 20 µl H₂O (LiChrosolv®, Merck), 40 µl Binding Buffer (10mM Tris-HCl, 2 M NaCl, 1mM EDTA, 0.1 % Tween 20, pH 7.6) and 5 µl Streptavidin Sepharose beads. All components were mixed on ice, pipetted into a deep well plate (Greiner Labortechnik, Kremsmünster, Österreich) and shaken for 5 min at RT. Using the vacuum pump, the beads containing the bound PCR products were trapped to the PSQ96 Sample Prep tool (Biotage AB, Uppsala, Sweden) and cleaned up in the following washing procedure:

- 3-5 sec in 70 % EtOH
- 3-5 sec in 0.2 M NaOH
- 5-10 sec in washing buffer (10 mM Tris acetate, pH 7.6)

Afterwards, the beads were released into a PSQ 96 Plate Low (Biotage AB) containing 39 µl annealing buffer (10 mM Tris acetate, 2 mM Mg acetate, pH 7.6) and 1 µl of the forward sequencing primer in each well. This plate was heated to 80°C for 2 min, slowly chilled to RT and placed into the Pyrosequencer PSQ96TM MA. The PSQ 96 reagent cartridge was loaded with nucleotides, enzyme and substrate mixtures according to the supplier's instructions (Pyro Gold Reagent, Biotage AB). The corresponding software (PSQ96 MA SNP) was started and the standard SNP analysis program was accomplished. The first 10-15 bp of the expected sequence were entered into the corresponding field, the SNP positions in this sequence were marked by X/Y. The pyrosequencing reaction was started afterwards. After the run, the software was changed from the SNP mode to the "analysis

quality" (AQ) mode and the SNP output was analyzed. The results were displayed in diagrams, calculated from the curve height as SNP frequencies.

For each of the 10 selected cultivars of the BIOSOL genotypes, the expression was checked in three biological replicates with three technical replicates for all five time points. In total, 150 cDNAs were synthesized and analyzed in three technical replicates including a genomic DNA, RNA and water as one positive and two negative controls (3x3x5x10).

2.2.15 ASSOCIATION ANALYSIS

For association studies, amplicons of several candidate genes were generated and sequenced. Amplicon lengths varied between 700 bp and 1100 bp. They primarily span long exon sequences and include in few cases some intron sequences. The amount of analyzed SNPs was dependent on amplicon length and evaluable sequence length. Due to the tetraploidy of *S. tuberosum*, five allelic states of one locus are possible to be identified. They can be classified into the two homozygous allelic states "XXXX" and "YYYY" or the three heterozygous types: XXXY (triplex/simplex), XXYY (duplex/duplex) and XYYY (simplex/triplex). For statistical analysis, the allelic states were transformed into digits as in the following example: AAAA-0, AAAG-1, AAGG-2, AGGG-3 and GGGG-4. Scoring of allelic states was done with the help of sequence alignments in the Lasergene software Seqman and by using NovoSNP. Additionally, the manually scored SNPs were re-checked with the help of the DaX Software. Statistical calculations were performed using SPSS software. Association analysis was carried out by using a univariate general linear model (GLM) for the GABICHIPS population. This model was corrected for the factor "origin" (Li et al., 2008). Means plots were compiled manually upon the descriptive statistics output of the GLM. Significant differences between allelic states were calculated by using the multiple-comparison post-hoc test LSD (Fisher's Least Significant Difference). For a statistical analysis of SNPs in the BIOSOL population, the phenotypic data was transformed into normalized data. Association was analyzed by using the one-way ANOVA as statistical test.

2.2.16 HAPLOTYPE CALCULATION

For the calculation of haplotypes, the software SATlotyper was used (Neigenfind et al., 2008). In general, 9 to 12 SNPs per candidate gene were selected. The selected SNPs were either a representative of a group of SNPs that were in linkage disequilibrium (LD) or single SNPs showing no co-segregation. LD groups were classified either upon an obvious co-segregation pattern or upon a correlation coefficient which was calculated by using a bivariate correlation procedure based on Pearson.

2.2.17 PROGRAMS AND SOFTWARE

Table 7: List of frequently used online tools, databases and software.

	Name (Citation)	Source (company/URL)
Online tool:	Primer3 (Rozen and Skaletsky, 2000)	http://frodo.wi.mit.edu/primer3/
	Net primer	http://www.premierbiosoft.com/netprimer/
	Emboss	http://emboss.bioinformatics.nl/
	Expasy	http://expasy.org/
	Multalign	http://multalin.toulouse.inra.fr/multalin/
	DNA Calculator	http://www.sigma-genosys.com/calc/DNACalc.asp
	Oligo Calc (Kibbe, 2007)	http://www.basic.northwestern.edu/biotools/oligocalc.html
	ChloroP, SignalP, TMHMM	http://www.cbs.dtu.dk/index.shtml
	MultiLoc2 (Höglund et al., 2006)	http://abi.inf.uni-tuebingen.de/Services/MultiLoc2
	Ammonium sulfate Calculator	http://www.encorbio.com/protocols/AM-SO4.htm Encor Biotechnology Inc., Florida
Database:	SGN	http://solgenomics.net/
	PGSC	Potato Genome Sequencing Consortium http://www.potatogenome.net/
	NCBI	National Center for Biotechnology Information http://www.ncbi.nlm.nih.gov/
	PoMaMo	http://www.gabipd.org/projects/Pomamo/
	Pfam	http://pfam.sanger.ac.uk/
	Uniprot	http://www.uniprot.org/
	Software:	SNP analysis
Sequence analysis		FinchTV
Alignments and phylogenetic trees		Lasergene, GeneDOC, ClustalX, Njplot
Haplotype modeling (Neigenfind et al., 2008)		SATlotyper
Statistics		SPSS

3 RESULTS

3.1 GLUCAN WATER DIKINASE - GENOMIC ORGANIZATION

The cDNA sequence (Y09533.1) of the α -glucan water dikinase (*GWD*) locus in *Solanum tuberosum* was published as *Ritte1* (*R1*) by Lorberth et al. in 1998. The gene consists of 4395 bp and forms a polypeptide with a length of 1464 amino acids. The genomic sequence and organization of *GWD* was unknown and has been established by using a BAC library screen as described in the following section of this thesis. In order to generate a probe for the BAC library screen, different primer combinations were either manually designed by taking advantage of the published cDNA sequence or taken from publications (Chen et al., 2001). They were tested on different genomic DNA templates, including the genomic DNA of the two heterozygous diploid cultivars P40 and P41 (Gebhardt et al., 1989, Leonards-Schippers et al., 1992,) and their offspring P6/210 (Leister et al., 1996). The DNA of hybrid P6/210 is the genomic background of the used BAC libraries and was used to increase the probe specificity. Two adequate primer pairs were obtained (#1: LS08-LS06, #2: LS21-LS20, Table 5). The corresponding amplicons were used as probe in several BAC library screens on both available BAC libraries BA and BC (Ballvora et al., 2002; 2007). In the hybridization experiment on the BC library, no signal was obtained for BAC probe #2, whereas probe #1 showed many unspecific hybridization signals. Similar unclear results of probe #1 were obtained for the BA library, but a single positive signal was obtained for probe #2. The spot corresponding to this signal was identified as BAC clone BA202-H17. A colony PCR with the probe primers confirmed this clone. BAC plasmid DNA was extracted and used in further PCR reactions to analyze whether *GWD* is included as full-length sequence. The primer pairs were designed on parts of the cDNA sequence which covered the 5' region including the start codon or the 3' region containing the stop codon (LS27-LS24; LS25-LS26; Table 5). For each combination a clear PCR signal on the BAC DNA was obtained. In order to analyze the orientation of the gene in the BAC plasmid, a vector primer and a very distal 5' (or 3') sequence primer were used. An approximately 5 kb fragment was obtained for the amplification of the 3' end (T3-LS25, Table 5). This shows that the genomic sequence of *GWD* is in 5'-3' orientation and located close to the vector sequence (ca. 5 kb).

The size of the genomic insert included in the BAC was determined via Pulsed Field Gel Electrophoresis (PFGE). The two digestion reactions with the restriction enzymes NotI and BamHI resulted in fragments of about 46 to 48 kb of size, respectively (Figure 8). The smaller fragments represented vector sequence. Thereupon, the BAC plasmid DNA was sent for sequencing to GATC Biotech (Koblenz). Custom sequencing of a shotgun-library prepared from the BAC plasmid with approximately 10 fold coverage revealed that the insert size of the genomic DNA in the BAC is even

smaller than estimated by PFGE. The sequencing results provided by the company displayed as largest contig a sequence of approximately 36 kb instead of 46 kb.

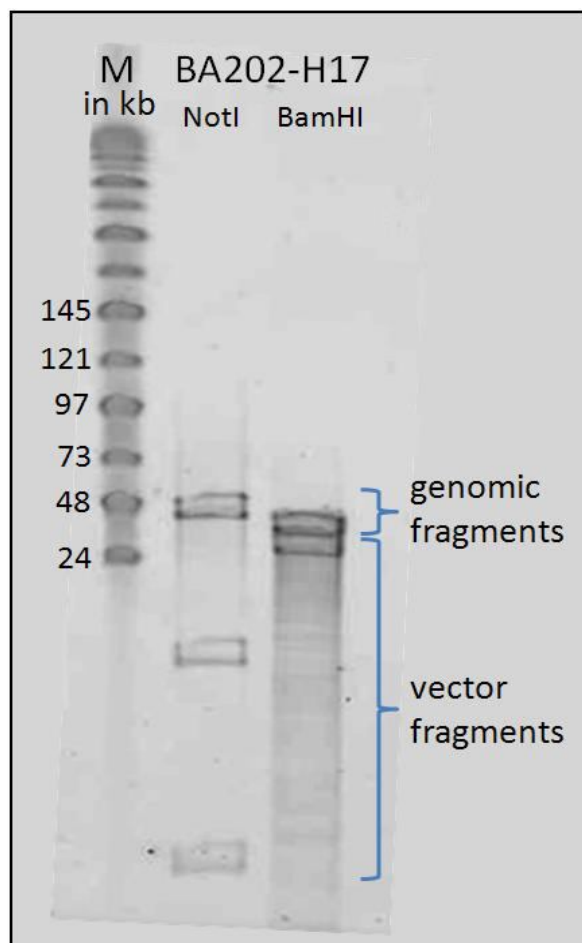


Figure 8: Insert size of BAC clone BA202-H17 estimated by PFGE.

All sequences were aligned using the software Seqman (DNASTAR Lasergene 8). In total, 11 contigs were assembled (Table 8), the complete vector sequence (pCLD04541) has been trimmed. The average length of the shotgun sequence reads ranged between 700 - 800 bp. Most of the genomic sequences were covered by up to 30 sequences of the shotgun library, only few parts of the sequence were represented by 10 to 15 sequences. Blast analysis of all assembled contigs in NCBI (<http://www.ncbi.nlm.nih.gov/>) revealed that all 10 smaller contigs were *E. coli* contaminations from the plasmid preparation. Only the largest contig (#13) contained the gene of interest and was further analyzed. The corresponding total BAC sequence of contig 13 is documented in the appendix as a data file (C.1).

Table 8: Obtained shotgun library data from GATC Biotech.

Name	Length in bp	Number of sequences	Blast result
Contig 13	35962	668	<i>GWD</i>
Contig 17	421	2	<i>E. coli</i> contamination
Contig 18	1417	2	<i>E. coli</i> contamination
Contig 21	1783	2	<i>E. coli</i> contamination
Contig 23	1820	2	<i>E. coli</i> contamination
Contig 25	1707	2	<i>E. coli</i> contamination
Contig 28	1519	2	<i>E. coli</i> contamination
Contig 33	202	2	<i>E. coli</i> contamination
Contig 36	1545	2	<i>E. coli</i> contamination
Contig 38	1147	2	<i>E. coli</i> contamination
Contig 39	1812	2	<i>E. coli</i> contamination

The structure of the whole contig is displayed in figure 9. The 36 kb insert harbours the complete full-length *GWD* sequence of about 16.7 kb. As indicated by previous PCR amplifications, the gene is located close to the 3' end of the BAC insertion. The 19 kb part upstream of the 5'UTR of *GWD* was analyzed for a putative promoter region by using Promoter 2.0 (<http://www.cbs.dtu.dk/>; Knudsen, 1999). A putative promoter region was predicted to be located approximately 1-2 kb upstream of the gene. Additional blast analyses were done on the remaining sequence of ~15 kb upstream of the putative promoter region, but did not result in the identification of further genes. Recent comparisons to the released potato genome sequence (PGSC, 2011) confirmed that in close proximity, there is no additional annotated gene.

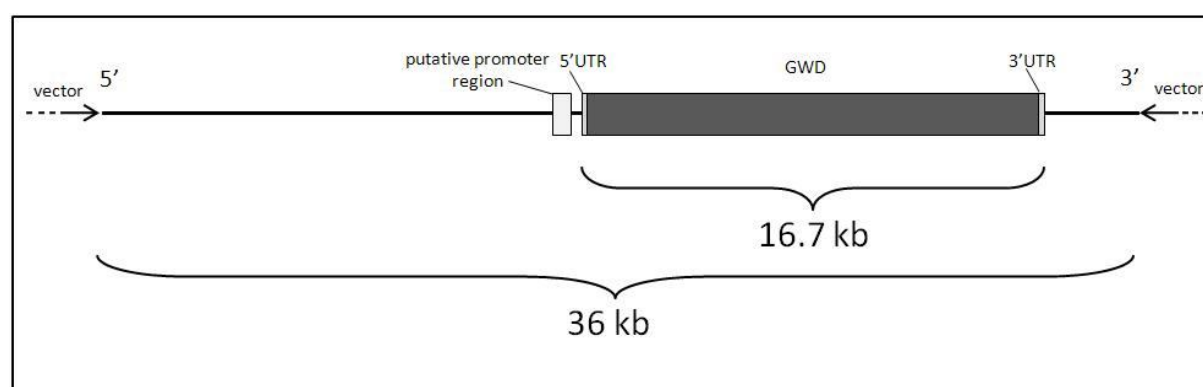


Figure 9: Length and structure of BAC clone BA202-H17 (contig 13).

The 36 kb genomic insert of the BAC contains the full-length sequence of *GWD* of 16.7 kb. A putative promoter region was predicted to be located 1-2 kb upstream of the 5'UTR.

With the help of the software Spidey (<http://www.ncbi.nlm.nih.gov/spidey/>), the published cDNA sequence Y09553.1 was aligned to the obtained genomic sequence and the exon-intron structure (genomic organization) was predicted (Figure 10). In total, *GWD* consists of 33 exons and 32 introns which range from 46 to 343 bp with an average size of 133 bp. A 33rd intron is located in the 5'UTR and may have an effect on gene expression (Chung et al. 2006) as it is located close to the start codon. About two thirds of the genomic sequence (15.7 kb) is represented by introns and one-third is represented by exons. This ratio of around 2:1 also prevails for the lengths of the introns and exons but the variation in length is generally much higher in introns compared to the exons. All detailed lengths of exons and introns are listed in the table B.1 in the appendix.

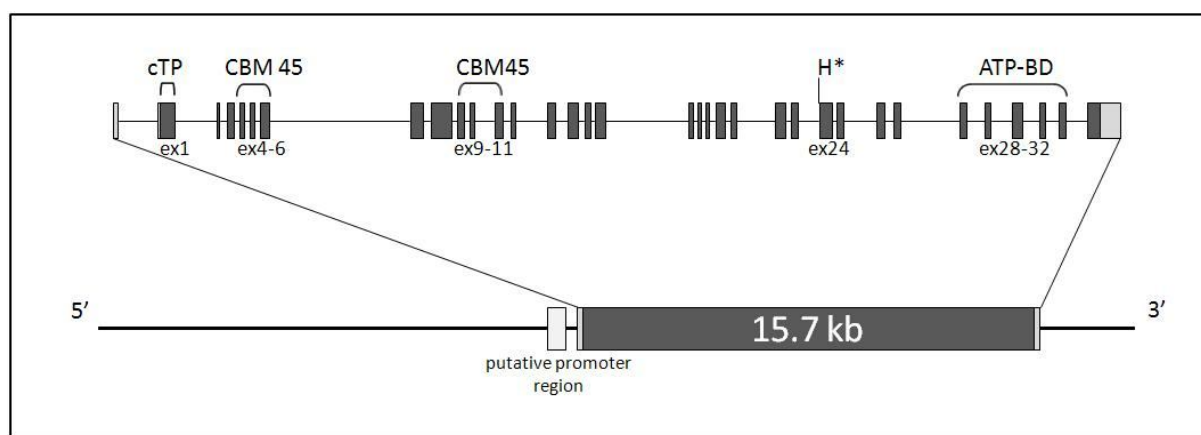


Figure 10: Genomic structure of *GWD* in *S. tuberosum*.

Predicted domains and catalytic sites (Mikkelsen et al., 2005) on amino acid level were assigned to the corresponding exon structure. The exon numbers are highlighted which correspond to a functional domain. cTP: chloroplast target peptide; CBM45: carbohydrate binding module family 45; H*: catalytic histidine; ATP-BD: nucleotide (ATP) binding domain.

3.1.1 FUNCTIONAL DOMAINS OF *GWD*

All characterized and important domains of *GWD* were assigned to the corresponding location in the coding sequence (Figure 10). First indications for functional domains were obtained from the published characterization of *GWD* from *A. thaliana* (Yu et al., 2001). The predicted nucleotide binding domain (ATP-BD; ATP as phosphate donor) in the 3' region was confirmed for potato *GWD* via blast analysis against the Pfam database (<http://pfam.sanger.ac.uk/>). It is similar to a binding domain of a pyruvate phosphate dikinase (PPDK; Mikkelsen et al., 2004). The chloroplast target peptide (cTP) was calculated by using the ChloroP prediction server (<http://www.cbs.dtu.dk/>; Emanuelsson et al., 1999). It is located in the 5' region of the first exon up to base 156 (amino acid 52). The catalytic histidine was previously assigned to position 1069 of the protein sequence by Mikkelsen et al. (2004) which equates on coding sequence to the very distal end of exon 24. For *A. thaliana*, a regulatory disulfide-bridge was predicted in close proximity to the catalytic histidine (Yu et al., 2001) which can be as well identified in potato in the following exon 25. A bipartite starch binding domain was analyzed Glaring et al. (2011) and designated as a member of the CBM45 family

(carbohydrate binding module family 45). This CBM covers a large part including exons 3-6 and exons 9-11. No domain has been declared as dimerization domain so far, although *GWD* is assumed to form homodimers (Mikkelsen et al., 2004). Additional blast analysis of the predicted protein sequence did not reveal any further functional domains.

3.1.2 STRUCTURAL COMPARISON OF *S. TUBEROSUM* AND *A. THALIANA* *GWD*

The comparison of the genomic sequence and organization of *GWD* from *S. tuberosum* (*StGWD*) to the published sequence and organization from *A. thaliana* (*AtGWD*) showed a high conservation between both genes (Figure 11). The genomic organization of both genes can be partitioned into similar blocks. At cDNA level, both genes are rather large and differ only slightly in length, 4.2 kb compared to 4.4 kb (Table 9). Both cDNA sequences show a similarity of 67.8%. Comparisons of the length and number of exons revealed that they are very consistent. *StGWD* possesses 33 exons, whereas *AtGWD* consists of 32 exons.

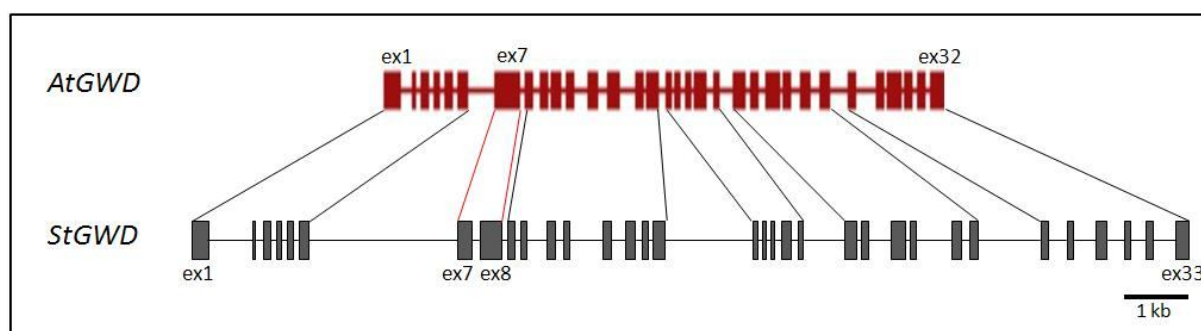


Figure 11: Structural comparisons of *StGWD* and *AtGWD*.

Introns are displayed as lines (red: *A. thaliana*, black: *S. tuberosum*). The exons are represented by boxes. Blocks of similar exon-intron distribution are linked by grey lines. Differences are marked by red lines.

In *S. tuberosum*, exon 7 and exon 8 form two separate exons, but are merged to one larger exon 7 in *A. thaliana*. Regarding the intron lengths, more variation can be observed in general. The entire intron length in *S. tuberosum* is higher, it is almost twice the length (15.6 kb) compared to *A. thaliana* (8.8 kb; Table 9). On the amino acid sequence level, both genes show a rather low identity of around 65.8%.

Table 9: Comparison of *StGWD* and *AtGWD* sequences.

	cDNA length in bp	Identity in %	gDNA length in bp	No of exons	Length of aa sequence	Identity of aa sequence in %
<i>AtGWD</i>	4200	67.8	8834	32	1399	65.8
<i>StGWD</i>	4395		15664	33	1464	

bp: base pair; aa: amino acid

3.1.3 COMPARISON OF THE *GWD* BAC SEQUENCE AND THE GENOMIC SEQUENCE OF *S. PHUREJA*

Blast analysis against the recently published *Solanum phureja* genomic sequence (PGSC, 2011) resulted in one clear superscaffold for *GWD* (PGSC0003DMB000000248), indicating that *GWD* is encoded at a single locus. The sequence deduced from this blast analysis and the genomic sequence from the BAC sequencing were compared by alignments with Needle software from Emboss (Figure C.8, Appendix) Both sequences showed 87 % identity in total, whereas 11.5 % were gaps. Most of the detected gaps were located in introns (Table 10). One single deletion is located in exon 1 which does not lead to a frame shift as it consists of 3bp resulting in one missing amino acid. All intronic insertions or deletions vary from 3 bp to 1390 bp. In one case, the deletion is located in the 5' splicing side of intron 23, but its functionality is presumed because the following base can compensate as a splicing site donor. In the middle of intron 23, there is a 145 bp variation in a highly repetitive sequence included in this region (3 bp simple sequence repeat). Neglecting the three major gaps (in intron 6, 23 and 25) which were caused by insertions or deletions in one of the sequences, the similarity increases to 97.5 % for the remaining sequence. In average, a SNP (single nucleotide polymorphism) between both sequences excluding indels can be detected every 50 bp.

Table 10: Comparison of the genomic sequence of *GWD* obtained from BAC sequencing with the released *S. phureja* sequence.

	Location	In/del in BAC seq	Length	Position
1	Exon 1	Deletion	3 bp	Close to splicing side
2	Intron 6	Insertion	1390 bp	Middle of intron
3	Intron 12	Deletion	3 bp	Middle of intron
4	Intron 16	Deletion	6 bp	Middle of intron
5	Intron 20	Deletion	13 bp	Middle of intron
6	Intron 21	Insertion	11 bp	Middle of intron
7	Intron 23	Deletion	3 bp	Within 5' splicing side
8	Intron 23	Deletion	145 bp repetitive sequence	Middle of intron
9	Intron 25	Insertion	218 bp	Middle of intron
10	Intron 27	Deletion	3 bp	Middle of intron
11	Intron 28	Deletion	13 bp	Middle of intron
12	Intron 32	Insertion	4 bp	Middle of intron

Sequences were aligned and all insertions and deletions were recorded. The designation insertion or deletion is referred to the BAC sequence, the sequence of *S. phureja* is considered as reference sequence.

3.1.4 CONSERVATION OF *GWD* AMONG DIFFERENT SPECIES

In order to obtain information about *GWD* sequences from other species for an analysis of the conservation, the cDNA sequence Y09553.1 was blasted against the nucleotide collection of NCBI. Further sequences were retrieved by peptide search on NCBI and on Uniprot (<http://www.uniprot.org/>). The blast resulted in 10 sequences which covered two monocots (*Oryza sativa* and *Triticum aestivum*) and eight dicots (Figure 12). The class of dicot sequences contained

members of the two different major groups of asterids and rosids. The group of asterids is only represented by the family of Solanaceae. For the rosids, members of different families were obtained (e.g. Brassicaceae, Salicaceae, Euphorbiaceae, Vitaceae, Fabaceae and Rutaceae). The sequence of *Nicotiana tabacum* is only partially available, but did not alter the output of the phylogenetic tree. The blast against all unigenes of the SGN database (<http://solgenomics.net/>) resulted in five large fragments of the tobacco *GWD* sequence which were combined manually. A phylogenetic tree was generated by using the software ClustalX and displayed with Njplot (Figure 12).

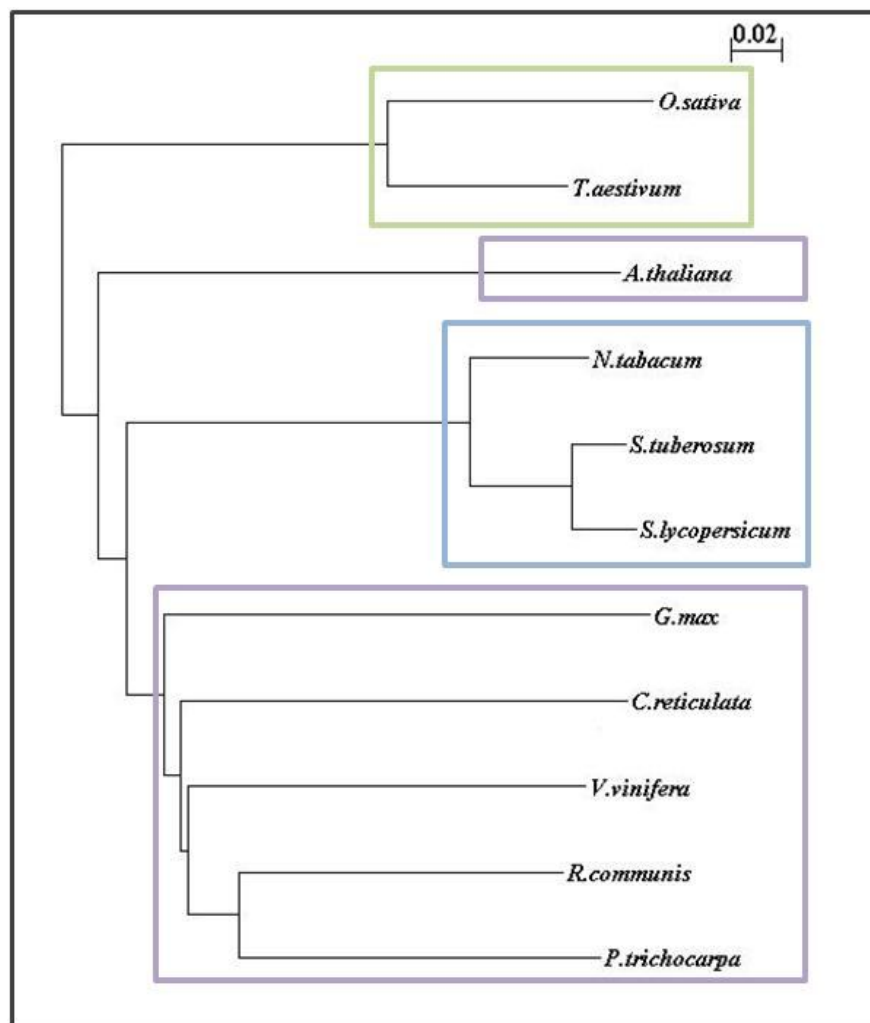
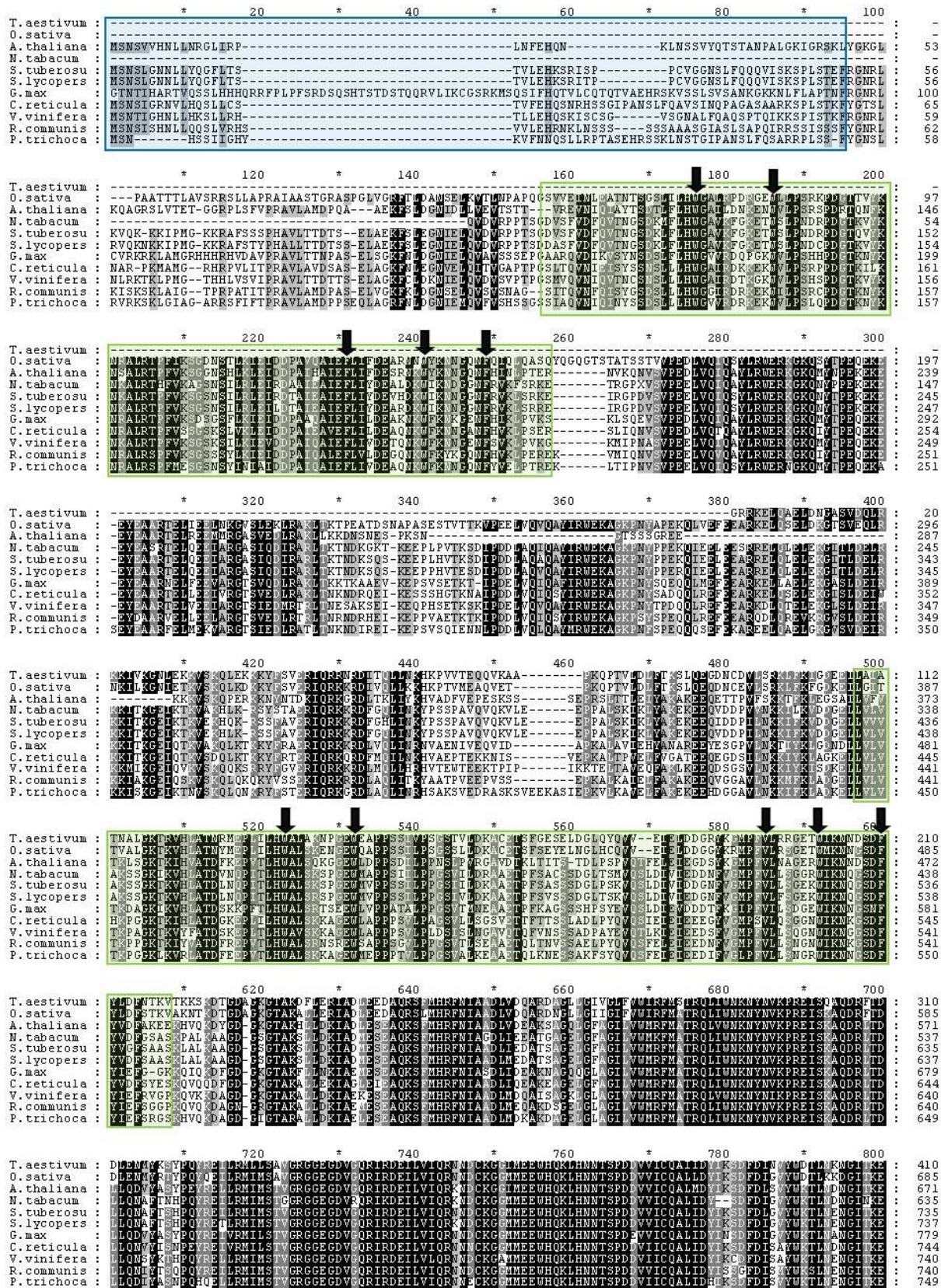


Figure 12: Phylogenetic tree of GWD deduced protein sequences

Retrieved from diverse blast analyses. The different phylogenetic classes are highlighted by coloured boxes. Green: monocots; violet: Rosids including different families (dicots); blue: Asterids including Solanaceae (dicots). Accessions: O.s. Q0DC10; T.a. D7PCT9; A.t. At1g1076; N.t. alignment of SGN-U423492,-U434521,-U421880,-U497595 and -U450019; S.t. Y095331; S.l. EU908574; G.m. AW133227; C.r. AY0940642; V.v. XM_002270449; R.c. XM_002527856; P.t. XM_002315643.

The two members of the monocot group (green box) form a distinct class compared to the remaining dicots. Within this latter group, the three sequences of the Solanaceae can be clearly considered as a distinct subgroup. Interestingly, the sequence of *GWD* from *A. thaliana* forms a separate group and displays a rather low similarity to the remaining sequences of the class of the Rosids.

Besides the phylogenetic tree, alignments of all 11 sequences were constructed with the help of ClustalX and subsequently modified with GeneDOC (Figure 13). The sequence lengths for all full-length sequences varied from 1399 amino acids (*A. thaliana*) to 1477 amino acids (*P. trichocarpa*). For *N. tabacum*, *T. aestivum*, *O. sativa* and *G. max*, no full-length sequences were available.



T.aestivum : RLLSYDRAIHSEPIFRSREDRGLLRDLGQYMRSLRAVHSGADLESALATCMGYSRSEGGGFVGVINPVNLSGCPDQLQVFDHVEERHVEAEPLLEGLL : 510
 O.sativa : RLLSYDRAIHSEPIFRSREDRGLLRDLGQYMRSLRAVHSGADLESALATCMGYSRSEGGGFVGVINPVNLSGCPDQLQVFDHVEERHVEAEPLLEGLL : 785
 A.thaliana : RLLSYDRAIHSEPIFRSREDRGLLRDLGQYMRSLRAVHSGADLESALATCMGYSRSEGGGFVGVINPVNLSGCPDQLQVFDHVEERHVEAEPLLEGLL : 771
 N.tabacum : RLLSYDRAIHSEPIFRSREDRGLLRDLGQYMRSLRAVHSGADLESALATCMGYSRTEGGGFVGVINPVNLSGCPDQLQVFDHVEERHVEAEPLLEGLL : 734
 S.tuberosus : RLLSYDRAIHSEPIFRSREDRGLLRDLGQYMRSLRAVHSGADLESALATCMGYSRTEGGGFVGVINPVNLSGCPDQLQVFDHVEERHVEAEPLLEGLL : 835
 S.lycopers : RLLSYDRAIHSEPIFRSREDRGLLRDLGQYMRSLRAVHSGADLESALATCMGYSRTEGGGFVGVINPVNLSGCPDQLQVFDHVEERHVEAEPLLEGLL : 837
 G.max : RLLSYDRAIHSEPIFRSREDRGLLRDLGQYMRSLRAVHSGADLESALATCMGYSRSEGGGFVGVINPVNLSGCPDQLQVFDHVEERHVEAEPLLEGLL : 879
 C.reticular : RLLSYDRAIHSEPIFRSREDRGLLRDLGQYMRSLRAVHSGADLESALATCMGYSRSEGGGFVGVINPVNLSGCPDQLQVFDHVEERHVEAEPLLEGLL : 844
 V.vinifera : RLLSYDRAIHSEPIFRSREDRGLLRDLGQYMRSLRAVHSGADLESALATCMGYSRSEGGGFVGVINPVNLSGCPDQLQVFDHVEERHVEAEPLLEGLL : 840
 R.communis : RLLSYDRAIHSEPIFRSREDRGLLRDLGQYMRSLRAVHSGADLESALATCMGYSRSEGGGFVGVINPVNLSGCPDQLQVFDHVEERHVEAEPLLEGLL : 840
 P.trichoca : RLLSYDRAIHSEPIFRSREDRGLLRDLGQYMRSLRAVHSGADLESALATCMGYSRSEGGGFVGVINPVNLSGCPDQLQVFDHVEERHVEAEPLLEGLL : 849

T.aestivum : EAQDELPLLLKSHDRDLRDLFLDIALSTVTRTAVERSYEEELNAGAEIMYFISLVLEMLALSDDMEDLYCLRGQVQALMSKSRDTHHALYARAVL : 610
 O.sativa : EAQDELPLLLKSHDRDLRDLFLDIALSTVTRTAVERSYEEELNAGAEIMYFISLVLEMLALSDDMEDLYCLRGQVQALMSKSRDTHHALYARAVL : 885
 A.thaliana : EAQDELPLLLKSHDRDLRDLFLDIALSTVTRTAVERSYEEELNAGAEIMYFISLVLEMLALSDDMEDLYCLRGQVQALMSKSRDTHHALYARAVL : 871
 N.tabacum : EAQDELPLLLKSHDRDLRDLFLDIALSTVTRTAVERSYEEELNAGAEIMYFISLVLEMLALSDDMEDLYCLRGQVQALMSKSRDTHHALYARAVL : 834
 S.tuberosus : EAQDELPLLLKSHDRDLRDLFLDIALSTVTRTAVERSYEEELNAGAEIMYFISLVLEMLALSDDMEDLYCLRGQVQALMSKSRDTHHALYARAVL : 935
 S.lycopers : EAQDELPLLLKSHDRDLRDLFLDIALSTVTRTAVERSYEEELNAGAEIMYFISLVLEMLALSDDMEDLYCLRGQVQALMSKSRDTHHALYARAVL : 937
 G.max : EAQDELPLLLKSHDRDLRDLFLDIALSTVTRTAVERSYEEELNAGAEIMYFISLVLEMLALSDDMEDLYCLRGQVQALMSKSRDTHHALYARAVL : 979
 C.reticular : EAQDELPLLLKSHDRDLRDLFLDIALSTVTRTAVERSYEEELNAGAEIMYFISLVLEMLALSDDMEDLYCLRGQVQALMSKSRDTHHALYARAVL : 944
 V.vinifera : EAQDELPLLLKSHDRDLRDLFLDIALSTVTRTAVERSYEEELNAGAEIMYFISLVLEMLALSDDMEDLYCLRGQVQALMSKSRDTHHALYARAVL : 940
 R.communis : EAQDELPLLLKSHDRDLRDLFLDIALSTVTRTAVERSYEEELNAGAEIMYFISLVLEMLALSDDMEDLYCLRGQVQALMSKSRDTHHALYARAVL : 940
 P.trichoca : EAQDELPLLLKSHDRDLRDLFLDIALSTVTRTAVERSYEEELNAGAEIMYFISLVLEMLALSDDMEDLYCLRGQVQALMSKSRDTHHALYARAVL : 949

T.aestivum : DRTRLALAKAEWHLQVLPQSAEYLGSLPVDQAVNFTTEIIRAGSAAALSLLNRDLPVLRQALGSGWQVISPVEVAVGWVWDELLVQMRITYR : 710
 O.sativa : DRTRLALAKAEWHLQVLPQSAEYLGSLPVDQAVNFTTEIIRAGSAAALSLLNRDLPVLRQALGSGWQVISPVEVAVGWVWDELLVQMRITYR : 985
 A.thaliana : DRTRLALAKAEWHLQVLPQSAEYLGSLPVDQAVNFTTEIIRAGSAAALSLLNRDLPVLRQALGSGWQVISPVEVAVGWVWDELLVQMRITYR : 971
 N.tabacum : DRTRLALAKAEWHLQVLPQSAEYLGSLPVDQAVNFTTEIIRAGSAAALSLLNRDLPVLRQALGSGWQVISPVEVAVGWVWDELLVQMRITYR : 934
 S.tuberosus : DRTRLALAKAEWHLQVLPQSAEYLGSLPVDQAVNFTTEIIRAGSAAALSLLNRDLPVLRQALGSGWQVISPVEVAVGWVWDELLVQMRITYR : 1035
 S.lycopers : DRTRLALAKAEWHLQVLPQSAEYLGSLPVDQAVNFTTEIIRAGSAAALSLLNRDLPVLRQALGSGWQVISPVEVAVGWVWDELLVQMRITYR : 1037
 G.max : DRTRLALAKAEWHLQVLPQSAEYLGSLPVDQAVNFTTEIIRAGSAAALSLLNRDLPVLRQALGSGWQVISPVEVAVGWVWDELLVQMRITYR : 1079
 C.reticular : DRTRLALAKAEWHLQVLPQSAEYLGSLPVDQAVNFTTEIIRAGSAAALSLLNRDLPVLRQALGSGWQVISPVEVAVGWVWDELLVQMRITYR : 1044
 V.vinifera : DRTRLALAKAEWHLQVLPQSAEYLGSLPVDQAVNFTTEIIRAGSAAALSLLNRDLPVLRQALGSGWQVISPVEVAVGWVWDELLVQMRITYR : 1040
 R.communis : DRTRLALAKAEWHLQVLPQSAEYLGSLPVDQAVNFTTEIIRAGSAAALSLLNRDLPVLRQALGSGWQVISPVEVAVGWVWDELLVQMRITYR : 1040
 P.trichoca : DRTRLALAKAEWHLQVLPQSAEYLGSLPVDQAVNFTTEIIRAGSAAALSLLNRDLPVLRQALGSGWQVISPVEVAVGWVWDELLVQMRITYR : 1049

T.aestivum : PTLIVARRVKGEEIIPDGAVLITDMPDVLSHVSVRARNKGVFATC'DPDMILADQAGKRIQLRPTSDIYVNSLLEGLDASSTNLTEGS-PSP : 809
 O.sativa : PTLIVARRVKGEEIIPDGAVLITDMPDVLSHVSVRARNKGVFATC'DPDMILADQAGKRIQLRPTSDIYVNSLLEGLDASSTNLTEGS-PSP : 1083
 A.thaliana : PTLIVARRVKGEEIIPDGAVLITDMPDVLSHVSVRARNKGVFATC'DPDMILADQAGKRIQLRPTSDIYVNSLLEGLDASSTNLTEGS-PSP : 1068
 N.tabacum : PTLIVARRVKGEEIIPDGAVLITDMPDVLSHVSVRARNKGVFATC'DPDMILADQAGKRIQLRPTSDIYVNSLLEGLDASSTNLTEGS-PSP : 1031
 S.tuberosus : PTLIVARRVKGEEIIPDGAVLITDMPDVLSHVSVRARNKGVFATC'DPDMILADQAGKRIQLRPTSDIYVNSLLEGLDASSTNLTEGS-PSP : 1132
 S.lycopers : PTLIVARRVKGEEIIPDGAVLITDMPDVLSHVSVRARNKGVFATC'DPDMILADQAGKRIQLRPTSDIYVNSLLEGLDASSTNLTEGS-PSP : 1134
 G.max : PTLIVARRVKGEEIIPDGAVLITDMPDVLSHVSVRARNKGVFATC'DPDMILADQAGKRIQLRPTSDIYVNSLLEGLDASSTNLTEGS-PSP : 1178
 C.reticular : PTLIVARRVKGEEIIPDGAVLITDMPDVLSHVSVRARNKGVFATC'DPDMILADQAGKRIQLRPTSDIYVNSLLEGLDASSTNLTEGS-PSP : 1144
 V.vinifera : PTLIVARRVKGEEIIPDGAVLITDMPDVLSHVSVRARNKGVFATC'DPDMILADQAGKRIQLRPTSDIYVNSLLEGLDASSTNLTEGS-PSP : 1139
 R.communis : PTLIVARRVKGEEIIPDGAVLITDMPDVLSHVSVRARNKGVFATC'DPDMILADQAGKRIQLRPTSDIYVNSLLEGLDASSTNLTEGS-PSP : 1139
 P.trichoca : PTLIVARRVKGEEIIPDGAVLITDMPDVLSHVSVRARNKGVFATC'DPDMILADQAGKRIQLRPTSDIYVNSLLEGLDASSTNLTEGS-PSP : 1148

T.aestivum : SLSVRRKPLGAYLSAEFSEEMVGAQRSMIAVLRGRVPSWVILPTSVAIFFGFEFVLSDEINQVARELQQLRRLSGDGFSAIIRITVLDLQAP : 909
 O.sativa : SLSVRRKPLGAYLSAEFSEEMVGAQRSMIAVLRGRVPSWVILPTSVAIFFGFEFVLSDEINQVARELQQLRRLSGDGFSAIIRITVLDLQAP : 1183
 A.thaliana : SLSVRRKPLGAYLSAEFSEEMVGAQRSMIAVLRGRVPSWVILPTSVAIFFGFEFVLSDEINQVARELQQLRRLSGDGFSAIIRITVLDLQAP : 1168
 N.tabacum : SLSVRRKPLGAYLSAEFSEEMVGAQRSMIAVLRGRVPSWVILPTSVAIFFGFEFVLSDEINQVARELQQLRRLSGDGFSAIIRITVLDLQAP : 1131
 S.tuberosus : SLSVRRKPLGAYLSAEFSEEMVGAQRSMIAVLRGRVPSWVILPTSVAIFFGFEFVLSDEINQVARELQQLRRLSGDGFSAIIRITVLDLQAP : 1232
 S.lycopers : SLSVRRKPLGAYLSAEFSEEMVGAQRSMIAVLRGRVPSWVILPTSVAIFFGFEFVLSDEINQVARELQQLRRLSGDGFSAIIRITVLDLQAP : 1234
 G.max : SLSVRRKPLGAYLSAEFSEEMVGAQRSMIAVLRGRVPSWVILPTSVAIFFGFEFVLSDEINQVARELQQLRRLSGDGFSAIIRITVLDLQAP : 1278
 C.reticular : SLSVRRKPLGAYLSAEFSEEMVGAQRSMIAVLRGRVPSWVILPTSVAIFFGFEFVLSDEINQVARELQQLRRLSGDGFSAIIRITVLDLQAP : 1244
 V.vinifera : SLSVRRKPLGAYLSAEFSEEMVGAQRSMIAVLRGRVPSWVILPTSVAIFFGFEFVLSDEINQVARELQQLRRLSGDGFSAIIRITVLDLQAP : 1239
 R.communis : SLSVRRKPLGAYLSAEFSEEMVGAQRSMIAVLRGRVPSWVILPTSVAIFFGFEFVLSDEINQVARELQQLRRLSGDGFSAIIRITVLDLQAP : 1238
 P.trichoca : SLSVRRKPLGAYLSAEFSEEMVGAQRSMIAVLRGRVPSWVILPTSVAIFFGFEFVLSDEINQVARELQQLRRLSGDGFSAIIRITVLDLQAP : 1246

T.aestivum : TPLVQELRKTQSSGMPMPGDEGDRWEQAWAIRVWASRWNERAYFSTRVRLDHDYLQMAVLVQEVINADYAFVHIHTNP :SGDSSEIYAEVVRGLG : 1009
 O.sativa : TPLVQELRKTQSSGMPMPGDEGDRWEQAWAIRVWASRWNERAYFSTRVRLDHDYLQMAVLVQEVINADYAFVHIHTNP :SGDSSEIYAEVVRGLG : 1283
 A.thaliana : TPLVQELRKTQSSGMPMPGDEGDRWEQAWAIRVWASRWNERAYFSTRVRLDHDYLQMAVLVQEVINADYAFVHIHTNP :SGDSSEIYAEVVRGLG : 1268
 N.tabacum : TPLVQELRKTQSSGMPMPGDEGDRWEQAWAIRVWASRWNERAYFSTRVRLDHDYLQMAVLVQEVINADYAFVHIHTNP :SGDSSEIYAEVVRGLG : 1231
 S.tuberosus : TPLVQELRKTQSSGMPMPGDEGDRWEQAWAIRVWASRWNERAYFSTRVRLDHDYLQMAVLVQEVINADYAFVHIHTNP :SGDSSEIYAEVVRGLG : 1332
 S.lycopers : TPLVQELRKTQSSGMPMPGDEGDRWEQAWAIRVWASRWNERAYFSTRVRLDHDYLQMAVLVQEVINADYAFVHIHTNP :SGDSSEIYAEVVRGLG : 1334
 G.max : TPLVQELRKTQSSGMPMPGDEGDRWEQAWAIRVWASRWNERAYFSTRVRLDHDYLQMAVLVQEVINADYAFVHIHTNP :SGDSSEIYAEVVRGLG : 1378
 C.reticular : TPLVQELRKTQSSGMPMPGDEGDRWEQAWAIRVWASRWNERAYFSTRVRLDHDYLQMAVLVQEVINADYAFVHIHTNP :SGDSSEIYAEVVRGLG : 1344
 V.vinifera : TPLVQELRKTQSSGMPMPGDEGDRWEQAWAIRVWASRWNERAYFSTRVRLDHDYLQMAVLVQEVINADYAFVHIHTNP :SGDSSEIYAEVVRGLG : 1339
 R.communis : TPLVQELRKTQSSGMPMPGDEGDRWEQAWAIRVWASRWNERAYFSTRVRLDHDYLQMAVLVQEVINADYAFVHIHTNP :SGDSSEIYAEVVRGLG : 1338
 P.trichoca : TPLVQELRKTQSSGMPMPGDEGDRWEQAWAIRVWASRWNERAYFSTRVRLDHDYLQMAVLVQEVINADYAFVHIHTNP :SGDSSEIYAEVVRGLG : 1346

T.aestivum : ETLVGAYPGRALSFCRKRDLNSPQVLYGYSRPIGLFIRSIIFRSDSNGEDLEGYACAGLYDSVPMDDEEVLDYTTDGLIDGFRNLSLSIARAG : 1109
 O.sativa : ETLVGAYPGRALSFCRKRDLNSPQVLYGYSRPIGLFIRSIIFRSDSNGEDLEGYACAGLYDSVPMDDEEVLDYTTDGLIDGFRNLSLSIARAG : 1383
 A.thaliana : ETLVGAYPGRALSFCRKRDLNSPQVLYGYSRPIGLFIRSIIFRSDSNGEDLEGYACAGLYDSVPMDDEEVLDYTTDGLIDGFRNLSLSIARAG : 1368
 N.tabacum : ETLVGAYPGRALSFCRKRDLNSPQVLYGYSRPIGLFIRSIIFRSDSNGEDLEGYACAGLYDSVPMDDEEVLDYTTDGLIDGFRNLSLSIARAG : 1331
 S.tuberosus : ETLVGAYPGRALSFCRKRDLNSPQVLYGYSRPIGLFIRSIIFRSDSNGEDLEGYACAGLYDSVPMDDEEVLDYTTDGLIDGFRNLSLSIARAG : 1432
 S.lycopers : ETLVGAYPGRALSFCRKRDLNSPQVLYGYSRPIGLFIRSIIFRSDSNGEDLEGYACAGLYDSVPMDDEEVLDYTTDGLIDGFRNLSLSIARAG : 1434
 G.max : ETLVGAYPGRALSFCRKRDLNSPQVLYGYSRPIGLFIRSIIFRSDSNGEDLEGYACAGLYDSVPMDDEEVLDYTTDGLIDGFRNLSLSIARAG : 1478
 C.reticular : ETLVGAYPGRALSFCRKRDLNSPQVLYGYSRPIGLFIRSIIFRSDSNGEDLEGYACAGLYDSVPMDDEEVLDYTTDGLIDGFRNLSLSIARAG : 1444
 V.vinifera : ETLVGAYPGRALSFCRKRDLNSPQVLYGYSRPIGLFIRSIIFRSDSNGEDLEGYACAGLYDSVPMDDEEVLDYTTDGLIDGFRNLSLSIARAG : 1439
 R.communis : ETLVGAYPGRALSFCRKRDLNSPQVLYGYSRPIGLFIRSIIFRSDSNGEDLEGYACAGLYDSVPMDDEEVLDYTTDGLIDGFRNLSLSIARAG : 1438
 P.trichoca : ETLVGAYPGRALSFCRKRDLNSPQVLYGYSRPIGLFIRSIIFRSDSNGEDLEGYACAGLYDSVPMDDEEVLDYTTDGLIDGFRNLSLSIARAG : 1446

T.aestivum : HAIIEELGSPQDIEGVVIRDKRIYVVQTRPQM : 1136
 O.sativa : HAIIEELGSPQDIEGVVIRDKRIYVVQTRPQM : 1414
 A.thaliana : HAIIEELGSPQDIEGVVIRDKRIYVVQTRPQM : 1399
 N.tabacum : HAIIEELGSPQDIEGVVIRDKRIYVVQTRPQM : 1362
 S.tuberosus : HAIIEELGSPQDIEGVVIRDKRIYVVQTRPQM : 1463
 S.lycopers : HAIIEELGSPQDIEGVVIRDKRIYVVQTRPQM : 1465
 G.max : HAIIEELGSPQDIEGVVIRDKRIYVVQTRPQM : 1509
 C.reticular : HAIIEELGSPQDIEGVVIRDKRIYVVQTRPQM : 1475
 V.vinifera : HAIIEELGSPQDIEGVVIRDKRIYVVQTRPQM : 1470
 R.communis : HAIIEELGSPQDIEGVVIRDKRIYVVQTRPQM : 1469
 P.trichoca : HAIIEELGSPQDIEGVVIRDKRIYVVQTRPQM : 1477

Figure 13: Sequence alignment of 11 GWD sequences from different species.

Amino acid residues which are identical in all sequences are highlighted by a black background, those conserved in most of the sequences are marked by a grey background. Conserved and annotated domains are indicated by boxes. Blue: Chloroplastic target peptide; light green: CBM 45; dark green: nucleotide binding domain. The black arrows highlight the five conserved aromatic amino acid residues which occur in both parts of the bipartite CBM. The catalytic histidine for the autophosphorylation is marked by a green arrow. T.aestivum: *Triticum aestivum*; O.sativa: *Oryza sativa*; A.thaliana: *Arabidopsis thaliana*; N.tabacum: *Nicotiana tabacum*; S.tuberosu: *Solanum tuberosum*; S.lycopers: *Solanum lycopersicum*; G.max: *Glycine max*; C.reticula: *Citrus reticulata*; V.vinifera: *Vitis vinifera*; R.communis: *Ricinus communis*; P.trichoca: *Populus trichocarpa*.

The alignment of the 11 amino acid sequences from the different species of various clades of the plant kingdom showed an overall high conservation. Especially the regions close to the C-terminus demonstrated strong preservation. From position 1326 to the C-terminal end, the nucleotide binding domain is almost identical throughout all analyzed species (green box). Upstream of this nucleotide binding domain, the catalytic histidine is located (green arrow). This part and the surrounding region are as well highly conserved. Interestingly, a large region upstream of the catalytic part shows only little variation without having a known function so far. In contrast to that, the part of the starch-binding domain CBM 45 seems to be less conserved (light green box), although all *GWD* homologs are supposed to bind to starch as their substrate. Within this bipartite starch binding domain, five conserved amino acids are highlighted (black arrows, Glaring et al., 2011) which are highly conserved between the species and occur in a similar pattern in both domains of the CBM. The region bearing the chloroplastic target peptide is least conserved.

3.1.5 ASSOCIATION ANALYSIS OF *GWD*

Due to the crucial position in the starch degradation pathway, *GWD* is an interesting candidate for an association analysis with starch and chip quality traits. By using an amplicon sequencing approach, an appropriate amplicon in an interesting region of *GWD* was found within the linker region of the bipartite starch-binding domain (CBM45). The corresponding sequencing primers are listed in table 5 of chapter 2 materials and methods (LS21-LS20). This amplified region overlaps with the amplicon used in Chen et al. 2001 for a potato molecular-function map for carbohydrate metabolism. A suitable amplicon of about 770 bp was obtained (Appendix A.6) and used for an association analysis in 218 genotypes of the GABICHIPS population (Li et al., 2008). It was confirmed that all four *GWD* alleles can be amplified with this primer combination because some SNPs displayed all five possible dosages. For the SNP analysis, all amplicons were sent for sequencing with the reverse primer and were afterwards scored manually for SNPs. In total, 30 SNPs were identified and tested for association with starch and chip quality traits. The calculated average SNP rate is around 1SNP/26bp. Calculations for association were performed with the statistical test univariate GLM (general linear models) considering the factor origin (Li et al, 2008). Out of the 30 analyzed SNPs, only those SNPs are shown which were significant at $p < 0.05$ (Table 11).

Table 11: Overview of associated SNPs of *GWD*.

SNP			Trait						Features				
#	LD group	From ATG	TY effect R ² (%)	TSC effect R ² (%)	TSY effect R ² (%)	CQA effect R ² (%)	CQS4 effect R ² (%)	CQS8 effect R ² (%)	X/Y	Ex/In	AA	True allele frequency	Apparent allele frequency
1	1	4446	ns	0.017 ↑ (T) 3.6	0.003 ↑ (T) 4.6	ns	ns	ns	A/T	In	-	6.6%	23.4%
3	1	4633	ns	0.017 ↑ (G) 3.6	0.003 ↑ (G) 4.6	ns	ns	ns	A/G	Ex 8	K / E	6.6%	23.4%
5	1	4774	ns	0.023 ↑ (G) 3.3	0.006 ↑ (G) 4.2	ns	ns	ns	C/G	Ex 8	P / A	6.6%	23.0%
9	1	4980	ns	0.023 ↑ (T) 3.3	0.006 ↑ (T) 4.2	ns	ns	ns	G/T	Ex 9	A / S	6.6%	23.0%
2	2	4620	ns	ns	0.033 ↓ (C) 2.8	0.017 ↓ (C) 3.9	ns	ns	A/C	Ex 8	-	7.1%	27.2%
4	2	4763	ns	ns	0.034 ↓ (C) 2.6	0.034 ↓ (C) 2.2	ns	ns	T/C	Ex 8	L / S	6.9%	26.1%
6	2	4795	ns	ns	0.030 ↓ (A) 2.9	0.023 ↓ (A) 2.5	ns	ns	C/A	Ex 8	L / M	9.2%	27.5%
7	-	4826	0.021 ↓ (G) 2.6	0.001 ↑ (G) 5.9	0.004 ↑ (G) 4.4	ns	ns	ns	A/G	Ex 8	D / G	2.6%	9.6%
8	-	4843	ns	ns	ns	ns	ns	0.033 ↑ (G) 2.4	A/G	Ex 8	K / E	6.6%	23.0%

The position of the SNP is counted from the ATG (ref.: BAC sequence), where 1 represents the adenine of the start codon. All SNPs were tested for association with the following traits: TY: tuber yield, TSC: tuber starch content, TSY: tuber starch yield, CQA: chip quality in autumn, CQS4 or CQS8 chip quality after storage at 4°C or 8°C, respectively. For each trait, the statistical test univariate GLM was applied with p<0.05. The feature part lists allele frequencies, exon or intron location and if the SNP leads to an aa exchange. All scored allele dosages are indicated. True allele frequency (872 alleles) and apparent allele frequency (218 genotypes) is given in %. The SNPs are sorted upon their classification into LD groups. R²: percent of the total variance explained; ns = not significant, ↑ positive effect, ↓ negative effect.

In total, nine of the 30 analyzed SNPs were associated. Several SNPs showed a high degree of co-segregation and were associated to a similar extent with the same traits. They were classified into LD (linkage disequilibrium) group 1 (LD1: SNPs 1, 3, 5 and 8) and LD group 2 (LD2: SNPs 2, 4 and 6). In LD1, the change from the common allele (reference allele of annotated sequence) to the variant allele leads to a positive association with the traits tuber starch content (TSC, % fresh weight) and tuber starch yield (TSY (dt/ha) = TSC x TY). The means plots for SNP A₄₄₄₆T are displayed in figure 14 as a representative for the LD group 1. The remaining SNPs of this LD group showed similar effects.

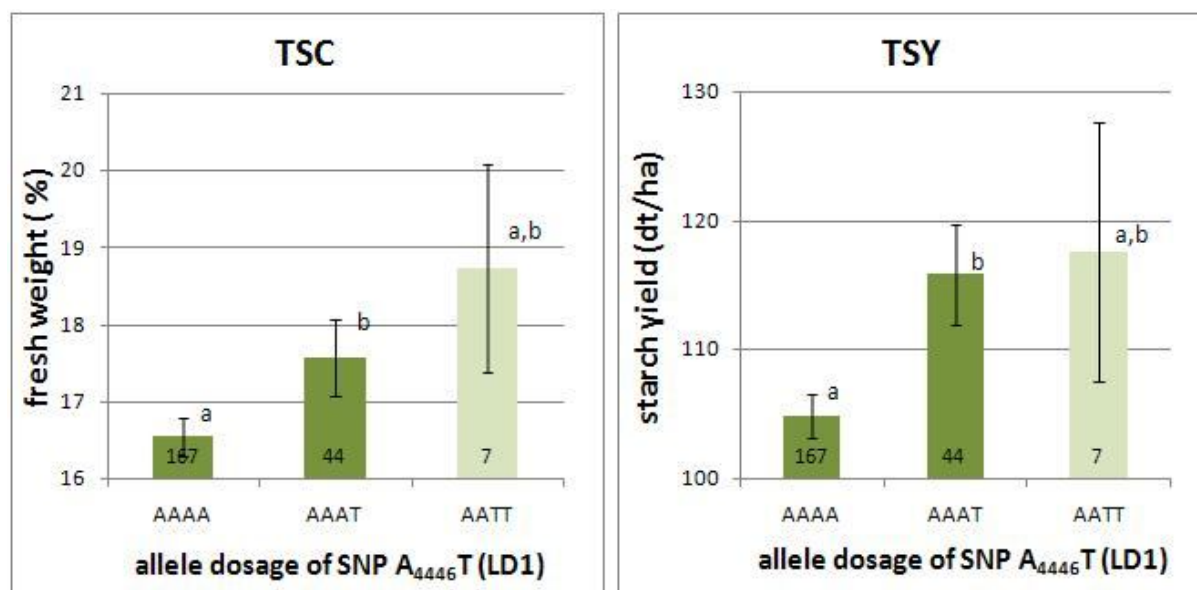


Figure 14: Means plots of SNP 1 (A₄₄₄₆T) of LD1 of GWD.

The X-axis represents different allelic states which were scored for the SNP. The Y-axis shows the phenotypic data for the two traits tuber starch content (TSC) and tuber starch yield (TSY). The number given in the lower part of the bar stands for the number of genotypes bearing the allelic state. Significant differences between the different groups (allelic states) are indicated by small letters. Different letters symbolize different groups, same letters stand for the same group and two letters indicate no clear classification to one of the groups. The bars in dark green represent more reliable values for the phenotypic means. Light green indicates mean values of only a small sample size, including a large error.

The presence of the allele "T" leads to higher tuber starch content and starch yield. Whether the presence of a second allele (a higher allelic dosage) further increases the phenotypic values cannot be concluded. A significant difference was only calculated between the homozygous reference allele ("AAAA") and the simplex dosage ("AAAT") of the "T" allele, due to only seven genotypes containing the SNP in duplex ("AATT"). The four SNPs of this LD group 1 explained around 3.5 % of the phenotypic variance of tuber starch content and up to 4.6 % of the variance of tuber starch yield. The LD1 allele is not very frequent in the population (23.4 %; Table 11). It is predominantly present in simplex which leads to a very low allele frequency of 6.6 %. LD group 2 showed similar results. The three SNPs are associated with the traits TSY and COA (chip quality in autumn). They explain around 2.8 % of the variance of tuber starch yield and around 3% of the phenotypic variance of COA. The apparent and the true allele frequency are low and the alleles predominantly occur in simplex (Table 11). On average, there are only three genotypes having the allele in duplex dosage and are therefore

not representative. For all three SNPs of LD2, the shift from the basic allele to the variant allele results in a negative association with both traits.

SNP 7 in exon 8 (position A₄₈₂₆G) is not in LD with any of the analyzed SNPs. It is associated with the three traits TY, TSC and TSY. The change from the common "A" allele to the "G" allele leads to a negative association with tuber yield (TY), but a positive association with TSC and TSY. The means plots are given exemplarily for TY and TSC in figure 15. As previously observed, no clear observation can be made for genotypes bearing two copies of the variant allele due to a small sample size. The SNP explains 2.9 % of the tuber yield variance, 5.9 % of the phenotypic variance of tuber starch content and 4.4 % of the variance of tuber starch yield. The calculated apparent allele frequency of the allele is just 2.6 %. Additionally, the variant allele occurs predominantly in simplex which leads to a low allele frequency of 9.6 % (Table 11). There are only two genotypes containing the allele in duplex (Figure 15) which results in a not representative mean of tuber yield and tuber starch content for this allelic state.

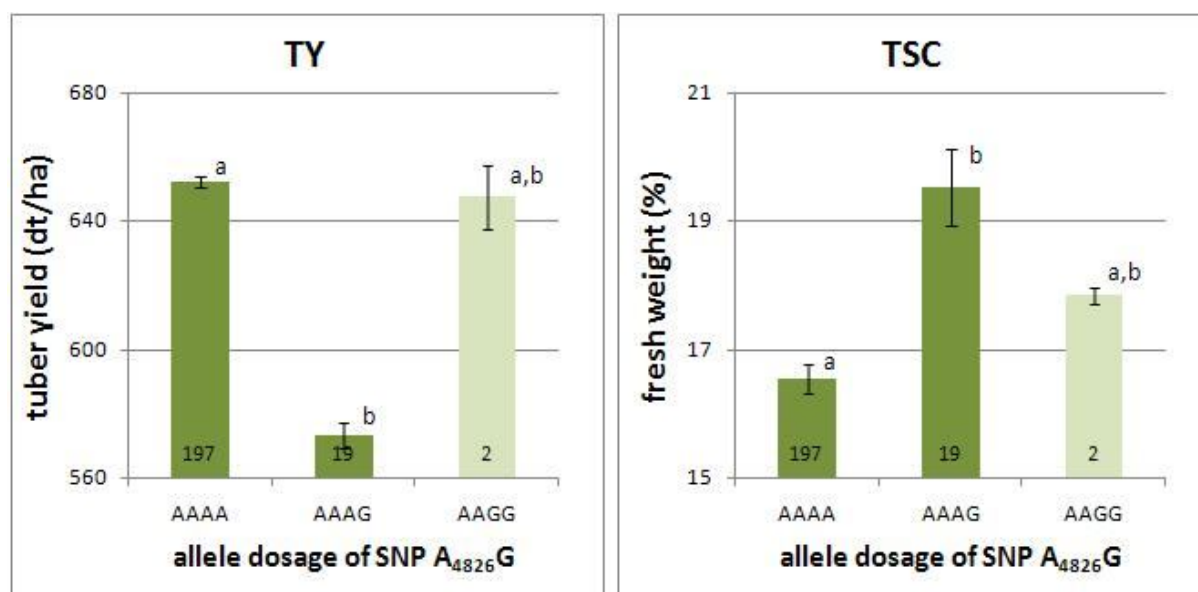


Figure 15: Means plots of SNP 7 (A₄₈₂₆G) of GWD.
Further explanations as described in Figure 14.

A moderate association was detected for SNP 8 with the trait COS8 (chip quality after storage at 8°C). The shift from the reference "A" allele to the "G" allele has a positive effect on the chip quality and explains 2.4 % of the phenotypic variance. This SNP is not in obvious LD with other SNPs.

3.1.6 GWD HAPLOTYPE MODELING

All 30 SNPs scored in the GWD amplicon of the previous association analysis were assigned into LD groups. In total, 26 of the scored SNPs could be classified into at least 8 LD groups, the remaining 4 SNPs form single groups. Of each group, one representative SNP was chosen for the calculation of

haplotypes with the software SATlotyper (Neigenfind et al., 2008). The used 12 SNPs are at the positions 4466, 4633, 4672, 4713, 4746, 4763, 4826, 4839, 4843, 4878, 4924 and 4974, whereas the blue SNPs are representatives of the associated SNPs from table 11. SATlotyper is an *in silico* tool to estimate haplotypes and calculating their possible frequency by combining different SNPs upon a certain algorithm.

The software estimated 25 possible haplotypes for *GWD* based the selected 12 SNPs. The top 10 of the haplotype models are shown in table 12 (frequency > 4.5 %). The remaining haplotypes are listed in the appendix (Table B.6). In general, the most frequent haplotype models (true allele frequency) refer to the most common ones (apparent allele frequency). There are many genotypes that contained the haplotype models only in simplex or duplex state. The most frequent haplotype (no 9) with 17.4 % frequency was identical to the haplotype of the obtained BAC sequence. It is present in all allelic states from simplex to quadruplex. The haplotype which corresponds to the published *S. phureja* sequence (PGSC) is still in the top 10 (haplotype number 20), but its frequency is already very low (4.7 %). For the remaining haplotypes, it cannot be confirmed that they are real due to missing sequence information.

Table 12: List of *StGWD* haplotype models based on 12 SNPs scored in 218 genotypes.

no	haplotype sequence	true haplotype frequency % (no)	apparent haplotype frequency % (no)	allelic states n/s/d/t/q
9	CAGTATAAATTG	17.4 (152)	45.0 (98)	120/55/34/7/2
21	TAGGATATAGTC	16.2 (141)	47.7 (104)	114/70/31/3/0
7	CAGTATAAAGTG	7.7 (67)	30.7 (67)	151/67/0/0/0
5	CAGTACAAAGTC	6.1 (53)	22.9 (50)	168/47/3/0/0
22	TAGGATATATTG	5.5 (48)	21.6 (47)	171/46/1/0/0
8	CAGTATAAATTC	4.9 (43)	19.3 (42)	176/41/1/0/0
12	CAGTGTAAGAC	4.8 (42)	17.9 (39)	179/36/3/0/0
13	CAGTGTAAGAG	4.7 (41)	17.9 (39)	179/37/2/0/0
20	TAGGATAAGGTC	4.7 (41)	17.0 (37)	181/33/4/0/0
16	CGGATAAAGTC	4.7 (41)	16.5 (36)	182/31/5/0/0

The two haplotypes which are either identical to the BAC sequence (designated as reference sequence) or to the PGSC sequence, are highlighted in grey. Red letter: associated reference SNP; Blue letter: associated SNP with positive effect. True allele frequency: Number of haplotypes among possible 872 haplotypes (4 x 218 genotypes); Apparent allele frequency: Number of genotypes containing the haplotype. Nulliplex (n), simplex (s), duplex (d), triplex (t) and quadruplex (q): Allelic states of haplotypes per genotype.

3.2 PHOSPHOGLUCAN WATER DIKINASE - GENOMIC ORGANIZATION

PWD (phosphoglucan water dikinase) has been initially identified as *GWD3* in *A. thaliana* (accession AY747068, Baunsgaard et al., 2005). Only little information was available in *S. tuberosum* about this gene. The assumed important role in the first phase of starch degradation (Kötting et al., 2005) makes it an interesting functional candidate for association analysis and structural characterization. To gain information about this gene in potato, blasts of the annotated cDNA sequence from *A. thaliana* were performed against the SGN database. A hit for a potato EST (CV431165) and additionally a tomato BAC sequence (C09HBa0226D21) were found. The 900 bp potato EST was used as template for the design of specific primers. An approximately 700 bp fragment was generated by PCR amplification (LS15-LS16; Table 5) on genomic DNA from P6/210 (section 3.1) and used as a *PWD* probe for the BAC library screen. For library BA, 11 positive BACs were identified of which 9 BACs could be validated by PCR using the initial probe primers. Instead of primer walking on all 9 positive BACs, the sequence of the tomato BAC was used to design further primers in more distal parts due to a high degree of synteny between tomato and potato (Tanksley et al., 1992). Six out of the 9 BACs were assumed to contain the putative full-length *PWD* gene basing on positive amplifications of primers which were designed approximately 20 kb up-and downstream of the EST. The sizes of the genomic fragment from the 6 positive BACs were estimated via PFGE (Figure 16). After digestion with the restriction enzyme *NotI*, all BACs showed the same bands for the vector fragments (three lower bands) and the upper band for the genomic fragment which ranged from 36 kDa to 95 kDa.

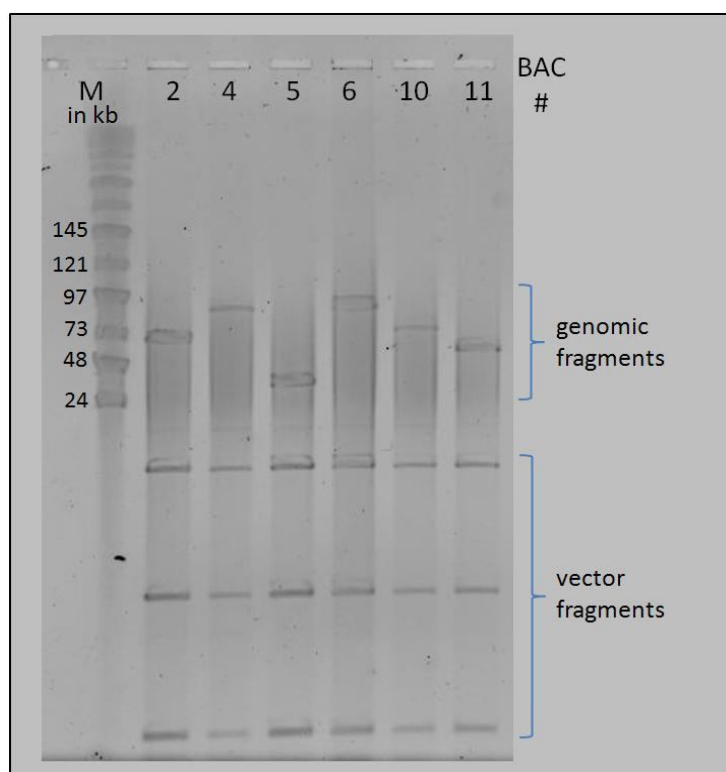


Figure 16: PFGE of *PWD* BACs digested with *NotI*.

Fragments sizes were estimated by comparison to the MidRange II Marker (M).

In order to obtain the whole genomic DNA sequence of *PWD* in potato, BAC # 4 was selected for sequencing due to its large genomic insert (Figure 16). The alignment of the potato *PWD* EST and the published *PWD* cDNA sequence of *A. thaliana* to the tomato BAC sequence indicated that the EST corresponded to the last exon. Several different primer pairs were designed on that part of the tomato BAC sequence which was upstream of the aligned EST (Table 5, Materials & Methods). PCRs were performed on the BAC sequence and additionally on cDNA of different genotypes (BIOSOL) and subsequently sequenced. All sequences were aligned in SeqMan (Lasergene) and assembled into genomic and cDNA sequence contigs. Although different genotypes were used as cDNA templates, all sequences were easily aligned and resulted in a cDNA sequence with only few polymorphisms. Comparisons of the deduced cDNA and genomic sequences by using Spidey software resulted in clear splicing sites. The genomic organization for *PWD* from *S. tuberosum* was generated and is shown in (Figure 17). The genomic sequence consists of 11364 bp. So far, no information about the length of the UTR could be retrieved. The length of the deduced cDNA sequence is 3609 bp, distributed on 19 exons. The exons vary from 53 bp up to 905 bp with an average exon size of 190 bp. Translation of the sequence leads to a 1202 amino acid sequence. The intron sequences show a broader range from 74 bp up to 1630 bp and hence represent more than two thirds of the whole genomic sequence. Detailed lengths of the exons and introns and the sequences are listed in appendix (B.2).

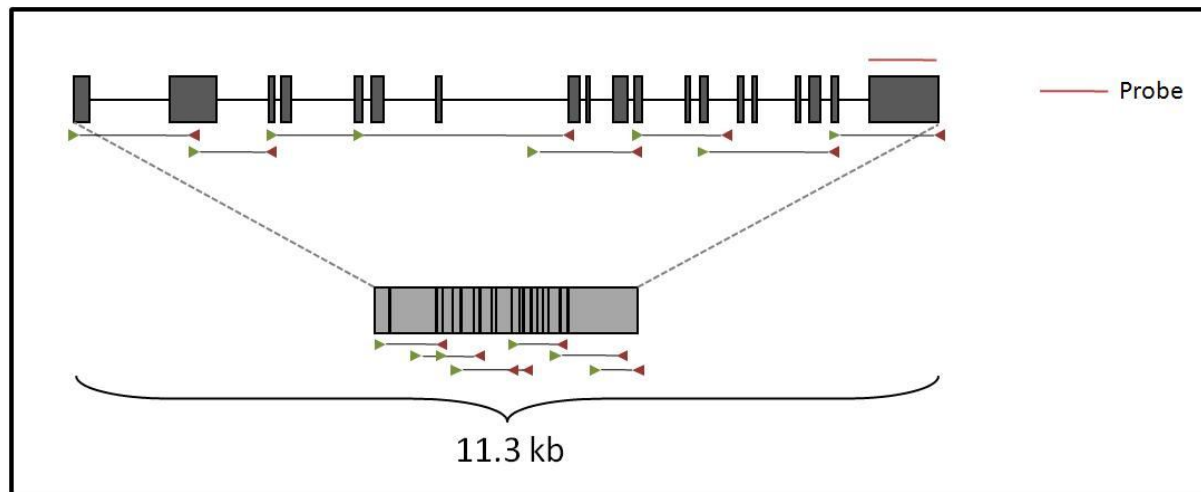


Figure 17: Genomic organization of *PWD* from *S. tuberosum*.

The probe for the BAC library screen is represented by a red line above exon 19. The upper part shows the intron-exon structure obtained from alignments of the genomic DNA and the cDNA sequence in Spidey. The lower part represents the deduced cDNA sequence. Below the sequences, the primers are indicated which were designed on the tomato BAC sequence and used for amplification and sequencing reactions on the potato sequences. The fragments are represented by black lines and the primer by arrows. Green arrow: forward primer, red arrow: reverse primer.

The tomato BAC sequence was annotated to be located on chromosome 9. Due to the release of the genomic sequence of *S. phureja*, it was possible to locate PoMaMo marker in the *PWD* superscaffold of the PGSC database and map them together to the long arm of chromosome 9 (Appendix A.4).

3.2.1 STRUCTURAL COMPARISON OF *PWD* FROM *S. TUBEROSUM* AND *A. THALIANA*

The comparison of the genomic sequence and structural organization of *PWD* from *S. tuberosum* (*StPWD*) to the published sequence from *A. thaliana* (*AtPWD*) shows an overall high conservation between both genes (Figure 18).

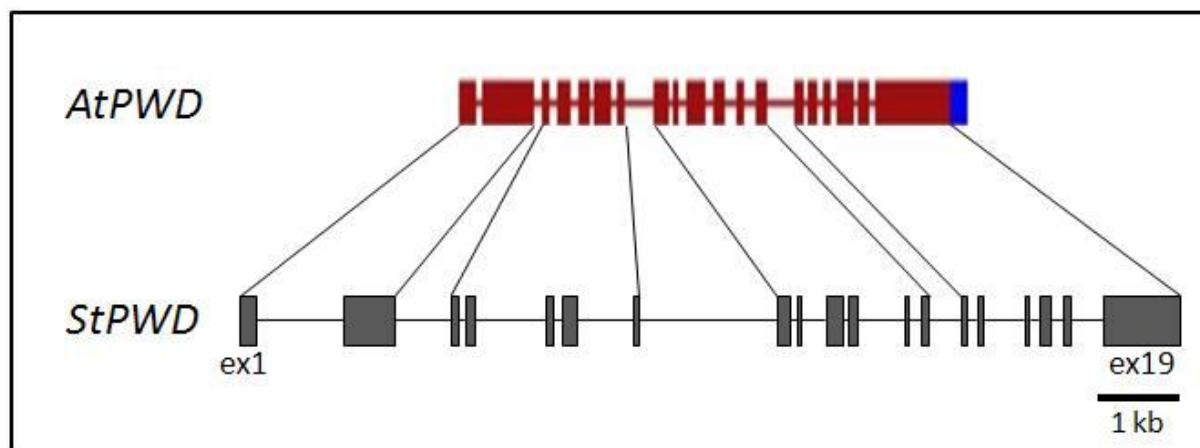


Figure 18: Structural comparison of *StPWD* and *AtPWD*.

Introns are displayed as straight lines (red: *A. thaliana*, black: *S. tuberosum*). The exons are represented by boxes. Blocks of similar exon-intron distribution are linked by grey lines.

The genomic organization is highly conserved. The length and the distribution of the 19 exons are very similar between both genes. Most differences can be pinpointed to the intron lengths which lead to the almost double length of the genomic sequence of *StPWD* (Table 13). At the cDNA level, both genes have the same length of around 3600 bp and are identical to 62.9 %. At the amino acid sequence level, they share an identity of around 59.8 %.

Table 13: Comparison of *StPWD* and *AtPWD* at sequence length level.

	cDNA length in bp	Identity in %	gDNA length in bp	No. of exons	Length of aa sequence	Identity of aa sequence in %
<i>AtPWD</i>	3591	62.9	5947	19	1196	59.8
<i>StPWD</i>	3609		11364	19	1202	

3.2.2 FUNCTIONAL DOMAINS OF *PWD*

Several domains of *PWD* have been described for the orthologue in *A. thaliana* by Kötting et al. (2005) and Baunsgaard et al. (2005). There are small discrepancies in the lengths of the different domains between both publications but the general organization is the same. In order to designate functional domains to the established *StPWD* sequence, several blasts against protein databases were done. Additionally, the known domains from *A. thaliana* were considered. Analysis of the amino acid sequence of *StPWD* using ChloroP predicted a chloroplast transit peptide (cTP) composed of 62 amino acids at the N-terminal part. It covers the whole exon 1 (Figure 19). For *A. thaliana*, the

sequence was predicted to consist of 51-52 amino acids. Blasts of the amino sequence of StPWD against the Pfam database revealed two domains which have been described for AtPWD. The first one is the carbohydrate binding domain 20 (CBM 20) which is covering the whole second exon (aa 90 to 172). Its length of 82 aa is within the range of the predicted CBM of AtPWD of either 71 aa or 100 aa. The second prominent domain is a nucleotide binding domain which is similar to a binding domain of the pyruvate phosphate dikinase (PPDK). It is located at position 879 to 1200 (entire exon 19) at the very C-terminal end and matches to AtPWD PPDK domain. The catalytic histidine is located at position 776 and surrounded by a small but conserved phosphohistidine domain of around 10 amino acids. This position corresponds to the very distal part of exon 15. The location of this particular part of the sequence is deduced from its high conservation among all characterized water dikinases. No further domains or active sites were detected.

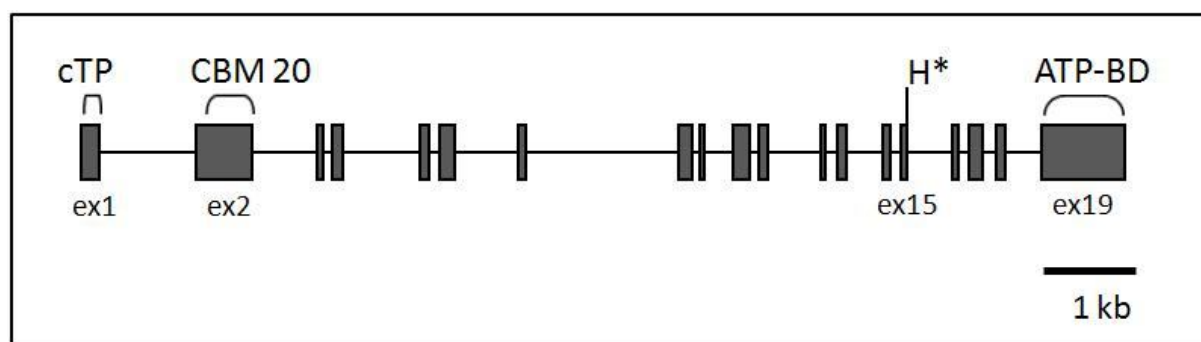


Figure 19: Functional domains of *StPWD*.

The domains are indicated to their corresponding position of cDNA sequence. cTP: chloroplast target peptide, CBM20: carbohydrate binding module of family 20; H*: catalytic histidine; ATP-BD: nucleotide binding domain

3.2.3 COMPARISON OF THE *PWD* BAC SEQUENCE AND THE GENOMIC SEQUENCE OF *S. PHUREJA*

Comparisons between the recently published *S. phureja* genomic sequence (PGSC, 2011) and the genomic sequence from the BAC sequencing of *PWD* were performed by alignments with the software Needle from Emboss explorer (<http://emboss.bioinformatics.nl/>). Both sequences showed an overall identity of 99.7 % with no gaps. Only few SNPs were detected (approximately 40 SNPs in 11.3 kb genomic sequence). The sequence of *S. phureja* and the sequence from P6/210 which was used for the BAC library construction were identical.

3.2.4 CONSERVATION OF *PWD* AMONG DIFFERENT SPECIES

Little is known about *PWD* in other species than *A. thaliana*. By using cDNA and protein sequences in blasts against several databases, further orthologues from different species were identified. For this analysis, the databases of SGN, NCBI and Uniprot were used and resulted in six sequences of orthologues. From the division of vascular seed plants, *PWD* orthologues were detected in the large family of Rosids (families Brassicaceae, Euphorbiaceae and Magnoliopsidae). An interesting orthologue was indentified in one of the mosses, in *Physcomitrella patens* (Bryophyta). Additionally, a tomato cDNA sequence was manually deduced by aligning the StPWD cDNA sequence with the software Spidey against the tomato BAC sequence (C09HBa0226D21) from previous analysis. The obtained cDNA was translated in to the corresponding amino acid sequence. The phylogenetic tree was compiled with ClustalX2 and displayed with Njplot (Figure 20). The sequence of *P. patens* forms a clear out-group in the phylogenetic tree. According to their classes, the remaining groups build small subgroups. Interestingly, the *PWD* sequence of *A. thaliana* did not form a clear clade with the remaining members of the Rosids (*R. communis* and *V. vinifera*) and shows less conservation than the two members of the monocotyledons (*O. sativa* and *Sorghum bicolor*).

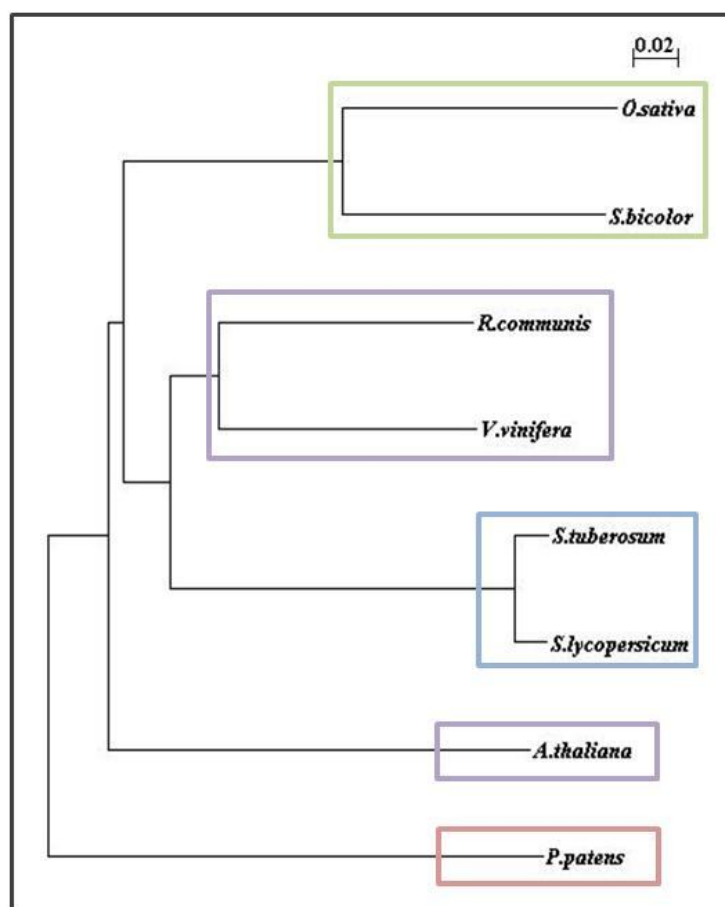
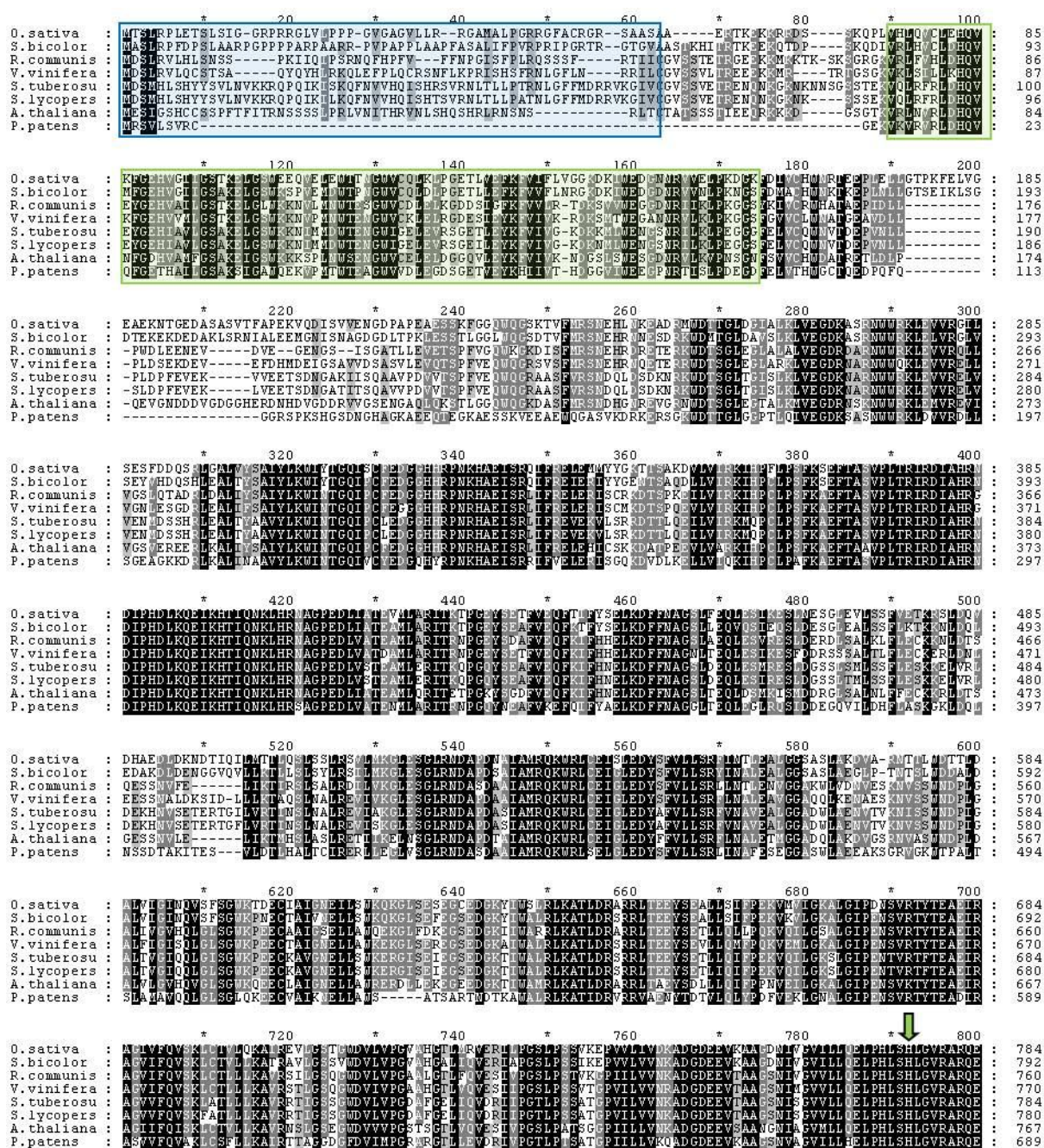


Figure 20: Phylogenetic tree of PWD deduced protein sequences.

Retrieved from diverse blast analyses. The different phylogenetic classes and groups are highlighted by coloured boxes. Red: mosses; Green: monocots; Dicots: Violet; Rosids: Blue; Asterids. Accessions: O.s. OS120297500; S.b. C5XZM3; R.c. XM_002518566; V.v. F61490; S.t. BAC BA19-F1; S.l. C09HBa00226D21; A.t. At5g26570; P.p. A9RDP4.

The alignment of the eight amino acid sequences was performed with ClustalX2 and afterwards modified by using GeneDOC (Figure 21). In general, the eight sequences show very high conservation among each other, although they were obtained from different phylogenetic classes. The CBM 20 domain (light green), the region around the catalytic histidine (green arrow) and the large nucleotide binding domain (dark green) are strongly conserved. They form clear and distinct blocks with less conserved parts in between. However, the chloroplast transit peptide shows the lowest conservation and does not seem to be present or strongly shortened in the sequence of *P. patens* (light blue). A large part between the starch binding domain and the catalytic domain shows a high degree of conservation (e.g. pos 310 to 480), although no function has been predicted or identified so far.



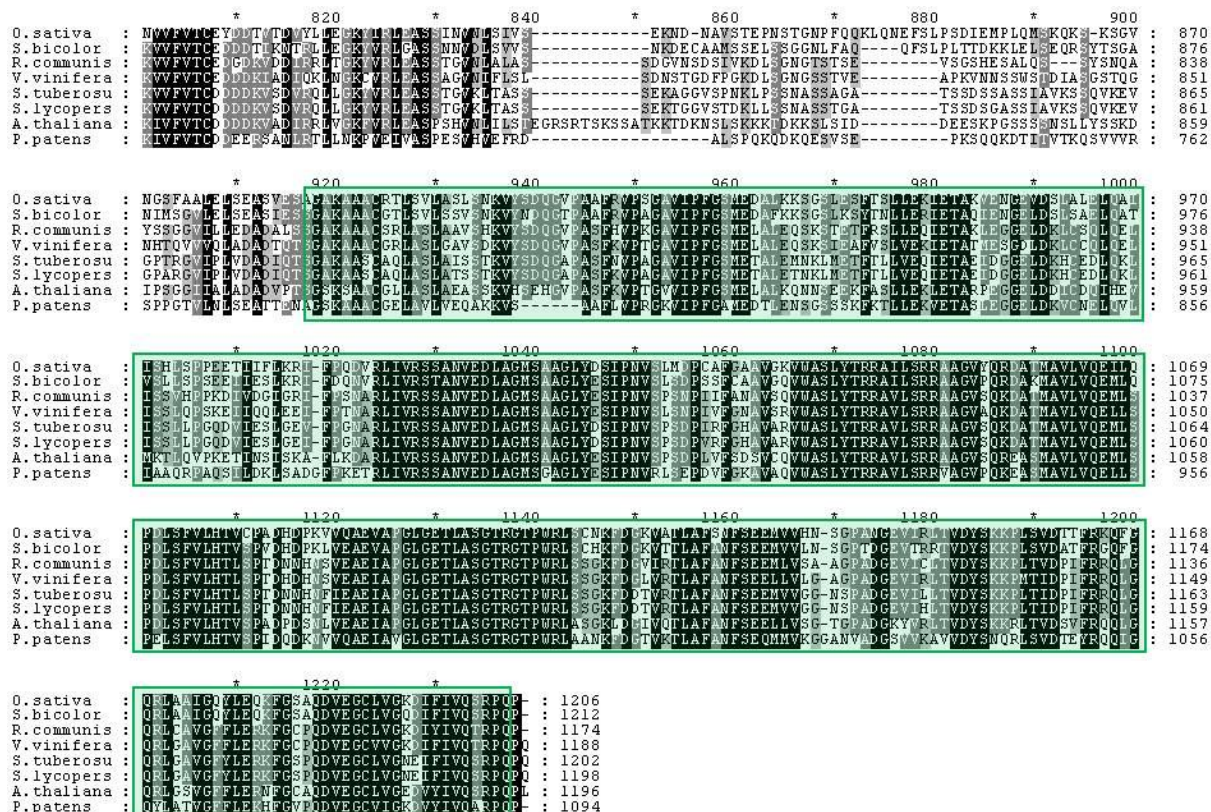


Figure 21: Alignment of eight amino acid sequences of different PWD orthologues.

Amino acid residues identical in all sequences are highlighted by a black background, those conserved in most of the sequences are marked by a grey background. Conserved and annotated domains are indicated by boxes. Blue: Chloroplastic target peptide; light green: CBM 20; dark green: nucleotide binding domain. The catalytic histidine is marked by a green arrow. *O. sativa*; *S. bicolor*; *R. communis*; *V. vinifera*; *S. tuberosu*; *S. tuberosum*; *S. lycopers*; *S. lycopersicum*; *A. thaliana*; *P. patens*.

3.2.5 ASSOCIATION ANALYSIS OF PWD

PWD is assumed to play a central role in the first steps of starch degradation (Kötting et al., 2005). With the help of an association analysis, the importance and influence of this gene on different starch and chip quality traits was analyzed. For the analysis, an adequate amplicon was constructed with the primer combination LS15-LS16 (Table 5). The 823 bp amplicon included the last and largest exon 19 and no intronic region of the *PWD* sequence. This region encodes the nucleotide binding domain. Altogether, fragments from 219 individuals from the GABICHIPS population (Li et al. 2008) were amplified and sequenced. In all evaluable sequences, 29 SNPs were scored and afterwards tested for association with traits for starch and chip quality. Calculations for association were performed with the statistical test univariate GLM considering the factor origin (Li et al., 2008). Out of the 29 analyzed SNPs, only those nine SNPs are displayed in table 14 which were significant for $p < 0.05$. The positions of the 29 SNPs in the *PWD* sequence are given in figure A.7 in the Appendix.

Table 14: Overview of associated SNPs of PWD.

SNP			Trait						Features				
#	LD group	From ATG	TY effect R ² (%)	TSC effect R ² (%)	TSY effect R ² (%)	CQA effect R ² (%)	CQS4 effect R ² (%)	CQS8 effect R ² (%)	X/Y	Ex/In	AA	True allele frequency	Apparent allele frequency
1	1	10547	ns	0.004 ↑ (C) 6.5	0.020 ↑ (C) 4.4	ns	ns	ns	T/C	Ex	M/T	49.0%	92.3%
3	1	10911	ns	0.006 ↑ (T) 7.1	0.035 ↑ (T) 5.0	ns	ns	ns	C/T	Ex	-	49.3%	92.8%
4	2	10916	ns	0.047 ↓ (A) 4.5	0.041 ↓ (A) 4.2	ns	ns	ns	C/A	Ex	T/K	21.5%	64.1%
5	2	10917	ns	ns	0.028 ↓ (G) 4.6	ns	ns	ns	A/G	Ex	-	22.5%	67.0%
6	2	10923	ns	0.018 ↓ (T) 4.6	0.039 ↓ (T) 3.6	ns	ns	ns	C/T	Ex	-	21.4%	63.6%
7	3	11112	0.042 ↓ (A) 2.2	ns	ns	0.001 ↑ (A) 4.5	0.0002 ↑ (A) 7.7	0.0003 ↑ (A) 5.7	G/A	Ex	-	1.2%	4.3%
8	3	11140	0.043 ↓ (A) 2.2	ns	ns	0.007 ↑ (A) 3.4	0.003 ↑ (A) 7.3	0.001 ↑ (A) 4.9	G/A	Ex	V/I	2.2%	7.6%
2	-	10758	ns	0.036 ↓ (A) 2.1	ns	0.00003 ↓ (A) 5.6	ns	0.004 ↓ (A) 4.7	G/A	Ex	-	3.8%	0.9%
9	-	11148	ns	0.004 ↑ (A) 3.9	0.006 ↑ (A) 2.5	ns	0.032 ↑ (A) 2.3	0.012 ↑ (A) 1.8	T/A	Ex	-	1.0%	3.8%

The position of the SNP is counted from the ATG (ref.: BAC sequence), where 1 represents the adenine of the start codon. All SNPs were tested for association with the following traits: TY: tuber yield, TSC: tuber starch content, TSY: tuber starch yield, CQA: chip quality in autumn, CQS4 or CQS8 chip quality after storage at 4°C or 8°C, respectively. For each trait, the statistical test univariate GLM was applied with p<0.05. The feature part lists allele frequencies, exon or intron location and if the SNP leads to an aa exchange. All scored allele dosages are indicated. True allele frequency (876 alleles) and apparent allele frequency (219 genotypes) is given in %. The SNPs are sorted upon their classification into LD groups. R²: percent of the total variance explained; ns = not significant, ↑ positive effect, ↓ negative effect.

Out of the nine associated SNPs, seven can be classified into three LD groups. SNP 1 and 3 form LD group 1 (LD1), SNP 4, 5 and 6 form group (LD2) and SNP 7 and 8 build LD group 3 (LD3). In LD group 1, the shift from the common allele (reference sequence) to the variant allele leads to an increase of TSC and explains up to 7 % of the phenotypic variation of this trait. Additionally, the two SNPs of this group are associated with higher TSY and explain up to 5 % of the variance. The SNPs from LD2 are associated with the same traits. But in this case, the shift from the common allele to the variant one is associated with a lower amount of TSY and TSC. Although the three SNPs are highly co-segregating, there is some variation in the degree of association which resulted in no association for SNP 5 with the trait TSC with $p < 0.05$. The third LD group shows high association with all chip quality traits (CQA, CQS4 and CQS8) and moderate association with tuber yield. The presence of the variant allele "A" leads to better chip quality and to inferior TY (Figure 22). The two SNPs explain between 3.4 % and 7.7 % of the variance of chip quality traits and 2.2 % of the variance for tuber yield. Whether the presence of a second variant allele ("AAGG") has an additional effect cannot be analyzed due to only two genotypes carrying the duplex state.

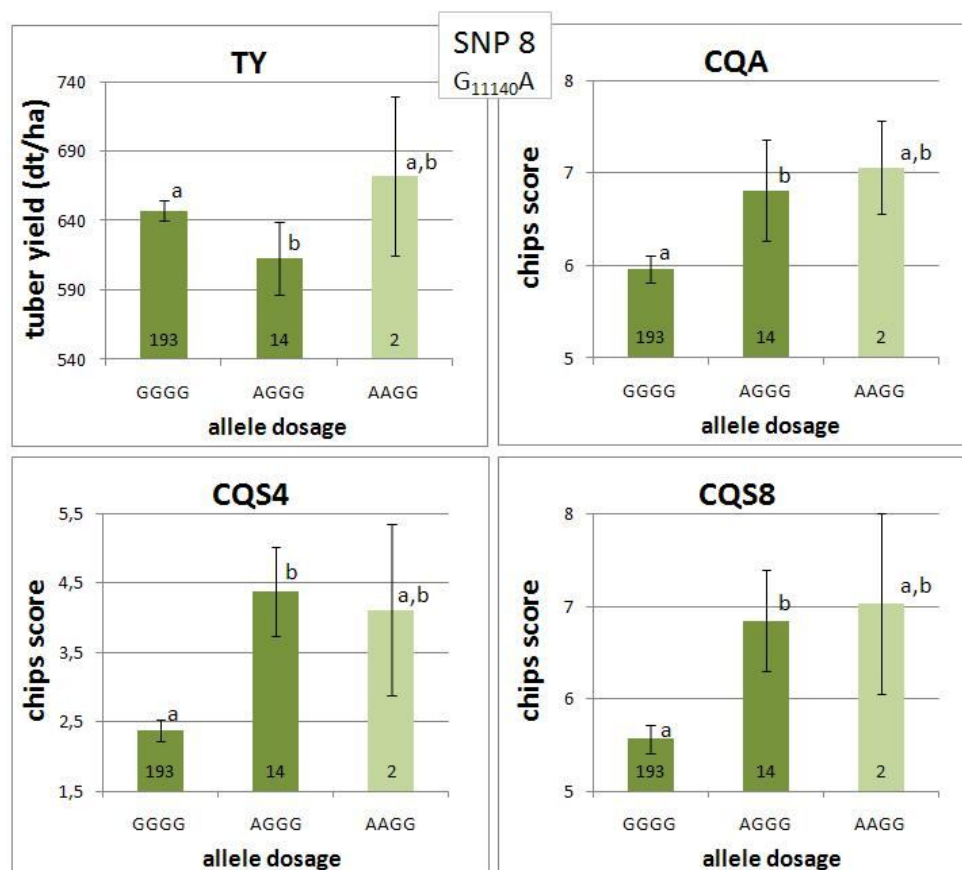


Figure 22: Means plots of SNP 8 (G₁₁₁₄₀A) of LD3 of PWD.

The X-axis represents the different allelic dosages which were scored for the SNP. The Y-axis shows the phenotypic data for the traits (TY, CQA, CQS4 and CQS8). The number given in the lower part of the bar stands for the number of genotypes bearing the allelic state. Significant differences between the different groups (allelic states) are indicated by small letters. Different letters symbolize different groups, same letters stand for the same group and two letters indicate no clear classification to one of the groups. The bars in dark green represent reliable values for the phenotypic means. Light green indicates mean values of only a small sample size, including a large error.

SNP 2 is not in LD with any of the analyzed SNPs. It is highly associated with the traits COA (0.00003) and CQS8 (0.004) and moderately associated with TSC (0.036). The presence of the variant allele ("A") leads to a significant decrease of the quality of the three traits. This SNP does not lead to an amino acid exchange and it is not very frequent in this population (3.8 %). SNP 9 is associated with the four traits TSC, TSY, CQS4 and CQS8. The change to the variant allele ("A") has for all 4 traits a positive effect. The SNP frequency is low (3.8 %) and explains between 2 % and 4 % of the phenotypic variation of the traits.

3.2.6 PWD HAPLOTYPE MODELING

All 29 SNPs scored in the PWD amplicon were grouped into LD groups. In total, 22 of the scored SNPs were classified into at least 8 groups, the remaining 7 SNPs did not co-segregate and formed single groups. From each LD group, one SNP was chosen for the calculation of haplotypes. Additionally, two associated SNPs and one non-associated SNP from the single groups were included. The 11 SNPs from this selection (SNPs 10547, 10629, 10689, 10758, 10761, 10815, 10923, 10932, 10962, 11140, and 11148) were used for the calculation with SATlotyper. The software calculated 17 possible haplotypes which are sorted according to their frequency in table 15. The 10 most frequent haplotype models are listed. The remaining low frequent haplotypes are listed in table B.7 of the appendix.

Table 15: List of *StP* PWD haplotype models based on 12 SNPs scored in 218 genotypes.

No	Haplotype Sequence	True haplotype frequency % (no)	Apparent haplotype frequency % (no)	Allelic states n/s/d/t/q
16	TTCGGGCGCGT	39.6 (331)	73.2 (153)	56/39/61/42/11
2	CCCGGACGTGT	14.1 (118)	41.6 (87)	122/63/18/5/1
12	CTCGGGTGTGT	10.2 (85)	33.5 (70)	139/58/9/3/0
5	CCGGGGTGTGT	7.9 (66)	26.8 (56)	153/48/6/2/0
10	CTCGGGCGTGT	5.4 (45)	20.1 (42)	167/39/3/0/0
15	TCCGGGCGCGT	5.1 (43)	18.2 (38)	171/33/5/0/0
14	TCCGGACGCGT	3.2 (27)	12.9 (27)	182/27/0/0/0
8	CTCGGACGTGT	3.0 (25)	10.5 (22)	187/19/3/0/0
17	TTCGGGCGTGT	2.9 (24)	9.1 (19)	191/16/2/0/1
3	CCCGGGCTTGT	1.8 (15)	6.7 (14)	195/13/1/0/0

The haplotype which is identical to the BAC sequence and to the PGSC sequence, is indicated in grey. Red letter: associated reference SNP; Blue letter: associated SNP with positive effect. True allele frequency: Number of haplotypes among possible 836 haplotypes (4 x 209 genotypes); Apparent allele frequency: Number of genotypes containing the haplotype. Nulliplex (n), simplex (s), duplex (d), triplex (t) and quadruplex (q): Allelic states of haplotypes per genotype.

In general, the most frequent haplotypes are also the most common ones regarding the frequency per genotype: There are many genotypes containing the haplotypes only in simplex or duplex state. The most frequent haplotype (number 16) shows a frequency of 39.6 % and is identical to the haplotype of the *PWD* BAC sequence and the published *S. phureja* sequence (PGSC, 2011). It is present in all allelic states from simplex to quadruplex in around 73 % of the genotypes of the whole population. For the remaining haplotypes, the existence has not been validated by now.

3.2.7 STRUCTURAL COMPARISON OF *GWD* AND *PWD*

For the calculation of the similarity between potato *GWD* and *PWD*, both amino acid sequences were analyzed with the software Needle. The sequences show an identity of around 19 % and a similarity of 32 % (Table 16). Further alignments were performed with ClustalX and GeneDOC software. The obtained alignment showed very few conserved regions (Appendix Figure A.11). The region containing the catalytic histidine and the nucleotide binding domain (C-terminus) are assumed to have similar functions in both genes and show a slightly higher similarity. Interestingly, by splitting the sequences in to two parts, one containing the N-terminus and the other one containing the C-terminus, the high conservation for the catalytic histidine and the nucleotide binding domain and the low conservation for the starch binding domains become more apparent. The level of identity for the C-terminus increases to 27 %, whereas the N-terminal part shows only 11.4 % of identity. Additionally, the genomic and cDNA sequences were compared. They show around 41 % and 43 % of identity.

Table 16: Comparison of *StGWD* and *StPWD*.

	cDNA length in bp	Identity	gDNA length in bp	Identity	No. of exons	Length of aa sequence	Identity (Similarity)
StGWD	4395	43.4 %	15664	41.2 %	33	1465	18.9 %
StPWD	3609		11378		19	1202	(31.7 %)

All obtained sequences (cDNA, genomic DNA and amino acid sequences) were compared according to their length and identity level.

3.3 *INVERTASE 6* - GENOMIC ORGANIZATION

Invertases play an important role in the conversion of sucrose into the active and usable monosaccharide glucose and fructose. *In silico* research resulted in the detection a new cDNA sequence of a putative 6th invertase (*Inv6*; G. Orjeda, personal communication). Based on this single sequence, further research and investigations were started with detailed *in silico* analysis. Blast analyses of the cDNA sequence (T12217) on the released *S. phureja* genome identified several putative superscaffolds. The obtained five best hits are listed in table 17. The first hit shows very high identity (97.7%) to the blasted cDNA sequence and has an e-value of 0. This superscaffold is located on chromosome 8. The remaining hits were identified to be the vacuolar invertase (*Pain-1*) with 70.3% of identity and several cell wall bound invertases (*InvGE/InvGF* and *InvCD111/InvCD141*). The cell wall bound invertases *InvGE* and *InvGF* corresponded to the identified superscaffold on chromosome 9 and *InvCD111* and *InvCD141* belong to the superscaffold on chromosome 10. A not further characterized β -fructofuranosidase was found on chromosome 6. All of them displayed a rather low level of identity (59 %) to the putative 6th invertase.

Table 17: Best hits of *Inv6* blasted against the *S. phureja* genome (PGSC, 2011).

Hit	Name	Identity %	e-value	Annotated gene	Chromosome
1	PGSC0003DMB000000201	97.7	0	"invertase"	8
2	PGSC0003DMB000000605	70.3	2.5e ⁻¹⁰⁹	<i>Pain-1</i>	3
3	PGSC0003DMB000000021	59.9	1.3e ⁻⁵⁴	<i>InvGE/InvGF</i>	9
4	PGSC0003DMB000000008	59.3	9.3e ⁻⁴⁷	<i>InvCD111/InvCD141</i>	10
5	PGSC0003DMB000000222	59.7	1.5e ⁻³²	β -fructofuranosidase, cell-wall isozyme	6

Due to the high degree of identity to the query sequence, the superscaffold 201 was selected for further analysis. Within the sequence of this superscaffold, several known markers from the genetic map of chromosome 8 of the PoMaMo database (<http://www.gabipd.org/projects/Pomamo/>) were identified via homology analysis. Four different STS (sequence-tagged site) markers at the position of around 60 cM were identified in the superscaffold sequence and indicated the location of the superscaffold in the lower distal part of a starch QTL (Appendix Figure A.3). Consequently, the genomic sequence of the putative 6th invertase (*Inv6*) was deduced from the superscaffold (Appendix C.5). The alignment in Spidey of the retrieved genomic sequence with the cDNA sequence resulted in the genomic organization of *Inv6* (Figure 23). The gene consists of 7 exons and 6 introns, whereas the second exon is a very small exon of only 9 bp (Appendix B.3). This small exon is highly conserved in all invertases of plant origin (Ramloch-Lorenz et al., 1993).

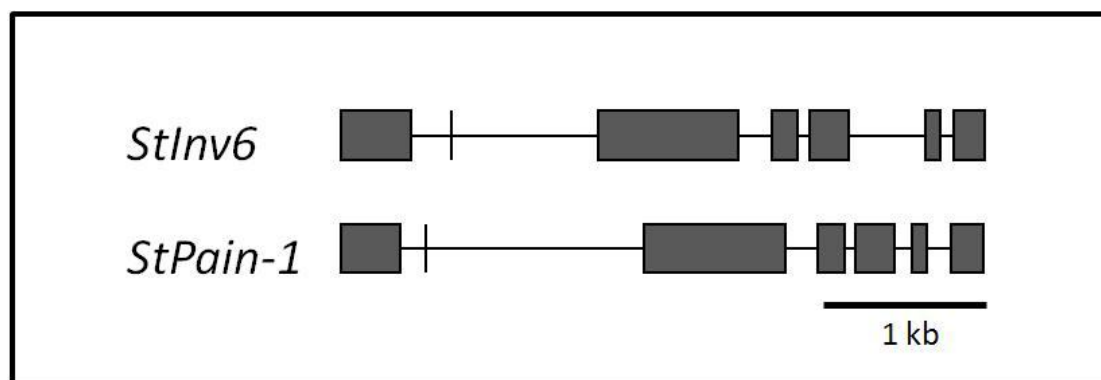


Figure 23: Genomic organization of *Inv6* and *Pain-1* of *S. tuberosum*.

The exon-intron structures of the two most similar invertase sequences obtained from the blast are displayed.

3.3.1 STRUCTURAL COMPARISON OF *INV6* AND *PAIN-1* OF *S. TUBEROSUM*

The highest level of identity of *Inv6* was identified for the vacuolar invertase *Pain-1* (70%, Table 17). A more detailed comparison between both invertases was performed. The results of the sequence comparison are listed in table 18. Both genes do not vary much in the length of the cDNA and amino acid sequences, but the identity is only 60 %. At the genomic sequence level, both genes share almost the same length and again a low sequence identity of 49.4 %.

Table 18: Comparison of *Inv6* and *Pain-1* from *S. tuberosum* at different sequence level.

	cDNA	Identity %	Genomic DNA	Identity %	AA	Identity % Similarity %
<i>Inv6</i>	1962	61.1	3965	49.4	653	56.7
<i>Pain-1</i>	1920		3951		639	71.3

For the comparison of the cDNA and genomic sequence, the identity of the sequences is stated in percent. The comparison of the amino acid sequences additionally includes the degree of similarity in percent between the different amino acid residues.

Some differences can be observed regarding the lengths of several introns (intron 2 and 5, Figure 23). Additionally, exon 1 is around 60 bp longer in *Inv6* compared to *Pain-1* and is the main reason for the difference of around 20 amino acids in the length of the peptide sequence. The remaining exons are highly conserved. The alignment of the amino acid sequence was done by using ClustalX and confirmed a high conservation between both genes on amino acid level (Figure 24). Related to the apparent differences in exon 1, the lowest conservation between both amino acid sequences matches to this N-terminal region (blue box). Analysis of this domain with SignalP (<http://www.cbs.dtu.dk/services/SignalP/>) resulted in similar outputs for both genes and indicated that they may contain a cleavage site for a signal peptide. For *Pain-1*, a putative cleavage site was predicted to be located between amino acid 51 and 52, whereas for *Inv6* the cleavage site was predicted to be downstream of the position of amino acid 59. Further evidence for the presence of a target signal was obtained by using the transmembrane prediction tool TMHMM

(<http://www.cbs.dtu.dk/services/TMHMM/>). Both N-terminal sequences were predicted to contain a transmembrane helix which indicated transport trough or at least a location within a membrane. No information about the type of membrane was predictable. Additional localization programs were used for further predictions (MultiLoc2 and pSort) and resulted in similar ambiguous localizations for both genes either in the vacuole or in the cytoplasm. Analysis of the protein sequence of *Inv6* and *Pain-1* by using the Pfam database revealed similar functional domains. The typical β -fructosidase motif (NDPNG, Sturm and Chrisples, 1990) and the catalytic domain (WECVD, Martin et al., 1987) were clearly identified in *Inv6* and reside in the N-terminal glycosyl hydrolase family 32 domain (light green box, Figure 24). This large domain forms a five bladed beta propeller (Alberto et al., 2004) and shows a high homology to the *Pain-1* sequence which posses the same domain. The C-terminal part of this glycosyl hydrolase domain (dark green box) shows also high conservation to the *Pain-1* sequence. Both domains are separated by a less conserved part (pos 460-530) with no characterized clear function.

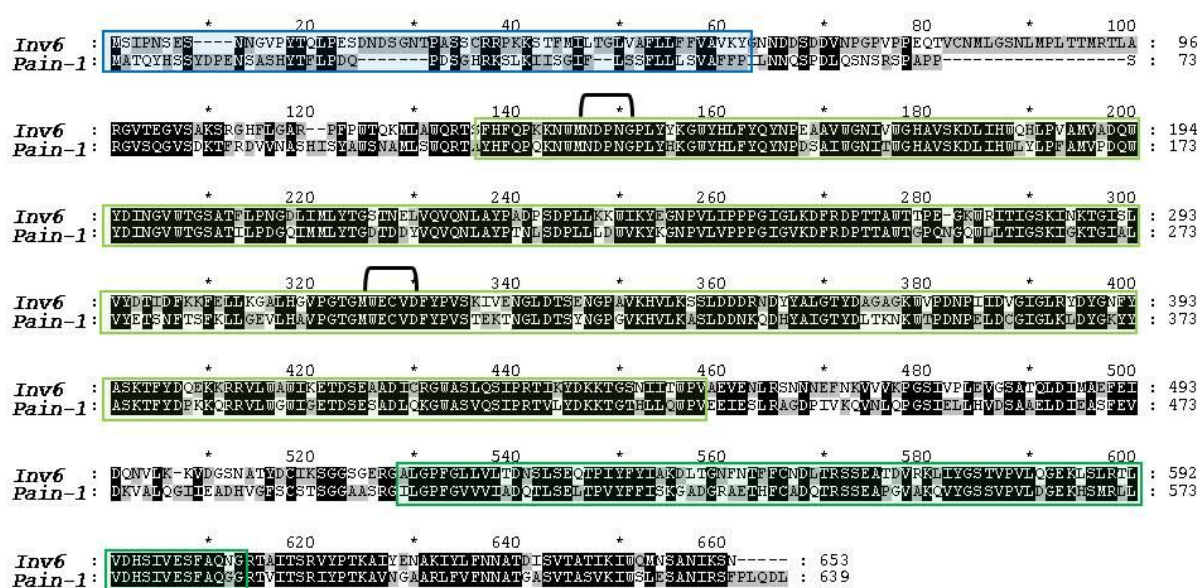


Figure 24: Amino acid sequence alignment of *Inv6* and *Pain-1* from *S. tuberosum*.

Characteristic domains with localization or catalytic function are indicated by coloured boxes. Blue box: putative signal peptide; light green box: N-terminal glycosyl hydrolase family 32 domain; dark green box: C-terminal glycosyl hydrolase domain. Important motifs are highlighted by brackets.

3.3.2 CONSERVATION OF *Inv6* AMONG DIFFERENT SPECIES

The 6th invertase has not been analyzed in other species so far. Only the sequence for the orthologue *Lin9* of *S. lycopersicum* has been annotated (Accession: AM050394). Blast analysis against the NCBI database resulted in several full-length and partial sequences from different species. Alignments and phylogenetic tree construction of this amino acid sequences were carried out using ClustalX. Sequences which showed higher similarity to *Pain-1*, were excluded in the following alignment (Figure 25). The remaining four full-length sequences belong to the Solanaceae (*S. tuberosum* and *S. lycopersicum*), the Rutaceae (*Citrus sinensis*) and the Cucurbitaceae (*Cucumis melo*). Additionally, a partial sequence of *Nicotiana tabacum* was added to the alignment.

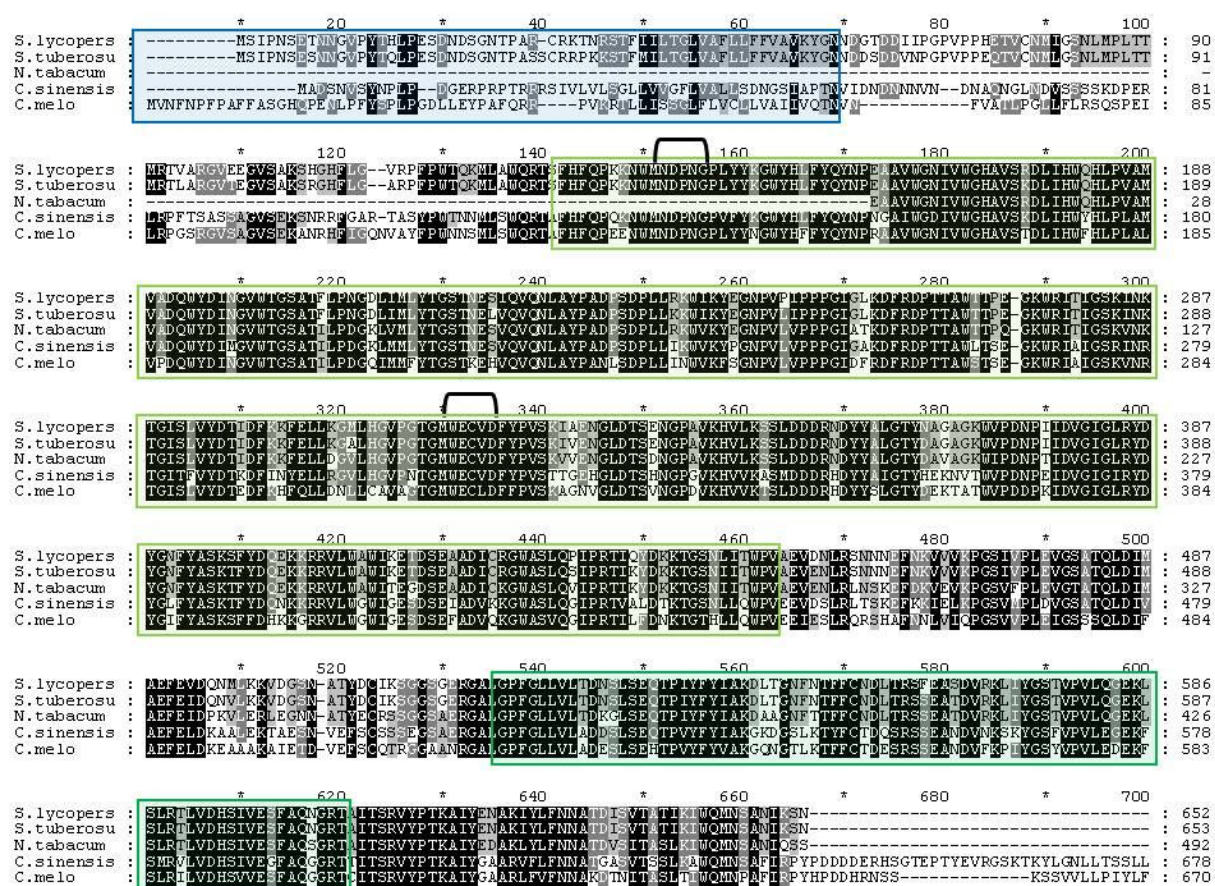


Figure 25: Alignment of the deduced amino acid sequences of different *Inv6* orthologues.

Amino acid residues identical in all sequences are highlighted by a black background, those conserved in most of the sequences are marked by a gray background. Conserved and annotated domains are indicated by coloured boxes. Blue: putative signal peptide; light green: N-terminal glycosyl hydrolase family 32 domain; dark green: C-terminal glycosyl hydrolase domain. Important and conserved motifs are highlighted by brackets. *S. lycopersicum* (AM050394); *S. tuberosum* (T12217); *N. tabacum* (HM022269); *C. sinensis* (AB276107); *C. melo* (FJ653659).

Disregarding the N-terminal part, the sequences show very high conservation. The two parts of the glycosyl hydrolase domains including the two catalytic domains form distinct blocks (light and dark green boxes, black brackets). The five amino acid sequences belong to the two phylogenetic clades asterids (family: Solanaceae) and rosids (families: Rutaceae and Cucurbitaceae). Within their clades, they show an even higher conservation (Figure 26).

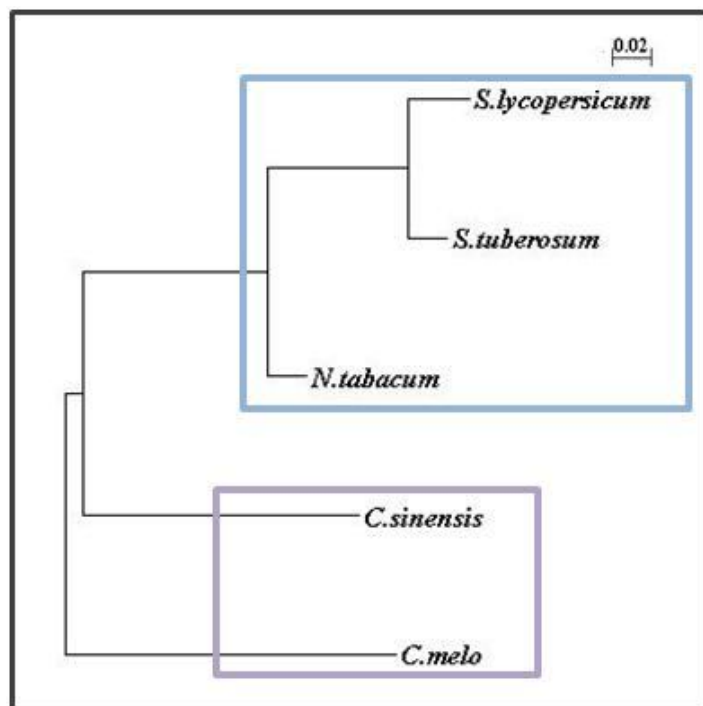


Figure 26: Phylogenetic tree of *Inv6* protein sequences retrieved from diverse blast analyses.
The different phylogenetic classes and groups are highlighted by coloured boxes. Violet: Rosids; Blue: Asterids.

3.3.3 ASSOCIATION ANALYSIS OF *Inv6* IN THE GABICHIPS POPULATION

Invertase 6 is a new functional candidate in the complex pathway of sugar metabolism. There is no knowledge about the role and location of this gene. Additionally, initial *in silico* mapping analysis located this gene in the distal part of a starch QTL and close to a sugar QTL on chromosome 8 (Appendix Figure A.3). To investigate whether the invertase contributes to the natural variation of starch and chip quality traits, an association analysis was performed. A gene-specific primer pair was designed on the sequences of exon 3 and exon 5 (LS183-LS173, MM Table 5). Of the amplified 1100bp fragment, around 500 bp of the sequence up to the first indel were used for the SNP analysis. In total, 17 SNPs and 1 indel (Appendix A.8) were analyzed in 210 individuals of the GABICHIPS population (Li et al., 2008). The calculated average SNP rate is around 1SNP/26bp of sequence. Calculations for association with starch and chip quality traits were performed with the statistical test univariate GLM considering the factor origin (Li et al, 2008). Of the 18 analyzed SNPs, only those 10 SNPS are shown in table 19 which were significant for $p < 0.05$.

Table 19: Overview of associated SNPs of *Inv6*.

SNP			Trait						Features				
#	LD group	From ATG	TY effect R ² (%)	TSC effect R ² (%)	TSY effect R ² (%)	CQA effect R ² (%)	CQS4 effect R ² (%)	CQS8 effect R ² (%)	X/Y	Ex/In	AA	True allele frequency	Apparent allele frequency
1	1	2065	ns	0.007 ↑ (A) 6.4	0.020 ↑ (A) 5.1	0.0001 ↑ (A) 7.1	0.005 ↑ (A) 6.8	0.002 ↑ (A) 5.8	G/A	Ex	-	51.6%	91.6%
2	1	2076	ns	0.0004 ↑ (T) 9.2	0.001 ↑ (T) 8.0	0.00006 ↑ (T) 7.8	0.002 ↑ (T) 7.7	0.0002 ↑ (T) 7.3	C/T	Ex	A / V	62.4%	94.7%
3	1	2272	ns	0.00005 ↑ (T) 11.0	0.004 ↑ (T) 6.8	0.000001 ↑ (T) 10.4	0.0002 ↑ (T) 9.7	0.00002 ↑ (T) 8.9	A/T	Ex	-	66.2%	95.2%
4	2	2116	ns	ns	ns	0.008 ↑ (A) 4.6	0.011 ↑ (A) 6.0	ns	G/A	Ex	-	35.7%	80.1%
8	2	2302	ns	ns	ns	0.027 ↑ (A) 3.7	0.034 ↑ (A) 4.8	ns	G/A	Ex	-	53.2%	80.5%
6	3	2255	0.034 ↓ (A) 2.9	0.00023 ↑ (A) 7.6	ns	0.024 ↑ (A) 2.5	0.013 ↑ (A) 4.0	0.049 ↑ (A) 2.2	G/A	Ex	A / T	13.6%	49.5%
9	3	2320	0.039 ↓ (C) 2.8	0.0003 ↑ (C) 7.2	ns	0.041 ↑ (C) 2.1	0.038 ↑ (C) 3.0	ns	T/C	Ex	-	13.0%	48.1%
4	-	2134	0.023 ↓ (G) 4.8	0.050 ↑ (G) 4.4	0.016 ↑ (G) 5.3	0.003 ↑ (G) 5.3	0.013 ↑ (G) 5.8	0.029 ↑ (G) 4.2	T/G	Ex	-	41.7%	85.3%
5	-	2182	0.027 ↓ (A) 3.9	0.00014 ↑ (A) 9.2	ns	0.003 ↑ (A) 4.6	0.026 ↑ (A) 4.3	0.01 ↑ (A) 4.0	G/A	Ex	-	16.2%	58.1%
10	-	2369	ns	ns	ns	0.020 ↑ (A) 3.3	ns	0.018 ↑ (A) 3.6	G/A	Ex	V / I	7.4%	21.5%

The position of the SNP is counted from the ATG (ref.: PGSC sequence). Traits: TY: tuber yield, TSC/TSY: tuber starch content/yield, CQA: chip quality in autumn, CQS4/CQS8 chip quality after storage at 4°C or 8°C. For each trait, the statistical test univariate GLM was applied ($p < 0.05$). The feature part lists allele frequencies, exon or intron location and if the SNP leads to an aa exchange. All scored allele dosages are indicated. True allele frequency (840 alleles) and apparent allele frequency (210 genotypes) is given in %. The SNPs are sorted upon their classification into LD groups. R²: percent of the total variance explained; ns = not significant, ↑ positive effect, ↓ negative effect.

Most of associated SNPs can be grouped into linkage disequilibrium groups as they are associated with similar traits and show analogous allele dosages. LD group 1 (LD1) consists of SNP 1, 2 and 7. They are highly associated with all tested starch and chip quality traits except the trait TY. The highest association was identified for CQA with p-values up to $1e^{-6}$. The SNPs of LD1 explain up to 10.4% of the phenotypic variance for this trait. The degree of association with chip quality decreases corresponding to increasing cold-storage treatment (CQS8 and CQS4), but it is highly associated, and 5.8 % to 9.7 % of the phenotypic variance can be explained. The shift from the reference allele (e.g. for SNP 2: "C") to the variant allele ("T") results in higher yield and better chip quality. The means plots for SNP C₂₀₇₆T show the higher chip quality in autumn (CQA) and after storage at 4 or 8°C (CQS4/CQS8) resulting from the presence and dosage of the variant allele "T" (Figure 27).

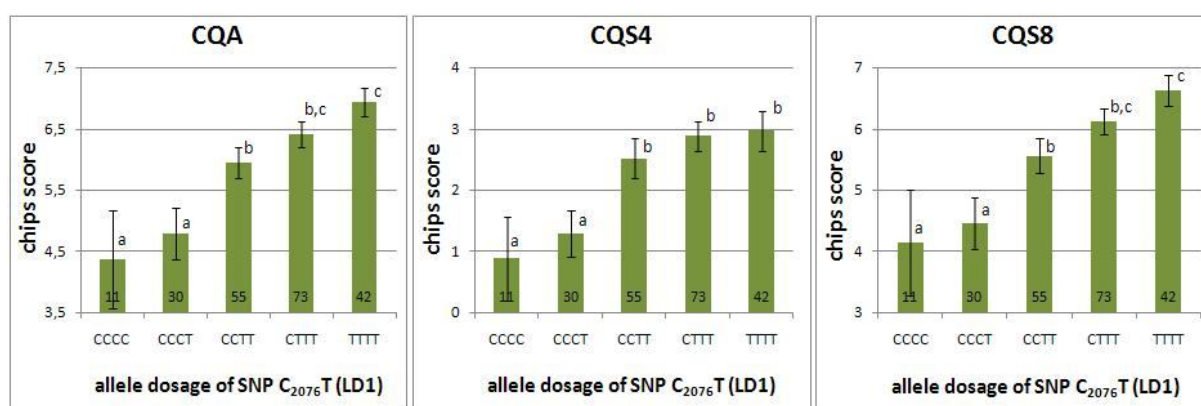


Figure 27: Means plots of SNP 2 (C₂₀₇₆T) of LD1 of *Inv6*.

The X-axis represents the different allelic dosages which were scored for the SNP. The Y-axis shows the phenotypic data for the traits (CQA, CQS4 and CQS8). The number given in the lower part of the bar stands for the number of genotypes. Significant differences between the different groups (allelic states) are indicated by small letters. Different letters symbolize different groups, same letters stand for the same group and two letters indicate no clear classification to one of the groups.

Besides chip quality, the degree of association of LD1 is as well very high with TSC and up to 11 % of the variance of this trait can be explained. The true allele frequency of LD1 is around 50-60 %. There are only few genotypes without this superior allele (10 %). The second LD group 2 (LD2) consists of the two SNPs 3 and 8. Both SNPs are moderately associated with chip quality traits and explain between 3.7 and 6.0 % of the phenotypic variance. LD group 3 consists of the two SNPs 6 and 9. They show moderate association with chip quality traits and high association with tuber starch yield. Additionally, they are slightly associated with tuber yield. The variant allele leads to a decrease. SNP 4 cannot be grouped into an obvious linkage disequilibrium group. It is moderately associated with all traits and explains in average 5 % of all phenotypic variations. A further interesting SNP is SNP 5 which is associated with all chip quality traits, tuber yield and tuber starch content. Similar to previous results, the presence of the variant allele leads to increased chip quality and starch content, but the tuber yield decreases. The last SNP 10 is a very low frequent allele and is slightly associated with chip quality traits (CQA and CQS8).

3.3.4 ANALYSIS OF THE DISTRIBUTION OF ASSOCIATED *Inv6* ALLELES IN THE BIOSOL POPULATION

The analyzed SNPs of *Inv6* show high association in the GABICHIPS population with chip quality traits at harvest time and with cold-storage treatment. Although the functional role of *Inv6* has not been characterized yet, it can be assumed that it is involved in the cleavage of sucrose into the reducing sugars glucose and fructose. To analyze whether *Inv6* is directly associated with the amount of reducing sugars, a further association analysis in a second population (BIOSOL) was performed. The population consists of 40 individuals which were pre-selected according to their chip quality. As phenotypic data and basis for the grouping into better and worse chipping genotypes, the amount of reducing sugars (glucose and fructose) at time point 0 (TP0; 0 weeks of cold-storage) was measured and used as phenotypic data for the association (M. Fischer, unpublished data). Figure 28 shows the amount of basic sugar level at TP0. Exemplarily, the three highly associated SNPs 2076 (LD1), 2116 (LD2) and 2134 are shown corresponding to their scored allele dosage. All of the SNPs are very frequent and present in good and bad chipping genotypes. No significant affect on low or high sugar content can be calculated upon the presence or absence of the highly associated SNPs. However, this data confirms the previously identified high frequency of these alleles in potato cultivars.

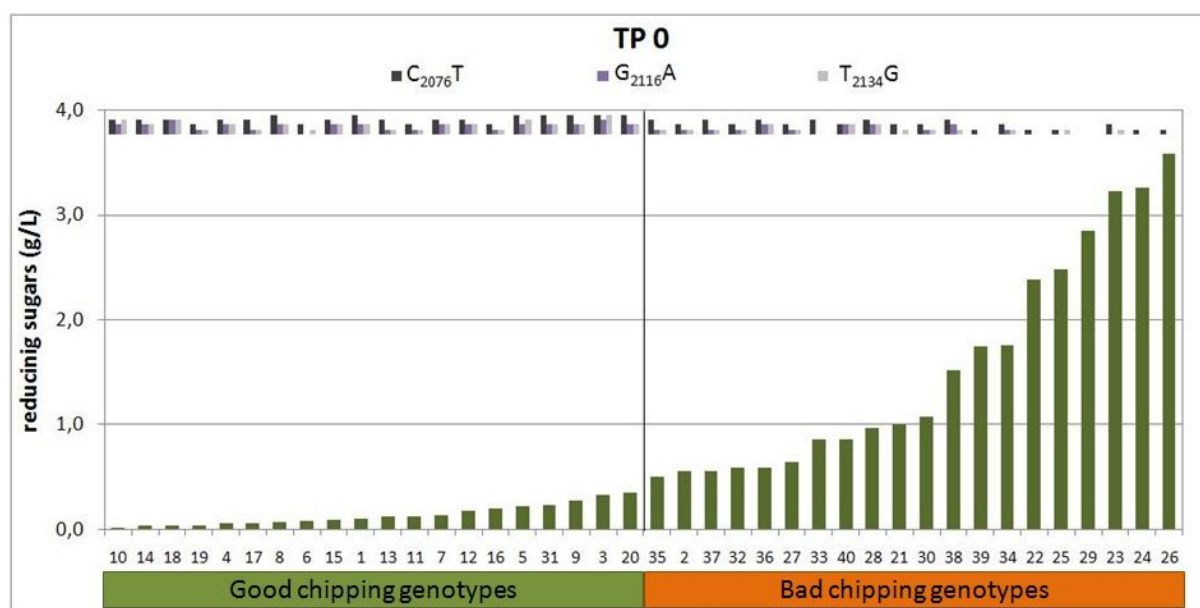


Figure 28: Presence of three associated *Inv6* SNPs in potato tubers without cold-storage.

The amount of reducing sugars (glucose and fructose) was measured in 40 different genotypes of the BIOSOL population at TP0 (without cold-storage). The genotypes are sorted by increasing total content of reducing sugars. The 20 genotypes with the lowest amount of reducing sugars are referred to the “good” chipping genotypes, whereas the 20 genotypes with the highest amount are grouped to the “bad” chipping genotypes.

For TP4 (4 weeks of cold-storage at 4°C), the amount of reducing sugars is the measured value of this time point subtracted by the value for TP0 to focus on the accumulation of reducing sugars upon cold-storage and to neglect the basic level of reducing sugars at TP0. The genotypes are sorted according to the reducing sugar content and classified into the two groups of good and bad chipping

genotypes (Figure 29). The same three SNPs were displayed as for TP0. The distribution did not change significantly. SNP 2076 can be found in 39 genotypes (20 good ones and 19 bad ones). Although SNP 2116 and SNP 2134 were identified with a lower frequency in the group of bad chipping genotypes, the differences are not significant.

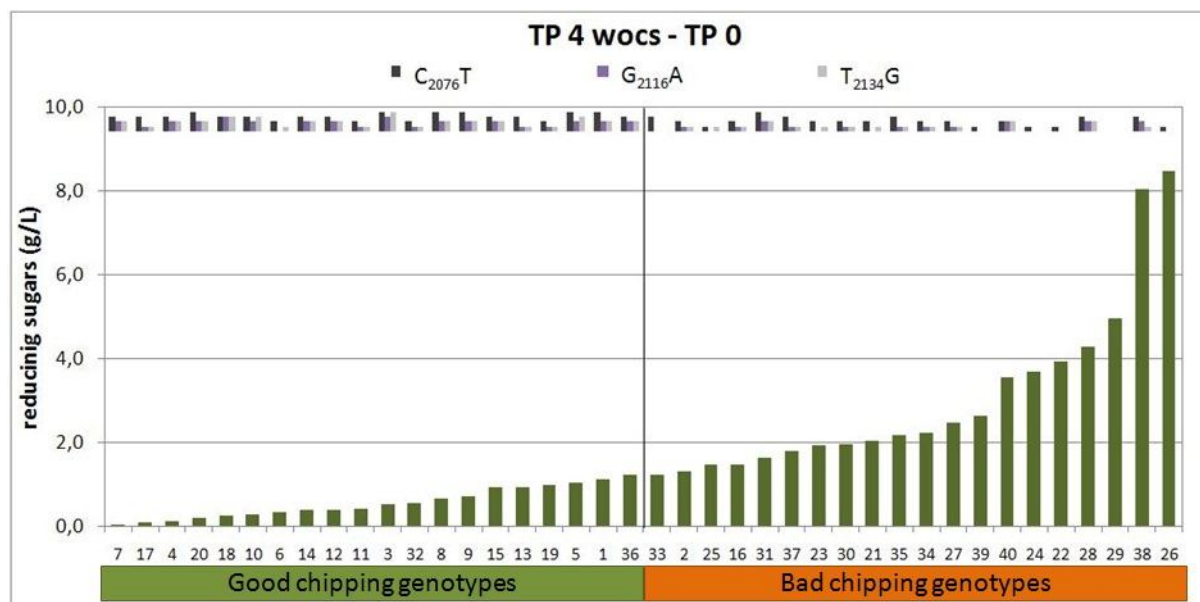


Figure 29: Presence of three associated *Inv6* SNPs in potato tubers after 4 weeks of cold-storage.

The amount of reducing sugars (glucose and fructose) of 40 different genotypes was measured in 40 different genotypes of the BIOSOL population at TP4 (4 weeks of cold-storage). The genotypes are sorted by increasing total content of reducing sugars. The 20 genotypes with the lowest amount of reducing sugars are referred to the “good” chipping genotypes, whereas the 20 genotypes with the highest amount are grouped to the “bad” chipping genotypes.

Due to non-detectable significant differences in the allele frequency (presence) between the group of good and bad chipping genotypes, the allele dosage was analyzed. Comparisons between the sum of good alleles in all good chipping cultivars and the sum of good alleles in the bad chipping cultivars for all SNPs (all associated SNPs in the GABICHIPS population, previous section 3.3.3) show that the superior alleles occur significantly more often in the group of good genotypes (Figure 30A). The significantly lower amount of reducing sugars in the good genotypes is in agreement with the presence of the associated SNPs in high allelic dosages (Figure 30B). Strong differences in the number of superior alleles can be observed for the very highly associated SNPs, whereas moderately associated SNPs (2182, 2255, 2302, 2320, and 2369) did not show differences (tested by χ^2 test for the goodness of fit). The same ratios between good and bad chipping genotypes can be observed for all remaining time points. In general, there is no large variance in the allelic dosages of all tested SNPs and therefore no effect on the increasing reducing sugar content upon cold-storage. Containing a high allele dosage of the associated SNPs increases the probability of belonging to the class of good genotypes.

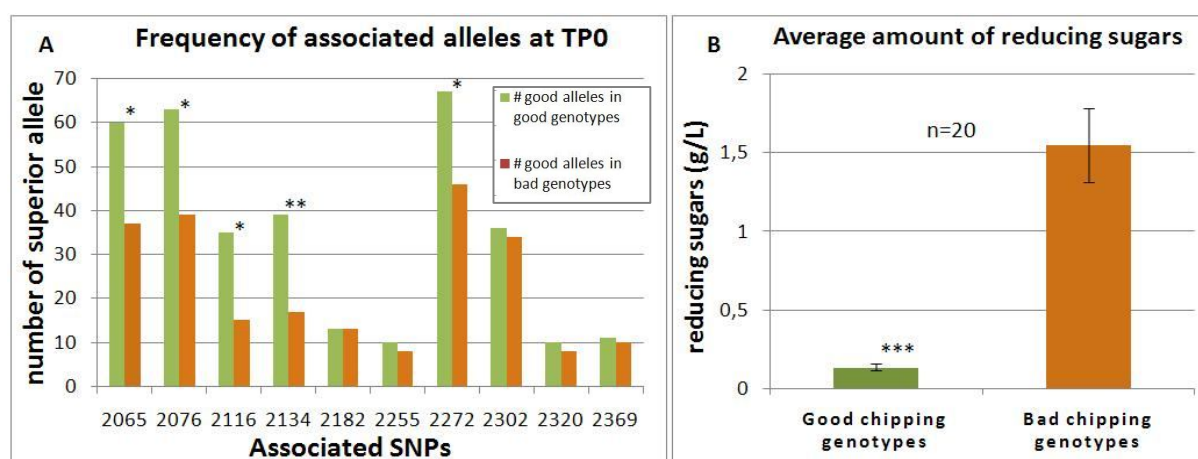


Figure 30: Effect of the allelic dosage on low and high amount of reducing sugars.

A: Sum of good and bad alleles of associated SNPs at TPO. **B:** Average amount of reducing sugars in good and bad chipping genotypes. A higher allele dosage of the good allele results in a lower sugar content.

3.3.5 *Inv6* HAPLOTYPE MODELING AND ASSOCIATION

All 19 SNPs scored in the *Inv6* amplicon of the association analysis in the GABICHIPS population were grouped into LD groups according to co-segregation. In total, 15 of the scored SNPs could be classified into at least 5 groups, the remaining 4 SNPs did not co-segregate and formed single groups. From each LD group, one representative SNP was chosen for the calculation of haplotypes. Additionally, all SNPs from the single groups were included. The resulting 9 SNPs from this selection were used for the calculation with SATlotyper (SNPs 2116, 2134, 2182, 2197, 2209, 2255, 2272, 2369 and 2371). The top 10 of the haplotype models are listed in table 20. The 10 remaining low frequent haplotypes are given in table B.8 of the appendix.

Table 20: List of *Inv6* haplotype models.

No	Haplotype Sequence	True haplotype frequency % (no)	Apparent haplotype frequency % (no)	Allelic states n/s/d/t/q
18	GT GG C GAGT	24.6 (206)	59.8 (125)	85/72/32/14/7
5	AGGG C GTGT	24.2 (202)	55.5 (116)	94/53/43/17/3
20	GTGG C GTGT	9.4 (79)	34.9 (73)	137/67/6/0/0
7	GGAG C ATGT	5.4 (45)	21.1 (44)	166/43/1/0/0
10	GGGG C GTGT	5.4 (45)	21.1 (44)	166/43/1/0/0
6	ATGG C GTGT	5.1 (43)	19.1 (40)	170/39/0/0/1
13	GTAG C TGA	3.6 (30)	12.9 (27)	183/24/3/0/0
12	GTAG C GAGT	3.3 (28)	13.4 (28)	182/28/0/0/0
2	AGGG C ATGT	3.0 (25)	11.5 (24)	185/23/2/0/0
11	GTAG C ATGT	3.0 (25)	12.0 (25)	185/25/0/0/0

The haplotype which is identical to the *S. phureja Inv6* sequence is highlighted in grey. Blue letter: associated reference SNP. Green letter: associated allele with positive effect.

In general, the haplotypes with the highest true frequency are also the most apparent haplotypes: They are present in most of the genotypes and occur in all allelic states. There are many genotypes containing the low frequent haplotypes only in simplex or duplex state. The most frequent haplotype (no 18) shows a true allele frequency of 24.6 %. It is identical to the reference *Inv6* sequence of *S. phureja* which was used for the association analysis and does not contain any of the positively associated SNPs. It is present in all allelic states from simplex to quadruplex in around 60 % of the genotypes of the whole population. Interestingly, haplotype 20 differs only in one position compared to haplotype 18. This haplotype is less frequent and is only present in the two allelic states simplex and duplex. For the remaining haplotypes, no evidence is available at present which confirms their existence. There are two haplotypes which contain 4 of the 6 positively associated SNPs (no 2 and 7) and would suggest being a desirable haplotype regarding to a positive effect on chip quality traits. No haplotype models were calculated which contained all positively associated SNPs.

In order to test haplotypes for association, additional haplotype calculations were performed including only one LD group containing three associated SNPs (LD1, Table 19). The five haplotype models are listed in table 21. The reference sequence (*Inv6* in *S. phureja*) corresponds to haplotype number 4 and contains all basic and inferior alleles (GCA). Interestingly, it is highly negatively associated with chip and starch quality traits. The most frequent haplotype model is ATT. This haplotype contains all positive associated SNPs, but the analysis did not reveal any association. Associations of haplotypes GTT and ACA in a positive way with TSC and TSY lead to the assumption that these three tested SNPs do not form a true LD group.

Table 21: List of haplotype models for LD1 of *Inv6*.

Haplotype No	Sequence	True haplotype frequency % (no)	Apparent haplotype frequency % (no)	Allelic states n/s/d/t/q	Association p-value	effect	R ²
3	ATT	34.3 (287)	66.0 (138)	85/72/32/14/7	ns		
5	GTT	28.3 (237)	62.2 (130)	94/53/43/17/3	TSC 0.001 TSY 0.002	↑ 7.3 % ↑ 6.7%	
4	GCA	19.5 (163)	35.4 (74)	137/67/6/0/0	TSC 0.0003 CQA 0.00001 CQS4 0.001 CQS8 0.0002	↓ 9.5 % ↓ 8.8 % ↓ 9.1 % ↓ 7.4 %	
1	ACA	14.2 (119)	51.2 (107)	166/43/1/0/0	TSC 0.003 TSY 0.008	↑ 5.3 % ↑ 3.8 %	
2	ACT	3.6 (30)	14.4 (30)	166/43/1/0/0	ns		

The haplotype which fits to the reference sequence (basic alleles), is indicated in grey. Blue letter: associated reference SNP; Green letter: associated SNP with positive effect. True allele frequency: Number of haplotypes among possible 840 haplotypes (4 x 210 genotypes); Apparent allele frequency: Number of genotypes containing the haplotype. Nulliplex (n), simplex (s), duplex (d), triplex (t) and quadruplex (q): Allelic states of haplotypes per genotype. Positive effect: ↑, negative effect: ↓.

Additional analysis of this LD group excluding one of the SNPs each revealed that only SNP 2076 and 2272 are truly co-segregating. The association analysis confirmed a negative association with the tested traits, if the haplotype combination CA (inferior allele of SNP 2076 and 2272) is present. A positive association is calculated for the haplotype combination TT (superior alleles, Table 22). Interestingly, these haplotype models are associated with the same traits as the single SNPs (Table 19) and the effect on the phenotypic variance was slightly higher.

Table 22: List of haplotype models for two SNPs of LD1 of *Inv6*.

No	Haplotype Sequence	Allelic states n/s/d/t/q	Association				
			TSC	TSY	COA	CQS4	CQS8
A	CA	51/81/46/22/10	0.0001 ↓ 11.1 %	0.004 ↓ 6.8 %	0.0001 ↓ 10.4 %	0.0004 ↓ 9.7 %	0.0002 ↓ 9.0 %
B	TT	13/30/55/72/40	0.001 ↑ 7.3 %	0.002 ↑ 6.7 %	0.0001 ↑ 7.4 %	0.002 ↑ 7.7 %	0.001 ↑ 6.9 %
C	CT	177/33/0/0/0	ns				

The haplotype which fits to the reference sequence (basic alleles), is indicated in grey. Blue letter: associated reference SNP; Green letter: associated SNP with positive effect.

3.3.6 EXPRESSION ANALYSIS OF *INV6*

A semi-quantitative expression analysis of *Inv6* in different potato tissues was performed. A primer pair (LS183-LS179; Table5) was specifically designed for exon 3 of *Inv6* and amplified a 500 bp fragment. Expression was detected in all aerial tissues tested except for old and young leaves (Figure 31). The strongest expression was identified in flower buds. The expression level in flowers, petals and anthers was of similar high intensity. In tuber tissue, the expression was generally very low. Remarkable is that there was almost no expression in tubers of a bad chipping genotype compared to a good chipping genotype after harvest which changed slightly after 4 weeks of cold-storage.



Figure 31: Semi-quantitative expression analysis of *Inv6*.

PCR was performed on cDNA from different tissues from a potato plant. The lower part displays the result for the housekeeping gene *EF1α* (Nicot et al., 2005).

3.4 LEUCINE AMINOPEPTIDASE - LAP

Recent comparative 2D proteomics approaches (M. Fischer, unpublished data) revealed significant differences in the presence of a leucine aminopeptidase (LAP) between good and bad chipping genotypes (Figure 32). Total protein extracts of a pool of the 5 best and a pool of the 5 worst chipping genotypes of the BIOSOL population were separated on 2D gels and compared with each other. The gel picture of the pooled 5 good cultivars shows three distinct spots which are completely missing in the pool of the bad cultivars. From mass spectrometry analysis (MALDI-TOF/TOF), a protein sequence was obtained and identified as leucine aminopeptidase (P31427). In the following chapter, *in silico* based analyses of the genomic organization and an association analysis of LAP with starch and chip quality traits will be described.

So far, it was assumed that LAP plays a role in development, as well as in wounding and pathogen defense reactions. For the solanaceous species tomato and potato, it was shown, that they possess one constitutive and one inducible LAP (Tu et al., 2003). LAP is no obvious functional candidate for starch and chip quality in potato tubers.

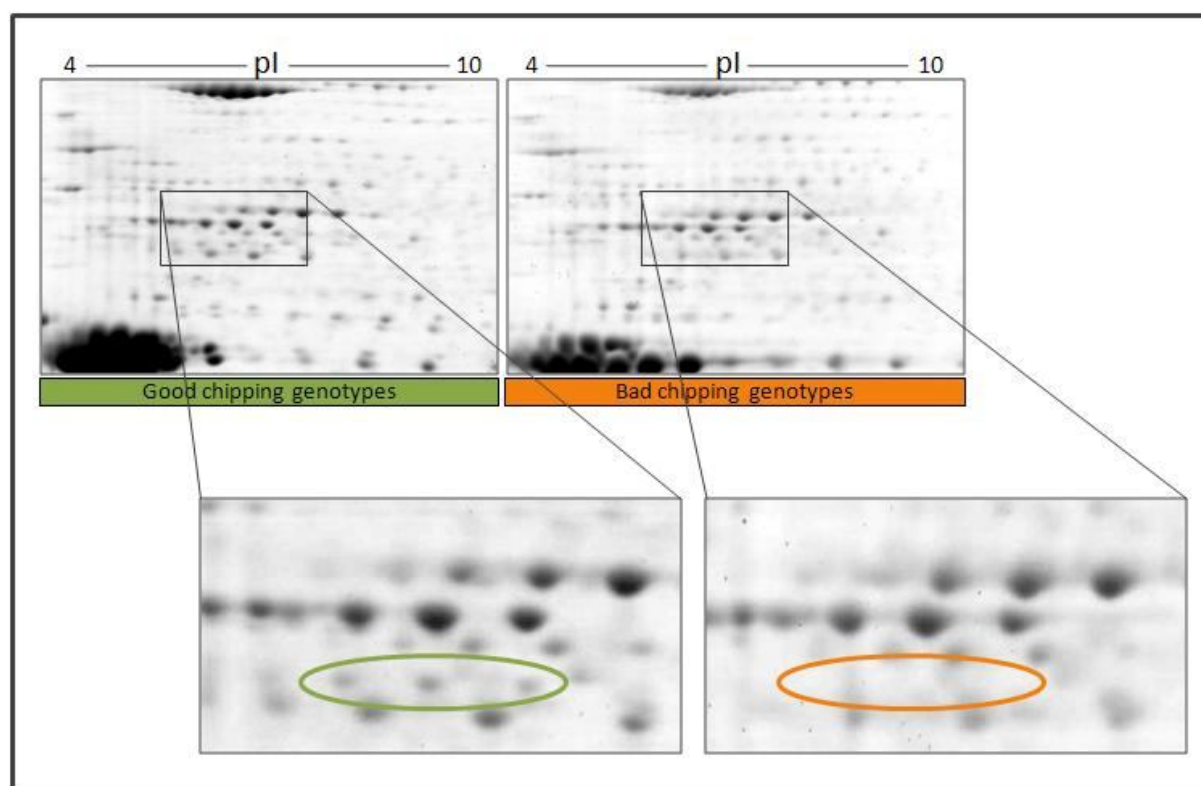


Figure 32: Comparison of 2D protein gels of good and bad chipping genotypes.

(Provided by M. Fischer, unpublished data). The two sets contain total protein extracts of five pooled genotypes each which were selected according to the highest or lowest amount of reducing sugars in a population of 40 genotypes (BIOSOL).

3.4.1 *LAP* - GENOMIC ORGANIZATION

Blast analysis of the deduced protein sequence was performed by using the Uniprot database. The identified protein sequence P31427 corresponded to the two cDNA sequences X77015 and X67845 (Hildmann et al., 1992). Both sequences are the only annotated sequences in NCBI for leucine aminopeptidase in *S. tuberosum*. Alignments of both sequences showed that they are identical, except X67845 that was not full-length. Consequently, the full-length sequence X77015 was used in Blasts searches against the released *S. phureja* genome. Four positive hits were obtained and analyzed according to the coverage rate and the identity towards the query sequence (Table 23).

Table 23: Best hits of *LAP* blasted against the *S. phureja* genome (PGSC, 2011).

Hit	Superscaffold	Identity %	e-value	Annotated gene in genome browser	Chromosome
1	PGSC0003DMB000000116	96.5	1.3e ⁻²⁹⁰	Neutral leucine aminopeptidase	12
2	PGSC0003DMB000000260	79.3	2.3e ⁻⁵²	-	5/3
3	PGSC0003DMB000000027	79.3	7.1e ⁻⁵²	Leucine aminopeptidase, chloroplastic	8/5
4	PGSC0003DMB000000068	79.0	1.1e ⁻⁵⁰	LAP17.1a protein	1

The first hit shows the highest identity and e-value. It is annotated as neutral leucine amino peptidase and located on chromosome 12. The 3rd hit is designated as *chloroplastic LAP*, but the analyzed superscaffold sequence contains poor sequence information (long N-stretches) and only short fragments of the middle part of the query sequence. No clear conclusion can be drawn whether it is a pseudogene or a real gene with limited sequence information in the published sequence. Additionally, the scaffold cannot be pinpointed to a single chromosome. The 4th hit is annotated as *LAP17.1a* and shows moderate identity to a very small fragment (350 bp) of the queried sequence X77015. Based on the retrieved data, further *in silico* analysis was performed on superscaffold 116 (Hit 1, Table 23).

Alignments of the cDNA sequence (X77015) to the genomic sequence of superscaffold 116 using Spidey resulted in two different regions of sequence similarity. Blast analysis in the PGSC database confirmed two independent matches to a region of around 17 kb. Although there is only one gene annotated in this region, *LAP* seems to be located in a tandem repeat at this locus (Figure 33). Both genomic organizations consist of 10 exons. The exon distribution and the sizes show high conservation between both sequences. More variation can be observed for the introns (detailed

lengths are listed in appendix Table B.4 and B.5) which leads to a double length for the gene in the 5' region (6.9 kb).

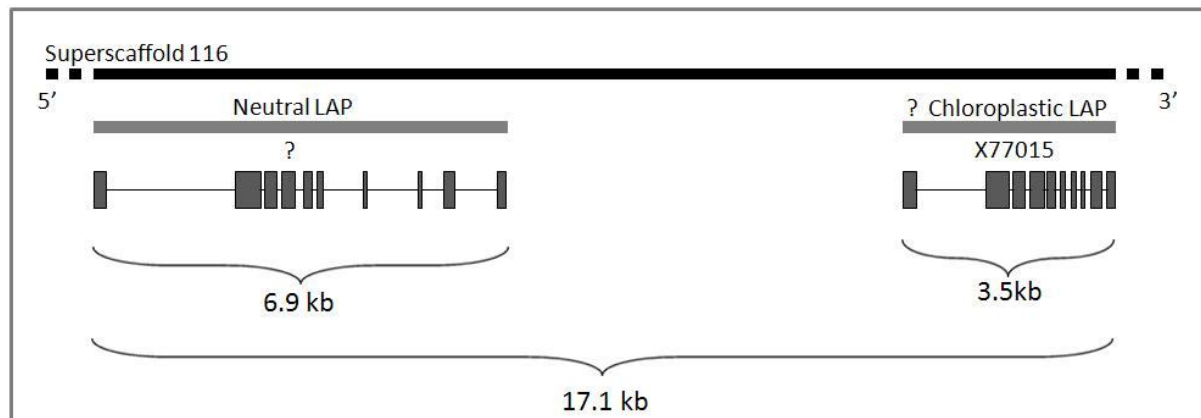


Figure 33: Genomic organization of the identified tandem repeat of LAP in *S. tuberosum*.

The blasted sequence X77015 matched to two regions within 17.1 kb in the genomic sequence of *S. phureja*. The identity to the 3' part was higher. No cDNA sequence for the 5' part is annotated for *S. tuberosum*.

The alignment of the potato LAP sequence X77015 resulted in a clear match to the 3' part which has not been annotated as a gene by now. To obtain more information, two tomato sequences were aligned to this particular genomic region. Sequence U50151 is annotated as *SILAP_A* with "A" representing "acidic", whereas AF510743 is designated as a *neutral leucine aminopeptidase* (*SILAP_N*). *SILAP_A* showed higher sequence similarity to the 3 prime part, whereas *SILAP_N* clearly matched to the 5' part and additionally to a higher degree to the annotated gene and transcript of *LAP_N*, indicating for the 5' part to correspond to the neutral *LAP*. The similarities of the four cDNA sequences were compared with each other and are given in table 24. Based on the comparison, it can be assumed that X77015 corresponds to the *acidic LAP* sequence of tomato (94.6%) and matches to the 3 prime part on the genomic sequence. In contrast to that, the deduced cDNA sequence from the first part of the tandem repeat shows more similarity (97.8%) to the sequence of a *neutral LAP* from tomato (AF510743).

Table 24: Comparison of the tomato and the potato LAP cDNA sequences.

	<i>S. tuberosum</i>	%	<i>S. lycopersicum</i>	
<i>Neutral LAP</i>	deduced cDNA from <i>S. phureja</i> sequence	97.8	AF510743	<i>Neutral LAP</i>
%	87.2	86.2	84.4	%
		85.3		
<i>Acidic LAP</i> (chloroplastic)	X77015	94.6	U50151	<i>Acidic LAP</i> (chloroplastic)
	<i>S. tuberosum</i>	%	<i>S. lycopersicum</i>	

The sequences were aligned and the level of identity was calculated by using the software Needle. The two tomato sequences were annotated on NCBI as U50151 (*LAP_A*) and AF510743 (*LAP_N*). For potato, sequence X77015 is annotated as *LAP* and was not further classified.

At amino acid level, the leucine aminopeptidases are annotated either as neutral or chloroplastic *LAP*, but at cDNA level, the designations are neutral and acidic. In the following, the obviously synonymously used name *chloroplastic LAP* will be uniformly renamed to *acidic LAP* (*LAP_A*) because the recent analyses identified them to be identical.

A comparison at amino acid sequence level reflects the high similarity and conservation among the four sequences from tomato and potato (Figure 34). Apart from that, it can be clearly seen that the four sequences form two separate groups of acidic and neutral LAPs. Throughout the whole sequence, few group characteristic differences regarding indels or group specific amino acid residues can be recognized.

```

      *           20           *           40           *           60           *           80           *           100
s1LAP_N : MAALRVSSAIAACSSSSSEPHSYPSIFTRKQSSPIMNSFSLVTEPLCSRRAKRMASISARDTLGLTHTNQSDAPKISFAARETLLVEMRGDILLVGAATERDLA : 100
stLAP_N : MAALRVSSAIAACSSSSSEPHSYPSIFTRKQSSPIMNSFSLVTEPLCSRRAKRMASISARDTLGLTHTNQSDAPKISFAARETLLVEMRGDILLVGAATERDLA : 100
stLAP_A : MAALRVSSAIAACSSSSSEPHSYPSIFTRKQSSPIMNSFSLVTEPLCSRRAKRMASISARDTLGLTHTNQSDAPKISFAARETLLVEMRGDILLVGAATERDLA : 100
s1LAP_A : MAALRVSSAIAACSSSSSEPHSYPSIFTRKQSSPIMNSFSLVTEPLCSRRAKRMASISARDTLGLTHTNQSDAPKISFAARETLLVEMRGDILLVGAATERDLA : 99

      *           120           *           140           *           160           *           180           *           200
s1LAP_N : RDCNSKFRNPLLCQLDSEHNGLLSAAASSEEDFSGKAGOSTILRLPGLGSRRTIALVGLGSEPTSSSTAARCLGEGAAAAAASQAQASNIATIALASTDGLSAEL : 200
stLAP_N : RDCNSKFRNPLLCQLDSEHNGLLSAAASSEEDFSGKAGOSTILRLPGLGSRRTIALVGLGSEPTSSSTAARCLGEGAAAAAASQAQASNIATIALASTDGLSAES : 200
stLAP_A : RDCNSKFRNPLLCQLDSEHNGLLSAAASSEEDFSGKAGOSTILRLPGLGSRRTIALVGLGSEPTSSSTAARCLGEGAAAAAASQAQASNIATIALASTDGLSAES : 197
s1LAP_A : RDCNSKFRNPLLCQLDSEHNGLLSAAASSEEDFSGKAGOSTILRLPGLGSRRTIALVGLGSEPTSSSTAARCLGEGAAAAAASQAQASNIATIALASTDGLSAES : 196

      *           220           *           240           *           260           *           280           *           300
s1LAP_N : KLSASASATITGAVLGTPEFNRPKSESKRPTILKSDLDLGLGTGPETERKIKRYAADVCAGVILGRELVNAPANVITPEAVLAAEAKKIASTYSDVFSANILLDV : 300
stLAP_N : KLSASASATITGAVLGTPEFNRPKSESKRPTILKSDLDLGLGTGPETERKIKRYAADVCAGVILGRELVNAPANVITPEAVLAAEAKKIASTYSDVFSANILLDV : 300
stLAP_A : KLSASASATITGAVLGTPEFNRPKSESKRPTILKSDLDLGLGTGPETERKIKRYAADVCAGVILGRELVNAPANVITPEAVLAAEAKKIASTYSDVFSANILLDV : 297
s1LAP_A : KLSASASATITGAVLGTPEFNRPKSESKRPTILKSDLDLGLGTGPETERKIKRYAADVCAGVILGRELVNAPANVITPEAVLAAEAKKIASTYSDVFSANILLDV : 296

      *           320           *           340           *           360           *           380           *           400
s1LAP_N : EQCKELKMGAYLAVAAAS-ANPAHFHLSYRPSSEGEHRRKIALVKGKLTDFDSGGYNIKRGPCCSIELMKDMGGAAAVLGAALKALGQIKRPAQVEVHFIVA : 399
stLAP_N : EQCKELKMGAYLAVAAAS-ANPAHFHLSYRPSSEGEHRRKIALVKGKLTDFDSGGYNIKRGPCCSIELMKDMGGAAAVLGAALKALGQIKRPAQVEVHFIVA : 399
stLAP_A : EQCKELKMGAYLAVAAAS-ANPAHFHLSYRPSSEGEHRRKIALVKGKLTDFDSGGYNIKRGPCCSIELMKDMGGAAAVLGAALKALGQIKRPAQVEVHFIVA : 397
s1LAP_A : EQCKELKMGAYLAVAAAS-ANPAHFHLSYRPSSEGEHRRKIALVKGKLTDFDSGGYNIKRGPCCSIELMKDMGGAAAVLGAALKALGQIKRPAQVEVHFIVA : 396

      *           420           *           440           *           460           *           480           *           500
s1LAP_N : ACENMISAGMRPGDIITASNGKTIEVNNTDABGRITLADALVYACNOGVKILDLATITGALVVALGFSVAGAFPTPDDLAKEVVVAASEVSGEKLWRLE : 499
stLAP_N : ACENMISAGMRPGDIITASNGKTIEVNNTDABGRITLADALVYACNOGVKILDLATITGALVVALGFSVAGAFPTPDDLAKEVVVAASEVSGEKLWRLE : 499
stLAP_A : ACENMISAGMRPGDIITASNGKTIEVNNTDABGRITLADALVYACNOGVKILDLATITGALVVALGFSVAGAFPTPDDLAKEVVVAASEVSGEKLWRLE : 497
s1LAP_A : ACENMISAGMRPGDIITASNGKTIEVNNTDABGRITLADALVYACNOGVKILDLATITGALVVALGFSVAGAFPTPDDLAKEVVVAASEVSGEKLWRLE : 496

      *           520           *           540           *           560           *
s1LAP_N : MEDSYWDSMKSGVADMVNTGPEGGGATTAALFLNQFVDEKVMQMLHLDLAGEVMSDRKKNATGFGVSTLVEWVLRNLTN : 577
stLAP_N : MEDSYWDSMKSGVADMVNTGPEGGGATTAALFLNQFVDEKVMQMLHLDLAGEVMSDRKKNATGFGVSTLVEWVLRNLTN : 577
stLAP_A : MEDSYWDSMKSGVADMVNTGPEGGGATTAALFLNQFVDEKVMQMLHLDLAGEVMSDRKKNATGFGVSTLVEWVLRNLTN : 574
s1LAP_A : MEDSYWDSMKSGVADMVNTGPEGGGATTAALFLNQFVDEKVMQMLHLDLAGEVMSDRKKNATGFGVSTLVEWVLRNLTN : 571

```

Figure 34: Amino acid sequence alignment of the LAP sequences of potato and tomato.

For *S. lycopersicum*, it was hypothesized that the *LAP_A* exists in two very similar isoforms (A1 and A2) and is organized in a tandem repeat (Ruiz-Rivero et al., 1998; Chao et al. 2000). The neutral LAP in tomato is supposed to be located at a different locus which has not been described and published by now. Both genes have not been designated to a certain chromosome. The findings in the released *S. phureja* sequence and additional blast analyses in the tomato database (SGN) indicate now that the tandem repeat is formed by the neutral LAP and the acidic LAP for both solanaceous species. The blast analyses of the two tomato sequences (*SILAP_A* and *SILAP_N*) against the sequence of *S. phureja* resulted in the same hits at superscaffold level and in the same chromosome match (Chr 12). Blasts against preliminary tomato chromosomes (SGN) resulted in two matches on tomato

chromosome 12. A comparison of the exact positions turned out that the positions are very similar to the identification of a tandem locus for the *S. phureja* sequence (Figure 33). Tomato *LAP_N* maps to the 5 prime part and the sequence of tomato *LAP_A* maps to the 3 prime part of the tandem repeat in the genomic sequences of both organisms.

3.4.2 CONSERVATION OF LAP AMONG DIFFERENT SPECIES

From the Uniprot database, several protein sequences were obtained by blasting the StLAP_A sequence. Identical output was obtained for the same blast analysis with StLAP_N sequence. In figure 35, the phylogenetic tree of all obtained 14 full-length sequences is displayed.

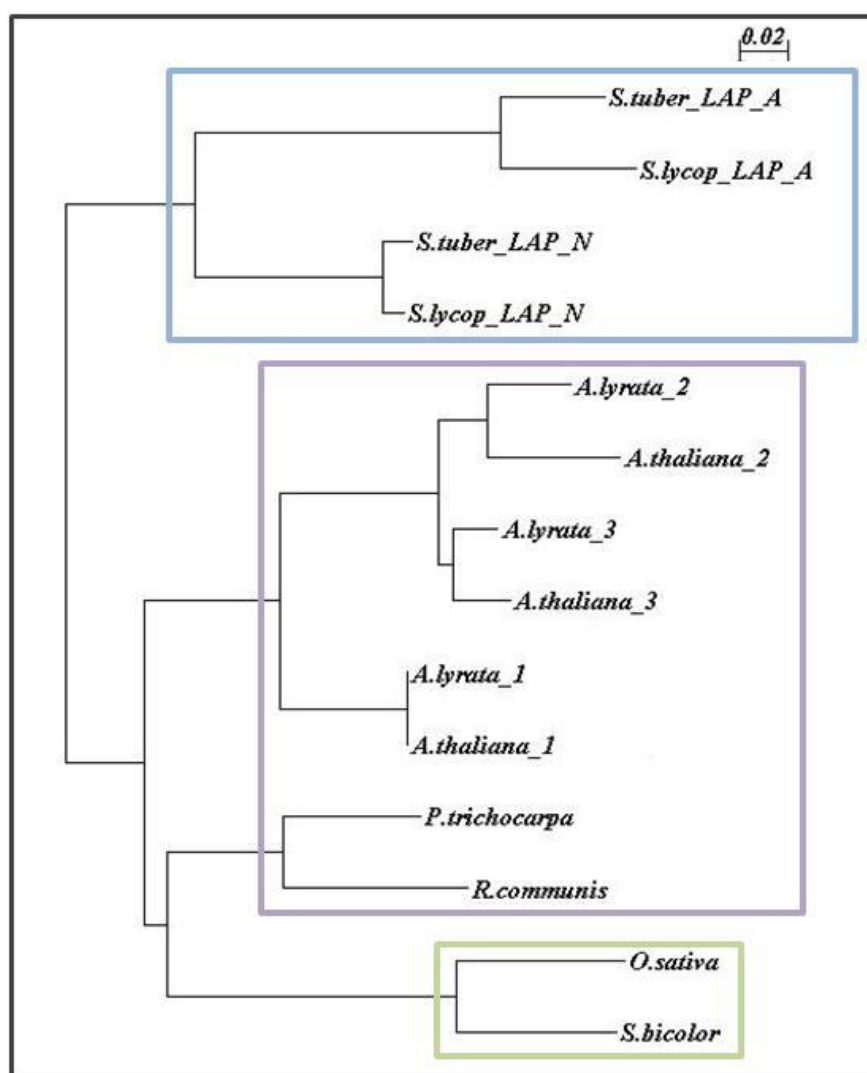
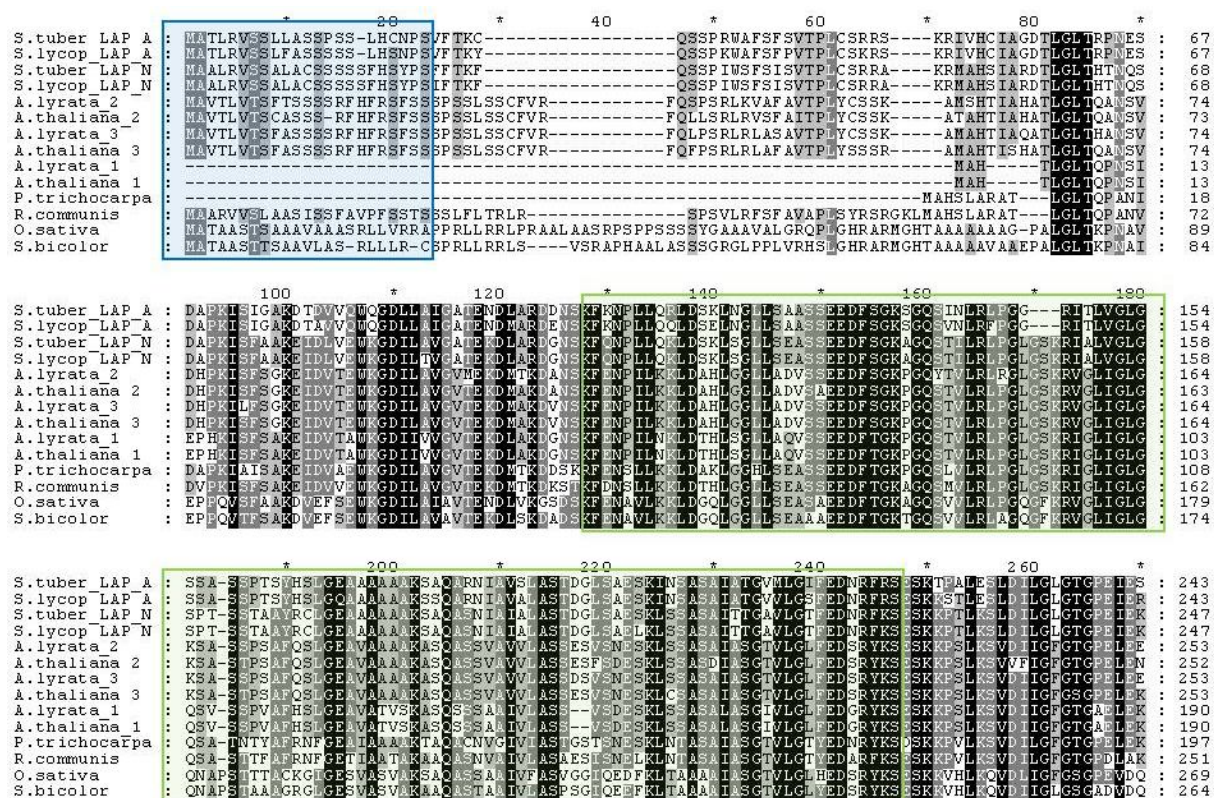


Figure 35: Phylogenetic tree of LAP sequences from different organisms.

Members of the same class are surrounded by boxes. Green: monocots; Violet: Rosids including different families (dicots); Blue: Asterids including only the family of Solanaceae (dicots). Accessions: *S. tuber_LAP_A*: X77015; *S. lycopersicon_LAP_A*: U50151; *S. tuber_LAP_N*: -; *S. lycopersicon_LAP_N*: AF510743; *A. lyrata_2*: D7MBK7; *A. thaliana_2*: Q8RX72; *A. lyrata_3*: D7MBK6; *A. thaliana_3*: Q944P7; *A. lyrata_1*: D7LHG7; *A. thaliana_1*: P30184; *P. trichocarpa*: B9IMD7; *R. communis*: B9STR1; *O. sativa*: Q6K669; *S. bicolor*: C5XUF6.

The acidic (LAP_A) and neutral LAPs (LAP_N) from potato and tomato form a distinct class from the other sequences. For *A. thaliana* and *A. lyrata*, three sequences were obtained which showed higher homology in the sequence alignment to the LAP_N family (Figure 36). No further LAP_A sequences were identified which indicates that the duplication and specialization event is unique for the solanaceous species.

Characteristic functional domains were identified via blast analysis against the Pfam database. The two major domains belong to the peptidase M17 family and consist of an N-terminal and a catalytic C-terminal domain. For the N-terminal domain, it has been predicted that they show more divergence compared to the C-terminal part (Matsui et al., 2006). The alignment of the 15 different sequences confirms these findings (light green box, Figure 36). The catalytic domain at the C-terminus spans a region of around 300 amino acids and shows an overall high conservation (dark green box). The catalytic important amino acids, lysine and arginine (Gu and Walling, 2002) are marked by green arrows. The region around the arginine shows almost no divergence, whereas the surrounding region of the lysine shows slightly less conservation for the LAP_A genes from potato and tomato. The most N-terminal part was analyzed with SIG-Pred and SignalP to predict the presence of putative signal peptides (http://bmbpcu36.leeds.ac.uk/prot_analysis/Signal.html). For both sequences, a signal cleavage site was calculated around position 20. No definite predictions were obtained about the localization of both genes but the two prediction servers SignalP and ChloroP indicated a location in the chloroplast.



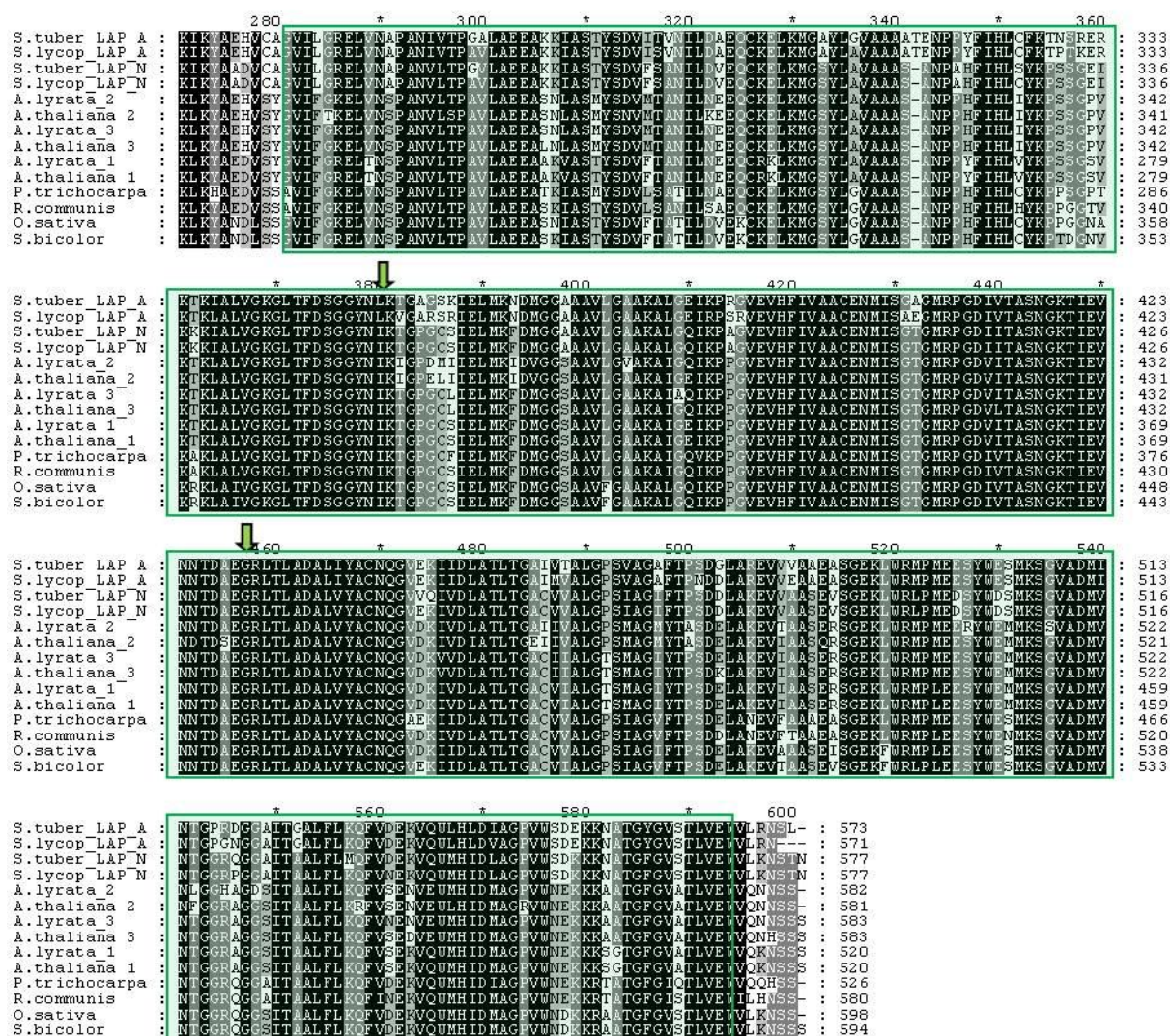


Figure 36: Alignment of leucine aminopeptidase peptide sequences of different species.

Amino acid residues identical in all sequences are highlighted by a black background, those conserved in most of the sequences are marked by a grey background. Conserved and annotated domains are indicated by coloured boxes. Blue box: putative cTP; light green box: N-terminal part of M17 peptidase domain; dark green box: C-terminal/catalytic part of M17 peptidase domain. Green arrow: catalytic amino acid.

3.4.3 ASSOCIATION ANALYSIS OF *LAP*

Previous results indicated *LAP_N* and *LAP_A* to be located in a tandem repeat on chromosome 12. By blasting several markers of the potato genetic map (PoMaMo) against the *S. phureja* superscaffolds, the superscaffold (116) containing *LAP* was identified with some markers (e.g. GP34, *AGPaseB-b*; Appendix Figure A.5). Based on these findings, the exact locus of the *LAP* tandem could be matched to a distal part on the long arm of chromosome 12. One of the identified markers (GP34) has been previously characterized as a flanking marker of a starch QTL (Schäfer-Pregl et al., 1998). Additionally, a sugar QTL was mapped to a locus in close proximity (Menéndez et al., 2002). Consequently, *LAP* can be considered as a positional candidate. To analyze whether *LAP* influences starch and chip quality, an association study in two populations was done.

3.4.3.1 ASSOCIATION ANALYSIS OF *LAP* IN THE GABICHIPS POPULATION

The short peptide sequences of the 2D proteome and MS-analysis showed in the blast analysis against Uniprot a higher identity to a sequence part of the acidic *LAP* (*LAP_A*). The sequence of the neutral *LAP* from tomato was obtained as second best hit. The position of the spots in the 2D gel matched to an approximate pI (isoelectric point) of 6 (Figure 31). Based on the calculated pIs which were estimated for *LAP_A* to be around 5.8 and for *LAP_N* to be 7.0, the spot was assigned to *LAP_A*. Due to very high sequence similarity at the coding sequence level between both *LAP* genes and additional rather short exons (Figure 33), it was challenging to obtain adequate gene-specific primers and an amplicon sequence without frame-shifts. In the end, it was possible to generate suitable gene-specific primers for a divergent region only within the coding sequence for *LAP_N* (L164-L165, Table 5). Due to the short physical distance between both *LAP* genes (ca. 7 kb, Figure 33), the frequency of recombination between both genes was assumed as rather low. Therefore, the primer combination for *LAP_N* was used to analyze whether the whole locus is associated and not one single gene.

The *LAP_N* primers were located in the large exon 2 and in intron 4. An 850 bp fragment was amplified in around 210 genotypes of the GABICHIPS population. In total, 25 SNPs and 2 indels were identified. Of the 25 SNPs, eleven were located upstream of both indels and showed a good basis for scoring. Downstream of the first indel, the sequence was frame shifted until the second indel appears. Although the second indel obviously compensated the first frame shift, the remaining 14 SNPs downstream of the two indels showed less reliability due to very apparent and strong co-segregation for many SNPs. For the analysis, the 11 upstream SNPs, six of the downstream SNPs and one indel were chosen (Appendix A.9). Association was tested with univariate GLM calculation in SPSS. The significance level was set to $p < 0.05$.

In total, four associated SNPs were identified (Table 25) which did not show any co-segregation among each other. Only one SNP was located in the exon sequence (SNP 1), the remaining SNPs were identified in intron sequence. SNP 1 (pos 2746) showed a very high association with the two traits starch content and starch yield ($p\text{-value} < 1e^{-6}$). It explains up to 17.5 % of the phenotypic variance of tuber starch content and around 4 % of starch yield. Additionally, this SNP is highly associated with chip quality traits (COA, CQS4 and CQS8). Between 4% and 10.5 % of the phenotypic variance can be explained by this SNP.

Table 25: Overview of associated SNPs of LAP.

SNP		Trait						Features				
#	From ATG	TY effect R ² (%)	TSC effect R ² (%)	TSY effect R ² (%)	CQA effect R ² (%)	CQS4 effect R ² (%)	CQS8 effect R ² (%)	X/Y	Ex/In	AA	True allele frequency	Apparent allele frequency
1	2746	ns	0.000000005 ↑ (A) 17.5	0.00000002 ↑ (A) 14.1	0.006 ↑ (A) 4.1	0.00003 ↑ (A) 10.5	0.0003 ↑ (A) 6.3	G/A	Ex	-	10.7%	34.0%
2	2810	0.047 ↓ (C) 2.6	ns	ns	0.042 ↓ (C) 2.1	0.038 ↓ (C) 3.0	0.018 ↓ (C) 2.6	T/C	In	-	1.6%	6.5%
3	2816	ns	ns	ns	ns	ns	0.043 ↑ (G) 3.4	A/G	In	-	13.6%	31.3%
4	3312	ns	0.001 ↑ (C) 7.4	0.003 ↑ (C) 6.1	ns	0.026 ↑ (C) 4.3	0.039 ↑ (C) 2.9	T/C	In	-	15.7%	49.0%

The position of the SNP is counted from the ATG (ref.: PGSC sequence). Traits: TY: tuber yield, TSC/TSY: tuber starch content/yield, CQA: chip quality in autumn, CQS4/CQS8 chip quality after storage at 4°C or 8°C. For each trait, the statistical test univariate GLM was applied with p<0.05. The feature part lists allele frequencies, exon or intron location and if the SNP leads to an aa exchange. All scored allele dosages are indicated. True allele frequency (856 alleles) and apparent allele frequency (214 genotypes) is given in %. R²: percent of the total variance explained; ns = not significant, ↑ positive effect, ↓ negative effect.

The means plots for four of the traits for this SNP are displayed in figure 37. The shift from the reference allele ("G") to the variant allele ("A") has a positive effect on all associated traits. Regarding the allele dosage, it can be assumed that only the status absence/presence of the allele has most likely an effect. In other words, a higher allele dosage of the superior allele does not lead to significant higher yield and quality. The statistical tests only provided significant differences between the homozygous common allele ("GGGG") and the simplex state of the variant allele ("AGGG"). The two other allelic states (duplex and triplex) are only present in 17 genotypes of the population and contain large variation (large error bars) which does not result in significant differences.

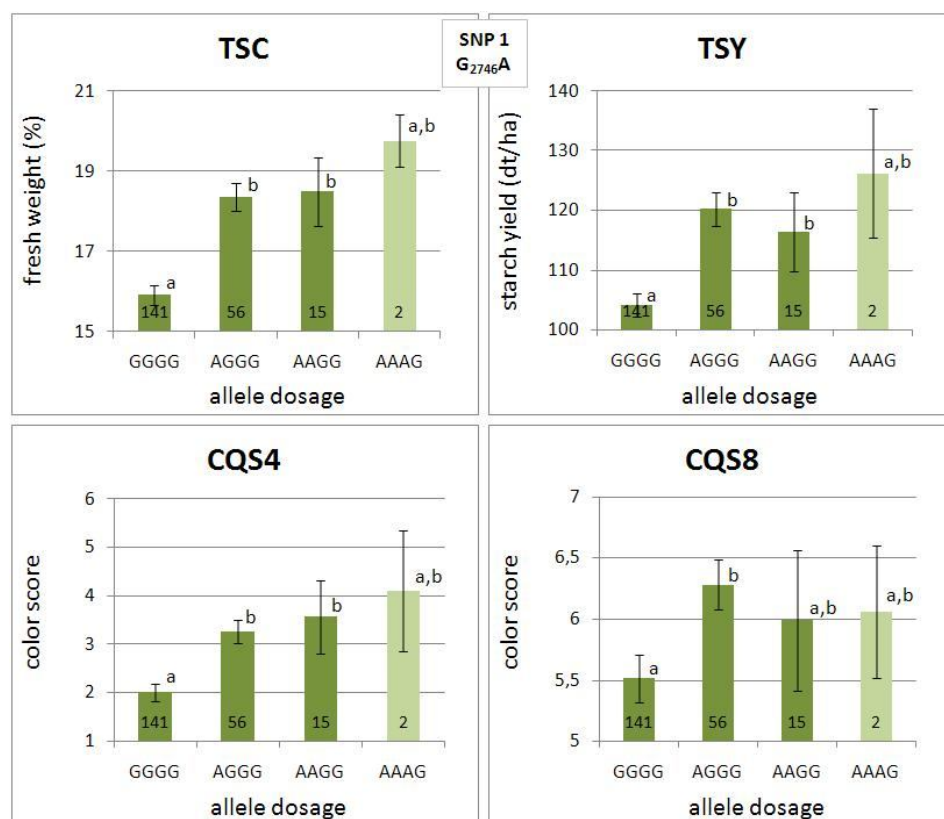


Figure 37: Means plots of SNP 1 (G₂₇₄₆A) of the LAP locus in *S. tuberosum*.

The X-axis represents the different allelic dosages which were scored for the SNP. The Y-axis shows the phenotypic data for the traits (TSC, TSY, CQS4 and CQS8). The number given in the lower part of the bar stands for the number of genotypes. Significant differences between the different groups (allelic states) are indicated by small letters. Different letters symbolize different groups, same letters stand for the same group and two letters indicate no clear classification to one of the groups.

SNP 2 is a SNP with very low frequency. It is negatively associated with tuber yield (TY) and chip quality (COA, CQS4 and CQS8). The presence of the variant allele explains around 2-3 % of the phenotypic variance. More frequent, but only associated with CQS8 in a very moderate way ($p=0.043$) is SNP 3. The shift to the superior allele explained around 3.4 % of the phenotypic variance. The 4th SNP is not very frequent, but it showed high association with starch quality traits (TSC and TSY). Up to 7% of the phenotypic variance was explained by this SNP. Better phenotypic data resulted from the presence of the superior allele, a higher dosage did not lead to any higher quality.

3.4.3.2 ANALYSIS OF THE DISTRIBUTION OF ASSOCIATED *LAP* ALLELES IN THE BIOSOL POPULATION

A second association analysis was performed in the BIOSOL population. SNP 1 (SNP G₂₇₄₆A) was scored in all 40 genotypes. Figure 38 shows the sugar data for TP0 which is defined as basic sugar level without cold-storage (section 3.3.4).

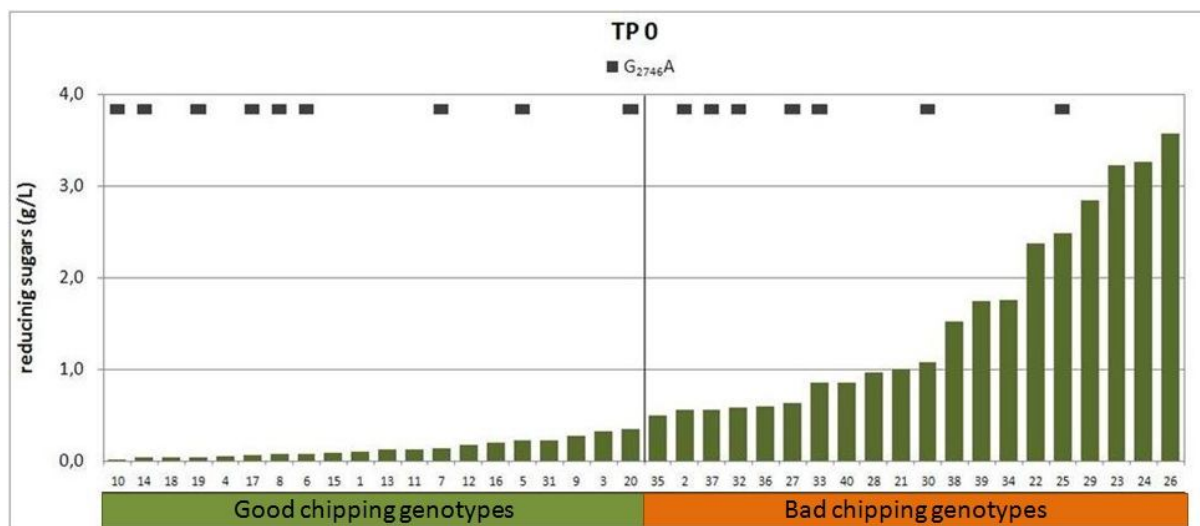


Figure 38: Presence of the associated *LAP* SNP 1 (G₂₇₄₆A) without cold-storage.

The amount of reducing sugars (glucose and fructose) was measured in 40 genotypes of the BIOSOL population at TP0 (without cold-storage). The genotypes are sorted by increasing total content of reducing sugars. The 20 genotypes with the lowest amount of reducing sugars are referred to the "good" chipping genotypes, whereas the 20 genotypes with the highest amount are grouped to the "bad" chipping genotypes.

The presence of the SNP is indicated as a small square above each genotype. The superior allele was identified in nine genotypes of the group of good chipping genotypes containing less reducing sugars and in seven genotypes of the group of bad chipping genotypes.

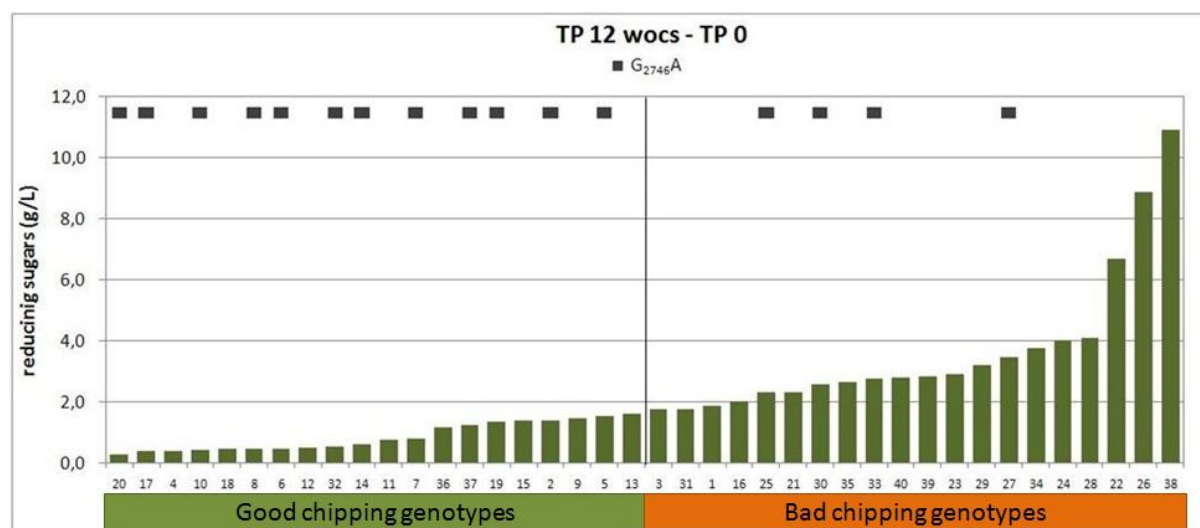


Figure 39: Presence of the associated *LAP* SNP 1 (G₂₇₄₆A) after 12 weeks of cold-storage.

The amount of reducing sugars (glucose and fructose) in potato tubers was measured in 40 genotypes of the BIOSOL population at TP12 (12 weeks of cold-storage). The genotypes are sorted according to increasing accumulation of reducing sugars. The 20 genotypes with the lowest amount of reducing sugars are referred to the "good" chipping genotypes, whereas the 20 genotypes with the highest amount are grouped to the "bad" chipping genotypes.

To focus on the accumulation of the reducing sugars, the basic sugar amount measured at TP0 was subtracted from the amount measured at TP12 (Figure 39). Analyzing the presence of the marker at TP 12 resulted in a higher frequency of the superior allele in the group of good genotypes. The ratio increased to 12 genotypes in the group of good genotypes bearing the superior allele of SNP 1 compared to 4 genotypes belonging to the group of bad chip quality genotypes. These findings suggest an association between the presence of the superior allele and the affiliation to the group of better genotypes.

Statistical calculations were performed to confirm the previous result. To obtain normally distributed sugar data, the data were transformed by using the log10 function. For all five time points, the amount of reducing sugars was significantly lower in genotypes containing the superior allele (Figure 40). Additionally, the level of significance is increasing with the time of cold-storage. The presence of the superior allele explains between 5% and 12 % of the phenotypic variance. These findings are in agreement with the association analysis in the GABICHIPS population. Obviously, the presence of the variant allele ("A") contributes to a higher starch yield, a better chip quality and to less accumulation of reducing sugars. The allele dosage does not show any additional effect in both tested populations.

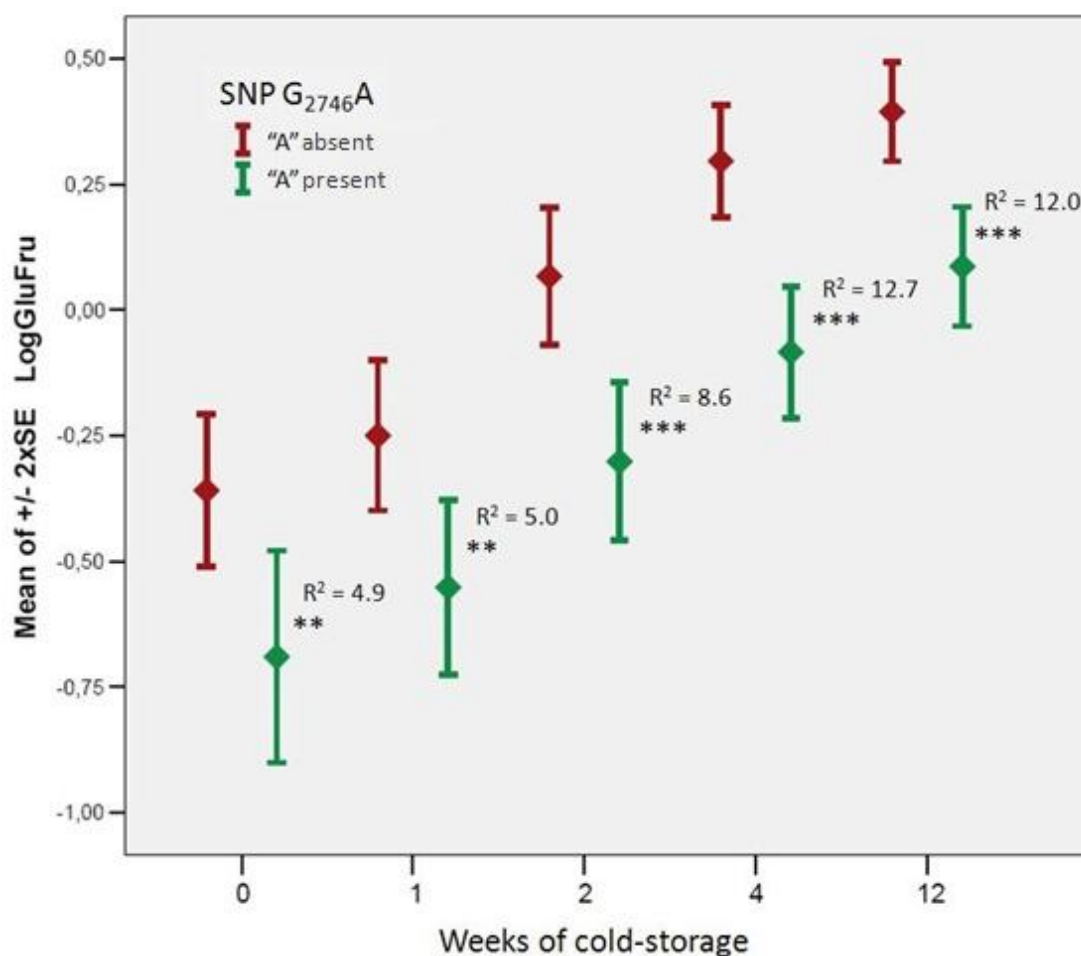


Figure 40: Relationship between sugar content and the presence of the superior LAPSNP1 (G₂₇₄₆A).

The presence of the superior allele "A" is associated with a lower amount of reducing sugars (mean values +/- SE) in the BIOSOL population. Significance levels are displayed by asterisks and the R² values are given in percentage.

3.5 ANALYSIS OF THE PUTATIVE INVERTASE INHIBITOR

For the starch degradation pathway, most of the directly involved enzymes have been identified and characterized. Besides these catalytic enzymes, there are regulatory proteins which regulate and act on the pathway enzymes. In this respect, a putative invertase inhibitor has been purified from potato tubers. After amino acid sequencing, it was designated as a member of the large family of Kunitz-type protease inhibitors (Glaczinski et al., 2002). Strong inhibitory effects on partially purified invertase from potato tubers were shown. Basing on the peptide sequence, degenerated primers were designed and several highly similar cDNA sequences were amplified and cloned. All of them belong to group C of the Kunitz-type inhibitors. Within group C, the sequences show at least 90% similarity (Heibges et al., 2003a), so it cannot be confirmed that the clones are distinct genes or allelic variants. Most of the clones were not full-length, but the estimated length for all of them amounts to 660 bp which corresponds to 220 amino acids (Glaczinski et al., 2002).

3.5.1 GENOMIC ORGANIZATION OF THE PUTATIVE INVERTASE INHIBITOR

The analysis was started by gathering all available cDNA sequences from members of Kunitz-type inhibitor group C which were obtained from previous and current work on the Kunitz-type inhibitors in this group (Glaczinski et al., 2002; Heibges et al., 2003a and 2003b; M. Fischer, unpublished data). In total, 30 cDNA sequences were identified and aligned using ClustalX software (Appendix Figure A.12). A phylogenetic tree was constructed (Figure 41) to classify the numerous cDNA sequences. Four subgroups which contain between 6 and 10 sequences each, can be clearly distinguished from each other. They are named after one prominent member of each subgroup (X56874, 1421, 1422 and 1423). The putative invertase inhibitor is assumed to belong to subgroup 1421 or 1422 (Glaczinski et al., 2002) which showed higher similarities between each other compared to the remaining two subgroups. In general, sequence similarities between the single subgroups are very high and vary, for instance, from 91.5 % for 1423 and X56874 to 97.1 % for 1421 and 1422. Within a subgroup, the cDNA sequences show similarities up to 97.4 % and 99.2% (subgroups 1421, 1422 and 1423) and around 93.9 % for subgroup X56874.

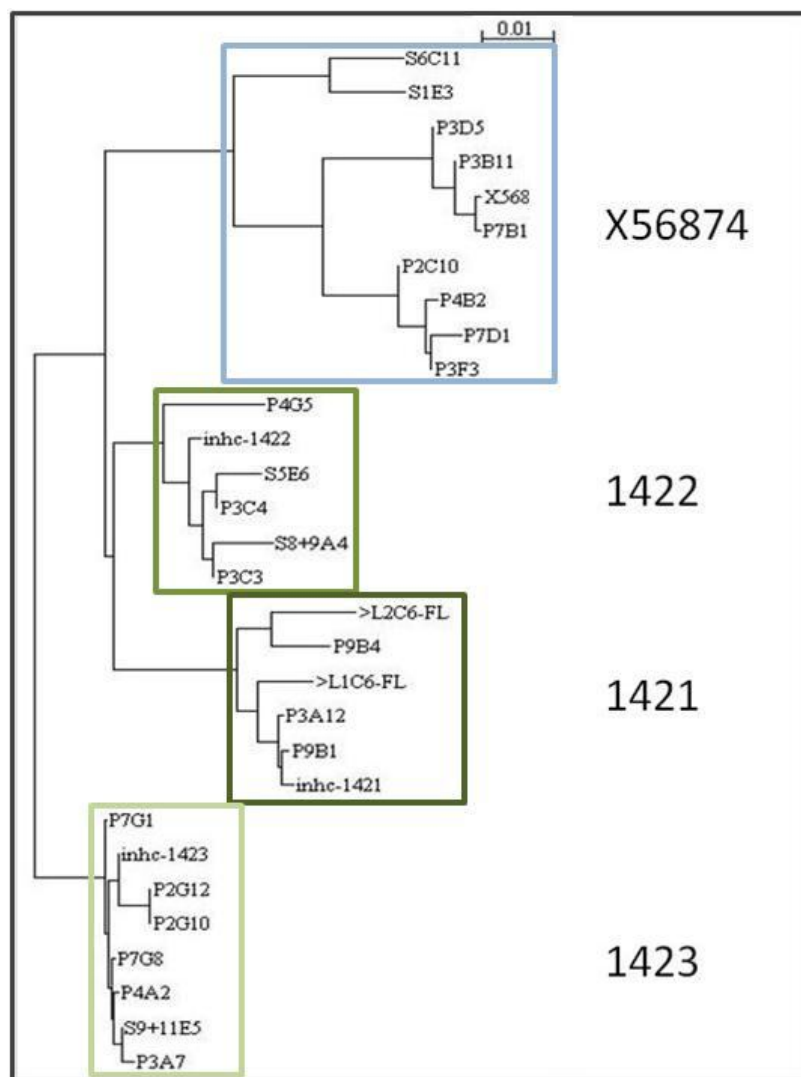


Figure 41: Phylogenetic tree of 30 cDNA sequences of group C Kunitz-type inhibitors.

The sequences can be grouped into 4 clades and were named after one of the members which have been characterized before (Glaczinski et al, 2002; Heibges et al, 2003a and 2003b; Yamagishi et al., 1991).

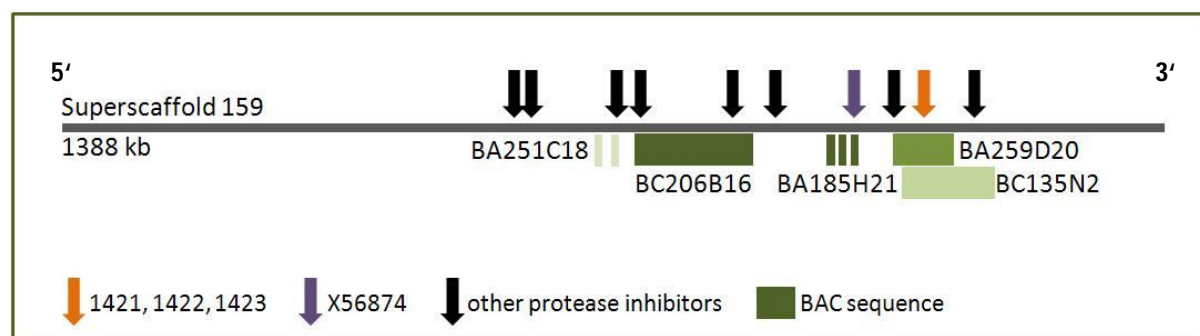
Four representative cDNA sequences (*1421*, *1422*, *1423* and *X56874*) were used for the blast analysis against the released genome sequence of *S. phureja* (PGSC, 2011). The blast hits obtained for different datasets of the PGSC blast server are listed in table 26. All four sequences matched to the same chromosome, superscaffold and scaffold. These sequences could be mapped to a central part close to the centromere on chromosome 3 and confirmed the findings of a previous study (Heibges et al., 2003a). Differences showed up when blasting the four subgroups against gene and transcript datasets of the PGSC blast server. The same matches were obtained for *1421*, *1422* and *1423*, but *X56874* was mapped to a different locus. Sequence comparisons revealed that the *S. phureja* sequence showed the highest identity to the *1421* cDNA sequence (99.7%). In the following part, this genomic locus will be referred to as *1421*. No further appropriate hits were retrieved which indicated a second or third locus for the remaining subgroups *1422* and *1423*.

Table 26: Kunitz-type protease inhibitor cDNA sequences blasted against different datasets of the *S. phureja* genome.

	Blast of 1421,1422,1423	Blast of X56874
Chromosome	3	3
Superscaffold	PGSC0003DMB000000159	PGSC0003DMB000000159
Scaffold	PGSC0003DMS000000374	PGSC0003DMS000000374
Gene	PGSC0003DMG200010146	PGSC0003DMG200010143
Transcript	PGSC0003DMT200026289	PGSC0003DMT200026285
Position in scaffold	1064103-1064874	964101-964893

The best hit for each of the representatives of the four subgroups are listed.

Further analysis of the 1.3 Mb superscaffold sequence revealed at least eight additional protease inhibitor sequences which are present in this sequence (Figure 42). Most of them were annotated as cysteine and Kunitz-type protease inhibitors. They were all located within 500 kb and are in close proximity to each other. The identified protease inhibitors showed high correspondence to members of group X56874, especially to the sequences S6C11 and S1E3 which form a distinct subgroup in the phylogenetic tree (Figure 41). The remaining 800 kb did not contain any further sequences with similarity to protease inhibitors. The sequences 1421, 1422 and 1423 matched to the same position within the scaffold (orange arrow), whereas X56874 is located approximately 100 bp upstream (violet arrow).

**Figure 42: Protease inhibitor cluster in superscaffold 159.**

Within 500 kb, 10 protease inhibitor sequences were identified. Several BAC sequences from a previous study (Odeny et al., 2010) matched to the same superscaffold. The superscaffold matches to a central region of chromosome 3 (Appendix A.1)

It is assumed that the Kunitz-type protease inhibitors are clustered on one chromosomal location in the potato genome (Heibges et al., 2003a; Appendix A.1). Several BAC clones were obtained by a BAC library screen in previous work in this group with specific protease inhibitor probes (Odeny et al., 2010). Five of them were identified to contain members of the group C Kunitz-type protease inhibitors. Via comparisons, it was possible to match 3 of the BAC contigs completely and 2 of the BACs partially to the obtained *S. phureja* superscaffold 159. These findings support the hypothesis of a clustered location because all available sequence information (BAC sequences and results from the *in silico* analysis) matched to a high degree only to this single locus in the genome.

3.5.2 CONSERVATION OF THE PUTATIVE INVERTASE INHIBITOR AMONG DIFFERENT SPECIES

Several blast analyses against different databases did not result in sequences with high similarity to the tested subgroups of the putative invertase inhibitor. Blast analyses against the SGN database resulted in a match with a gene located on chromosome 3. Although it was annotated as a cysteine protease inhibitor, direct amino acid sequence comparisons showed only 67 % identity. In the database of NCBI, the best hit showed around 82 % sequence identity to the cDNA sequences of 1421, 1422, 1423 and X56874 and corresponded to a not-annotated sequence from *S. lycopersicum*. At the protein level, the degree of identity between the sequences decreased to 65 % (77 % of similarity). The putative sequence homologs were not confirmed in other databases. Further low identity hits (less than 50%) were obtained from other solanaceous species, for example from *Nicotiana glauca*. It can be assumed that this particular group of protease inhibitors either shows only low conservation or may be unique for *S. tuberosum* or at least limited to solanaceous species.

3.5.3 EXPRESSION ANALYSIS OF THE PUTATIVE INVERTASE INHIBITOR

Expression analysis has been performed for the three major groups A, B and C of the Kunitz-type protease inhibitors (Turrà et al., 2008). They showed the highest expression in underground tissues like potato tubers, stolons and roots. There is no data available yet to proof that there are any differences in the expression level between the subgroups 1421, 1422 and 1423. By using the alignment of all group C sequences (Appendix Figure A.12), subgroup-specific primer pairs were designed (Table 5, section 2.1.7). They are mainly based on subgroup-specific SNPs and on a 6 bp indel in the vacuolar target peptide at the N-terminal part.

3.5.3.1 SEMI-QUANTITATIVE EXPRESSION IN DIFFERENT POTATO TISSUES

On the basis of subgroup-specific primers, the expression pattern was analyzed for the three subgroups 1421, 1422 and 1423. PCR amplifications were performed on cDNA from different tissues, extracted from seven aerial parts of a potato plant and from potato tuber (Figure 43). The expression for all three subgroups was confined to the tuber. Solely for subgroup 1422, a weak expression was detected in flower buds.

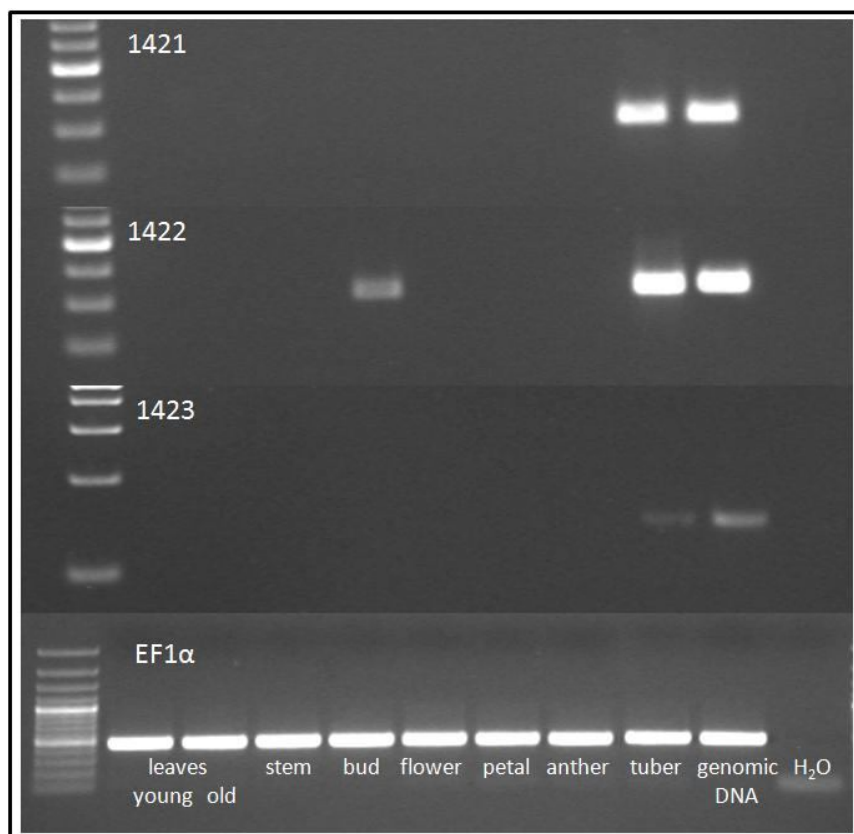


Figure 43: Semi-quantitative expression analysis of subgroups 1421, 1422 and 1423.

PCR was performed with subgroup specific primers (Table 5) on cDNA from different tissues from a potato plant. The bottom panel displays the result for the housekeeping gene EF1 α (Nicot et al., 2005).

3.5.3.2 CULTIVAR-DEPENDENT EXPRESSION OF THE PUTATIVE INVERTASE INHIBITOR

Further analysis was done at the genomic level. The subgroup-specific primers were used to screen the genomic sequences of the 40 BIOSOL genotypes for the presence of the three subgroups 1421, 1422 and 1423. The specific primer pairs are listed in chapter 2 materials and methods (Table 5). As a positive control, the cloned sequences of 1422 and 1423 from previous work (Heibges et al., 2003a and 2003b) were used to guarantee for the specificity of the primers. The specific forward primer for 1421 is located within the sequence for the vacuolar signal which was not included in the available 1421 cDNA clone and therefore could not be used as a positive control. Consequently, all amplified sequences of subgroup 1421 were sequenced and the specificity for this subgroup was proven by analyzing all sequences for unique SNPs of subgroup 1421.

Figure 44 shows the presence of the three subgroups in the 40 genotypes of the BIOSOL population. The subgroup 1421 was amplified in 38 out of the 40 genotypes, whereas 1422 and 1423 were only detectable in some cultivars. A signal for subgroup 1422 was observed in 23 genotypes and almost equally distributed between the 20 good and 20 bad chipping cultivars. The amplification of subgroup 1423 resulted in 8 positive genotypes which were predominantly members of the group of

bad chipping cultivars. The expression pattern of subgroups 1422 and 1423 seemed to alternate to a certain extent. One exception is cultivar 32 (blue arrow), in which all three subgroups were amplified.

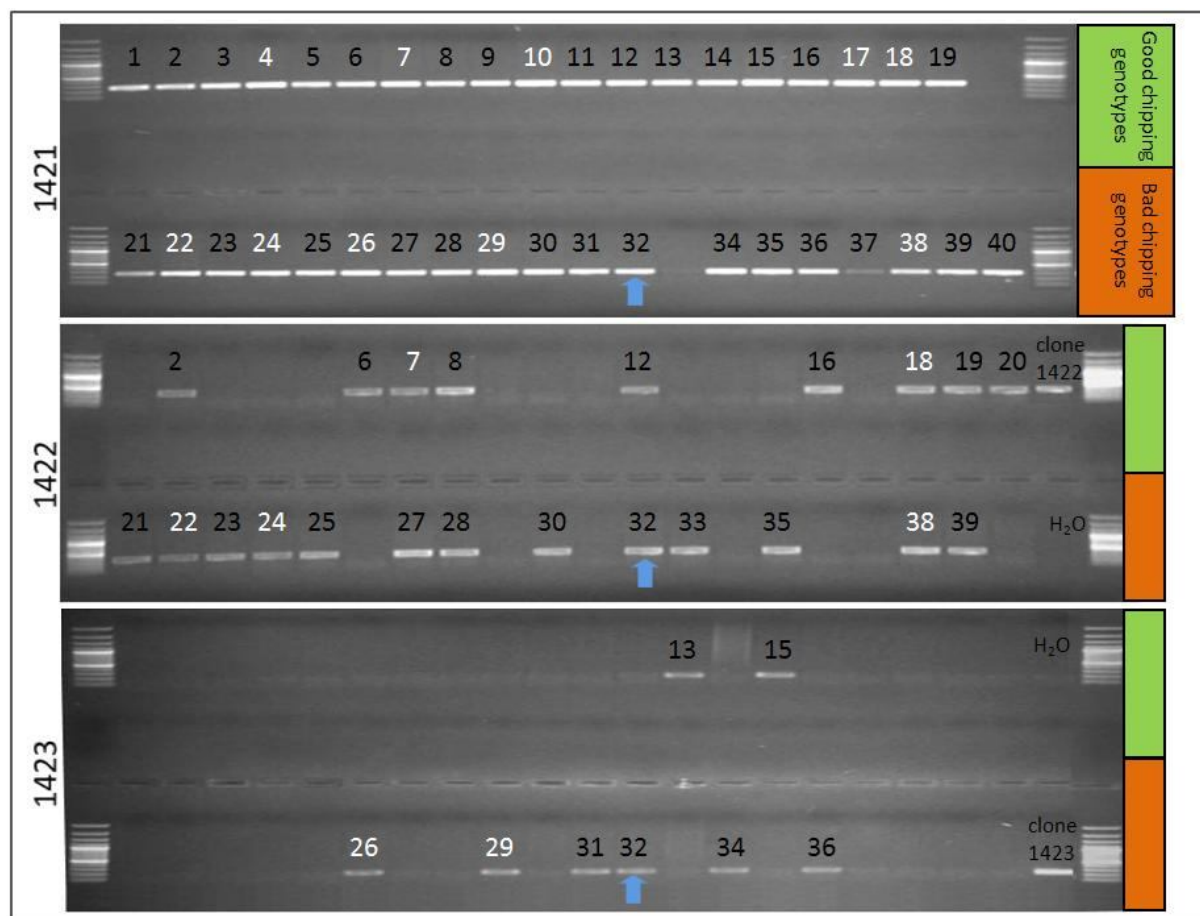


Figure 44: Presence of the three inhibitor subgroups 1421, 1422 and 1423 at genomic DNA level.

PCR was performed with subgroup specific primers (Table 5) on genomic DNA of the 40 genotypes of the BIOSOL population. As a control for the specificity of the primers, the cloned cDNA sequence was used. For subgroup 1421, the specificity was tested by amplicon sequencing. The genotypes names are indicated by their numerical name. Only those genotypes are marked which show amplification with the subgroup specific primers. The five best and worst chipping genotypes are highlighted in white.

3.5.3.3 EXPRESSION ANALYSIS ON TRANSCRIPT LEVEL

Previous expression studies showed a clear restriction of the three subgroups 1421, 1422 and 1423 to potato tubers. The tubers are predominantly stored at low temperatures to avoid sprouting and diseases. To analyze whether the cold-storage has an effect on the expression level of these three subgroups, tubers of genotypes from the BIOSOL population were stored for a period of 12 weeks in the cold and samples were taken at 5 different time points (TP0, TP1, TP2, TP4 and TP12, while the number corresponds to the week of cold-storage). These genotypes were phenotyped according to their basic amount of reducing sugars directly after harvest without cold treatment (TP0). Based on the sugar data (M. Fischer, unpublished data), five genotypes containing the lowest level (4, 7, 10, 17 and 18) and five genotypes containing the highest level of sugars were chosen (22, 24, 26, 29 and 38). For the transcriptional expression analysis, cDNA was synthesized from RNA extracted from

these selected potato tubers. Extractions were made from three biological replicates per genotype and per time point. The analysis was performed using the pyrosequencing approach (Ronaghi et al., 1996). As the sequence template, a 335 bp amplicon was created with primers amplifying all sequences from the group C of the Kunitz-type protease inhibitors. For the pyrosequencing reaction, a sequencing primer was designed which specifically detects one SNP for subgroup 1421 and two SNPs for subgroup 1423 (Table 5). No specific sequencing-primer could be designed for subgroup 1422, as none of the SNPs in the amplified region was unique for 1422. Consequently, a sequencing primer was used which analyzed 1421 along with 1422.

The expression of subgroup 1421 is shown in figure 45. No differences between the group of good and bad chipping genotypes could be identified which would fit to the grouping according to the measured sugar levels. Regarding all tested genotypes, variation in the expression levels occurred between 10-30% and therefore a certain response to the exposure to the cold is shown. From the relative expression data, no common trend can be concluded.

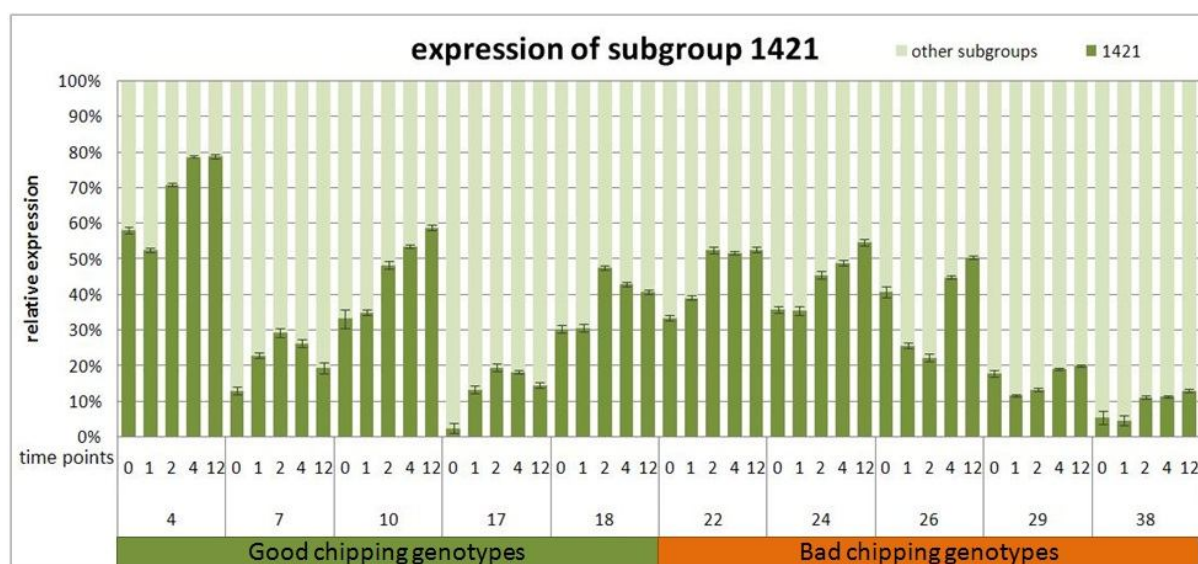


Figure 45: Analysis of the transcript level of subgroup 1421 during cold-storage.

The expression levels were analyzed by using a pyrosequencing approach in the five best and five worst chipping genotypes of the BIOSOL population. Samples were analyzed after 0, 1, 2, 4 and 12 weeks of cold-storage. One bar represents the average of nine replicates (3 technical replicates of 3 biological replicates) per genotype.

The expression analysis on transcript level for subgroup 1423 supported the previous findings of a limited occurrence in the tested genotypes (section 3.5.3.2, Figure 44). This subgroup was only encoded in the two cultivars 26 and 29 which belong to the group of bad chipping genotypes. For the remaining eight genotypes, amplification on genomic DNA and detection of subgroup-specific on cDNA level was not possible. In figure 46, the relative expression for subgroup 1423 against the remaining subgroups of group C is illustrated. Upon cold-storage, the relative expression level of

subgroup 1423 increased slightly and differed between 5 and 10 % at the most. No clear response to the storage in the cold was recorded.

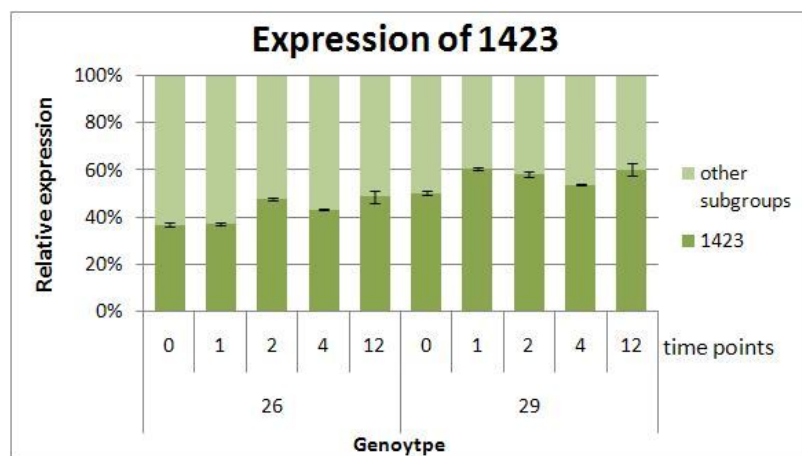


Figure 46: Analysis of the transcript level of subgroup 1423 during cold-storage.

The expression levels were analyzed by using a pyrosequencing approach. Transcript levels were detected in two genotypes (26 and 29) of ten. Samples were analyzed after 0, 1, 2, 4 and 12 weeks of cold-storage. One bar represents the average of nine replicates (3 technical replicates of 3 biological replicates) per genotype.

The third pyrosequencing test included both subgroups 1421 and 1422 because no specific SNP for subgroup 1422 was present in the amplicon. To estimate the relative expression level of subgroup 1422, figure 47 shows the values obtained for both subgroups in light green, whereas in dark green, the expression level of subgroup 1421 is displayed (Figure 45). The difference between both subgroups (visible as remaining parts of light green bars) is the deduced relative expression for subgroup 1422. For genotypes 4 and 26, it can be concluded, that 1422 was not expressed. These findings are in agreement with the previously obtained amplification results on genomic DNA (Figure 44). For the genotypes 10, 17 and 29, the absence of subgroup 1421 was hypothesized as well, but identified at expression level. Probably a further subgroup with very high similarity was amplified with the sequencing primer for 1421 and 1422 which influences the expression level in some of the tested genotypes. Neglecting the values of the five genotypes which did not contain 1422 on genomic level (genotypes 4, 10, 17, 26 and 29), the relative expression level varied between 5 % and 25 % for the remaining genotypes (7, 18, 22, 24 and 38). The variation was slightly higher in the three genotypes of the bad chipping genotypes.

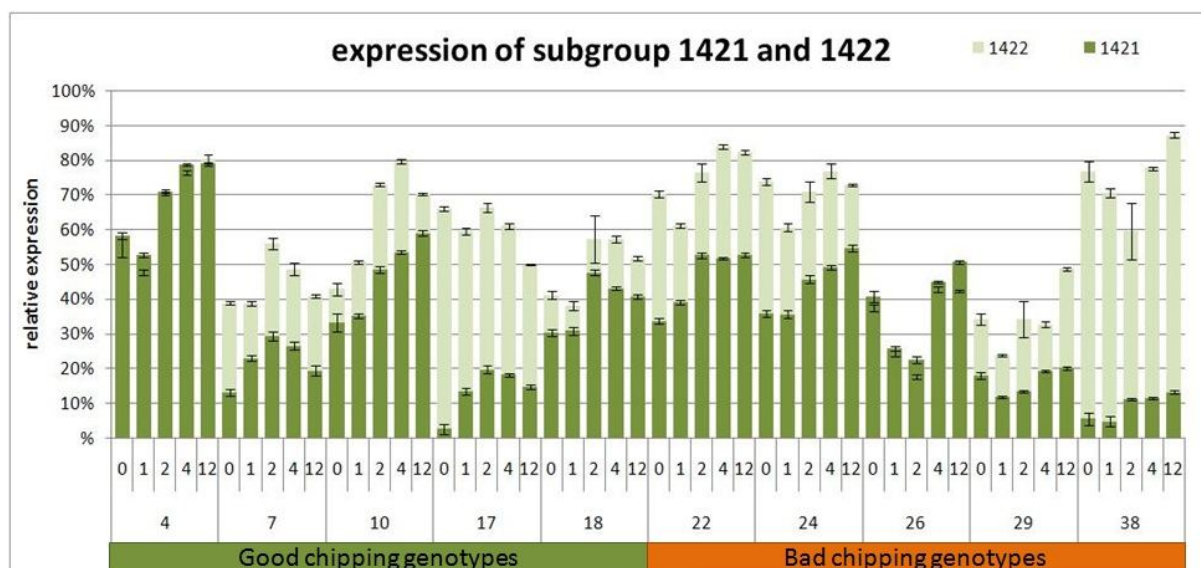


Figure 47: Calculated expression level of subgroup 1422 during cold-storage.

The expression level has to be subtracted from the combined value for 1421 and 1422 (light green bars) and the expression for 1422 (dark green bars). One bar represents the average of nine replicates (3 technical replicates of 3 biological replicates) per genotype.

3.5.4 ASSOCIATION ANALYSIS OF THE PUTATIVE INVERTASE INHIBITOR

For the Kunitz-type protease inhibitors of group C, little is known about their true function. For native protein extracts of this subgroup from potato tubers, it was shown that they inhibit the activity of invertase (Glaczinski et al., 2002). Invertase converts sucrose into reducing sugars and is associated with starch and chip quality traits (Li et al., 2005; 2008). To investigate whether members of group C contribute to the traits mentioned above, an association analysis in the GABICHIPS population was performed. The sequence of the superscaffold 159 (Table 26) was used to find a unique region in close proximity to the locus of 1421 to avoid the amplification of the very similar loci of other protease inhibitors (Figure 41; Table 26). An 1150 bp fragment was obtained with the primer pair LS170-LS171 (Table 5). The forward sequencing of the amplicon turned out to be difficult, due to repetitive parts and long stretches of the same base in the UTRs up- and downstream of this locus. Consequently, an internal reverse primer (starting at the stop codon of 1421) was used for sequencing (LS65, Table 5). The reverse sequenced amplicons showed that there was allelic variation of 1421, and characteristic SNPs for either subgroup 1421 or subgroup 1422 were detected. No characteristic SNP for subgroup 1423 was identified.

In total, 15 SNPs and one indel were identified (Appendix A.10 and A.12) and analyzed by using Univariate GLM with $p=0.05$. The retrieved nine associated SNPs and the associated indel are shown in table 27. All of them were highly significantly associated with the traits tuber starch content (TSC) and tuber starch yield (TSY).

Table 27: Overview of associated SNPs of 1421.

SNP		Trait						Features				
#	From ATG	TY effect R ² (%)	TSC effect R ² (%)	TSY effect R ² (%)	CQA effect R ² (%)	CQS4 effect R ² (%)	CQS8 effect R ² (%)	X/Y	Ex/In	AA	True allele frequency	Apparent allele frequency
1	112	ns	0.00006 ↑ (+) 11.2	0.00002 ↑ (+) 11.6	ns	ns	ns	6 bp indel	Ex	DD	20.0%	60.1%
2	206	ns	0.003 ↑ (G) 6.5	0.0002 ↑ (G) 8.7	ns	ns	ns	C/G	Ex	G / A	10.1%	33.8%
3	261	ns	0.0001 ↑ (G) 9.5	0.0003 ↑ (G) 8.3	ns	ns	ns	A/G	Ex	-	20.3%	60.9%
4	334	ns	0.00007 ↑ (T) 9.9	0.00003 ↑ (T) 10.2	ns	ns	ns	G/T	Ex	D / Y	12.4%	39.6%
5	352	0.028 ↓ (G) 4.0	0.006 ↑ (G) 5.8	0.001 ↑ (G) 7.7	ns	ns	ns	C/G	Ex	L / V	14.4%	44.0%
6	362	ns	0.003 ↑ (G) 6.5	0.0001 ↑ (G) 8.9	ns	ns	ns	C/G	Ex	A / G	13.4%	41.1%
7	391	0.047 ↓ (T) 4.2	0.035 ↑ (T) 4.9	0.002 ↑ (T) 7.7	ns	ns	ns	A/T	Ex	T / S	13.5%	42.5%
8	395	ns	0.001 ↑ (G) 7.2	0.00009 ↑ (G) 9.4	ns	ns	ns	A/G	Ex	K / R	2.1%	6.7%
9	396	ns	0.006 ↑ (A) 5.8	0.001 ↑ (A) 7.8	ns	ns	ns	G/A	Ex	-	5.3%	7.7%
10	430	ns	0.002 ↑ (C) ↓ (A) 11.3	0.003 ↑ (C) ↓ (A) 10.5	ns	ns	ns	G/C/A	Ex	D / H / N	A 7.3% C 14.5%	A 28.0% C 46.8%

The position of the SNP is counted from the ATG (ref.: PGSC sequence). Traits: TY: tuber yield, TSC/TSY: tuber starch content/yield, CQA: chip quality in autumn, CQS4/CQS8 chip quality after storage at 4°C or 8°C. For each trait, the statistical test univariate GLM was applied with p<0.05. The feature part lists allele frequencies, exon or intron location and if the SNP leads to an aa exchange. All scored allele dosages are indicated. True allele frequency (828 alleles) and apparent allele frequency (207 genotypes) is given in %. The SNPs are sorted upon their classification into LD groups. R²: percent of the total variance explained; ns = not significant, ↑ positive effect, ↓ negative effect.

They explained between 4.9 % and 11.6 % of the phenotypic variance of the two traits. Additionally, the two SNPs 5 and 7 (position 352 and 391) showed low association with a negative effect on tuber yield. Most of the SNPs could be clearly grouped into one large LD group based on their apparent co-segregation. Interestingly, only few SNPs contained the triplex allelic state of the associated allele, whereas the quadruplex state was clearly identified (Figure 48). The presence of one superior allele ("AAAG") did not result in significantly higher TSC and TSY, whereas a higher allelic dosage (duplex "AAGG" or quadruplex "GGGG") significantly increased the starch content and starch yield.

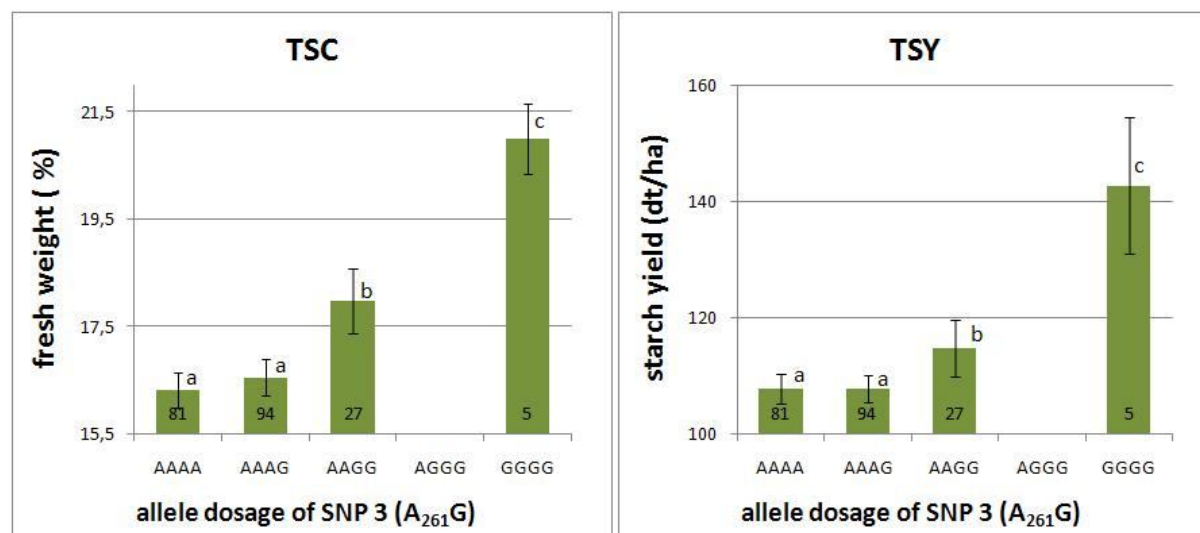


Figure 48: Means plots of SNP 3 (A₂₆₁G) of the putative invertase inhibitor.

The X-axis represents the different allelic states. The Y-axis shows the phenotypic data for the traits TSC and TSY. Significant differences between the different groups (allelic states) are indicated by small letters. Different letters symbolize different groups, same letters stand for the same group and two letters indicate no clear classification to one of the groups.

All reference alleles (identical to locus 1421 in *S. phureja*) belong to subgroup 1421, whereas all associated variant alleles belong to subgroup 1422. The only exception from the co-segregation pattern is SNP 8 (position 395) which was identified to be characteristic for the member P4G5 of subgroup 1422. It shows a low frequency of around 2%, but has a very high association and explains up to 9 % of the phenotypic variance. SNP 10 is a triple SNP and explains up to 11.3 % of the phenotypic variation of tuber starch content. Interestingly, the variant allele "C" (↑) is present in subgroup 1422 and fits to the co-segregation pattern of the other SNPs, but the variant allele "A" (↓) has not been identified before and is therefore not characteristic for one of the subgroups. No characteristic SNP for subgroup 1423 was detected and therefore was not included in the association analysis.

3.5.5 HAPLOTYPE MODELING AND ASSOCIATION

Previous results indicated that the subgroups 1421, 1422 and 1423 are allelic variants. As several sequences of Kunitz-type protease inhibitors have been cloned (Heibges et al., 2003a and 2003b),

possible haplotypes can already be concluded. Now, the SNPs of the association analysis were used to calculate haplotypes and compare them with the cloned cDNA sequences. The calculation was performed with the software SATlotyper and included eight associated SNPs (excluding the indel and the triple-SNP). All 15 haplotype models obtained are listed in table 28. The most frequent haplotype model is identical to subgroup 1421 (no 2). It is assumed to be present in 81 % of the genotypes, whereas the previous amplifications in the BIOSOL population resulted in an allele frequency of 95 % (Figure 43). The haplotype model number 14 corresponds to subgroup 1422. Its apparent allele frequency is clearly lower (14%) compared to the amplification results for 1422 (55 %; Figure 43). A haplotype model which is identical to subgroup 1422 member P4G5, is haplotype number 15. An association analysis using the haplotype models showed association for the two haplotypes corresponding to the known true haplotypes 1422 and P4G5. An additional haplotype model (no 11) showed high association with TSY and TSY. This haplotype model shared 6 out of 8 identical positively associated SNPs. The remaining haplotypes mainly contained less than 4 of the associated SNPs and were more similar to the not associated haplotype which corresponds to subgroup 1421.

Table 28: List of haplotype models for the putative invertase inhibitor.

No	Haplotype Sequence	True haplotype frequency % (no)	Apparent haplotype frequency % (no)	Allelic states n/s/d/t/q	Haplotype association
2	CAGCCAAG	60.6 (502)	80.7 (167)	40/24/26/42/75	-
13	GAGCCAAG	5.4 (45)	21.3 (44)	132/43/1/0/0	-
4	CAGCCTAG	5.3 (44)	21.3 (44)	163/44/0/0/0	-
7	CGGCCAAG	5.1 (42)	20.3 (42)	165/42/0/0/0	-
14	GGTGGTAG	4.1 (34)	14.0 (29)	178/27/0/1/1	TSC 0.047 3.7 % ↑ TSY 0.021 4.4 % ↑
9	CGTCCAAG	3.6 (30)	13.5 (28)	179/26/2/0/0	-
6	CAGGCAAG	3.5 (29)	14.0 (29)	178/29/0/0/0	-
8	CGGGGAAG	2.8 (23)	11.1 (23)	184/23/0/0/0	-
5	CAGCGAAG	2.7 (22)	10.6 (22)	185/22/0/0/0	-
12	CGTGGTAG	2.7 (22)	10.6 (22)	185/22/0/0/0	-
1	CAGCCAAA	1.3 (11)	5.3 (11)	196/11/0/0/0	-
3	CAGCCAGG	0.8 (7)	3.4 (7)	200/7/0/0/0	-
10	CGTGCTAG	0.8 (7)	3.4 (7)	200/7/0/0/0	-
11	CGTGGAGA	0.6 (5)	2.4 (5)	202/5/0/0/0	TSC 0.000 5.8% ↑ TSY 0.001 5.1% ↑
15	GGTGGTGA	0.6 (5)	1.4 (3)	204/2/0/1/0	TSC 0.006 4.8% ↑ TSY 0.003 5.3% ↑

The haplotype models which fit to the cloned cDNA sequences (1421, 1422 and P4G5) are highlighted in grey. Blue letter: associated reference SNP; Green letter: associated SNP with positive effect. True allele frequency: Number of haplotypes among possible 828 haplotypes (4 x 207 genotypes); Apparent allele frequency: Number of genotypes containing the haplotype. Nulliplex (n), simplex (s), duplex (d), triplex (t) and quadruplex (q): Allelic states of haplotypes per genotype. Positive effect: ↑, negative effect: ↓.

3.5.6 FUNCTIONAL CHARACTERIZATION OF THE PUTATIVE INVERTASE INHIBITOR

To investigate whether an interaction exists between the inhibitor candidates and acid invertase (*PAIN-1*), different interaction assays were set up. An initial yeast-2-hybrid assay did not show any interaction, but this may be due to problems in the expression or transformation steps. Invertase and inhibitor specific antibodies did not detect proteins of the expected size in performed western blots.

The subsequent experiment was based on heterologous expression in *Pichia pastoris*. In total, three invertase alleles (Draffehn et al., 2010) and three inhibitor alleles were selected for the analysis (Table 29). The inhibitor alleles represent all three subgroups 1421, 1422 and 1423 of Kunitz-type protease inhibitors group C. All sequences were cloned into a secreting expression vector (pPICZαA, Invitrogen). The inserts were cloned in frame with the two tags c-myc and histidin₆ which add approximately 2.5 kDa to the C-Terminus of the final protein. The N-terminal fused secretion signal is supposed to be cleaved and does not contribute to the size.

Table 29: Invertase and inhibitor alleles used for cloning and heterologous expression.

Candidate gene	Association	Reference
<i>Pain-1-Sa</i>	Associated allele with the traits TSC, TSY, CQA and CQS4	Draffehn et al., 2010
<i>Pain-1-Sb</i>	Basic allele	Draffehn et al., 2010
<i>Pain-1-Da</i>	Associated allele with the traits TSC, TSY, CQA and CQS4	Draffehn et al., 2010
1421	Basic allele	Glaczinski et al., 2002
1422	Associated allele with the traits TSC and TSY	Glaczinski et al., 2002
1423	Not known	Heibges et al., 2003a and 2003b

Two of the cloned invertase alleles showed association with starch and chips quality traits (Li et al, 2008). The association of 1422 with the traits TSC and TSY was shown in the previous section.

The six constructs were transformed into strain GS115 of *P. pastoris*. A high amount of expressed proteins was detected after three days of growing and optimized upon these results. The supernatant of the medium including the secreted protein was harvested and used for initial western blots and preliminary interaction tests. The c-myc antibody detected the three invertases with an approximate size of 90 kDa and the three inhibitors with a size of around 24 kDa (Figure 49). The specific antibodies for the invertases and inhibitors showed the same result, respectively. The same bands were obtained for protein extracts which were partially purified by ammonium sulfate precipitation or by affinity purification (HiTrap affinity columns, ÄKTA prime chromatography system). The latter purification was possible due to the fused histidine tag to all tested proteins.

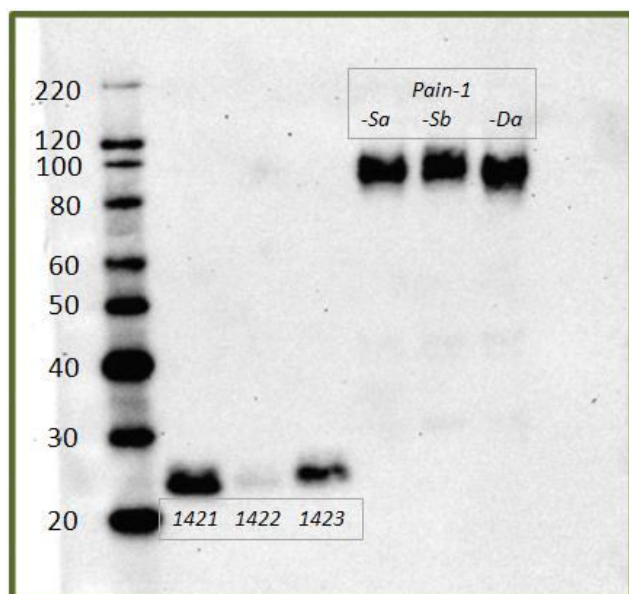


Figure 49: Western Blot of purified invertase and inhibitor alleles.
The bands were detected by using a c-myc tag antibody.

3.5.7 ACTIVITY ASSAYS

3.5.7.1 ACTIVITY ASSAY FOR HETEROLOGOUSLY EXPRESSED INVERTASE

The activity of the different invertase alleles (Pain-1-Sa/-Sb/-Da) was tested in a modified protocol of Zrenner et al. (1996). The amount of produced glucose was used to determine the activity of the invertase. Figure 50 shows that all tested invertase alleles were active and converted sucrose into glucose and fructose. As a control, almost no glucose can be detected, when invertase was heat inactivated (STOP).

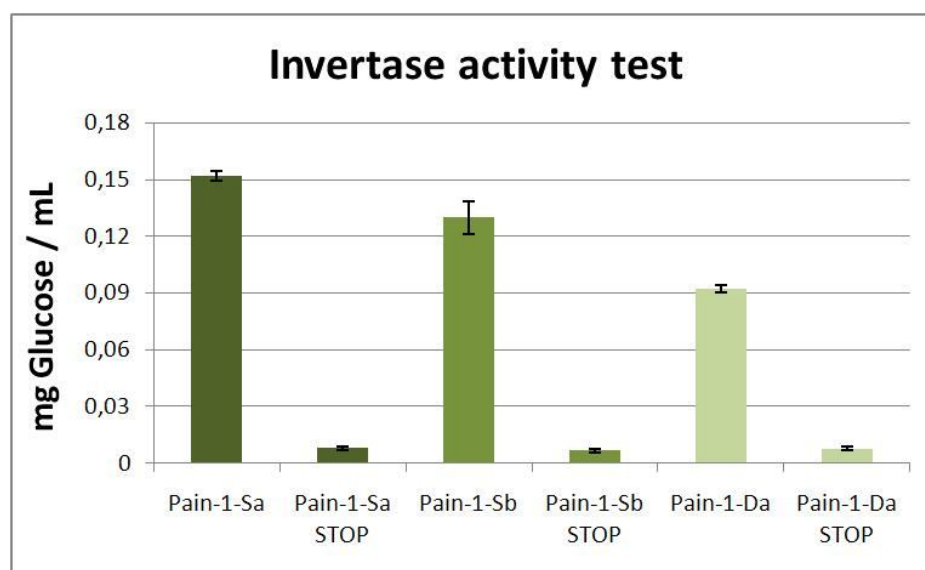


Figure 50: Analysis of invertase activity.

Invertase activity was estimated from the amount of glucose produced after incubation of heterologously expressed invertase with sucrose.

3.5.7.2 ACTIVITY ASSAY FOR HETEROLOGOUSLY EXPRESSED PUTATIVE INVERTASE INHIBITOR

The activity analysis of the three inhibitor alleles was tested by a trypsin inhibition assay (Heibges et al., 2003b). The three purified inhibitors were applied in the molar ratios 1 nM inhibitor : 3.5 nM trypsin and 1 : 1.75. The presence of the three inhibitors resulted in a lower trypsin activity compared to untreated trypsin (Figure 51). The doubling of the molar concentration of the inhibitors resulted in a stronger inhibition of trypsin.

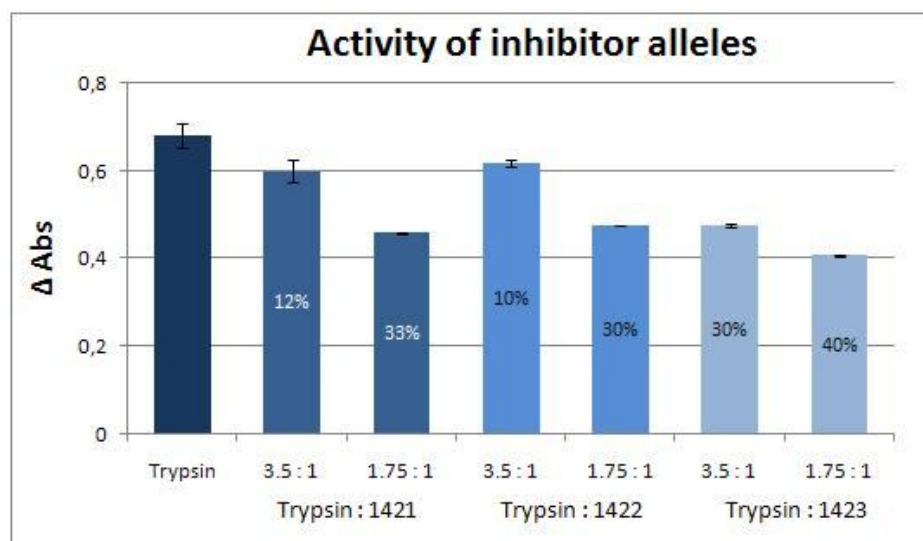


Figure 51: Inhibitor activity of the heterologously expressed Kunitz-type protease inhibitors. The three inhibitor alleles were tested in two molar ratios for trypsin inhibition.

3.5.7.3 INVERTASE INHIBITION ASSAY

The inhibition assay was performed by using equal molar amounts (1 x inv/1 x inh) of invertase and inhibitor in all possible combinations. Additionally, the molar ratio 1nM invertase : 5 nM inhibitor was tested (1/5). No significant inhibition can be detected for any of the combinations (Figure 52).

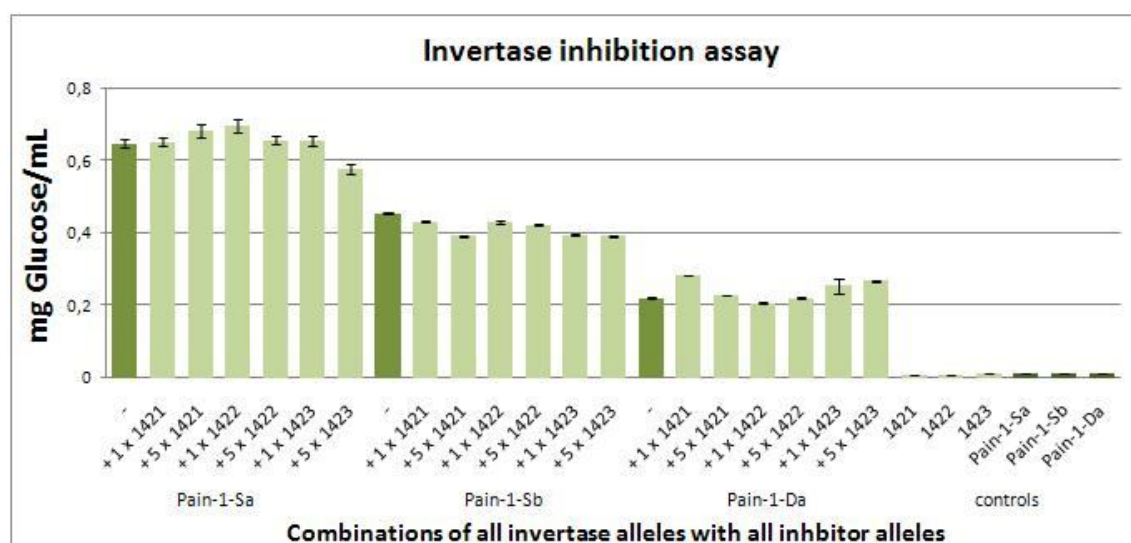


Figure 52: Inhibition assay of invertase alleles.

The three invertase alleles were incubated with an equimolar or a fivefold molar amount of inhibitor. Inhibition was tested between all three invertase and inhibitor alleles.

3.5.7.4 INHIBITION ASSAY OF NATIVE INVERTASE

In order to simulate similar conditions to the invertase inhibition assay of Glaczinski et al. (2003), native invertase was partially purified from potato tubers. The tubers were stored in the cold for 2 weeks and were assumed to contain a high amount of invertase (Bagnaresi et al., 2008). The native invertase was incubated with an equimolar amount of inhibitor protein and analyzed for inhibition (Figure 53). No decrease in the amount of produced glucose was detected for native invertase in the presence of any inhibitor protein.

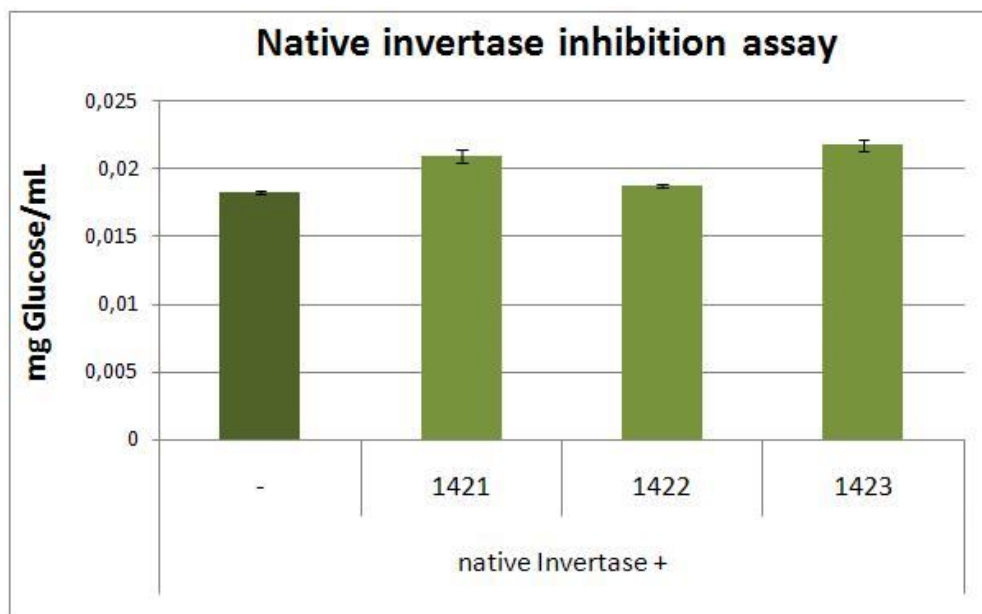


Figure 53: Analysis of purified putative invertase inhibitor on partially purified native invertase.

4 DISCUSSION

4.1 CANDIDATE GENES FOR SNP BASED ASSOCIATION ANALYSIS: FORWARD AND REVERSE APPROACHES

Candidate genes for association analyses with desired quality traits are predominantly functional and positional candidates (Byrne and McMullen, 1996). In other words, they are selected based on their known functional role in a pathway and/or their location within or close to a QTL (Pflieger et al., 2001). Dealing with quantitative traits like tuber starch content and chip quality directly implies to consider functional candidates of the carbohydrate metabolism which is among the best studied pathways in plants (Frommer and Sonnewald, 1995). Carbohydrates are synthesized during photosynthesis and predominately stored for long and short term in form of starch granules in plastids. Mobilization of starch and the conversion into active products is supposed to occur via the starch degradation pathway. Association studies on strong functional candidates of this pathway have already been reported. For instance, highly significant associations for vacuolar and cell-wall invertases and starch phosphorylases with tuber starch and processing qualities were demonstrated (Li et al., 2005; Li et al., 2008). Due to the complexity and the high number of enzymes involved in the starch degradation pathway, the influence of other functional candidates remains to be analyzed. In the following section, several functional candidates and their contribution to starch and chip quality traits will be discussed. Additionally, it will be focussed on alternative approaches to obtain novel candidates for association analysis. On the one hand, we used an *in silico* approach based on the availability of the genomic sequence of *S. phureja* and on the other hand a powerful and state-of-the-art “-omics” approach to identify and select new candidates whose involvement in the plant carbohydrate metabolism was not described before.

4.2 FORWARD CANDIDATE GENE APPROACH

The standard candidate gene approach can be considered as a straight forward approach. It is primarily based on functionally characterized genes (Byrne and McMullen, 1996). Most of these candidate genes have been identified in mutated plants which were generated by anti-sense constructs or T-DNA insertion knock-out lines. These altered plants sometimes show a very drastic phenotype, especially when important genes are completely abolished or on the contrary over-expressed. Analysis of plants displaying a desired mutant phenotype often identifies functional candidates which are involved in important pathways of plant metabolism. For several mutant loci, it

has been shown that they are in close vicinity of a QTL (Robertson, 1985; Koornneef et al., 1998) and fulfil a role as functional and as well as positional candidates.

4.2.1 GLUCAN WATER DIKINASE - A FUNCTIONAL CANDIDATE IN STARCH BREAKDOWN

The alpha-glucan water dikinase is assumed to initiate the first step of the starch degradation pathway by phosphorylation of glucose molecules in the amylopectin chains of the starch granule (Lorberth et al., 1998) and are therefore considered as strong functional candidate. It is hypothesized that the introduction of charged phosphates renders the dense starch granule more accessible for further degrading enzymes (Blennow et al., 2000; Yu et al., 2001). Abolishing the function of this enzyme via mutations or anti-sense constructs resulted in a starch excess phenotype accompanied by a lowered accumulation of reducing sugars (Lorberth et al., 1998) which coincide directly with desired characteristics of a potato tuber.

4.2.1.1 *GWD* IS A LARGE AND HIGHLY CONSERVED GENE

For the analysis of a candidate gene, certain basic information needs to be retrieved. For *GWD*, this implied to determine the genomic sequence by a BAC library screen using a probe generated on the annotated cDNA sequence (Y09533). Additionally, this method was used to interpret whether *GWD* is encoded by a single gene or whether it belongs to a gene family. Since only one positive BAC (BA202H17) was obtained, it was concluded that *GWD* is a single locus gene. Subsequent *in silico* blast analyses on the recently released *S. phureja* sequence (PGSC, 2011) confirmed the findings of a single locus for *GWD* on chromosome 5.

The genomic organization of *GWD* was established by aligning the available cDNA sequence to the genomic sequence. In total, 33 exons were identified. Comparing these data to the genomic organization of *GWD* from *A. thaliana*, a high level of conservation was detected. Both genes exhibit a similar organization (Figure 11) and a high number of exons (32 exons in *A. thaliana* vs. 33 exons in *S. tuberosum*). One exception is the separation of exon 7 and 8 in potato *GWD* which are fused in *A. thaliana* to one larger exon (exon 7). This region is assumed to be part of the linker region of the bipartite carbohydrate binding module 45 (CBM45; Glaring et al., 2011). Orthologs of *GWD* can be identified in many different plant species (Figure 12). The C-terminal nucleotide binding domain and the region surrounding the catalytic histidine displayed a considerably high level of conservation between different monocots and dicots (Figure 13). In contrast to this, rather low conservation was identified for the N-terminal part which contains the putative chloroplast target peptide and the carbohydrate binding module. For the CBM 45, it was recently shown that it is a low-affinity binding

domain which was defined as a frequent characteristic of plastidial starch enzymes (Glaring et al., 2011). Low binding affinity to the substrate is hypothesized to facilitate reversible binding and hence enables a dynamic regulation of the enzyme (Christiansen et al., 2009). Additionally, the rather low level of conservation refers to the entire carbohydrate binding domain, but putting a main focus on single amino acids, a noteworthy conservation was observed for five aromatic amino acids (tryptophan (W) and phenylalanine (F), highlighted by black arrows, Figure 13) which occur in similar repetitive pattern in both domains of the bipartite CBM. Reducing the major binding capacity to these 10 amino acid, the remaining less conserved part of the CBM including the linker region is considered on the one hand to be involved in the proper folding and on the other hand to provide specificity for the different starch types that were found in different plants (Hejazi et al., 2008). Although GWD acts theoretically on the same substrate which is commonly named starch, it has to be distinguished between different types of starch allomorphs which may require specific folding of the binding domain or the entire enzyme (Gallant et al., 1997). Besides the annotated functional domains, a high level of conservation was identified for a large and functionally uncharacterized region located between the catalytic histidine and CBM 45. Although *in silico* prediction programs did not suggest any functional role, the high degree of conservation is noteworthy and needs to be further characterized. One possible speculation about this region might be that it harbours an interaction domain for homodimer formation as GWD might be part of a protein complex (Mikkelsen et al., 2004). Besides GWD homodimers, recent hypotheses point to various protein complexes and large enzyme assemblies of starch metabolic enzymes (Tetlow, 2006).

4.2.1.2 *GWD* IS PREDOMINANTLY ASSOCIATED WITH STARCH QUALITY TRAITS

Due to the availability of the genomic sequence, it was possible to generate an amplicon for the association analysis which covers a large and important region of the candidate gene. The chosen amplicon was located within the linker region of the bipartite starch binding domain (CBM 45) and covered exons 7 and 8. Association analysis was performed in the GABICHIPS population which has been phenotyped for tuber starch and chip quality traits (Li et al., 2008). Interestingly, most of the associations identified for *GWD* were found with tuber starch content and tuber starch yield. The presence of superior alleles explained between 3.3 % and 5.9 % of the phenotypic variance of these traits (Table 11). In general, rather low association was identified with chip quality traits.

GWD represents a strong functional candidate gene in starch breakdown and its possible contribution to the natural variation of starch content in tubers was clearly shown by the association analysis. The presence of a superior allele (LD group 1; Table 11) leads to significantly higher tuber starch content and tuber starch yield. Whether the allele dosage has an additional effect, remains

inconclusive and needs to be tested in a different or larger population. In the GABICHIPS population, 44 genotypes were identified to contain the superior allele of LD group 1 in simplex and only 7 genotypes in duplex condition (Figure 14). Due to the small amount of duplex genotypes, it cannot be concluded whether the presence of two superior alleles leads to even more starch, although a slight trend can be observed (Figure 14). The same was observed for SNP A₄₈₂₆G (SNP 7, Table 11), whilst the variant allele "G" is a rather low frequent allele and occurs only in two genotypes in duplex state. The variant allele showed a slightly negative association with tuber yield but a positive association with tuber starch content and tuber starch yield. This SNP explained around 6 % of the phenotypic variation and causes a non-synonymous amino acid exchange. Only moderate positive associations with chip quality traits were detected for a single SNP (SNP 8) which was not in LD with any of the other SNPs. In summary, this association data suggests a rather low contribution of *GWD* to the natural variation of chip quality traits, but a significant effect on tuber starch quality traits.

Regarding the influence of superior *GWD* alleles to tuber starch content, it can be speculated that the associated alleles cause amino acid exchanges in the linker region of the starch binding domain which may lead to an altered conformation of the protein and result in a less efficient binding to the starch granule (e.g. SNP 5: aa exchange P/A). Consequently, the degree of phosphorylation would be reduced and starch will be less degraded. For *GWD*, many of the associated SNPs cause non-synonymous amino acid exchanges (Table 11). Moreover, most of them are in LD and may affect collectively the protein conformation. Besides direct effects of the identified SNPs on the folding capacity of the starch binding domain, it is possible that these SNPs are in linkage disequilibrium with SNPs residing in other functional domains. For instance, non-synonymous amino exchanges resulting from allelic variation in the nucleotide binding domain may lead to a reduced capacity to bind to ATP which is needed as phosphate donor in the phosphorylation reaction. As a consequence, the level of phosphorylation will be reduced and lead to higher starch content due to reduced accessibility of the starch granule for degrading enzymes (Kötting et al., 2005). Based on the complexity of carbohydrate metabolism and other sources for sucrose which may compensate partially for lowered starch degradation, it may be concluded that the effect of *GWD* on chip quality traits is balanced and not measurable. Additionally, it was hypothesized that phosphorylation is needed for starch degradation as well as for starch biosynthesis (Nielsen et al., 1994). The identified superior *GWD* alleles for tuber starch content and tuber starch yield may facilitate the synthesis of starch and therefore result in higher tuber starch content. In this respect, no direct effect on the chip quality would be expected.

However, the effects of anti-sense repression and subsequent complete silencing of this gene are likely to result in a more drastic and obvious phenotype than natural variants. For example, the complete knock-out of *GWD* in *A. thaliana* caused in addition to a starch excess phenotype a

retarded growth phenotype (Yu et al., 2001; Kötting et al., 2005). Considering the crucial role of GWD as an initiator of starch breakdown, it is unlikely that disadvantageous pleiotropic natural variations within this key gene (e.g. impaired starch degradation capacity resulting in impeded growth and development) are maintained. In summary, for marker-assisted breeding, interesting alleles were identified which fit in to a set of markers to screen for appropriate parents in a breeding population.

4.2.2 PHOSPHOGLUCAN WATER DIKINASE – NUMBER 2 IN STARCH DEGRADATION

Phosphoglucan water dikinase (*PWD*) was selected as a second functional candidate. It has been independently identified by two groups as *GWD3* in *A. thaliana* (Kötting et al., 2005; Baunsgaard et al., 2005). Similar to *GWD*, the functional characterization resulted in a dikinase reaction type which phosphorylates mainly the C³ position of amylopectin (Kötting et al., 2005). The authors concluded that *PWD* acts directly downstream of *GWD* as pre-phosphorylated starch was determined as the substrate needed for *PWD*. On the one hand, the *A. thaliana* mutant of *PWD* displayed a clear starch excess phenotype which was on the other hand not as strong as observed for *GWD*. Additionally, the synthesis of starch during the day was slightly reduced (Baunsgaard et al., 2005). This weaker starch excess phenotype was accompanied by a less severe retarded growth phenotype (Kötting et al., 2005; Baunsgaard et al., 2005). Due to the interesting phenotype with a certain similarity to *GWD* and the essential position in the starch degradation pathway, the Phosphoglucan water dikinase was selected as a functional candidate for an association analysis.

4.2.2.1 *PWD* IS LOCATED CLOSE TO A CHIP QUALITY AND STARCH QTL ON CHROMOSOME 9

PWD has only been described in *A. thaliana* (*AtPWD*), for potato no sequence annotation was available. Via blast analysis of the *AtPWD* cDNA, a putative 900 bp EST (CV431165) of potato *PWD* (*StPWD*) was obtained. Interestingly, this cDNA was annotated as “after-cooking darkening *S. tuberosum* cDNA clone 55202” (Flinn et al., 2004, direct submission to NCBI). Based on this sequence information, a BAC library (Ballvora et al., 2002) was screened and resulted in six positive BACs. Due to additional *in silico* studies in different databases, a tomato BAC sequence (C09HBa0226D21) was retrieved which served as initial template for primer design. Via PCR amplifications on the six potato BACs, the presence of the full-length genomic sequence of *PWD* was confirmed for all of them. Instead of tedious primer walking, further primer pairs were directly designed on the tomato BAC sequence and were used for the amplification and sequencing of the genomic and the coding sequence of potato *PWD*. The generated cDNA sequence was later compared to a recently submitted cDNA (GU045560) and was identified to be identical (Grabowska et al., 2010, direct submission to

NCBI). This sequence confirmed that we identified and characterized the correct potato *PWD* cDNA sequence. Additional information was deduced from the tomato BAC which was mapped to the long arm of tomato chromosome 9 which is syntenic with potato chromosome 9 (Gebhardt et al., 1991). After the release of the genomic sequence of *S. phureja* (PGSC, 2011), it was possible to map the non-coding SSR Marker STM3023b to the same superscaffold as *StPWD*. For this marker, an association with chip colour and tuber starch content was published by Li et al. (2008). On the basis of the close proximity to an associated marker and the relevant role in the starch degradation pathway, *PWD* was considered as functional and positional candidate gene.

4.2.2.2 *PWD* IS HIGHLY CONSERVED

Although only a low degree of similarity was identified for the amino acid sequence of *StPWD* (~60%) to the annotated *A. thaliana* amino acid sequence, the genomic organizations of both *PWD* sequences displayed remarkably high conservation (Figure 18). Both genes contain 19 exons and show similar lengths on entire cDNA level (Table 13). By screening databases, only few homologous sequences from other species were identified, that indicates that only little research was done on plant *PWDs* so far. The comparison of the *PWD* amino acid sequences showed highly conserved blocks, mainly representing the annotated functional domains (nucleotide binding domain, catalytic histidine). As previously shown for *GWD*, the starch binding domain displays more divergence which may be due to different types of starch present in the different species (Buléon et al., 1998). Additionally, the prediction of a chloroplast target peptide at the N-terminal part supports the putative chloroplastic location for *PWD*.

4.2.2.3 *PWD* SHOWS HIGH ASSOCIATION WITH STARCH AND CHIP QUALITY TRAITS

The initially identified EST sequence of *PWD* was characterized as the largest and last exon of the coding sequence. After the assignment of the putative functional domains, it turned out that this part of the sequence comprises the important nucleotide binding domain and was therefore selected for an association analysis of *PWD* with starch and chip quality traits. For *PWD*, several highly significantly associated SNPs were identified (Table 14). The highest association was found with chip quality traits for two SNPs in LD (LD3). For instance, these SNPs explained up to 7.7 % of the phenotypic variance for chip quality after storage at 4°C. The presence of the superior SNP alleles lead to a significantly higher chip quality score (Figure 22). Interestingly, these SNPs are associated in an opposite effect with tuber yield. These negative correlations were observed as well for other candidate genes and may result from a certain degree of limitation within a plant by combining too

many desired traits. In other words, increasing tuber starch content and chip quality may be at the expense of tuber yield. Although the presence of one of these superior alleles already improved some of the quality traits, no conclusions can be drawn on the effect of the allele dosage as only two genotypes carried the superior allele in duplex state. Further association studies in a larger population would be necessary to state the influence of the allele dosage. Besides the strong association with chip quality traits, associations with tuber starch content and tuber starch yield were found for LD group 3 (SNP 7 and SNP 8). They explained between 4.1 % and 7.1 % of the phenotypic variance (Table 14). In general, *PWD* showed higher sequence variation and higher contribution to the tested traits compared to *GWD* especially to chips quality traits. The reason for these observations might rely on the more important role of *GWD* in the initiation of the starch breakdown, whilst *PWD* acts only downstream of *GWD* and is less regulated. This is supported by the fact that *PWD* knock-out mutants showed an overall less severe phenotype (Kötting et al., 2005) and are not impaired or retarded in development or growth. Phosphorylation events mediated by *PWD* are probably an additional effect but not crucial and therefore more variation of the enzymatic performance is admitted. So far, *PWD* has not been described to play a role in starch synthesis. Consequently, it does not have an ambivalent function and allelic variants may contribute more to the end products of starch degradation and may have a stronger effect on chips quality traits compared to *GWD*.

4.3 CANDIDATE RECRUITMENT BASED ON *IN SILICO* ANALYSES - DISCOVERY OF INVERTASE 6

In general, a candidate gene approach relies on functional genes, whose elementary role in a metabolic pathway has been analyzed. A more novel and subtle approach is to test non-described homologs of well-characterized functional candidates. Blast and similarity searches can be used to find these homologs or isoforms which cannot be recognized by functional analysis due to a lack of obvious mutant phenotypes. Additionally, the availability of various databases and the release of the *S. phureja* genome sequence facilitate to combine *in silico* analysis tools and a broad range of data to map genes to distinct loci and directly link them to QTLs.

Invertases were considered as interesting candidates for such an *in silico* approach for various reasons. First of all, the group of neutral invertases has only been rudimentary characterized in *A. thaliana* and in few other organisms (Sturm, 1999). So far, neutral invertase has not been identified in *S. tuberosum*. Additionally, for vacuolar and cell wall-bound invertases it has been shown that they are highly associated with starch and chip quality traits (Li et al., 2005; Li et al., 2008; Draffehn et al., 2010). Invertases catalyze the irreversible hydrolysis of sucrose into the reducing sugars glucose and

fructose. They are part of the sink-source metabolism and strongly connected to sucrose signaling and the starch degradation pathway. The identification of new invertase homologs would raise questions about its classification, its conservation and its contribution to tuber quality traits.

The 6th invertase (*Inv6*) is a result of an *in silico* analysis of the potato genome sequence for new invertase sequences (Gisella Orjeda, unpublished data) and was used as an example in this thesis to link the standard candidate gene approach to a more novel one. In other words, this optimization of candidate gene selection is possible due to the tremendous gain of sequence information by the release of the potato genome sequence (PGSC, 2011) and improved *in silico* analysis tools.

For *Inv6*, it was possible to pinpoint the genomic localization to the lower arm of chromosome 8 by combining the released genomic sequence of *S. phureja* and by blasting markers with known positions in the PoMaMo maps against it to identify those which are in the same scaffold as the candidate gene (Table 17; Appendix Figure A.3). Analysis of the locus and its position on the map revealed, that *Inv6* co-localizes with a tuber starch QTL and is additionally in close proximity to a sugar QTL (Schäfer-Pregl et al., 1998; Menéndez et al., 2002). Therefore, *Inv6* was considered as a strong positional candidate for starch quality traits.

4.3.1 INV6 SHOWS STRONG HOMOLGY TO THE VACUOLAR INVERTASE PAIN-1

Comparison of the *Inv6* cDNA sequence with all known invertases from *S. tuberosum* clearly identified the highest similarity to the vacuolar invertase *Pain-1*. Detailed analyses resulted in an almost identical peptide sequence length and in a very similar genomic organization (Figure 23) for both invertases (Draffehn et al., 2010; Liu et al., 2010). In contrast to the high structural conservation, the amino acid sequences showed only 56.7 % identity (71.3 % similarity). In the sequence alignment (Figure 24), the uppermost conservation was observed for the annotated catalytic glycosyl hydrolase domains and the corresponding characteristic and important motifs within. The N-terminus exhibited the lowest conservation which may presume different subcellular localizations for both proteins. Motif prediction programs predetermined clear cleavage sites of putative signal peptides for *Inv6* as well as for *Pain-1*, but they also suggested an identical localization despite low sequence similarity. For both invertases, localization was assumed either in the vacuole or in the cytoplasm. In addition, one program even predicted a transmembrane helix in the N-terminal part of *Inv6* and *Pain-1* that would imply a membrane-bound localization. Eventually, *Pain-1* and *Inv6* are bound to the membrane of the vacuole (tonoplast-located) and act either in one or in both compartments. Despite the availability of various prediction tools which are based on different algorithms and databases, it was not possible to obtain an unambiguous localization for *Pain-1* and

Inv6. Interestingly, Pain-1 is used as one of the reference peptides for the vacuolar target peptide in various prediction tools (e.g. PSORT, Horton et al., 2007) without being completely proven to be restricted only to the vacuole. One single study on the subcellular localization of vacuolar invertase was done by Bhaskar et al. (2009) and resulted in Pain-1-GFP signals in the cytoplasm, the ER and the vacuole. To state whether Inv6 and Pain-1 have an identical or different location, further localization experiments are required.

The search for putative Inv6 homologs in other species resulted in only few sequences. Most of them displayed a higher similarity to Pain-1 and were consequently excluded from the Inv6 peptide sequence alignment (Figure 25; Appendix A.13). Despite a comparison of few sequences retrieved from very different plant families, the overall conservation at amino acid level was very high. Therefore, it can be assumed that this 6th invertase is most likely present in many different families of the eudicotyledons and not unique for solanaceous species. For further information of Inv6 in other species, more genomic sequence data and adequate databases and research tools are needed.

4.3.2 *INV6* IS HIGHLY ASSOCIATED WITH STARCH AND CHIP QUALITY TRAITS

The genomic structure and the knowledge about characteristic and important functional domains of *Inv6* which were deduced from *Pain-1*, helped to choose a region for an association analysis on starch and chip quality traits. Genomic fragments were amplified, including a large part of the N-terminal part of the glycosyl hydrolase domain which surround the catalytic motif WECVD (Ji et al., 2005). Within this region, ten associated SNPs were identified (Table 19). Most of them showed high association with multiple traits or in one case with all six tested traits (SNP 4, position 2134). As discussed above, the effect of the association on tuber yield was always opposed to the other traits (section 4.4.3). The SNP with the strongest association (SNP 7) was in LD with two other SNPs (SNP 1 and 2). These 3 SNPs explained up to 11% of the phenotypic variance for tuber starch content and up to 10.4 % of the variance for chip quality traits. It was observed that the degree of association decreased with increasing cold-treatment (CQA: 0.000001, CQS8: 0.00002, CQS4; 0.0002). These highly frequent SNPs were present in all possible allelic states and significant differences were detected between some of them (Figure 27). For example, for CQA and CQS8 the two homozygous (quadruplex states) and the duplex state are significantly different from each other, implying that a higher dosage of the superior alleles has a supplementary positive effect. Due to the strong associations, all SNPs were subsequently analyzed in the 40 cultivars of the BIOSOL population which was phenotyped for reducing sugar content (glucose and fructose) and their accumulation upon cold-storage. This analysis was performed to observe whether *Inv6* alleles have a direct effect on the amount of reducing sugars. No clear effect on the basic sugar amount (time point 0 = no cold

storage) was detectable for the presence or absence status of the superior alleles because the superior alleles were detectable in almost all genotypes. Regarding the allelic dosage of the associated alleles, clear differences were observed (Figure 28 and 29) supporting the previous assumption that the dosage of superior alleles has a clear effect on the amount on reducing sugars. The association analysis in the BIOSOL population confirmed that the presence of the superior alleles in higher dosages is significantly higher in the good chipping genotypes compared to the bad chipping genotypes (Figure 30). Regarding the accumulation of reducing sugars upon cold storage, it was observed that the ratio for highly associated alleles (number of superior alleles in good chipping genotypes : number of superior alleles in bad chipping genotypes) changes only slightly in the other time points (1, 2, 4 and 12 weeks of cold-storage). It was concluded that certain superior alleles of *Inv6* contribute clearly to higher starch content and therefore to better chip quality due to a lower amount of reducing sugars. These low sugar conditions are maintained to certain degree during cold-storage. As *Inv6* showed high association with the analyzed traits, this enzyme of the invertase family contributes most likely to carbohydrate metabolism and in detail to the amount of reducing sugar accumulation.

4.3.3 *INV6* SHOWS ONLY LOW EXPRESSION IN TUBERS

To get an impression of the function of *Inv6*, a semi-quantitative expression analysis was done. *Inv6* showed very strong expression in flowers and flower buds. No expression was detected in leaves which was in accordance to recent findings of Liu et al. (2010). Only a low expression level was detected in tubers (Figure 31) which rather slightly increased upon cold-storage when compared to *Pain-1* transcript induction in the cold (Zrenner et al., 1996; Liu et al., 2010). Although the expression pattern gives a first indication for tissue specificity, further investigations are needed to reveal the role of *Inv6* in the potato carbohydrate metabolism. Additionally, it has to be demonstrated that this locus encodes a functional protein which can convert sucrose and where it is located exactly in the cell. Further insights might be obtained by analyzing whether *Inv6* displays a mutant phenotype by specific anti-sense constructs. Due to the high similarity to *Pain-1*, it would be interesting to identify whether they can compensate for each other or even interact. Finally, it remains to be revised, whether the location within a starch QTL may also be the reason for the high degree of association or vice versa, if this is the gene which is responsible for the starch QTL.

4.3.4 HAPLOTYPE MODELING IS A POWERFUL TOOL, BUT REMAINS ARTIFICIAL

Haplotype calculation with the software SATlotyper (Neigenfind et al., 2008) was executed for all analyzed "forward" candidate genes (*GWD*, *PWD* and *Inv6*). In general, the reference sequence which

either corresponded to the sequence of *S. phureja* or to a sequence obtained from the BAC sequencing, was always calculated as one of the most frequent haplotypes (Table 12, 15 and 20). In the case of *GWD*, the BAC sequence and the sequence of *S. phureja* represented different haplotypes, but both of them were calculated and confirmed with the software. Interestingly, the frequency of the haplotype corresponding to the sequence of *S. phureja* was only 4.7 % and would have been considered as an artefact without the knowledge of its true existence. However, it can be seen that SNP scoring errors directly lead to new haplotype models (Sattarzadeh et al., 2006) because most of the haplotype models only differ in one or two SNP positions and show a low frequency (e.g. Table 15: haplotypes no 14 and 15, Table 20: haplotypes no 5 and 6). It is noteworthy that the calculations never resulted in a haplotype which contained all superior alleles. The calculated haplotypes mainly displayed only half of the superior alleles (*Inv6*: 4 out of 6, *GWD*: 1 out of 4, *PWD*: 3 out of 5).

A more detailed haplotype modeling including association studies was done for three associated SNPs from *Inv6* which were assumed to form a LD group based on co-segregation and associations with identical traits (Table 19; SNPs 1, 2 and 7). But the haplotype model which contained all three superior alleles, did not show any association (Table 21; ATT). Association was only obtained for other haplotype compositions which consisted either of the superior allele of SNP 1 or of the two superior alleles of SNP 2 and SNP 7 (Table 21 and 22). This indicates that only SNP 2 and 7 were in true LD. It can be observed that the presence of all three SNP alleles or a combination of SNP 1 with one of the SNPs from the LD group abolishes the positive effect of each other and results in no association. It can be concluded that the haplotype modeling only gives indication about possible haplotypes; the existence has to be proven by cloning approaches.

4.4 REVERSE CANDIDATE GENE APPROACH

In general, candidate gene approaches are performed on functional genes known to be involved in metabolic pathways or on positional gene which co-localize with a QTL. A novel approach is to test homologs of well characterized functional candidates which are obtained for example of *in silico* analyses. In the following section, an even more progressive approach to identify candidate genes will be discussed which can be considered as a reverse approach. Disregarding the knowledge of function and position, candidates are identified upon different protein levels between a group of cases and controls which can be easily detected on 2D protein gels. These differences in protein intensities are reproducible and give an image of the genome translated in any environment or stress (de Vienne et al., 1999).

4.4.1 LAP - COMPARATIVE PROTEOMICS REVEALED A NOVEL CANDIDATE GENE

New and broad data on natural variation is more and more obtained from various “-omics” approaches (Urbany et al., 2011). Populations comprising a set of “cases” and “controls” for a certain trait of interest are compared at transcript, metabolic or protein level. A 2D proteomics assay of the BIOSOL population revealed several differentially expressed proteins between two phenotypically different groups. The groups were classified upon the amount and the accumulation of reducing sugars which were measured at different time points including 0 to 12 weeks of cold-storage. Among obvious candidate genes which have been expected due to their known functional role, some unpredictable and non-pathway related candidates were obtained. One interesting candidate is the leucine aminopeptidase (*LAP*) which was present in the pool of five superior chipping genotypes (“cases”) and absent in the pool of five inferior chipping genotypes (“controls”; Figure 32). The peptide sequence was deduced by MALDI-TOF and identified to share more homologies to the *LAP* which is designated as acidic or chloroplastic *LAP* (*LAP_A*). This protein was initially identified after wounding events in potato (Hildmann et al., 1992) and is assumed to be unique for solanaceous plants. It was shown that *LAP_A* is induced after wounding by jasmonic acid or as a response to pathogen attack (Pautot et al., 1993; Gu et al., 1996). Further induction of *LAP_A* was identified upon water deficiency and high salinity (Chao et al., 1999). The neutral *LAP* (*LAP_N*) was identified as constitutively expressed gene in various plant species and is assumed to play a role in the proteolytic events in common cell maintenance (Chao et al., 2000). Regarding the functional role of *LAP*, pictured by previous studies, no clear link between the amount of reducing sugars and the presence in better chipping genotypes can be seen.

4.4.2 *LAP* IS ASSOCIATED WITH STARCH AND CHIP QUALITY TRAITS

On 2D gels, *LAP* was only identified in the group of five genotypes with low sugar accumulation (good processing genotypes). In order to analyze an effect on starch and chip quality traits, an association analysis in the GABICHIPS population was performed. Indeed, several SNPs were detected in the association analysis which are associated with starch and chip quality traits. Very strong association ($p = 5E^{-9}$) was obtained for one SNP (position 2746) which explained in this population 17.5 % of the phenotypic variation for tuber starch content and up to 10.5 % of chip quality after storage at 4°C. This SNP does not lead directly to an amino acid exchange. It can be speculated that either modifications in the mRNA sequence may influence the stability of the mRNA itself or changes the folding ability of a pre-mature protein (Chamary et al., 2006; Nackley et al., 2006; Kimchi-Sarfaty et al., 2007). Alternatively, this SNP may be in linkage disequilibrium with other SNPs that lead to non-synonymous amino acid exchanges and may influence protein folding or activity. These exchanges

may be located either in one of the functional and crucial catalytic sites and domains of LAP or even in the promoter region which regulates the expression level. This associated SNP was additionally analyzed in the BIOSOL population regarding the traits amount of reducing sugars and the corresponding accumulation upon cold-storage. This particular highly associated allele was significantly more frequent in the genotypes possessing less reducing sugars (Figure 38 and 39). Statistical calculations showed that the amount of reducing sugars was significantly lower in genotypes containing the superior allele (Figure 40). Regarding the accumulation of sugars upon different weeks of cold-storage, the level of significance increased (from $p < 0.01$ to $p < 0.001$). Without having a clear function in carbohydrate metabolism, interesting and strong associations were found for starch and chip quality traits for LAP and can be used as diagnostic SNP marker.

4.4.3 STRUCTURAL ANALYSIS OF *LAP* HELPS TO SPECULATE ABOUT THE FUNCTIONAL ROLE

In order to get a more detailed picture of *LAP* and its possible function in the carbohydrate metabolism of potato tubers, the genomic location and structural organization of the leucine aminopeptidase in *S. tuberosum* was investigated. First of all, *LAP* was identified to be located in a tandem repeat on chromosome 12 (Appendix Figure A.5). With the help of *in silico* blast analysis of known markers from the PoMaMo database, the exact position was pinpointed to a region at the very distal part of a starch QTL and in close proximity to a sugar QTL (Schäfer-Pregl et al., 1998; Menéndez et al., 2002). This adds the attribute positional candidate to a proteomics derived candidate. The particular case of a tandem repeat organization is likely evolved from gene duplication and may indicate a specification for a certain role.

From protein databases, information was obtained about conserved catalytically domains and putative localization signals. The catalytic domain can be classified as M17 peptidase family (Matsui et al., 2006). They form hexameric exopeptidases and mainly hydrolyze peptide and amino acyl substrates with N-terminal leucine residues (Herbers et al., 1994). The regions of the catalytic domains display a very high level of conservation, although the amino acid sequences originate from species of very divergent plant groups. The prediction servers indicated a putative localization to the chloroplast which shows rather low conservation. The prediction of a chloroplast target peptide is in agreement with recent *in vitro* studies on the localization of *LAP_A* and *LAP_N* done by Narváez-Vasquez et al. (2008). They hypothesized that it is not the primary sequence which defines a transit signal, but the structure. Regarding the putative location in the chloroplast and the peptidase activity, it can be assumed that the *LAPs* are either directly involved as post-transcriptional regulators of metabolic enzymes of the carbohydrate pathways or indirectly and act as part of signal cascades and regulate a general protein turn-over upon certain stresses and signals. For instance, it was shown

that *LAP_A* can be induced by wound signals like methyl-ester jasmonate and systemine. By Kovac and Ravnkar (1998) it was shown that sucrose and jasmonic acid interact in some events during plant development and affect photosynthesis pigmentation and root development. Chao et al. (1999) showed induction of *LAP_A* upon osmotic stresses as water deficiency and salinity. Regarding the fact that cold-storage is considered as well as a form of osmotic stress, it is hypothesized that *LAP_A* mediates the accumulation of free amino acids to compensate for the osmotic stress in the vacuole. Similar to elevated levels of reducing sugars upon cold-storage, an increase in free amino acids was observed in plants adopting for cold hardiness (Draper, 1975)

For the strongly associated leucine aminopeptidase, no direct functional role in the starch metabolism has been identified. However, the possibility to detect regulatory enzymes by comparative proteome analysis is assumed to be a powerful and reliable tool (de Vienne et al., 1999). These enzymes mainly act as transcriptional or post-translational modifier or regulate degradation events and are unlikely to be considered as functional candidates. Further functional analyses of LAP are necessary to give a clear insight in the way of being part in these complex traits. Additionally, it has to be considered that this protein is located close to a QTL and whether the association is due to fortuitous genetic linkage or to an actual physiological relationship remains to be analyzed (de Vienne et al., 1999).

4.5 PUTATIVE INVERTASE INHIBITOR – TO BE OR NOT TO BE AN INVERTASE INHIBITOR

For the putative invertase inhibitor, a whole candidate gene approach will be discussed. This inhibitor was shown to be a member of the Kunitz-type protease inhibitors (KTI) which were described to have a functional role in the inhibition of proteolytic enzymes like trypsin, papain, cathepsin and further proteases (Glaczinski et al., 2002). Protein extracts from potato tubers containing the putative invertase inhibitor inhibited soluble invertase completely and displayed a weak trypsin inhibition capacity (Glaczinski et al., 2002; Heibges et al., 2003b). Due to the identification of the not yet fully characterized putative regulator (inhibitor) for one of the key players in starch and chip quality (vacuolar invertase), further structural and functional characterizations of this putative invertase inhibitor were done in this thesis. Additionally, it was analyzed whether the inhibitor shows association with starch and chip quality traits due to the assumption of being a regulator of invertase. Although different approaches were followed, no clear evidence was obtained whether the putative invertase inhibitor is a true inhibitor for invertase or not. The characteristic differences from other KTIs (Heibges et al., 2003a) and whether the putative invertase inhibitor has an effect on the starch and sugar content in potato tubers will be discussed below.

4.5.1 THE PUTATIVE INVERTASE INHIBITOR BELONG TO A CLUSTER OF KUNITZ-TYPE PROTEASE INHIBITORS ON CHROMOSOME 3

Numerous cDNA sequences have been obtained in previous studies which are assumed to encode the putative inhibitor or variants of it (Glaczinski et al., 2002; Heibges et al., 2003a and 2003b; M. Fischer, unpublished data). The alignment of these sequences resulted in three major subgroups which were named after one of their members (1421, 1422 and 1423; Figure 41). These sequences showed around 90 % sequence similarity to a fourth subgroup which consists of cDNA sequences known to inhibit trypsin (subgroup X56874; Yamagishi et al., 1991). Despite the high sequence similarity to this trypsin inhibitor, it was concluded that the cDNA sequences of the subgroups 1421, 1422 and 1423 are most likely not conventional trypsin inhibitors because Heibges et al. (2003b) showed only a weak trypsin inhibition capacity for several members of these subgroups. Additionally, the deduced peptide sequences do not contain the characteristic trypsin inhibition motif PVR (Suh et al., 1990). Instead, the different sequences display variants like PVM or PVV indicating at least a reduced trypsin inhibition activity. Further proteases like chymotrypsin, papain and subtilisin were excluded as well as putative targets of these subgroups (Heibges et al., 2003b).

Insights about the genomic localization were obtained by a blast analysis of all putative invertase inhibitor cDNA sequences against the *S. phureja* genome (PGSC, 2011). For the three subgroups 1421, 1422 and 1423 only one single genomic locus was identified, whereas X56874 matched to a different locus only 100 kb upstream of the first hit. Detailed analyses of the surrounding genomic region showed besides the trypsin inhibitor X56874 the presence of further cysteine or serine protease inhibitors. Some of them were identified as members of the groups A and B of the Kunitz-type protease inhibitors. As it was already found by Heibges et al. (2003a) and Odeny et al. (2010), the group of protease inhibitors forms a cluster on chromosome 3.

4.5.2 THE cDNA SEQUENCES OF THE PUTATIVE INVERTASE INHIBITOR ARE ALLELIC VARIANTS

The match to one single locus for all cDNA sequences of 1421, 1422 and 1423 indicated that the subgroups are in fact allelic variants. The genomic sequence was identical to the sequences of subgroup 1421. This indicated that this group is most likely the basic (reference) and most common allele. Within the subgroups 1421, 1422 and 1423, the cDNA sequences showed between 97.4 % and 99.2 % of sequence identity which supports the hypothesis of allelic variants instead of distinct genes. Besides the identification of the single genomic locus, different expression analyses were done to verify the three subgroups 1421, 1422 and 1423 as alleles. Subgroup specific primers were generated on the basis of SNPs. Amplification of genomic DNA resulted in 38 out of 40 genotypes containing subgroup 1421 (Figure 44), supporting the assumption of representing the most common

and standard allele although it was tested in a rather small population. Subgroup 1422 was identified in 22 genotypes and subgroup 1423 could be amplified in eight of 40 genotypes. Additional support was obtained by the pyrosequencing approach including ten genotypes of the BIOSOL population (Figure 45-47). Characteristic SNPs of subgroup 1421 were identified in all tested genotypes, whereas characteristic SNPs for 1423 were identified in only two genotypes which is identical to the result from the amplification of the genotypes on genomic level. The relative expression for subgroup 1422 (Figure 47) displayed eight genotypes containing the characteristic SNP, although only five of the genotypes were assumed to contain this allele on genomic level. These differences may result from amplification errors and will be discussed in detail in the transcriptional expression analysis (section 4.5.3). Final evidence for dealing with allelic variants was obtained from the association analysis and the resulting haplotype calculation. Three of the haplotype models were identical to one cDNA sequence of 1421 and two sequences of subgroup 1422.

4.5.3 THE PUTATIVE INVERTASE INHIBITOR SHOWS AN UNIQUE EXPRESSION PROFILE IN TUBERS

The search for sequences with a high level of similarity to the putative invertase inhibitor alleles in other species than *S. tuberosum* did not result in clear orthologues. Sequences with a relatively low level of similarity were identified in the genomic sequence of tomato assuming that these genes encode most likely common protease inhibitors, but not an orthologue to the putative invertase inhibitor. Regarding the fact that the family of Kunitz-type protease inhibitors is large and highly variable, it can be concluded that the few similar genes identified, belong to the family of Kunitz-type protease inhibitors, but do not encode a putative invertase inhibitor. Although the species within the family of Solanaceae are highly conserved on the gene level and share a high degree of synteny among their chromosomes, it is likely that each species evolved also a set of specialized genes (Bonierbale et al., 1988; Gebhardt et al., 1991; Ronning et al., 2003). Especially, the formation of an edible tuber from stolon tissue is unique in *S. tuberosum* and therefore a characteristic set of proteins and genes is needed for this complex process of morphological change (Artschwager et al., 1924; Stiekema et al., 1988).

A semi-quantitative expression analysis in different tissues of a potato plant resulted in an almost exclusive expression in potato tubers (Figure 43). These findings support the hypothesis that the putative invertase inhibitor sequences have developed uniquely in *S. tuberosum* and have a specific function in the specialized storage and vegetative propagation capacity of the tuber or in the tuber formation. Especially for Kunitz-type protease inhibitors which are located on chromosome 3, it was shown that on the one hand potato contains around twice the number of Kunitz-type protease inhibitors compared to tomato and on the other hand, some of these sequences do not display a

homologous sequence in tomato (PGSC, 2011). Among these duplicated and specialized variants, it is possible that a Kunitz-type invertase inhibitor is a specialization event in potato. The very faint expression of *1422* in floral buds may indicate that either some of the allelic variants are differentially expressed and regulated or give hints about a contribution to developmental events and maturation.

Besides the interesting result of the confined expression of the allelic variants to the potato tuber tissue, it was analyzed whether cold-storage has an effect on the transcriptional expression level. For vacuolar invertase, the putative target of the inhibitor, it is known that the expression increases upon cold-storage (Zrenner et al., 1996; Bagnaresi et al., 2008; Draffehn, 2010). The basis for the expression analysis of the putative invertase inhibitor alleles was the cold-storage experiment of ten selected genotypes of the BIOSOL population (section 2.2.1). The relative expression was measured using cDNA from tubers which were stored for 0, 1, 2, 4 and 12 weeks at 4°C by using the pyrosequencing approach (Ronaghi et al., 1996). The advantage of this method is that it can be easily distinguished between the different inhibitor alleles by designing sequencing primers on unique SNPs for each subgroup. One sequencing primer was generated to include a specific SNP for subgroup 1421 and two SNPs which are unique for subgroup 1423. The evaluation of the pyrosequencing data showed for subgroup 1421 a clear response to cold storage in both directions (Figure 45). The level of relative expression varied between 10 % and 30 %, but no uniform trend was observed for the tested genotypes. Five genotypes of each group with either with very low or very high sugar accumulation were included in the analysis. No correlation between the increase or decrease of the expression levels and the classification either to the group with low sugar accumulation or to the group with high sugar accumulation was identified.

Subgroup *1423* turned out to be expressed in only two of the ten selected genotypes. These findings are identical to the amplification results with specific primers at genomic level (section 4.5.2; Figure 44). The transcript level of subgroup *1423* varied only between 5 and 10 %. Regarding the error between biological and technical replicates, the variation does not seem to result from cold-storage and *1423* might not be transcriptionally regulated upon cold treatment.

The analysis of the expression of subgroup *1422* was done based on subtraction of expression values of *1421* (Figure 47) because no unique SNP for this subgroup was present in the amplified region. The sequencing primer was assumed to amplify subgroups *1421* and *1422*. Knowing the relative expression level of *1421* from the previous experiment helped to calculate an approximate expression level for *1422*. In general, the relative expression level for *1422* varied between 5% and 25%. For most of the cold stored genotypes, a slight decrease of the expression level was detected for the first and second week of cold-storage. This estimated expression level contrasts to the one of

invertase which increases within the first two weeks of cold-storage (Zrenner et al., 1996; Bagnaresi et al., 2008; Draffehn, 2010). In general, the calculation of the relative expression seemed to be problematic because a putative expression was calculated for three genotypes (10, 17 and 29) which were assumed to not contain this allele (Figure 44). The expression levels in the genotypes 10 and 29 were very low and can be considered as errors from two different sequencing primers in the pyrosequencing approach which may bind with different affinities. For genotype 17, the expression levels are most likely true measurements. Probably, a third allele besides the two alleles 1421 and 1422 was amplified and leads to this detectable expression. Further indications for a third allele were obtained in the association analysis and will be discussed in the following section. However, it turned out that the calculatory way is not specific enough due to many unknown factors and interferences. It is necessary to design a new amplicon which includes one of the rare SNPs specific for allele 1422 to obtain clear data for this allele.

These data give a clear idea about the putative invertase inhibitor to be confined to tuber tissue and to show a rough response at least for some of the alleles to cold treatment. Therefore, a role in one of the tuber specific functions like long-term starch storage or in sprouting and further developmental steps can be assumed.

4.5.4 THE PUTATIVE INVERTASE INHIBITOR IS ASSOCIATED WITH STARCH QUALITY TRAITS

The identification of one single locus in the genomic sequence of *S. phureja* (PGSC, 2011) and the previously discussed results facilitated to characterize the different cDNA sequences as allelic variants of the putative invertase inhibitor. The knowledge of existing allelic variants and the assumed function as an inhibitor of vacuolar invertase raises the question whether a contribution of the natural variation of the putative inhibitor to starch and chip quality traits can be identified. Upon this hypothesis, an association analysis for a genomic fragment was done which corresponded to the 1421 locus. The analysis resulted in clear associations of several SNPs and one indel with the two traits tuber starch content (TSC) and tuber starch yield (TSY; Table 27). The SNPs explained between 5.8 % and 11.6 % of the phenotypic variance. No association was detected for any of the chip quality traits. Interestingly, most of the associated SNPs formed a clear LD group, as they predominantly show similar associations due to a high degree of co-segregation and similar allele frequencies. Small discrepancies in p-values or frequencies between SNPs that are in LD may result from inconclusive sequencing data and therefore mis-scoring (Draffehn et al., 2010).

It is noteworthy that the presence of a 6 bp indel showed high positive association with starch yield and starch quality. On amino acid level, the 6 bp insertion lead to two additional aspartic acids (DD)

in the protein sequence. Whether the presence of the two amino acids has an effect on the function of the protein remains to be elucidated. It can only be speculated that the additional amino acids may alter the conformation of the protein. Both additional amino acids reside in the targeting peptide and therefore may impede protein transport and result in lower protein abundance. In this respect, it is important to mention that these two additional amino acids are located close to the assumed splicing site of the putative vacuole target peptide and may have an effect on the proper cleavage of the signal and the subsequent transport. The lower amount of putative invertase inhibitor would lead to higher invertase activity and the subsequent accumulation of reducing sugars. Due to an association with higher starch yield and starch content, an increased level of end products in form of reducing sugars may act either as a negative feedback signal to the starch degradation pathway or as a positive signal for an increased starch synthesis (Müller-Röber et al., 1990). In both scenarios, this would lead to higher tuber starch content. However, this indel is characteristic for the associated superior allele and can be easily detected by amplification or sequencing experiments and is therefore a useful marker to analyze for the presence of the superior allele 1422 of the putative invertase inhibitor.

The associated SNPs were used for an *in silico* haplotype modeling approach which resulted in 15 haplotype models (Table 28). Three of the haplotype models were identical to the known cDNA clones of the putative invertase inhibitor (Glaczinski et al., 2002; Heibges et al., 2003a and 2003b). Certainly, the basic allele which corresponds to clone 1421 was obtained by this *in silico* approach and represented the most frequent haplotype. This fits to the amplification result of allele 1421 on genomic DNA with subgroup specific primers in almost all analyzed genotypes (Figure 44). An association analysis of this haplotype did not result in any association and supports the classification as reference and common haplotype. On the contrary, a second haplotype model which was identical to the cloned cDNA sequence 1422, was found to be associated with tuber starch content and tuber starch yield. It contained 6 out of 8 associated SNPs. The calculated true allele frequency was rather low (4.1%), but the presence of this haplotype was proven by its amplification with specific primers and by the existing cDNA clone (Figure 44; Glaczinski et al., 2002). Interesting conclusions can be drawn from the calculated haplotype number 15. It is identical to the cDNA clone P4G5 which belongs to subgroup 1422. This haplotype model contains all eight associated SNPs and shows a slightly higher association compared to the haplotype model identical to 1422. Initially, it was assumed that this clone is based only on amplification or sequencing errors, but the detection in the association analysis confirmed the existence. The low frequency of this haplotype model fits to the presence of only one cloned cDNA sequence out of 20 other cloned cDNAs from Heibges et al. (2003). It is amazing that it was amplified in the previous study and cloned at all, since only two genotypes were analyzed and its overall haplotype frequency in the GABICHIPS population is also

very low (3 genotypes out of 207, Table 28). Additionally, this haplotype might be one of the reasons for the detection of putative 1422 transcripts in genotypes (Figure 47) that did not contain 1422 SNPs at the genomic level and were assumed not to encode this allele. Due to the fact that haplotype *P4G5* has initially been classified as a member of subgroup 1422, it can now be assumed that it is a further true haplotype besides 1421, 1422 and 1423. In order to analyze this hypothesis, a pyrosequencing approach based on the identified SNP of the association analysis should result in transcript level which were not assignable before. Besides the identification of the new allele *P4G5*, allele 1423 was not discovered in this SNP analysis and the haplotype modeling. In the previous discussions of this section, the existence of this subgroup 1423 was confirmed by experiments with 1423 specific primers on genomic DNA and by reliable results from an *in silico* blast analysis. Although the primers for the association analysis were designed to amplify all alleles of this locus, the absence of 1423 specific SNPs may indicate that allele 1423 was not amplified. The genomic primers for the association analysis were selected according to their specificity for this single locus and were supposed to discriminate only between the highly similar Kunitz-type protease inhibitor sequences (for instance the trypsin protease *X56874* or other serine protease inhibitors) which are all located in a small cluster on chromosome 3 (section 4.5.1). Probably, the designed primers included SNPs which are linked to the known SNPs of subgroup 1423 and therefore did not amplify this allele. Additional indications for a missed amplification of one allele can be concluded from the fact that only few of the associated SNPs show all possible allele dosages. Hereby, the critical point is that genotypes were identified to contain the SNPs in nulliplex, simplex, duplex or quadruplex, but no genotype was identified to contain clearly a triplex status. It is not known if it is directly predictable from an obtained trace files whether all alleles are clearly amplified or not. Therefore, the missing of the triplex state can only be assumed as an indication that not all alleles are amplified and no characteristic SNP of subgroup 1423 was found. Consequences on the association obtained for allele 1422 are considered low because it was shown that the subgroup 1423 is not a high frequent allele (8 out of 40 cultivars) and therefore would only affect the scoring in only few genotypes. A possible approach to identify putative SNPs in the regions of the primer binding site would be the sequencing of these regions by primer walking on genomic DNA of different genotypes which are known to contain subgroup 1423.

No association was identified with one of the chip quality traits. This association analysis did not suggest an obvious contribution of the putative invertase inhibitor to the amount of reducing sugars which are one of the major factors for a bad chipping quality. However, no association at all was found for the Kunitz-type cluster (StKI) locus in the same population in a previous study (Li et al., 2008). This may be due either to the use of SSCP markers instead of SNP markers or that the analyzed fragments belong to members of the Kunitz-type protease inhibitor groups A and B and

therefore were specific for a different locus. The association with starch content and starch yield therefore indicates that group C contains inhibitors which are different from the remaining protease inhibitors and developed the capacity to inhibit non-proteolytic enzymes. Some of these non-proteolytic enzyme targets are assumed to be involved in the carbohydrate metabolism. The exact way of the contribution to a better starch content and starch yield of the group C Kunitz type inhibitors remains to be analyzed.

4.5.5 IN VITRO INHIBITION ASSAYS DO NOT PROVE INVERTASE INHIBITION

The final step in the analysis of the putative invertase inhibitor was to show a direct interaction and inhibition of invertase (the putative non proteolytic target) by using heterologously expressed and highly purified protein. The indication for inhibition of invertase was obtained from a purified invertase inhibitor fraction from potato tubers (Glaczinski et al., 2002). In further studies, a heterologous expression of the putative inhibitor was carried out in *E. coli* and *Saccharomyces pombe*, (Glaczinski, unpublished; Heibges, 2001) but resulted either in insoluble protein due to inclusion bodies or in protein which did not show inhibition of invertase protein extracts from potato tubers due to an improper experimental set up. Here, a heterologous expression system was used that is based on the yeast *Pichia pastoris*. This system has been successfully used for the expression of Kunitz-type protease inhibitors before (Wagner et al., 1992). In addition, it is well suited for the expression of plant proteins because the way of glycosylation is assumed to be more similar to the glycosylation pattern in plants (Cregg et al., 1993; Yesilirmak and Sayers, 2009). Another advantage of the *P. pastoris* system is the use of a secretory vector which is supposed to facilitate the purification and to avoid the formation of inclusion bodies. Besides the secretion signal, a c-myc epitope (c-myc-tag) and a polyhistidine-tag (His₆-tag) for the purification via anion-chromatography were cloned in frame to the 3' part of the gene. For the expression of the putative invertase inhibitor, the cloned cDNA sequences of the allelic variants 1421, 1422 and 1423 were used that encode the mature protein without the putative vacuolar target peptide (Heibges et al., 2003a and 2003b). In addition to heterologously expressed inhibitor proteins, different full-length alleles of the vacuolar invertase were cloned and expressed in the same system. The set of vacuolar invertases comprised one allele (*Pain-1-Sa*) which was associated with starch and chip quality traits and a non-associated allele (*Pain-1-Sb*; Draffehn, 2010). Both alleles have been cloned from the genotype Satina. The third allele was an associated allele from the cultivar Diana (*Pain-1-Da*; Draffehn, 2010) which showed as well association with starch and chip quality traits. The proteins were successfully purified via anion chromatography (His-tag) and detected by immunodetection (anti c-myc) on western blots. The putative invertase inhibitors displayed an approximate size of around 24 kDa each (Figure 49) and

were identical to previous findings (Glaczinski et al., 2002). For the three different alleles of the vacuolar invertase a size of approximately 90 kDa was estimated from the western blot. For most of acid (vacuolar) invertases the size of mature proteins is predicted to range between 55 and 77 kDa (Sturm, 1999). Approximately three kDa are explained by the two tags which are fused to the protein. A further reason for the increase in size may be that the vacuolar invertase protein is strongly glycosylated in *P. pastoris*. Difficulties may also result from an altered separation and mobility pattern due to the glycosylation or hindered unfolding in the SDS gel. However, enzymatic activity of all three invertase alleles was shown by measuring the amount of produced glucose (Figure 50). For the three putative inhibitor alleles, it can be concluded as well that active protein was purified: a trypsin inhibition test displayed a reduction in trypsin activity between 30 % and 40 % (Figure 51). It was shown before that 1421, 1422 and 1423 display a weak but measurable inhibition capacity towards trypsin (Heibges et al., 2003b). The lowered inhibition capacity on trypsin is assumed to be based on a modification in one peptide of the trypsin-motif, as discussed above and described in Heibges et al. (2003b). Due to limitations in the amount of inhibitor protein, it was not possible to incubate the same molar amount of inhibitor with trypsin. An equimolar amount of trypsin and inhibitor might have resulted in a slightly stronger inhibition, but no complete inhibition was expected. Despite of active putative invertase inhibitor protein, the invertase inhibition test did not show any reduction in invertase activity of all three tested invertase variants (Pain-1-Sa, -Sb and -Da) and on native invertase extracts (Figure 52 and 53). No differences were found, when the protein amount of the putative invertase inhibitor was 5 fold increased. It can be hypothesized that despite changing the expression system, the experimental conditions were not improved to confirm the inhibition of invertase as it was suggested. Bracho and Whitaker (1990) described that the correct acidic conditions need to be maintained for the inhibitor to avoid the loss of inhibition capacity. For the purification via anion-chromatography, it is a prerequisite to change the pH slightly into the basic conditions to cause the release of the His-tagged protein from the charged column. Additionally, it is possible that within the set up of the inhibition assay a disturbing factor is present that prevents the formation of the complex of invertase and inhibitor which has been described as a very stable complex (Pressey, 1967). Besides disturbing factors and unfavourable conditions, it may be that further molecules or factors are needed to establish the complex which are absent in the artificial set up of the expression, purification and interaction assay. Finally, it cannot be excluded that the initially purified invertase inhibitor by Glaczinski et al. (2002) contained either contaminations of another small protein described as invertase inhibitor or vice versa, the identified peptide sequence of the Kunitz-type invertase inhibitor was a contamination in the fraction containing the other inhibitor (Liu et al., 2010; Greiner et al., 1999).

The inhibition of vacuolar invertase by this putative inhibitor of the Kunitz-type family was neither proven nor disproven. The results obtained for the analysis of the putative invertase inhibitor indicated that this inhibitor is unique for *S. tuberosum* and encoded at a single locus. The previously identified numerous cDNA sequences (Glaczinski et al., 2002; Heibges et al., 2003a and 2003b) are shown to be allelic variants and not different inhibitor genes. The expression restricted to potato tubers suggests a putative role in carbohydrate storage or vegetative propagation. Additionally, a putative contribution to tuber starch content was shown by the association analysis. Whether this is due to inhibition of invertase or another enzyme remains to be analyzed. An interaction assay would be an adequate approach to identify interaction partners. One possible approach would be a modified version of tandem affinity purification (TAP) by binding the putative invertase inhibitor via its His-tag to a column and identify protein complexes after adding native protein extracts from potato tubers with mass-spectrometry approaches afterwards.

4.6 CANDIDATE GENE APPROACH AND ASSOCIATION ANALYSES ARE POWERFUL TOOLS

By using the candidate gene approach, several interesting associations of different functional, positional and non-obvious candidates with starch and chip quality traits were obtained. The associated SNPs identified may serve as diagnostic markers for marker-assisted selection. By an early selection for superior alleles in the germplasm of the breeding population, a certain number of breeding steps will be saved (Milbourne, 2007). Besides functional candidates, the candidate gene approach can be used to identify the contribution of genes which are not primarily and obviously involved in important pathways. By making use of transcriptomic and proteomic approaches, it is possible to identify new candidate genes which are differentially expressed or present in a cases and control population which is selected for a certain quality trait. In this study, a putative post-transcriptionally regulatory gene (*LAP*) was identified to contribute to the variation of starch and chips quality that would have never been selected as a candidate gene without knowing the striking differences in the expression pattern. Especially these regulatory genes are supposed to have a stronger influence on complex traits than structural genes (Wallace, 1963).

Limitations of the candidate gene approach are due to a too small or a not well phenotypically characterized population. Additionally, it can be concluded for the subsequent marker-assisted selection that it is not possible to select for many different traits. For instance, improved tuber starch content and chip quality may be at the expense of tuber yield. In the optimization of a parental crossing population, several limitations should be expected and the selection of diagnostic marker should be optimized for a small and similar set of traits.

LITERATURE

- Alberto, F., Bignon, C., Sulzenbacher, G., Henrissat, B., and Czjzek, M.** (2004). The Three-Dimensional Structure of Invertase (Beta-Fructosidase) from *Thermotoga maritima* Reveals a Bimodular Arrangement and an Evolutionary Relationship between Retaining and Inverting Glycosidases. *Journal of Biological Chemistry* **279**, 18903-18910.
- Alonso-Blanco, C., Aarts, M.G.M., Bentsink, L., Keurentjes, J.J.B., Reymond, M., Vreugdenhil, D., and Koornneef, M.** (2009). What Has Natural Variation Taught Us About Plant Development, Physiology, and Adaptation? *The Plant Cell Online* **21**, 1877-1896.
- Arai, M., Mori, H., and Imaseki, H.** (1992). Cloning and Sequence of cDNAs for an Intracellular Acid Invertase from Etiolated Hypocotyls of Mung Bean and Expression of the Gene During Growth of Seedlings. *Plant and Cell Physiology* **33**, 245-252.
- Artschwager, E.F.** (1924). Studies on the Potato Tuber. *Journal of Agricultural Research* **27**, 187-190.
- Bagnaresi, P., Moschella, A., Beretta, O., Vitulli, F., Ranalli, P., and Perata, P.** (2008). Heterologous Microarray Experiments Allow the Identification of the Early Events Associated with Potato Tuber Cold Sweetening. *BMC Genomics* **9**, 176.
- Ball, S.G., and Morell, M.K.** (2003). From Bacterial Glycogen to Starch: Understanding the Biogenesis of the Plant Starch Granule. *Annual Review of Plant Biology* **54**, 207-233.
- Ballvora, A., Ercolano, M.R., Weiß, W., Meksem, K., Bormann, C.A., Oberhagemann, P., Salamini, F., and Gebhardt, C.** (2002). The R1 Gene for Potato Resistance to Late Blight (*Phytophthora infestans*) Belongs to the Leucine Zipper/Nbs/Lrr Class of Plant Resistance Genes. *The Plant Journal* **30**, 361-371.
- Ballvora, A., Jöcker, A., Viehover, P., Ishihara, H., Paal, J., Meksem, K., Bruggmann, R., Schoof, H., Weisshaar, B., and Gebhardt, C.** (2007). Comparative Sequence Analysis of Solanum and Arabidopsis in a Hot Spot for Pathogen Resistance on Potato Chromosome V Reveals a Patchwork of Conserved and Rapidly Evolving Genome Segments. *BMC Genomics* **8**, 112.
- Baunsgaard, L., Lutken, H., Mikkelsen, R., Glaring, M.A., Pham, T.T., and Blennow, A.** (2005). A Novel Isoform of Glucan, Water Dikinase Phosphorylates Pre-Phosphorylated Alpha-Glucans and Is Involved in Starch Degradation in Arabidopsis. *Plant J* **41**, 595-605.
- Bernard, P., and Couturier, M.** (1992). Cell Killing by the F Plasmid ccdB Protein Involves Poisoning of DNA-Topoisomerase II Complexes. *Journal of Molecular Biology* **226**, 735-745.
- Bhaskar, P.B., Venkateshwaran, M., Wu, L., Ané, J.-M., and Jiang, J.** (2009). *Agrobacterium*-Mediated Transient Gene Expression and Silencing: A Rapid Tool for Functional Gene Assay in Potato. *PLoS ONE* **4**, e5812.
- Blennow, A., Bay-Smidt, A.M., Olsen, C.E., and Moller, B.L.** (2000). The Distribution of Covalently Bound Phosphate in the Starch Granule in Relation to Starch Crystallinity. *Int J Biol Macromol* **27**, 211-218.

- Bonierbale, M.W., Plaisted, R.L., and Tanksley, S.D.** (1988). RFLP Maps Based on a Common Set of Clones Reveal Modes of Chromosomal Evolution in Potato and Tomato. *Genetics* **120**, 1095-1103.
- Bracho, G.E., and Whitaker, J.R.** (1990). Characteristics of the Inhibition of Potato (*Solanum tuberosum*) Invertase by an Endogenous Proteinaceous Inhibitor in Potatoes. *Plant Physiology* **92**, 381-385.
- Brummell, D.A., Chen, R.K.Y., Harris, J.C., Zhang, H., Hamiaux, C., Kralicek, A.V., and McKenzie, M.J.** (2011). Induction of Vacuolar Invertase Inhibitor mRNA in Potato Tubers Contributes to Cold-Induced Sweetening Resistance and Includes Spliced Hybrid mRNA Variants. *Journal of Experimental Botany* **62**, 3519-3534.
- Buléon, A., Colonna, P., Planchot, V., and Ball, S.** (1998). Starch Granules: Structure and Biosynthesis. *International Journal of Biological Macromolecules* **23**, 85-112.
- Byrne, P.F., and McMullen, M.D.** (1996). Defining Genes for Agricultural Traits: QTL Analysis and the Candidate Gene Approach. *Probe* **7**, 24-27.
- Callis, J., and Vierstra, R.D.** (2000). Protein Degradation in Signaling. *Current Opinion in Plant Biology* **3**, 381-386.
- Calvin, N.M., and Hanawalt, P.C.** (1988). High-Efficiency Transformation of Bacterial-Cells by Electroporation. *Journal of Bacteriology* **170**, 2796-2801.
- Camardella, L., Carratore, V., Ciardiello, M.A., Servillo, L., Balestrieri, C., and Giovane, A.** (2000). Kiwi Protein Inhibitor of Pectin Methyltransferase. *European Journal of Biochemistry* **267**, 4561-4565.
- Chamary, J.V., Parmley, J.L., and Hurst, L.D.** (2006). Hearing Silence: Non-Neutral Evolution at Synonymous Sites in Mammals. *Nat Rev Genet* **7**, 98-108.
- Chao, W.S., Gu, Y.-Q., Pautot, V.R., Bray, E.A., and Walling, L.L.** (1999). Leucine Aminopeptidase RNAs, Proteins, and Activities Increase in Response to Water Deficit, Salinity, and the Wound Signals Systemin, Methyl Jasmonate, and Abscisic Acid. *Plant Physiology* **120**, 979-992.
- Chao, W.S., Pautot, V.R., Holzer, F.M., and Walling, L.L.** (2000). Leucine Aminopeptidases: The Ubiquity of Lap-N and the Specificity of Lap-A. *Planta* **210**, 563-573.
- Chen, X., Salamini, F., and Gebhardt, C.** (2001). A Potato Molecular-Function Map for Carbohydrate Metabolism and Transport. *TAG Theoretical and Applied Genetics* **102**, 284-295.
- Christiansen, C., Hachem, M.A., Glaring, M.A., Viksjö-Nielsen, A., Sigurskjold, B.W., Svensson, B., and Blennow, A.** (2009). A CBM20 Low-Affinity Starch-Binding Domain from Glucan, Water Dikinase. *FEBS letters* **583**, 1159-1163.
- Chung, B., Simons, C., Firth, A., Brown, C., and Hellens, R.** (2006). Effect of 5'UTR Introns on Gene Expression in *Arabidopsis thaliana*. *BMC Genomics* **7**, 120.
- Collard, B.C.Y., and Mackill, D.J.** (2008). Marker-Assisted Selection: An Approach for Precision Plant Breeding in the Twenty-First Century. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**, 557-572.

- Cregg, J.M., Vedvick, T.S., and Raschke, W.C.** (1993). Recent Advances in the Expression of Foreign Genes in *Pichia pastoris*. *Nat Biotech* **11**, 905-910.
- de Vienne, D., Leonardi, A.s., Damerval, C., and Zivy, M.** (1999). Genetics of Proteome Variation for QTL Characterization: Application to Drought-Stress Responses in Maize. *Journal of Experimental Botany* **50**, 303-309.
- Deschamps, P., Haferkamp, I., d'Hulst, C., Neuhaus, H.E., and Ball, S.G.** (2008). The Relocation of Starch Metabolism to Chloroplasts: When, Why and How. *Trends in Plant Science* **13**, 574-582.
- Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K., and Mattick, J.S.** (1991). 'Touchdown' PCR to Circumvent Spurious Priming During Gene Amplification. *Nucleic Acids Research* **19**, 4008.
- Douches, D.S., and Freyre, R.** (1994). Identification of Genetic-Factors Influencing Chip Color in Diploid Potato (*Solanum spp.*). *American Potato Journal* **71**, 581-590.
- Draffehn, A., Meller, S., Li, L., and Gebhardt, C.** (2010). Natural Diversity of Potato (*Solanum tuberosum*) Invertases. *BMC Plant Biology* **10**, 271.
- Draffehn, A.M.** (2010). Structural and Functional Characterization of Natural Alleles of Potato (*Solanum tuberosum* L.) Invertases Associated with Tuber Quality Traits. In Inaugural – Dissertation zur Erlangung des Doktorgrades der Mathematisch-Naturwissenschaftlichen Fakultät der Universität zu Köln. (Köln: Universität zu Köln).
- Draper, S.R.** (1975). Amino Acid Changes Associated with the Development of Cold Hardiness in Perennial Ryegrass. *Journal of the Science of Food and Agriculture* **26**, 1171-1176.
- Duwenig, E., Steup, M., Willmitzer, L., and Kossmann, J.** (1997). Antisense Inhibition of Cytosolic Phosphorylase in Potato Plants (*Solanum tuberosum* L.) Affects Tuber Sprouting and Flower Formation with Only Little Impact on Carbohydrate Metabolism. *Plant J* **12**, 323-333.
- Edner, C., Li, J., Albrecht, T., Mahlow, S., Hejazi, M., Hussain, H., Kaplan, F., Guy, C., Smith, S.M., Steup, M., and Ritte, G.** (2007). Glucan, Water Dikinase Activity Stimulates Breakdown of Starch Granules by Plastidial Beta-Amylases. *Plant Physiol* **145**, 17-28.
- Ellis, R.P., Cochrane, M.P., Dale, M.F.B., Duffus, C.M., Lynn, A., Morrison, I.M., Prentice, R.D.M., Swanston, J.S., and Tiller, S.A.** (1998). Starch Production and Industrial Use. *Journal of the Science of Food and Agriculture* **77**, 289-311.
- Emanuelsson, O., Nielsen, H., and von Heijne, G.** (1999). ChloroP, a Neural Network-Based Method for Predicting Chloroplast Transit Peptides and Their Cleavage Sites. *Protein science : a publication of the Protein Society* **8**, 978-984.
- EUROPOTATO.** The European Cultivated Potato Database, <http://www.europotato.org/>
- FAO.** Crops Statistics Database, <Http://faostat.fao.org>.
- Fettke, J., Eckermann, N., Kotting, O., Ritte, G., and Steup, M.** (2006). Novel Starch-Related Enzymes and Carbohydrates. *Cell Mol Biol (Noisy-le-grand)* **52 Suppl**, OL883-904.

- Fettke, J., Hejazi, M., Smirnova, J., Hoechel, E., Stage, M., and Steup, M.** (2009). Eukaryotic Starch Degradation: Integration of Plastidial and Cytosolic Pathways. *Journal of Experimental Botany* **60**, 2907-2922.
- Flinn, B., Rothwell, C., Sardana, R., Griffiths, R., Lague, M.D., De Koeyer, D., Audy, P., Goyer, C., Li, X.-Q., Wang-Pruski, G., and Regan, S.** (2004). Generation of ESTs from Mature Tubers Following 6 Months Storage at 9 Degrees Celsius, and 95% Relative Humidity. unpublished, NCBI direct submission.
- Frommer, W.B., and Sonnewald, U.** (1995). Molecular Analysis of Carbon Partitioning in Solanaceous Species. *Journal of Experimental Botany* **46**, 587-607.
- Gallant, D.J., Bouchet, B., and Baldwin, P.M.** (1997). Microscopy of Starch: Evidence of a New Level of Granule Organization. *Carbohydrate Polymers* **32**, 177-191.
- Gebhardt, C., Ritter, E., Debener, T., Schachtschabel, U., Walkemeier, B., Uhrig, H., and Salamini, F.** (1989). RFLP Analysis and Linkage Mapping in *Solanum tuberosum*. *TAG Theoretical and Applied Genetics* **78**, 65-75.
- Gebhardt, C., Ritter, E., Barone, A., Debener, T., Walkemeier, B., Schachtschabel, U., Kaufmann, H., Thompson, R.D., Bonierbale, M.W., Ganai, M.W., Tanksley, S.D., and Salamini, F.** (1991). RFLP Maps of Potato and Their Alignment with the Homoeologous Tomato Genome. *TAG Theoretical and Applied Genetics* **83**, 49-57.
- Gebhardt, C., and Valkonen, J.P.T.** (2001). Organization of Genes Controlling Disease Resistance in the Potato Genome. *Annual Review of Phytopathology* **39**, 79-102.
- Gebhardt, C., Lörz, H., and Wenzel, G.** (2005). Potato Genetics: Molecular Maps and More Molecular Marker Systems in Plant Breeding and Crop Improvement (Springer Berlin Heidelberg), pp. 215-227.
- Geigenberger, P.** (2003). Regulation of Sucrose to Starch Conversion in Growing Potato Tubers. *Journal of Experimental Botany* **54**, 457-465.
- Geiger, D.R., and Servaites, J.C.** (2001). Starch and Starch Grains. In *Encyclopaedia of Life Sciences*. Macmillan Reference Ltd.
- Glaczinski, H., Heibges, A., Salamini, F., and Gebhardt, C.** (2002). Members of the Kunitz-Type Protease Inhibitor Gene Family of Potato Inhibit Soluble Tuber Invertase in Vitro. *Potato Research* **45**, 163-176.
- Glaring, M.A., Zygadlo, A., Thorneycroft, D., Schulz, A., Smith, S.M., Blennow, A., and Baunsgaard, L.** (2007). An Extra-Plastidial Alpha-Glucan, Water Dikinase from Arabidopsis Phosphorylates Amylopectin in Vitro and Is Not Necessary for Transient Starch Degradation. *J Exp Bot* **58**, 3949-3960.
- Glaring, M.A., Baumann, M.J., Hachem, M.A., Nakai, H., Nakai, N., Santelia, D., Sigurskjold, B.W., Zeeman, S.C., Blennow, A., and Svensson, B.** (2011). Starch-Binding Domains in the CBM45 Family - Low-Affinity Domains from Glucan, Water Dikinase and α -Amylase Involved in Plastidial Starch Metabolism. *FEBS Journal*, **278**, 1175-1185.
- Grabowska, A., Sitnicka, D., and Orzechowski, S.** (2010). StGWD3, the Third Glucan/Water Dikinase in Potato: Novel Protein Related to Starch Metabolism. NCBI direct submission.

- Greiner, S., Rausch, T., Sonnewald, U., and Herbers, K.** (1999). Ectopic Expression of a Tobacco Invertase Inhibitor Homolog Prevents Cold-Induced Sweetening of Potato Tubers. *Nature Biotechnology* **17**, 708-711.
- Gu, Y.Q., Pautot, V., Holzer, F.M., and Walling, L.L.** (1996). A Complex Array of Proteins Related to the Multimeric Leucine Aminopeptidase of Tomato. *Plant Physiology* **110**, 1257-1266.
- Gu, Y.-Q., and Walling, L.L.** (2002). Identification of Residues Critical for Activity of the Wound-Induced Leucine Aminopeptidase (Lap-A) of Tomato. *European Journal of Biochemistry* **269**, 1630-1640.
- Guillén, D., Sanchez, S., and Rodriguez-Sanoja, R.** (2010). Carbohydrate-Binding Domains: Multiplicity of Biological Roles. *Applied Microbiology and Biotechnology* **85**, 1241-1249.
- Hawkes, J.G.** (1990). The Potato: Evolution, Biodiversity and Genetic Resources. . *American Journal of Potato Research* **67**, 733-735.
- Heibges, A.** (2001). Strukturelle Und Funktionelle Analyse Von Drei Gruppen Von Kunitz Typ Enzyminhibitoren Aus Kartoffelknollen (*Solanum Tuberosum L.*). In Mathematisch-Naturwissenschaftlichen Fakultät der Universität zu Köln (Köln: Universität zu Köln).
- Heibges, A., Glaczinski, H., Ballvora, A., Salamini, F., and Gebhardt, C.** (2003a). Structural Diversity and Organization of Three Gene Families for Kunitz-Type Enzyme Inhibitors from Potato Tubers (*Solanum Tuberosum L.*). *Mol Genet Genomics* **269**, 526-534.
- Heibges, A., Salamini, F., and Gebhardt, C.** (2003b). Functional Comparison of Homologous Members of Three Groups of Kunitz-Type Enzyme Inhibitors from Potato Tubers (*Solanum Tuberosum L.*). *Mol Genet Genomics* **269**, 535-541.
- Hejazi, M., Fettke, J., Haebel, S., Edner, C., Paris, O., Frohberg, C., Steup, M., and Ritte, G.** (2008). Glucan, Water Dikinase Phosphorylates Crystalline Maltodextrins and Thereby Initiates Solubilization. *Plant J.*
- Hendriks, J.H.M., Kolbe, A., Gibon, Y., Stitt, M., and Geigenberger, P.** (2003). ADP-Glucose Pyrophosphorylase Is Activated by Posttranslational Redox-Modification in Response to Light and to Sugars in Leaves of Arabidopsis and Other Plant Species. *Plant Physiology* **133**, 838-849.
- Herbers, K., Prat, S., and Willmitzer, L.** (1994). Functional-Analysis of a Leucine Aminopeptidase from *Solanum tuberosum L.* *Planta* **194**, 230-240.
- Hijmans, R.J., and Spooner, D.M.** (2001). Geographic Distribution of Wild Potato Species. *American Journal of Botany* **88**, 2101-2112.
- Hildmann, T., Ebnet, M., Penacortes, H., Sanchezserrano, J.J., Willmitzer, L., and Prat, S.** (1992). General Roles of Abscisic and Jasmonic Acids in Gene Activation as a Result of Mechanical Wounding. *Plant Cell* **4**, 1157-1170.
- Höglund, A., Dönnes, P., Blum, T., Adolph, H.-W., and Kohlbacher, O.** (2006). MultiLOC: Prediction of Protein Subcellular Localization Using N-Terminal Targeting Sequences, Sequence Motifs, and Amino Acid Composition. *Bioinformatics*.

- Horton, P., Park, K.-J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J., and Nakai, K.** (2007). Wolf Psort: Protein Localization Predictor. *Nucleic Acids Research* **35**, W585-W587.
- Huang, L.-F.** (2006). Molecular Analysis of an Acid Invertase Gene Family in Arabidopsis. In University of Florida (Florida: University of Florida).
- Ji, X., Van den Ende, W., Van Laere, A., Cheng, S., and Bennett, J.** (2005). Structure, Evolution, and Expression of the Two Invertase Gene Families of Rice. *Journal of Molecular Evolution* **60**, 615-634.
- Kelly, R.B., Cozzarelli, N.R., Deutscher, M.P., Lehman, I.R., and Kornberg, A.** (1970). Enzymatic Synthesis of Deoxyribonucleic Acid. *Journal of Biological Chemistry* **245**, 39-45.
- Kibbe, W.A.** (2007). OligoCalc: An Online Oligonucleotide Properties Calculator. *Nucleic Acids Research* **35**, W43-W46.
- Kimchi-Sarfaty, C., Oh, J.M., Kim, I.-W., Sauna, Z.E., Calcagno, A.M., Ambudkar, S.V., and Gottesman, M.M.** (2007). A "Silent" Polymorphism in the MDR1 Gene Changes Substrate Specificity. *Science* **315**, 525-528.
- Knudsen, S.** (1999). Promoter2.0: For the Recognition of PolII Promoter Sequences. *Bioinformatics* **15**, 356-361.
- Koornneef, M., Alonso-Blanco, C., Peeters, A.J.M., and Soppe, W.** (1998). Genetic Control of Flowering Time in Arabidopsis. In *Annual Review of Plant Physiology and Plant Molecular Biology*, pp. 345-370.
- Kötting, O., Pusch, K., Tiessen, A., Geigenberger, P., Steup, M., and Ritte, G.** (2005). Identification of a Novel Enzyme Required for Starch Metabolism in Arabidopsis Leaves. The Phosphoglucan, Water Dikinase. *Plant Physiol* **137**, 242-252.
- Kötting, O., Santelia, D., Edner, C., Eicke, S., Marthaler, T., Gentry, M.S., Comparot-Moss, S., Chen, J., Smith, A.M., Steup, M., Ritte, G., and Zeeman, S.C.** (2009). Starch-Excess4 Is a Laforin-Like Phosphoglucan Phosphatase Required for Starch Degradation in *Arabidopsis thaliana*. *Plant Cell* **21**, 334-346.
- Kötting, O., Kossmann, J., Zeeman, S.C., and Lloyd, J.R.** (2010). Regulation of Starch Metabolism: The Age of Enlightenment? *Current Opinion in Plant Biology* **13**, 321-329.
- Kovac, M., and Ravnkar, M.** (1998). Sucrose and Jasmonic Acid Interact in Photosynthetic Pigment Metabolism and Development of Potato (*Solanum tuberosum* L. Cv. *Sante*) Grown in Vitro. *Plant Growth Regulation* **24**, 101-107.
- Kunitz, M.** (1945). Crystallization of a Trypsin Inhibitor from Soybean. *Science* **101**, 668-669.
- Leister, D., Ballvora, A., Salamini, F., and Gebhardt, C.** (1996). A PCR-Based Approach for Isolating Pathogen Resistance Genes from Potato with Potential for Wide Application in Plants. *Nature Genetics* **14**, 421-429.
- Leonards-Schippers, C., Gieffers, W., Salamini, F., and Gebhardt, C.** (1992). The R1 Gene Conferring Race-Specific Resistance to *Phytophthora infestans* in Potato Is Located on Potato Chromosome-V. *Molecular & General Genetics* **233**, 278-283.

- Leszczynski, W., and Lisinska, G.** (1988). Influence of Nitrogen-Fertilization on Chemical-Composition of Potato-Tubers. *Food Chemistry* **28**, 45-52.
- Li, L., Strahwald, J., Hofferbert, H., Lubeck, J., Tacke, E., Junghans, H., Wunder, J., and Gebhardt, C.** (2005). DNA Variation at the Invertase Locus *InvGE/GF* Is Associated with Tuber Quality Traits in Populations of Potato Breeding Clones. *Genetics* **170**, 813.
- Li, L., Paulo, M.J., Strahwald, J., Lubeck, J., Hofferbert, H.R., Tacke, E., Junghans, H., Wunder, J., Draffehn, A., van Eeuwijk, F., and Gebhardt, C.** (2008). Natural DNA Variation at Candidate Loci Is Associated with Potato Chip Color, Tuber Starch Content, Yield and Starch Yield. *Theor Appl Genet.*
- Liu, X., Song, B., Zhang, H., Li, X.-Q., Xie, C., and Liu, J.** (2010). Cloning and Molecular Characterization of Putative Invertase Inhibitor Genes and Their Possible Contributions to Cold-Induced Sweetening of Potato Tubers. *Molecular Genetics and Genomics* **284**, 147-159.
- Lorberth, R., Ritte, G., Willmitzer, L., and Kossmann, J.** (1998). Inhibition of a Starch-Granule-Bound Protein Leads to Modified Starch and Repression of Cold Sweetening. *Nat Biotech* **16**, 473-477.
- Maniatis, T., Sambrook, J., and Fritsch, E.F.** (1989). *Molecular Cloning : A Laboratory Manual.* (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory).
- Matsui, M., Fowler, J.H., and Walling, L.L.** (2006). Leucine Aminopeptidases: Diversity in Structure and Function. *Biological Chemistry* **387**, 1535-1544.
- Menéndez, C.M., Ritter, E., Schafer-Pregl, R., Walkemeier, B., Kalde, A., Salamini, F., and Gebhardt, C.** (2002). Cold-Sweetening in Diploid Potato. Mapping QTL and Candidate Genes. *Genetics* **162**, 1423 - 1434.
- Mikkelsen, R., Baunsgaard, L., and Blennow, A.** (2004). Functional Characterization of Alpha-Glucan, Water Dikinase, the Starch Phosphorylating Enzyme. *Biochem J* **377**, 525-532.
- Mikkelsen, R., Mutenda, K.E., Mant, A., Schurmann, P., and Blennow, A.** (2005). Alpha-Glucan, Water Dikinase (Gwd): A Plastidic Enzyme with Redox-Regulated and Coordinated Catalytic Activity and Binding Affinity. *Proc Natl Acad Sci U S A* **102**, 1785-1790.
- Mikkelsen, R., and Blennow, A.** (2005). Functional Domain Organization of the Potato Alpha-Glucan, Water Dikinase (GWD): Evidence for Separate Site Catalysis as Revealed by Limited Proteolysis and Deletion Mutants. *Biochem J* **385**, 355-361.
- Mikkelsen, R., Suszkiewicz, K., and Blennow, A.** (2006). A Novel Type Carbohydrate-Binding Module Identified in Alpha-Glucan, Water Dikinases Is Specific for Regulated Plastidial Starch Metabolism. *Biochemistry* **45**, 4674-4682.
- Milbourne, D., Pande, B., Bryan, G.J., and Kole, C.** (2007). Potato. In *Genome Mapping and Molecular Breeding in Plants: Pulses, Sugar and Tuber Crops*, C. Kole, ed (Springer Berlin Heidelberg), pp. 205-236.
- Mottram, D.S., Wedzicha, B.L., and Dodson, A.T.** (2002). Food Chemistry: Acryl amide Is Formed in the Maillard Reaction. *Nature* **419**, 448-449.

- Mouille, G., Maddelein, M.L., Libessart, N., Talaga, P., Decq, A., Delrue, B., and Ball, S.** (1996). Preamylopectin Processing: A Mandatory Step for Starch Biosynthesis in Plants. *Plant Cell* **8**, 1353-1366.
- Müller-Röber, B.T., Kossmann, J., Hannah, L.C., Willmitzer, L., and Sonnewald, U.** (1990). One of 2 Different ADP-Glucose Pyrophosphorylase Genes from Potato Responds Strongly to Elevated Levels of Sucrose. *Molecular & General Genetics* **224**, 136-146.
- Müller-Thurgau, H.** (1882). Über Zuckerrückbildung in Pflanzenteilen in Folge Niedriger Temperatur. *Landwirtschaftliches Jahrbuch der Schweiz* **11**, 1229-1238.
- Murayama, S., and Handa, H.** (2007). Genes for Alkaline/Neutral Invertase in Rice: Alkaline/Neutral Invertases Are Located in Plant Mitochondria and Also in Plastids. *Planta* **225**, 1193-1203.
- Nackley, A.G., Shabalina, S.A., Tchivileva, I.E., Satterfield, K., Korchynskyi, O., Makarov, S.S., Maixner, W., and Diatchenko, L.** (2006). Human Catechol-O-Methyltransferase Haplotypes Modulate Protein Expression by Altering mRNA Secondary Structure. *Science* **314**, 1930-1933.
- Narvaez-Vasquez, J., Tu, C.J., Park, S.Y., and Walling, L.L.** (2008). Targeting and Localization of Wound-Inducible Leucine Aminopeptidase a in Tomato Leaves. *Planta* **227**, 341-351.
- Neigenfind, J., Gyetvai, G., Basekow, R., Diehl, S., Achenbach, U., Gebhardt, C., Selbig, J., and Kersten, B.** (2008). Haplotype Inference from Unphased SNP Data in Heterozygous Polyploids Based on Sat. *BMC Genomics* **9**.
- Nicot, N., Hausman, J.-F., Hoffmann, L., and Evers, D.** (2005). Housekeeping Gene Selection for Real-Time RT-PCR Normalization in Potato During Biotic and Abiotic Stress. *J. Exp. Bot.* **56**, 2907-2914.
- Nielsen, T.H., Wischmann, B., Enevoldsen, K., and Moller, B.L.** (1994). Starch Phosphorylation in Potato Tubers Proceeds Concurrently with De Novo Biosynthesis of Starch. *Plant Physiology* **105**, 111-117.
- Niittylä, T., Messerli, G., Trevisan, M., Chen, J., Smith, A.M., and Zeeman, S.C.** (2004). A Previously Unknown Maltose Transporter Essential for Starch Degradation in Leaves. *Science* **303**, 87-89.
- Nonis, A., Ruperti, B., Falchi, R., Casatta, E., Thamasebi Enferadi, S., and Vizzotto, G.** (2007). Differential Expression and Regulation of a Neutral Invertase Encoding Gene from Peach (*Prunus persica*): Evidence for a Role in Fruit Development. *Physiologia Plantarum* **129**, 436-446.
- Nonis, A., Ruperti, B., Pierasco, A., Canaguier, A., Adam-Blondon, A.-F.O., Di Gaspero, G., and Vizzotto, G.** (2008). Neutral Invertases in Grapevine and Comparative Analysis with *Arabidopsis*, Poplar and Rice. *Planta* **229**, 129-142.
- Odeny, D.A., Stich, B., and Gebhardt, C.** (2010). Physical Organization of Mixed Protease Inhibitor Gene Clusters, Coordinated Expression and Association with Resistance to Late Blight at the StKI Locus on Potato Chromosome III. *Plant Cell and Environment* **33**, 2149-2161.

- Pautot, V., Holzer, F.M., Reisch, B., and Walling, L.L.** (1993). Leucine Aminopeptidase - an Inducible Component of the Defense Response in *Lycopersicon esculentum* (Tomato). Proceedings of the National Academy of Sciences of the United States of America **90**, 9906-9910.
- Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H.** (2011). SignalP 4.0: Discriminating Signal Peptides from Transmembrane Regions. Nat Meth **8**, 785-786.
- Pflieger, S., Lefebvre, V., and Causee, M.** (2001). The Candidate Gene Approach in Plant Genetics: A Review. Molecular Breeding **7**, 275-291.
- PGSC, The Potato Genome Sequencing Consortium.** (2011). Genome Sequence and Analysis of the Tuber Crop Potato. Nature **475**, 189-U194.
- Pluthero, F.G.** (1993). Rapid Purification of High-Activity Taq DNA Polymerase. Nucleic Acids Research **21**, 4850-4851.
- Preiss, J., Cress, D., Hutny, J., Morell, M., Bloom, M., Okita, T., and Anderson, J.** (1989). Regulation of Starch Synthesis. In Biocatalysis in Agricultural Biotechnology (American Chemical Society), pp. 84-92.
- Pressey, R., and Shaw, R.** (1966). Effect of Temperature on Invertase, Invertase Inhibitor, and Sugars in Potato Tubers. Plant Physiol. **41**, 1657-1661.
- Pressey, R.** (1967). Invertase Inhibitor from Potatoes: Purification, Characterization, and Reactivity with Plant Invertases. Plant Physiol. **42**, 1780-1786.
- Ramloch-Lorenz, K., Knudsen, S., and Sturm, A.** (1993). Molecular Characterization of the Gene for Carrot Cell Wall β -Fructosidase. The Plant Journal **4**, 545-554.
- Rausch, T., and Greiner, S.** (2004). Plant Protein Inhibitors of Invertases. Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics **1696**, 253-261.
- Ritte, G., Lorberth, R., and Steup, M.** (2000). Reversible Binding of the Starch-Related R1 Protein to the Surface of Transitory Starch Granules. The Plant Journal **21**, 387-391.
- Ritte, G., Lloyd, J.R., Eckermann, N., Rottmann, A., Kossmann, J., and Steup, M.** (2002). The Starch-Related R1 Protein Is an Alpha-Glucan, Water Dikinase. Proceedings of the National Academy of Sciences **99**, 7166-7171.
- Ritte, G., Heydenreich, M., Mahlow, S., Haebel, S., Kotting, O., and Steup, M.** (2006). Phosphorylation of C6- and C3-Positions of Glycosyl Residues in Starch Is Catalysed by Distinct Dikinases. FEBS Lett **580**, 4872-4876.
- Ritter, E., Debener, T., Barone, A., Salamini, F., and Gebhardt, C.** (1991). RFLP Mapping on Potato Chromosomes of Two Genes Controlling Extreme Resistance to Potato Virus X (PVX). Mol Gen Genet **227**, 81-85.
- Robertson, D.S.** (1985). A Possible Technique for Isolating Genomic DNA for Quantitative Traits in Plants. J. Theor. Bio **117**, 1-10.
- Roe, M.A., Faulks, R.M., and Belsten, J.L.** (1990). Role of Reducing Sugars and Amino Acids in Fry Colour of Chips from Potatoes Grown under Different Nitrogen Regimes. Journal of the science of food and agriculture **52**, 207-214.

- Roitsch, T., and González, M.-C.** (2004). Function and Regulation of Plant Invertases: Sweet Sensations. *Trends in Plant Science* **9**, 606-613.
- Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlén, M., and Nyrén, P.** (1996). Real-Time DNA Sequencing Using Detection of Pyrophosphate Release. *Analytical Biochemistry* **242**, 84-89.
- Ronning, C.M., Stegalkina, S.S., Ascenzi, R.A., Bougri, O., Hart, A.L., Utterbach, T.R., Vanaken, S.E., Riedmuller, S.B., White, J.A., Cho, J., Perteza, G.M., Lee, Y., Karamycheva, S., Sultana, R., Tsai, J., Quackenbush, J., Griffiths, H.M., Restrepo, S., Smart, C.D., Fry, W.E., van der Hoeven, R., Tanksley, S., Zhang, P., Jin, H., Yamamoto, M.L., Baker, B.J., and Buell, C.R.** (2003). Comparative Analyses of Potato Expressed Sequence Tag Libraries. *Plant Physiology* **131**, 419-429.
- Rosahl, S., Schmidt, R., Schell, J., and Willmitzer, L.** (1986). Isolation and Characterization of a Gene from *Solanum tuberosum*; Encoding Patatin, the Major Storage Protein of Potato Tubers. *Molecular and General Genetics MGG* **203**, 214-220.
- Rozen, S., and Skaletsky, H.** (2000). Primer3 on the WWW for General Users and for Biologist Programmers. *Methods in molecular biology (Clifton, N.J.)* **132**, 365-386.
- Ruiz-Rivero, O.J., and Prat, S.** (1998). A -308 deletion of the tomato LAP promoters is able to direct flower-specific and MeJA-induced expression in transgenic plants. *Plant Molecular Biology* **36**,
- Sattarzadeh, A., Achenbach, U., Lubeck, J., Strahwald, J., Tacke, E., Hofferbert, H.R., Rotstein, T., and 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. *Mol Breed* **4**, 301 - 312.
- Schäfer-Pregl, R., Ritter, E., Concilio, L., Hesselbach, J., Lovatti, L., Walkemeier, B., Thelen, H., Salamini, F., and Gebhardt, C.** (1998). Analysis of Quantitative Trait Loci (QTLs) and Quantitative Trait Alleles (QTAs) for Potato Tuber Yield and Starch Content. *Theoretical and Applied Genetics* **97**, 834-846.
- Schwartz, D.C., and Cantor, C.R.** (1984). Separation of Yeast Chromosome-Sized DNAs by Pulsed Field Gradient Gel Electrophoresis. *Cell* **37**, 67-75.
- Schwimmer, S., Makower, R.U., and Rorem, E.S.** (1961). Invertase and Invertase Inhibitor in Potato. *Plant Physiology* **36**, 313-316.
- Shallenberger, R.S., Smith, O., and Treadway, R.H.** (1959). Role of Sugars in the Browning Reaction in Potato Chips. *J. Agric. Food Chem.* **7**, 274-277.
- Shewry, P.R.** (1999). *Enzyme Inhibitors of Seeds: Types and Properties.* (Amsterdam:: R. Kluwer Academic Publishers).
- Smith, A.M., Zeeman, S.C., Thorneycroft, D., and Smith, S.M.** (2003). Starch Mobilization in Leaves. *J Exp Bot* **54**, 577-583.
- Smith, A.M., Zeeman, S.C., and Smith, S.M.** (2005). Starch Degradation. *Annu Rev Plant Biol* **56**, 73-98.

- Spooner, D.M., McLean, K., Ramsay, G., Waugh, R., and Bryan, G.J.** (2005). A Single Domestication for Potato Based on Multilocus Amplified Fragment Length Polymorphism Genotyping. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 14694-14699.
- Stadler, R.H., Blank, I., Varga, N., Robert, F., Hau, J., Guy, P.A., Robert, M.-C., and Riediker, S.** (2002). Food Chemistry: Acryl amide from Maillard Reaction Products. *Nature* **419**, 449-450.
- Stiekema, W.J., Heidekamp, F., Dirkse, W.G., Beckum, J., Haan, P., Bosch, C.T., and Louwse, J.D.** (1988). Molecular Cloning and Analysis of Four Potato Tuber mRNAs. *Plant Molecular Biology* **11**, 255-269.
- Sträter, N., Sun, L., Kantrowitz, E.R., and Lipscomb, W.N.** (1999). A Bicarbonate Ion as a General Base in the Mechanism of Peptide Hydrolysis by Zinc Leucine Aminopeptidase. *Proceedings of the National Academy of Sciences* **96**, 11151-11155.
- Sturm, A., and Chrispeels, M.J.** (1990). cDNA Cloning of Carrot Extracellular Beta-Fructosidase and Its Expression in Response to Wounding and Bacterial-Infection. *Plant Cell* **2**, 1107-1119.
- Sturm, A., Hess, D., Lee, H.-S., and Lienhard, S.** (1999). Neutral Invertase Is a Novel Type of Sucrose-Cleaving Enzyme. *Physiologia Plantarum* **107**, 159-165.
- Suh, S.-G., Peterson, J.E., Stiekema, W.J., and Hannapel, D.J.** (1990). Purification and Characterization of the 22-Kilodalton Potato Tuber Proteins. *Plant Physiol.* **94**, 40-45.
- Svendsen, I., Hejgaard, J, Mundy, J.** (1986). Complete Amino Acid Sequence of the Alpha-Amylase/Subtilisin Inhibitor from Barley Carlsberg Research Communications **51**, 43-50.
- Tang, G.-Q., Lüscher, M., and Sturm, A.** (1999). Antisense Repression of Vacuolar and Cell Wall Invertase in Transgenic Carrot Alters Early Plant Development and Sucrose Partitioning. *The Plant Cell Online* **11**, 177-190.
- Tanksley, S.** (1983). Molecular Markers in Plant Breeding. *Plant Molecular Biology Reporter* **1**, 3-8.
- Tetlow, I.J.** (2006). Understanding Storage Starch Biosynthesis in Plants: A Means to Quality Improvement. *Canadian Journal of Botany* **84**, 1167-1185.
- Tu, C.J., Park, S.Y., and Walling, L.L.** (2003). Isolation and Characterization of the Neutral Leucine Aminopeptidase (LAPn) of Tomato. *Plant Physiology* **132**, 243-255.
- Turgeon, R.** (1989). The Sink-Source Transition in Leaves. *Annual Review of Plant Physiology and Plant Molecular Biology* **40**, 119-138.
- Turrà, D., Bellin, D., Lorito, M., and Gebhardt, C.** (2009). Genotype-Dependent Expression of Specific Members of Potato Protease Inhibitor Gene Families in Different Tissues and in Response to Wounding and Nematode Infection. *Journal of Plant Physiology* **166**, 762-774.

- Urbany, C., Stich, B., Schmidt, L., Simon, L., Berding, H., Junghans, H., Niehoff, K.-H., Braun, A., Tacke, E., Hofferbert, H.-R., Lubeck, J., Strahwald, J., and Gebhardt, C.** (2011). Association Genetics in *Solanum Tuberosum* Provides New Insights into Potato Tuber Bruising and Enzymatic Tissue Discoloration. *BMC Genomics* **12**, 7.
- Vidal, M.** (1997). *The Reverse Two-Hybrid System*. (New York: Oxford University Press).
- Vignal, A., Milan, D., SanCristobal, M., and Eggen, A.** (2002). A Review on SNP and Other Types of Molecular Markers and Their Use in Animal Genetics. *Genetics Selection Evolution* **34**, 275-305.
- Wagner, S.L., Siegel, R.S., Vedvick, T.S., Raschke, W.C., and Vannostrand, W.E.** (1992). High-Level Expression, Purification, and Characterization of the Kunitz-Type Protease Inhibitor Domain of Protease Nexin-2/Amyloid Beta-Protein Precursor. *Biochemical and Biophysical Research Communications* **186**, 1138-1145.
- Wallace, B.** (1963). The Annual Invitation Lecture. Genetic Diversity, Genetic Uniformity, and Heterosis. *Canadian Journal of Genetics and Cytology* **5**, 239-253.
- Wang, H., Kohalmi, S.E., and Cutler, A.J.** (1996). An Improved Method for Polymerase Chain Reaction Using Whole Yeast Cells. *Analytical biochemistry* **237**, 145-146.
- Weber, A., Servaites, J.C., Geiger, D.R., Kofler, H., Hille, D., Groner, F., Hebbeker, U., and Flügge, U.I.** (2000). Identification, Purification, and Molecular Cloning of a Putative Plastidic Glucose Translocator. *Plant Cell* **12**, 787-801.
- Weil, M., Krausgrill, S., Schuster, A., and Rausch, T.** (1994). A 17-Kda *Nicotiana tabacum* Cell-Wall Peptide Acts as an in-Vitro Inhibitor of the Cell-Wall Isoform of Acid Invertase. *Planta* **193**, 438-445.
- Wulff, E., Torres, S., and Vigil, E.** (2002). Protocol for DNA Extraction from Potato Tubers. *Plant Molecular Biology Reporter* **20**, 187-187.
- Yamagishi, K., Mitsumori, C., and Kikuta, Y.** (1991). Nucleotide Sequence of a cDNA Encoding the Putative Trypsin Inhibitor in Potato Tuber. *Plant Molecular Biology* **17**, 287-288.
- Yesilirmak, F., and Sayers, Z.** (2009). Heterologous Expression of Plant Genes. *International Journal of Plant Genomics* **2009**.
- Yu, T.S., Kofler, H., Hausler, R.E., Hille, D., Flügge, U.I., Zeeman, S.C., Smith, A.M., Kossmann, J., Lloyd, J., Ritte, G., Steup, M., Lue, W.L., Chen, J., and Weber, A.** (2001). The Arabidopsis Sex1 Mutant Is Defective in the R1 Protein, a General Regulator of Starch Degradation in Plants, and Not in the Chloroplast Hexose Transporter. *Plant Cell* **13**, 1907-1918.
- Zeeman, S.C., Smith, S.M., and Smith, A.M.** (2004). The Breakdown of Starch in Leaves. *New Phytologist* **163**, 247-261.
- Zeeman, S.C., Smith, S.M., and Smith, A.M.** (2007). The Diurnal Metabolism of Leaf Starch. *Biochem J* **401**, 13-28.
- Zrenner, R., Schüler, K., and Sonnewald, U.** (1996). Soluble Acid Invertase Determines the Hexose-to-Sucrose Ratio in Cold-Stored Potato Tubers. *Planta* **198**, 246-252.

ABSTRACT

The development of diagnostic markers by exploiting the natural allelic variation of candidate genes is a powerful tool in tetraploid potato cultivars for precision breeding regarding complex agronomic traits. Multiple genetic and environmental factors are the basis for important quality traits like starch and processing characteristics. The latter can rely on the balance between tuber starch and sugars that affect potato chip quality due to enzymatic reactions at low temperature (cold-sweetening). The major aim was to screen for superior natural alleles in an interesting set of candidate genes.

A candidate gene approach for starch and chip quality traits mainly targets functional candidates involved in carbohydrate metabolism. First, an α -glucan water dikinase (*GWD*) and a phosphoglucan water dikinase (*PWD*) were analyzed which trigger the initiation of starch breakdown in *Solanum tuberosum*. Second, as a post-transcriptional regulatory factor of vacuolar invertase, a putative invertase inhibitor was studied for its contribution to tuber quality traits and its functional role. In order to identify new candidates which do not operate in the starch pathway, *in silico* homology studies with well characterized genes on the released *Solanum phureja* genome and “-omics” approaches were used as an experimental basis. A hitherto uncharacterized 6th invertase and a leucine aminopeptidase (*LAP*) were obtained as third and fourth candidate genes.

In all candidate genes, SNPs (single nucleotide polymorphisms) having an effect on starch and chip quality traits were identified. The functional candidates *GWD* and *PWD* showed moderate association with all tested tuber quality traits and explained between 4 % and 7 % of the phenotypic variance. The 6th invertase (*Inv6*) showed high associations and explained up to 10 % and 11 % of the phenotypic variation of chip quality and starch quality, respectively. The role of *Inv6* remains to be elucidated, as semi-quantitative expression studies showed rather low expression in tubers and no strong increase upon cold treatment. Interestingly, *LAP* showed the highest association with tuber quality traits and explained up to 17 % of starch variance and 10 % of variance of chip quality after cold-storage. *LAP* has been selected as a candidate gene based on a proteomics analysis showing strong differences at protein level between good and bad processing genotypes. Finally, superior alleles of the putative invertase inhibitor were identified which explained 10 % of the phenotypic variance of starch quality traits, whereas no association was detected for chip quality traits. Functional analysis revealed a tuber-specific expression pattern. *In vitro* inhibition assays of heterologously expressed invertase and inhibitor did not confirm an inhibition of invertase activity.

Although natural DNA variants in the different candidate genes showed associations with starch and chip quality traits, their functional role remains to be identified. However, the candidate gene approach is a powerful tool to identify diagnostic molecular markers within functional genes which will facilitate marker-assisted selection and increase breeding efficiency.

ZUSAMMENFASSUNG

Die Entwicklung von diagnostischen molekularen Markern ist eine viel versprechende Methode, um bei den komplexen und agronomisch sehr bedeutsamen Qualitätsmerkmalen der Kartoffel präzise Züchtungserfolge zu erzielen. Die Entwicklung dieser Marker beruht auf natürlich auftretenden Varianten/Allelen in Kandidatengen, welche wichtige Stoffwechselschritte katalysieren. Kartoffeln dienen als Grundnahrungsmittel und als wichtiger Lieferant für Stärke. Kältelagerung von Kartoffeln minimiert die Ausbreitung von Pathogenen und vermeidet Keimung, jedoch bewirken niedrige Temperaturen auch, dass Stärke in reduzierende Zucker (Glukose und Fruktose) umgewandelt wird, was in der weiteren Verarbeitung der Knollen störend wirkt und somit unerwünscht ist. Die Zielsetzung dieser Arbeit lag darin, in einem einzigartigen Set von neuen Kandidatengen natürliche überlegene Allele für verbesserte Stärke- und Verarbeitungseigenschaften zu identifizieren.

Die Standardprozedur für die Auswahl von Kandidatengen in Bezug auf Stärke- und Verarbeitungsmerkmale basiert auf funktionellen Kandidaten, die maßgeblich am Kohlenhydratstoffwechsel beteiligt sind. So wurden eine α -Glucan Wasser Dikinase (*GWD*) und eine Phosphoglucan Wasser Dikinase (*PWD*) analysiert, die den Stärkeabbau in *Solanum tuberosum* initiieren. Zudem wurde ein möglicher Invertase-Inhibitor auf seinen Einfluss auf die Knollenqualität und seine Funktionsweise als regulierender Faktor der vacuolären Invertase hin untersucht. Um jedoch auch neue Kandidatengene zu identifizieren, die nicht unmittelbar im Stärke-Stoffwechselweg aktiv sind, wurden *in silico* Homologiestudien von bereits charakterisierten Genen in der veröffentlichten *Solanum phureja* Genomsequenz und vergleichende Proteomanalysen ("*-omics*") durchgeführt. Eine bisher uncharakterisierte sechste Invertase und eine Leucin-Aminopeptidase (*LAP*) konnten auf diese Weise als weitere Kandidatengene gefunden werden.

In allen Kandidatengen wurden SNPs (*single nucleotide polymorphisms*) identifiziert, die mit Stärkegehalt und Verarbeitungsqualität assoziiert sind. Die funktionellen Kandidaten *GWD* und *PWD* zeigen eine Assoziation mit allen getesteten Merkmalen und zwischen vier und sieben Prozent der phänotypischen Varianz kann erklärt werden. Bei der sechsten Invertase (*Inv6*) wurde ein etwas höherer Assoziationsgrad nachgewiesen, der zwischen 10 und 11 % der phänotypischen Variation erklärt. Da jedoch die semi-quantitative Expressionsanalyse vergleichsweise geringe Expressionen in den Kartoffelknollen zeigte und zudem kein starker Anstieg in der Expression durch Kältelagerung zu verzeichnen war, ist die genaue Wirkungsweise der *Inv6* auf verbesserte Knolleneigenschaften noch zu untersuchen. Bemerkenswerterweise wies *LAP* die stärkste Assoziation mit Knollenqualität auf und erklärt bis zu 17 % der Variation des Stärkegehalts und 10 % der Variationen der Verarbeitungseigenschaften nach Kältelagerung. Basierend auf der Proteom-Analyse konnten deutliche Expressionsunterschiede auf Proteinebene zwischen guten und schlechten Verarbeitungssorten

festgestellt werden. Für den putativen Invertase-Inhibitor konnten ebenfalls positiv assoziierte Allele identifiziert werden, die bis zu 10 % der phänotypischen Varianz der Stärkemenge erklären, hingegen auf die Verarbeitungseigenschaft keinen Einfluss haben. Eine funktionelle Analyse zeigte zwar eine auf das Knollengewebe begrenzte Expression, heterolog exprimiertes Protein führte jedoch zu keiner Inhibierung der Invertase.

Obgleich die natürlichen Varianten der Kandidatengene deutliche Assoziationen mit Stärkegehalt und Verarbeitungsmerkmalen aufweisen, bleibt ihre genaue funktionelle Rolle noch zu klären. Der Ansatz, verschiedenste Kandidatengene zu analysieren ist eine bedeutende Methode, um eine Vielfalt an diagnostischen Markern zu erhalten, die die Marker-gestützte Selektion in Kreuzungspopulationen künftig vereinfachen und somit die Zuchteffizienz signifikant steigern können.

APPENDIX

A. List of figures:

- A.1 *1421* locus on the genetic map of chromosome 3.
- A.2 *GWD* locus on the genetic map of chromosome 5.
- A.3 *Inv6* locus on the genetic map of chromosome 8.
- A.4 *PWD* locus on the genetic map of chromosome 9.
- A.5 *LAP* locus on the genetic map of chromosome 12.
- A.6 Amplicon sequence and SNP positions of *GWD*.
- A.7 Amplicon sequence and SNP positions of *PWD*.
- A.8 Amplicon sequence and SNP positions of *Inv6*.
- A.9 Amplicon sequence and SNP positions of *LAP*.
- A.10 Amplicon sequence and SNP positions of *1421*.
- A.11 Alignment of *GWD* and *PWD* protein sequences.
- A.12 Alignment of all Kunitz-type protease inhibitor group C cDNA sequences.
- A.13 Phylogenetic tree of invertase sequences of different species.

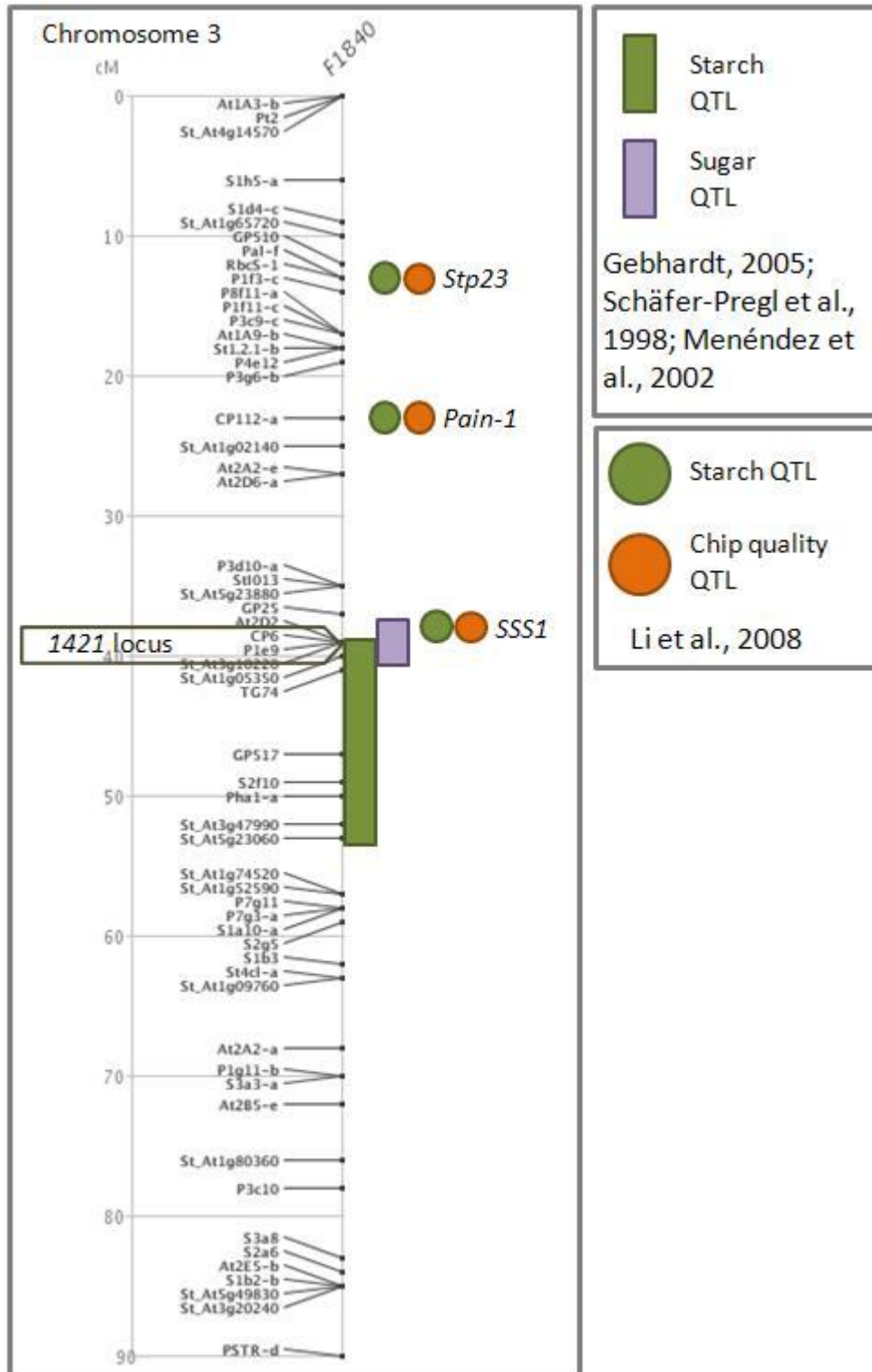
B. List of tables:

- B.1 Detailed exon and intron lengths of *GWD*.
- B.2 Detailed exon and intron lengths of *PWD*.
- B.3 Detailed exon and intron lengths of *Inv6*.
- B.4 Detailed exon and intron lengths of *LAP_N*.
- B.5 Detailed exon and intron lengths of *LAP_A*.
- B.6 Continued list of calculated haplotypes of *GWD*.
- B.7 Continued list of calculated haplotypes of *PWD*.
- B.8 Continued list of calculated haplotypes of *Inv6*.

C. Supplemental sequence data provided on CD:

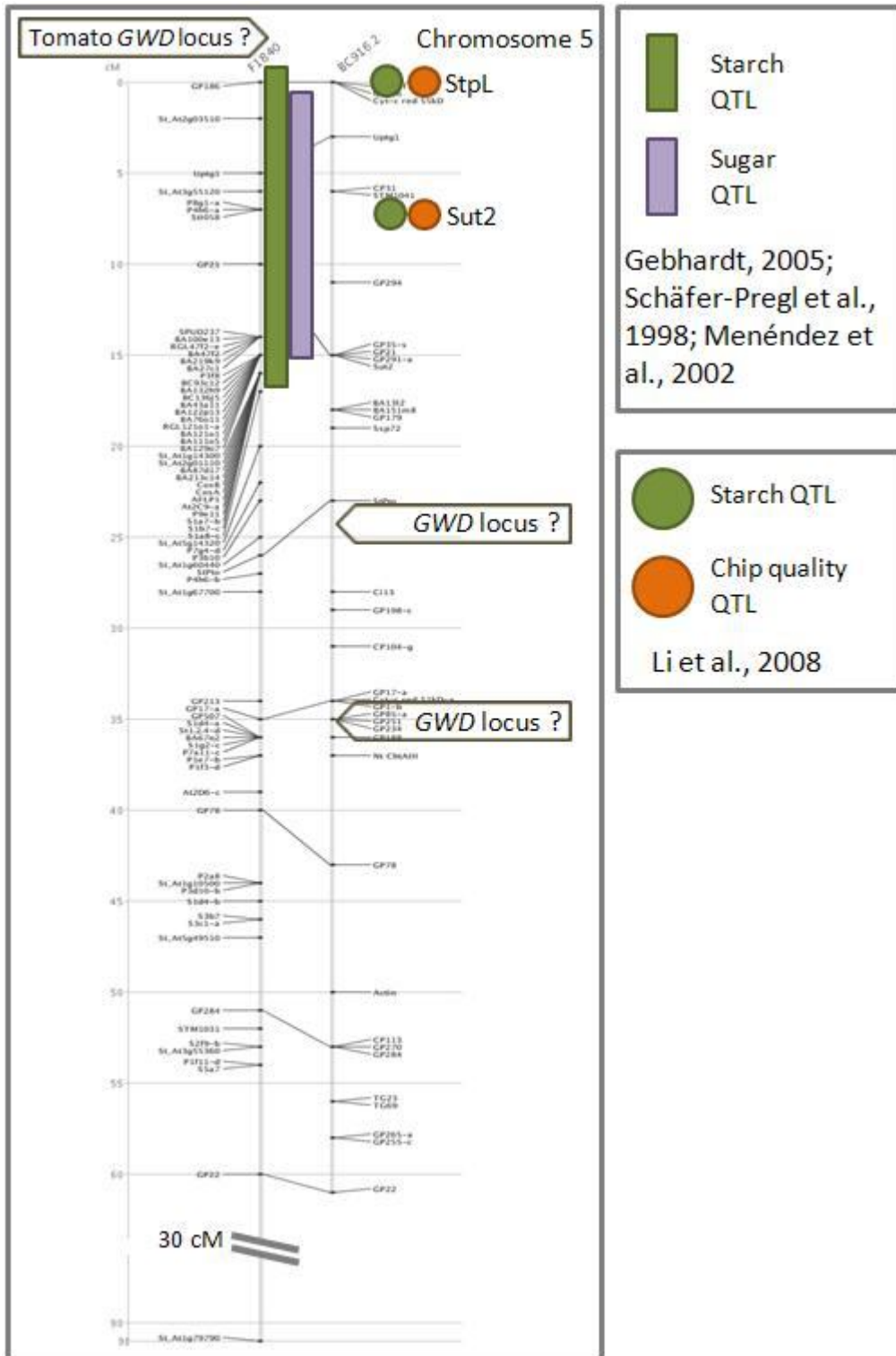
- C.1 Genomic sequence of contig 13(BAC BA202-H17).
- C.2 Genomic sequence of *GWD* (BAC BA202-H17).
- C.3 Genomic sequence of *PWD* (BAC BA19-F1).
- C.4 cDNA and protein sequence of *PWD*.
- C.5 Genomic sequence of *Inv6*.
- C.6 cDNA and protein sequence of *Inv6*.
- C.7 Genomic sequences of the *LAP_N* and *LAP_A* tandem.
- C.8 cDNA and protein sequences of *LAP_N* and *LAP_A*.
- C.9 Alignment of the *GWD* BAC sequence and the *GWD* sequence of *S. phureja*.

A.1 1421 locus on the genetic map of chromosome 3.



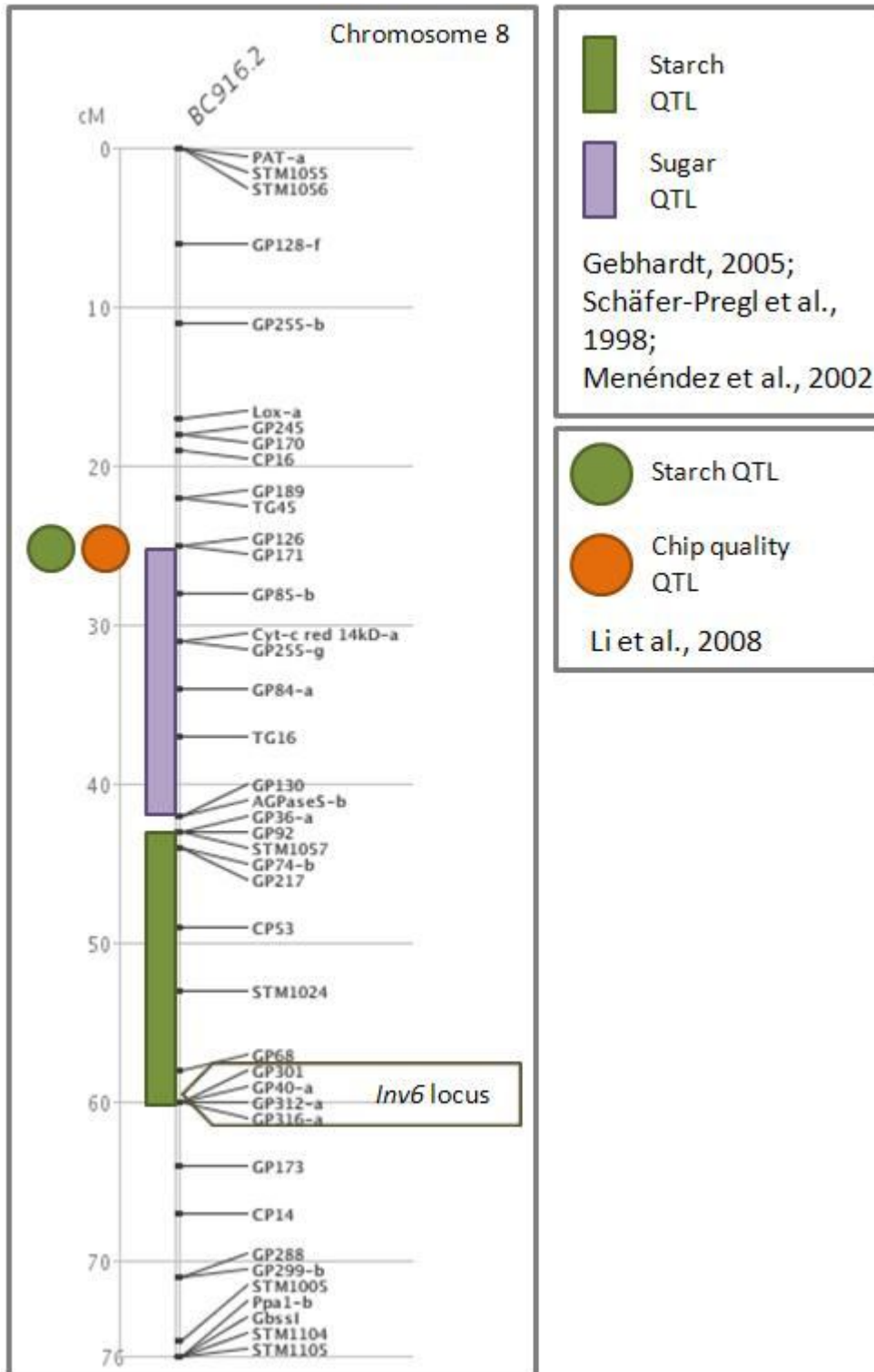
Genetic map of chromosome 3. The markers of the genetic map of the PoMaMo database were used to map the candidate gene to an approximate position on chromosome 3. Different starch and sugar QTL and several associated other candidate genes from literature were added.

A.2 *GWD* locus on the genetic map of chromosome 5.

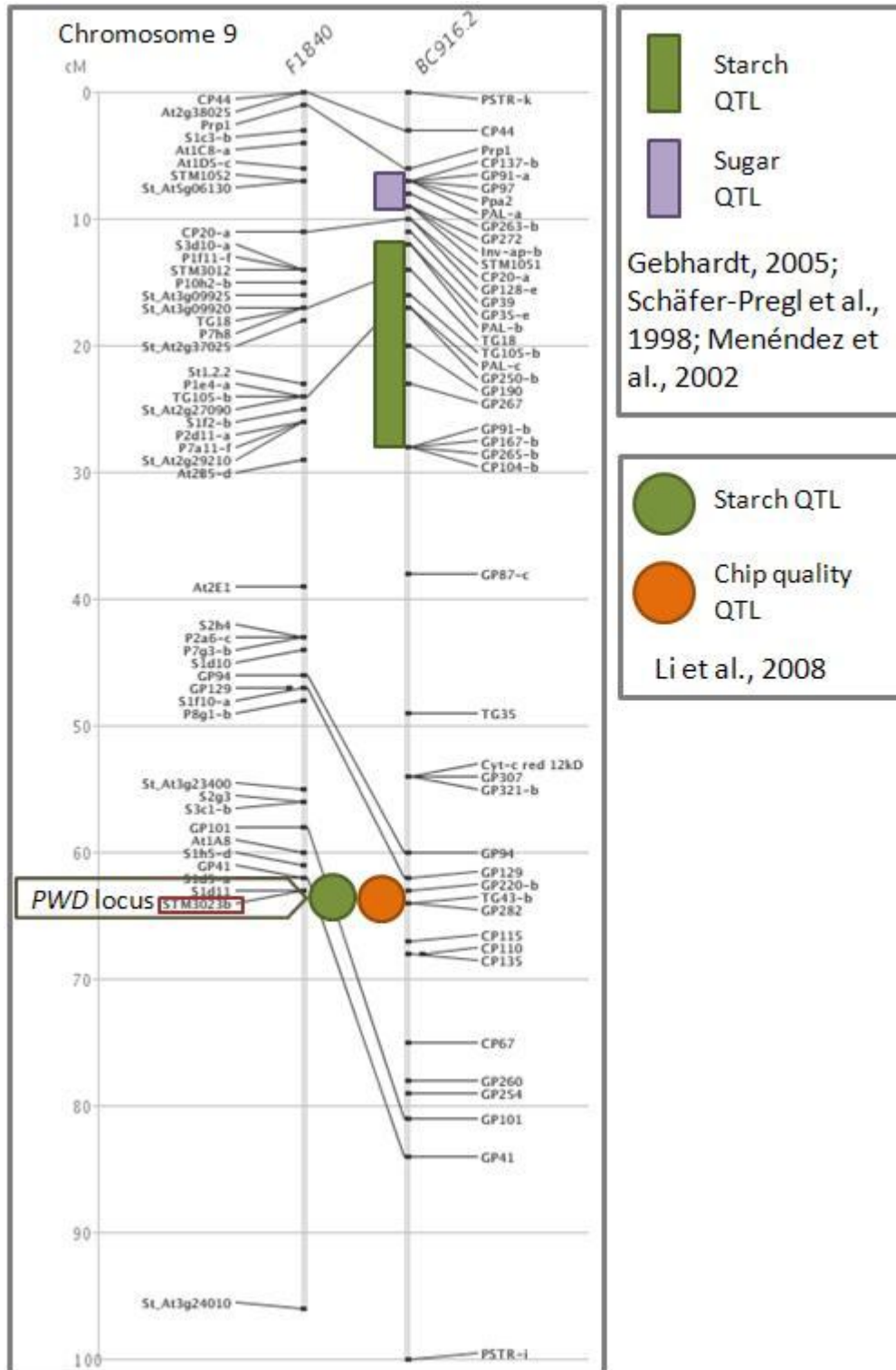


Genetic map of chromosome 5. The markers of the genetic map of the PoMaMo database and data from the preliminary tomato chromosome 5 were used to map the candidate gene *GWD*. The position remains inconclusive as different results were obtained for both databases. Different starch and sugar QTL from literature were added.

A.3 *Inv6* locus on the genetic map of chromosome 8.

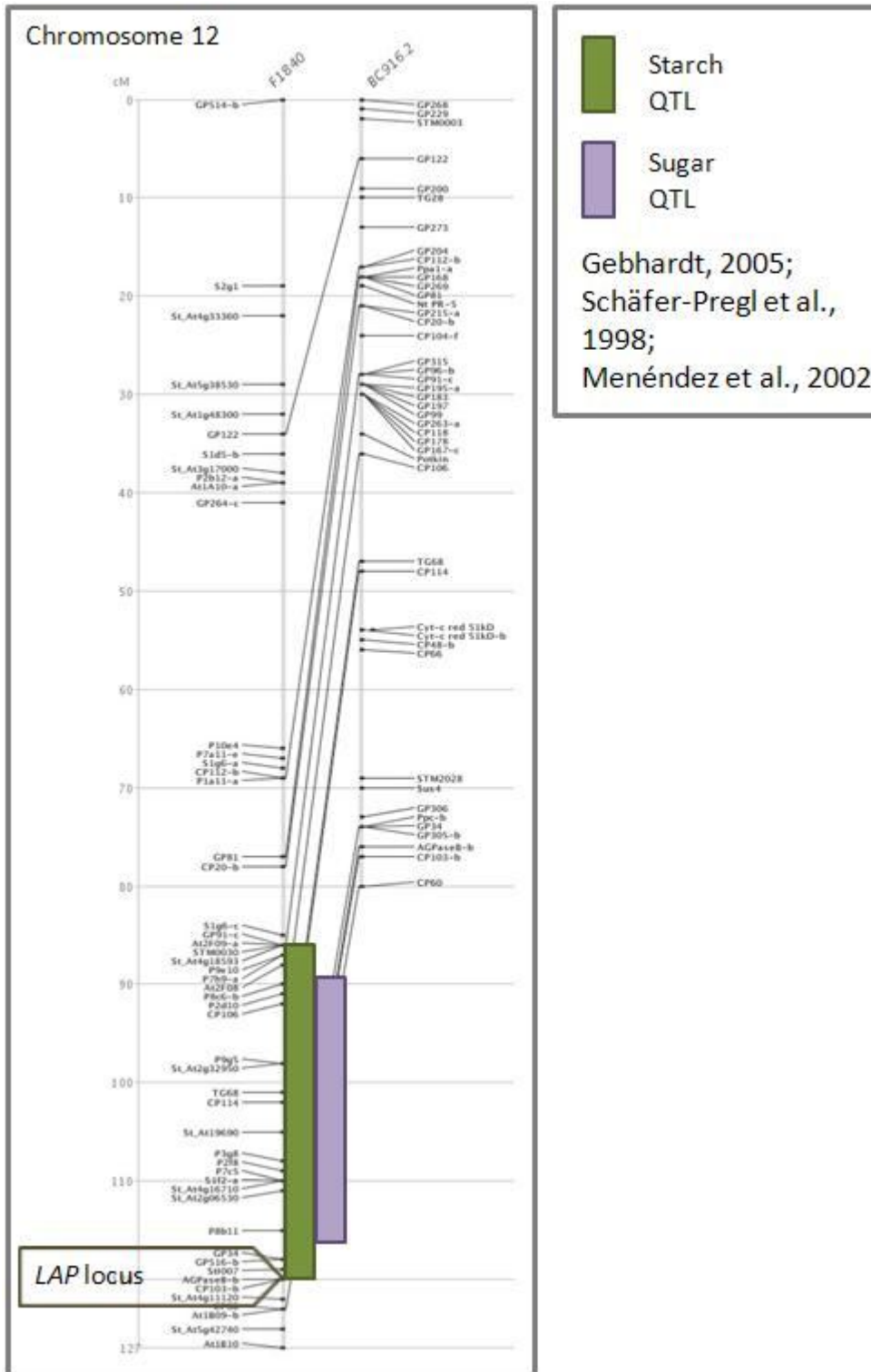


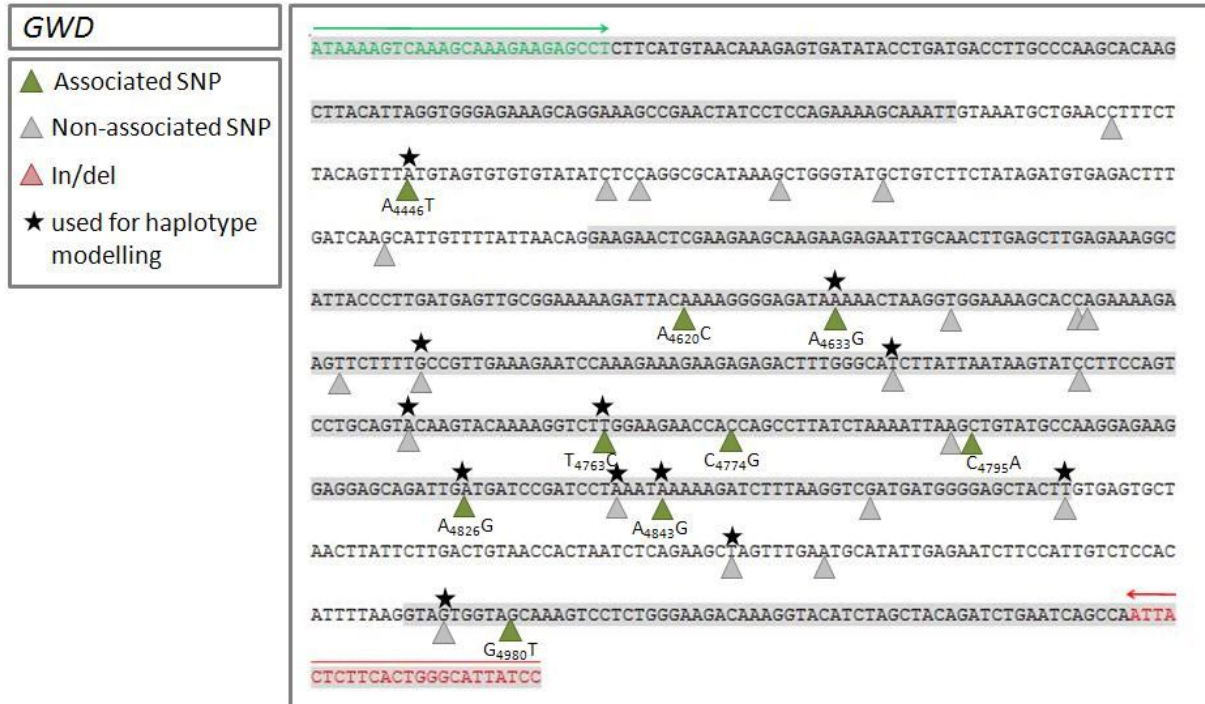
A.4 *PWD* locus on the genetic map of chromosome 9.



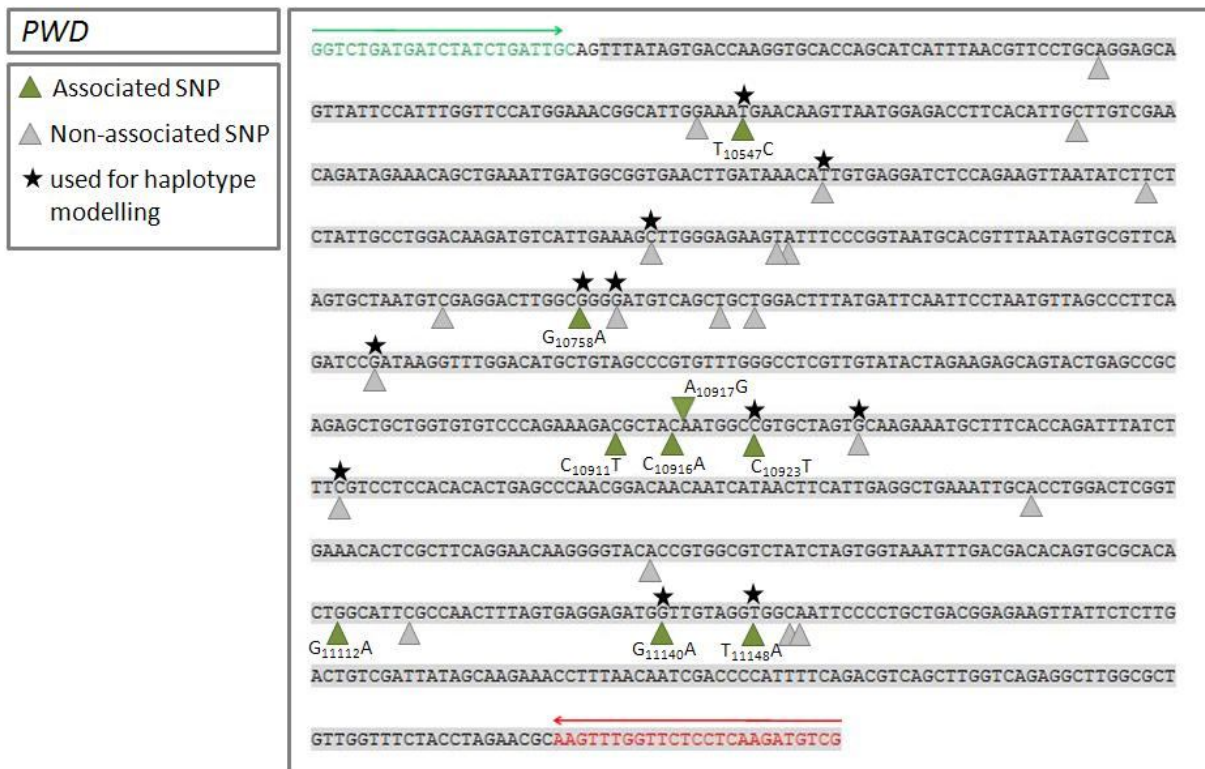
The red box indicates a marker which was identified to be associated with starch and chips quality traits. It is located within the same superscaffold which was identified for *PWD*.

A.5 *LAP* locus on the genetic map of chromosome 12.



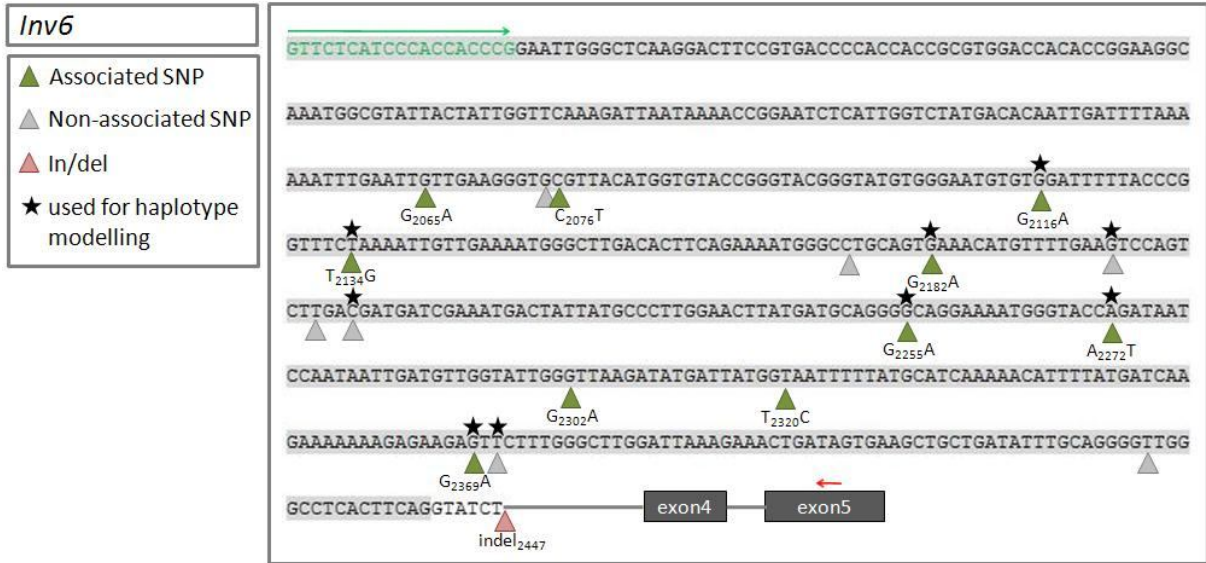
A.6 Amplicon sequence and SNP positions of *GWD*.

The part of the genomic sequence of *GWD* is displayed which was analyzed in the association analysis. The sequencing primers are indicated by a green and red arrow. Exons sequences are highlighted in grey. All SNPs and indels are marked by a triangle. The SNP positions of the associated SNPs are given. The SNPs used for the haplotype modeling display an asterisk.

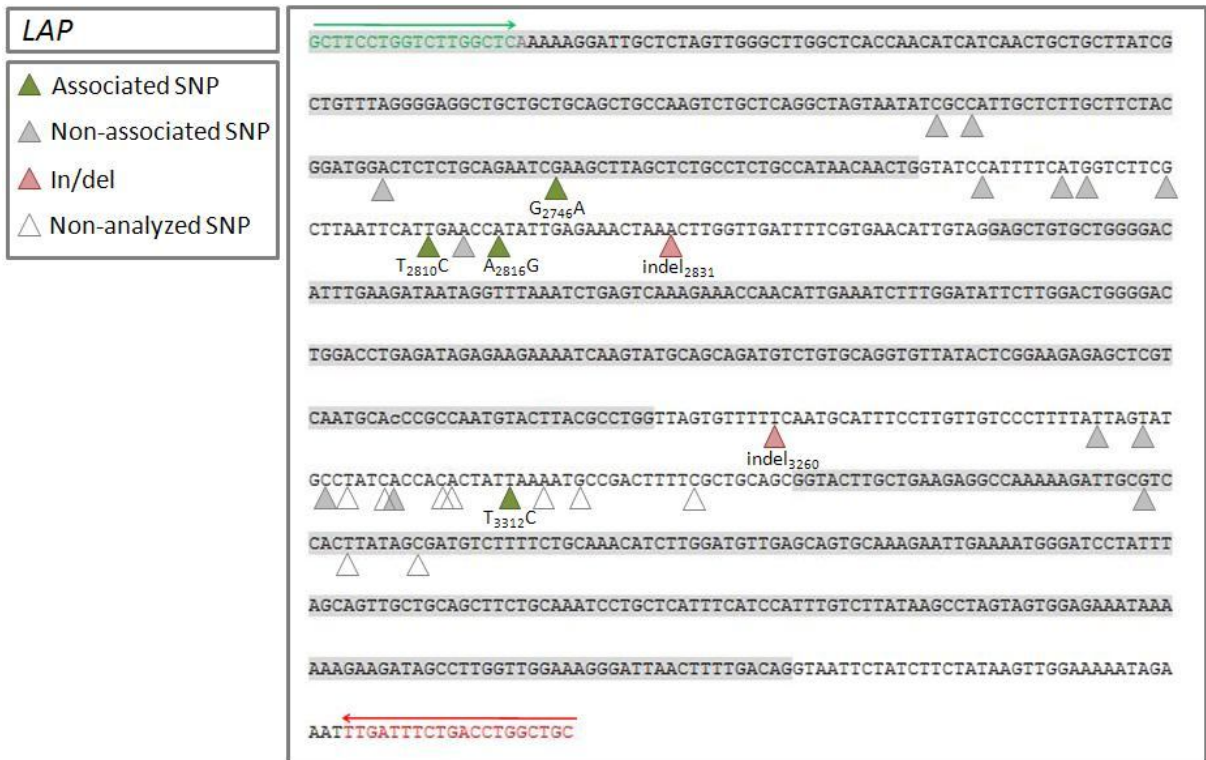
A.7 Amplicon sequence and SNP positions of *PWD*.

For a detailed description: Figure A.6

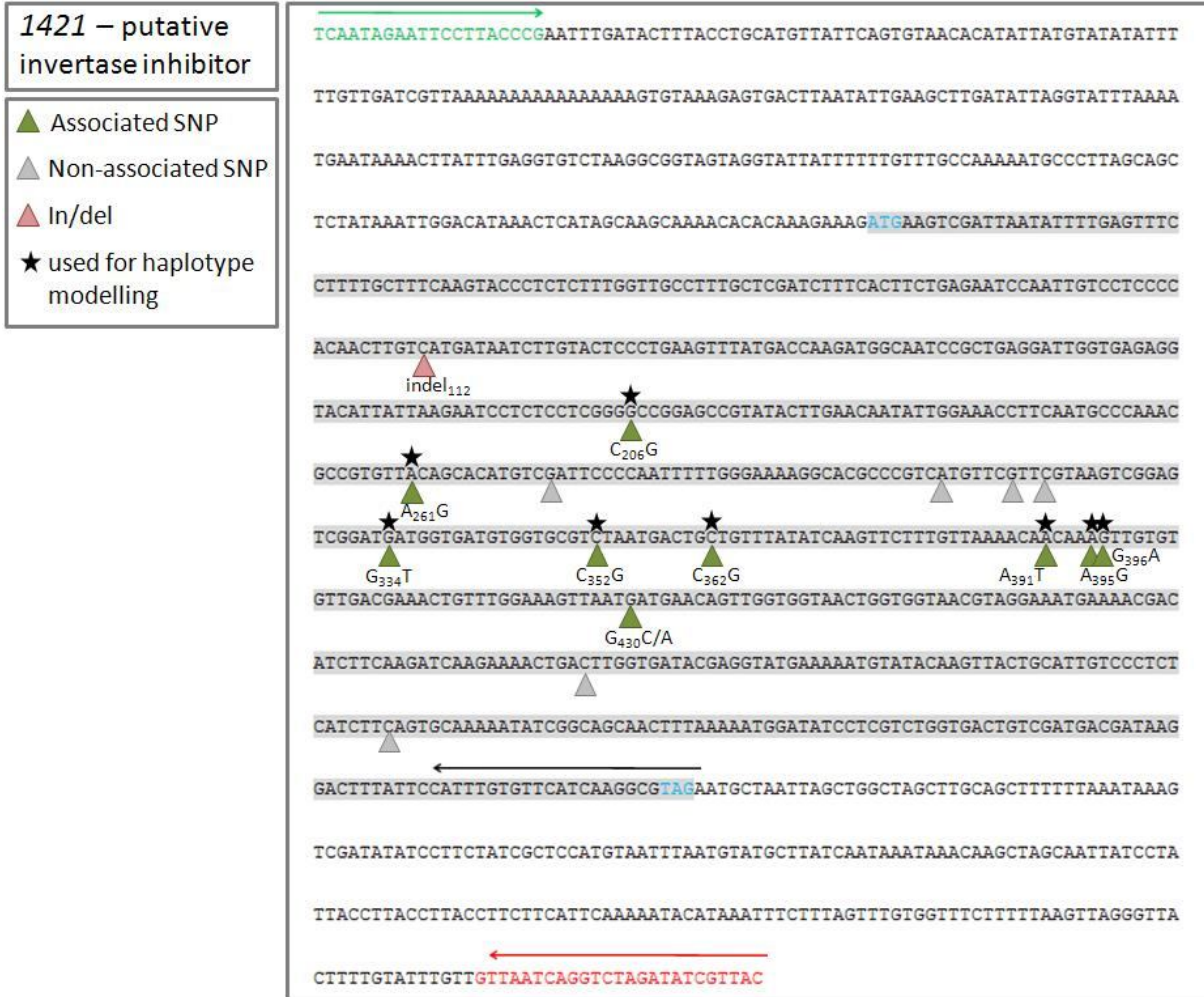
A.8 Amplicon sequence and SNP positions of *Inv6*.



A.9 Amplicon sequence and SNP positions of *LAP*.



A.10 Amplicon sequence and SNP positions of 1421.



For a general description: Figure A.6. The black arrow shows the internal reverse sequencing primer.

A.11 Alignment of GWD and PWD protein sequences.

```

StGWD : MSNSLGNLLVQGRITSTVLESHKSRISPPCVGGNSLFCQDVIKSKSPISNEFRNRIKVVQKRIKPMKKRAFSSSPHAILITDTSELAEKFSLENIIELOV : 100
StPWD : -MDSMELSHRYNS-VLVNKKKCPQIKISKQFNVVHQTSHRSVRNLITLLEPRNLCGFMORRVRKGLVLC-----VSSVREINQNKGRNKSSSSSTKVL : 90

StGWD : DVKDPISGDVSHVDFROVTVNSDQLFLHAGAVKFGKEWNLSPNDRPDDGTAVYKNAKRTIPKVKSSNSIDRLRIRDTAHEAHEFLIYDEVHDKWIKNNSEN : 200
StPWD : QLRERLDHGVVGEHIAVLGSADELGSFKKNIMMDWII-----ENGWICELEVRSGETLEKREVIIVKQKMLVINGSNRILKTP-----EGGE : 173

StGWD : ERVLSRKEIRGPDVSVPEELVHOSYLREKRGKQNYDPEKEEYEAARTELOEETIARGASTQDIRRRLTKTNDKSSCKEPELHVTSQIPDDLAQAQ : 300
StPWD : RRLVQGVN-----VLDPEVALLPLDPEVEKVVETSDNGAKIISQAVVDPVVTSPEVEQWQCRASPEVRSNDQLSDKN-----RQVITSLGTGISL : 261

StGWD : AVLRWEKACAPNVPPEKQVPESEARRELQLEKPKGTLDELKPKITKSKIKVEKHQKSSFAVERIQKRRDFGHANVYSSPAVQVQKVLSSPPA : 400
StPWD : KLVGGKNAKRWTRKLEVVRELVVENMDSHRLSALTYAAVYLKWNITLQIPLCEDGGHRR-----NHAETSRLLREVEKVLRRRDTILQELLV : 353

StGWD : LSKIKLYAKEKEEIDDPMLNKVFRKVDGELIVVAKSSGKIKVHATDANCPITLHWALSKSPGEMVPPSSILPPGSIILLDKAAETPFSASSSDGLP : 500
StPWD : LRKMQPCLSPKASRTASVPLTIRLDIAH-----RNLIPHDFKCKIKHILQN----- : 400

StGWD : SKVQSLDIVIDDGNFVGMPPVLLSGEKKIKNQCSDFYVGFSAASKIALKAGDGGSCAKSKLTKTADMESEAKKSEPHRNTAADLTEDATSAGELGFAG : 600
StPWD : -----KTHRNAGPELDVSTEAMLERIKQPEQYSEAFVQEKIFHNLKQFFNAGSL : 452

StGWD : ILVMMRFMATRQLIWNKNYVVKPREISKACDRITDLLQNAFVSHPMYREIIRRMMSIVGRGGECVQRIKDELIVVQRNNDQKGMEEWHQKIHNTS : 700
StPWD : -----DEQESMPESLDGSSLSMLSSFLSKKELVPRDEKHNVSEITERTEILVR-----TNSLNALREVLAKGLESQRNDAR : 526

StGWD : PDDVYICCALIDYIKSDFDLGVVWKTANVNTKERLISYDVAHSENFPRDQKGLLRDLGHVQILKAMHSGADLESANANCMGYKTEGEGFMVWQ : 800
StPWD : DASLAMPQKWR-----LDELISYEDYAVLLSRVFNVAVALG-----ADWIAENVTKNISWNPFC-----ALTVGIC : 591

StGWD : INPVSCLPSPGFQDLHFVLDHVEDKNEVITLLEAREELRPLLKNNRLKOLLDLIATDSTVRAVERGYEETNNANPEKIMYFISLVLENLALS : 900
StPWD : QLGSCWKE-----SSQK-----AVENBLSKFRGLSSLEGESE----- : 625

StGWD : DDNEDLVYCLKGNQALSMNGEDNHQAFEAQAVIDRHLALASKAEVYHRLQPSAVYCSITLGVQWAGNITLLEILIRACSAASLSSLNRPDPVRR : 1000
StPWD : -----DCKTIFALRLKADRSRLTEYSEITLLQEFPEKVIDLCKSLGTEPNTIATTEAEIRACVVVQVSKDAILLKKVRR : 704

StGWD : HANLGSQCLISVENVVYVWVDEMSVQNEIYEKPTLVAKSVKPEKIDBDG---AVAIKTPDMPVIVSHVSVRANRQVQPAHCFWENLDAQAKE : 1096
StPWD : IIGSSGMDVIVPEPAFELLIOVERITIPGLPSSATGIVILVVKADCDSEVTAAGSNIQWVLLQELPHLSHLVVRARQKVVFRCDDEKQKSDVQRL : 804

StGWD : ERMLDKPKPESDITLYSEVNEIELQSSNVLVEAEISATLRLVKKQFGCCYAFSADEFTSEMVG-----DKSRNTAVL : 1167
StPWD : SKVRLPEASSVVKLIDASSSEKAGGVSPNKLPSNASSAGATSSDSSASSLAVKSSQVKEVGPTRGVVPLVDADITQSGAKAASCAQIASTATSTRVYS : 904

StGWD : KGVVSSVGLHTSVALPFCVPSKMLSDDDNQC---VAKELDKTKKISECD--FSAMGSRITVLDGAPAOVLVKLEKKG : 1244
StPWD : DQCAPASINVPAGAVIPFGSMETALEMNKIMETPILLVQVETAEIDGSELDKHCEDLRLKSSILPQDVIETSLGEVFPFGNARIVRSSANVSDLAGMS : 1004

StGWD : GSEMPWPGDEGKRWEQAWMAKKWASKWERAVYSTKVRLDHDYLQMAVLVQETINADYAVLIHTINPSSGQDSEIYAVWVNGICEHTVGAYPERA : 1343
StPWD : HALEYDSIEINVSDEIRFGHAVRVAWASLYTRRAVLSRPRAGVSKDATMAVLVQEMISFDLSFVLHTLSEPTDNNFNPIAEIAPGLIGETLRSGRTEP : 1104

StGWD : LSFICKKKQLNSPQVLGYPSPKIGLEIKRSHIRSDSNCEDDTSCYAGAGLYDSVPMDSSEKVVLDYSSDPLITDGNFRQITLSNITARACHATEELYGSPQ : 1443
StPWD : WRLSSCKFDT-----VRIIAPANFSPEMVVCGNS-----PADGCVILLVDYSKPLNIIPIFRROLGQRLFAVSEVYLERKFGSPQ : 1181

StGWD : DTEGVVRDGRIVVQTRPQM- : 1463
StPWD : DVEGCLVQNEIFVQSRPQPQ : 1202

```

Alignment of GWD and PWD protein sequences. Conserved and annotated domains are indicated by boxes. Blue: Chloroplastic target peptide; light green: carbohydrate binding domains; dark green: nucleotide binding domain. The catalytic histidine is marked by an arrow.

A.12 Alignment of all Kunitz-type protease inhibitor group C cDNA sequences.

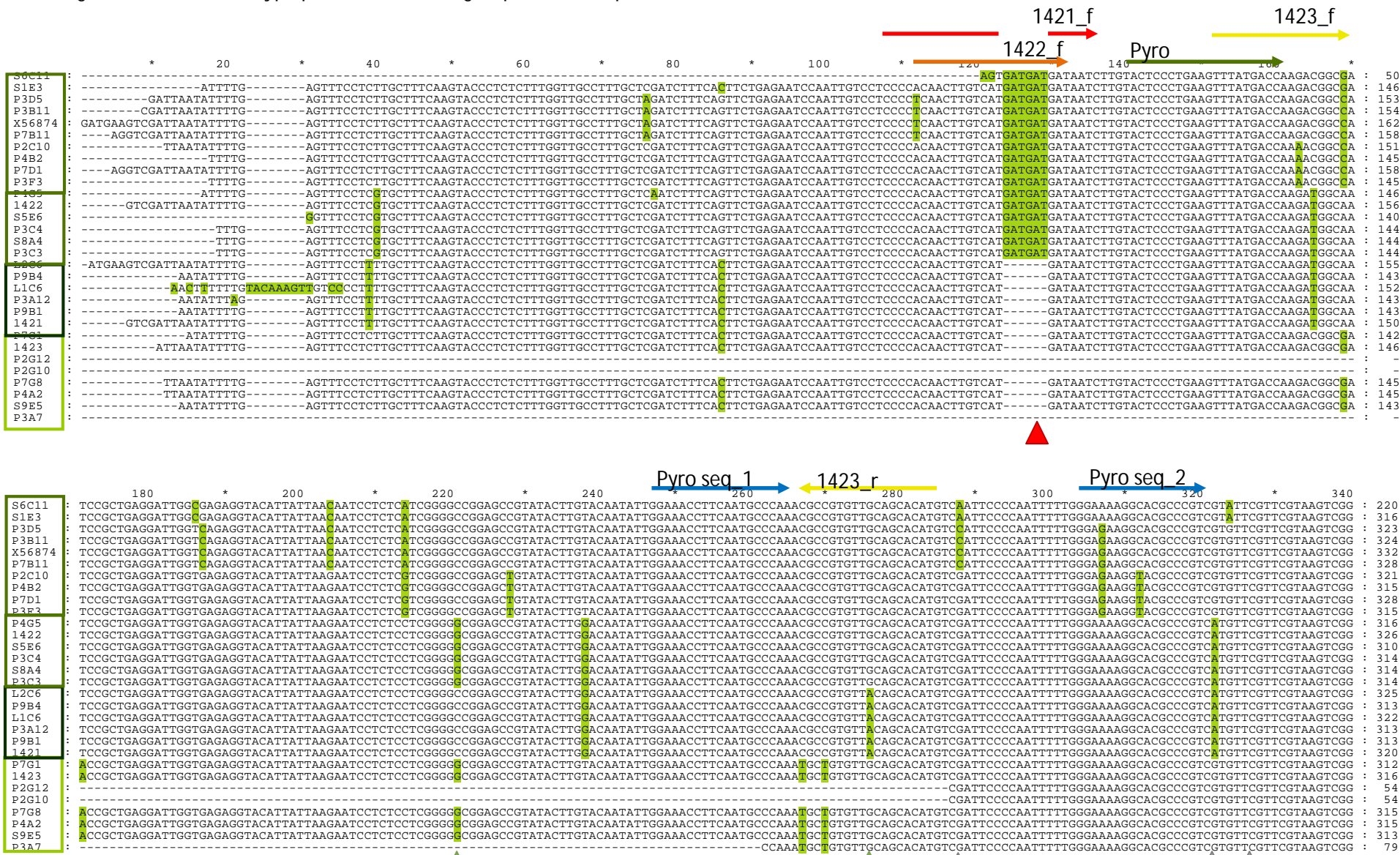
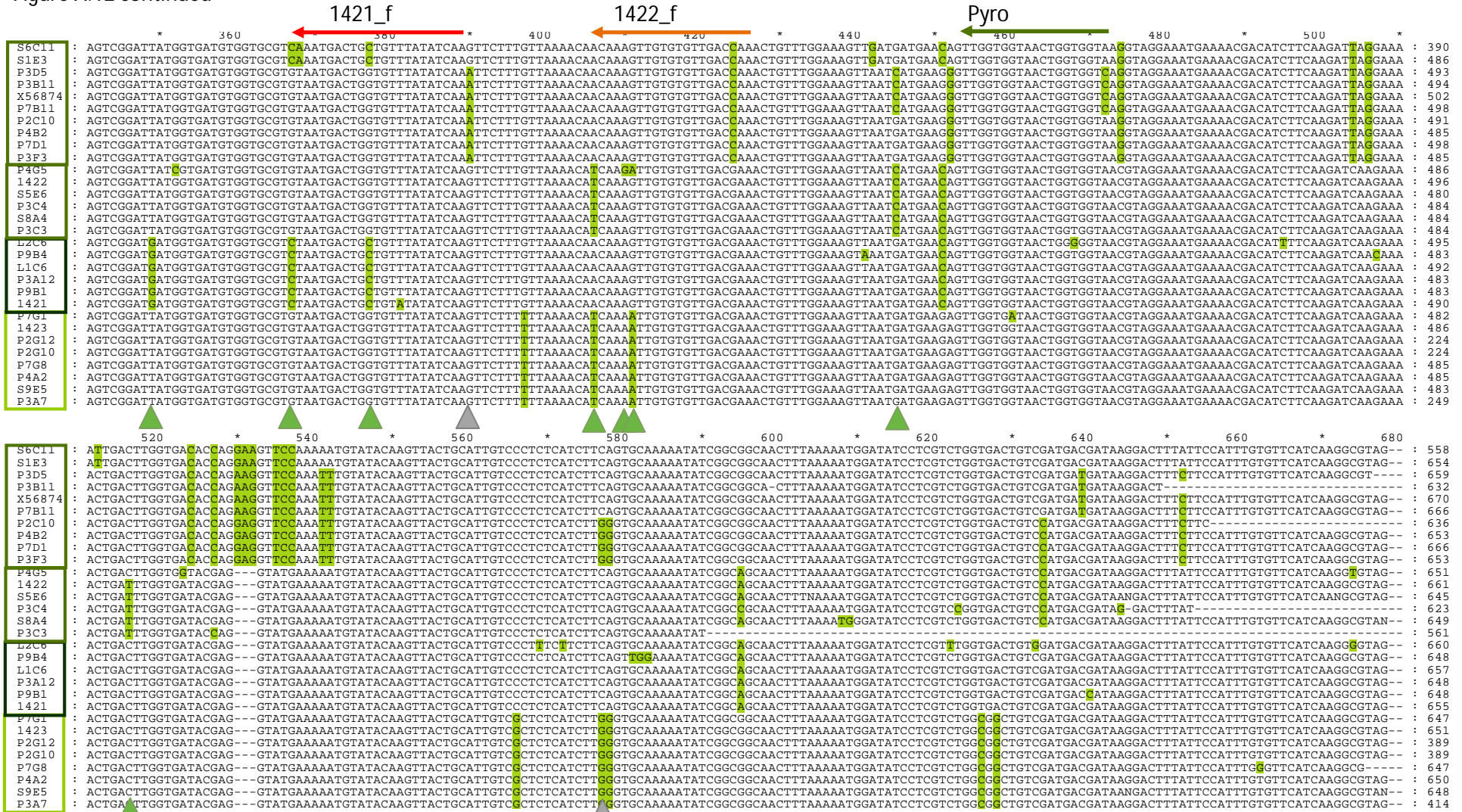
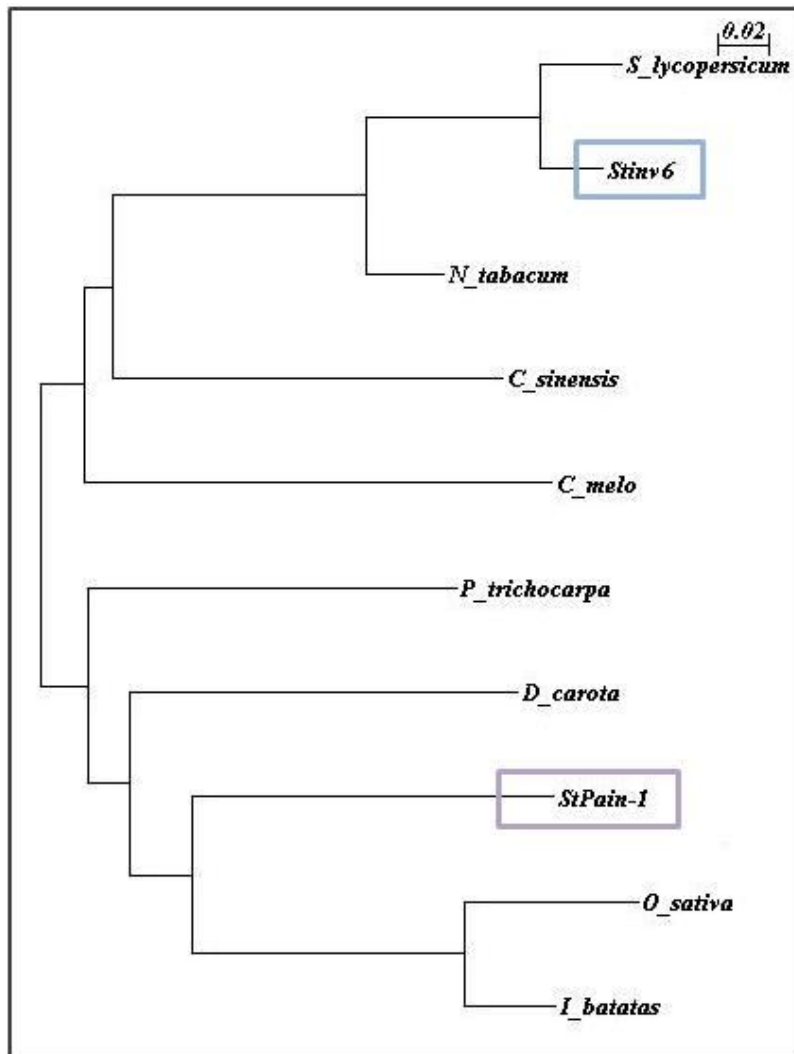


Figure A.12 continued



A.13 Phylogenetic tree of invertase sequences of different species.



Phylogenetic tree of invertase sequences which were obtained from various blast analyses. One half of the sequences showed higher similarity to the sequence of Pain-1, whereas on half was grouped into one class with Inv6. The latter one was further analyzed in the thesis.

B.1 Detailed exon and intron lengths of *GWD*

#	Exon length in bp	Intron length in bp
5'UTR_1	82	-
5'UTR_2	22	657
1	252	715
2	46	115
3	108	96
4	79	90
5	93	81
6	154	2365
7	225	118
8	343	92
9	109	112
10	98	320
11	140	127
12	90	530
13	133	211
14	163	108
15	91	74
16	186	1398
17	63	74
18	63	79
19	59	93
20	153	102
21	87	651
22	174	84
23	107	357
24	216	80
25	99	562
26	138	136
27	125	998
28	112	291
29	99	354
30	174	273
31	94	230
32	113	357
33	205	-
3'UTR	352	-
Σ	4847	11930
Σ total	16777	
w/o UTRs	4391	11273
Σ total	15664	

B.2 Detailed exon and intron lengths of *PWD*

#	Exon length in bp	Intron length in bp
1	212	1013
2	616	667
3	87	74
4	132	816
5	126	121
6	179	679
7	85	1630
8	168	89
9	53	283
10	209	88
11	119	571
12	70	134
13	119	381
14	87	108
15	87	486
16	77	101
17	166	143
18	112	385
19	905	-
Σ	3609	7769
Σ total	11378	

B.3 Detailed exon and intron lengths of *Inv6*

#	Exon length in bp	Intron length in bp
1	424	252
2	9	897
3	857	214
4	162	79
5	236	480
6	88	80
7	186	-
Σ	1962	2002
Σ total	3964	

B.4 Detailed exon and intron lengths of *LAP_N*

#	Exon length in bp	Intron length in bp
1	216	2146
2	415	81
3	196	87
4	225	145
5	127	97
6	96	667
7	73	863
8	70	364
9	179	716
10	137	-
Σ	1734	5166
Σ total	6900	

B.5 Detailed exon and intron lengths of *LAP_A*

#	Exon length in bp	Intron length in bp
1	216	1017
2	406	75
3	195	91
4	229	70
5	127	86
6	76	114
7	92	104
8	61	91
9	188	99
10	128	-
Σ	1718	1747
Σ total	3465	

B.6 Continued list of table 12 (3.1.6) of calculated haplotypes of *GWD*

No	Haplotype Sequence	True haplotype frequency % (no)	Apparent haplotype frequency % (no)	Allelic states n/s/d/t/q
19	TAGGATAAAGTC	12,8	3,7	191/26/1/0/1
6	CAGTATAAAGAC	11,9	3,4	192/22/4/0/0
15	CATTATAAATTG	9,6	2,5	197/20/1/0/0
14	CATTATAAAGTC	8,3	2,5	200/14/4/0/0
11	CAGTATGAAGAG	6,4	1,7	204/13/1/0/0
24	TAGTATAAATTG	6,0	1,5	205/13/0/0/0
25	TATGATATAGTC	6,0	1,5	205/13/0/0/0
2	CAGGATAAGGTG	5,5	1,4	206/12/0/0/0
17	CGGTATAAAGTC	4,1	1,0	209/9/0/0/0
18	CGTGATAAAGTC	3,7	0,9	210/8/0/0/0
23	TAGGATGTAGTC	3,7	0,9	210/8/0/0/0
4	CAGTACAAAGAC	3,2	0,8	211/7/0/0/0
1	CAGGATAAAGAG	2,8	0,7	212/6/0/0/0
10	CAGTATAAGGTC	1,8	0,5	212/6/0/0/0
3	CAGGGTAAAGTC	0,9	0,2	216/2/0/0/0

Blue letter: associated SNP with positive effect. True allele frequency: Number of haplotypes among possible 872 haplotypes (4 x 218 genotypes) in %; Apparent allele frequency: Number of genotypes containing the haplotype in %. Simplex (s), duplex (d), triplex (t) and quadruplex (q): Allelic states of haplotypes per genotype

B.7 Continued list of table 15 (3.2.6) of calculated haplotypes of *PWD*

No	Haplotype Sequence	True haplotype frequency % (no)	Apparent haplotype frequency % (no)	Allelic states n/s/d/t/q
7	CTCGAGTGTGT	1,7	6,7	195/14/0/0/0
9	CTCGGGCGTAT	1,2	4,8	199/10/0/0/0
1	CCCAGACGTGT	1,0	3,8	201/8/0/0/0
6	CTCGAGCGTAT	1,0	3,3	202/6/1/0/0
11	CTCGGGTGCGA	1,0	3,8	201/8/0/0/0
4	CCGGGATGTGT	0,7	2,9	203/6/0/0/0
13	CTGGGGCGCGT	0,4	1,4	206/3/0/0/0

Blue letter: associated SNP with positive effect. True allele frequency: Number of haplotypes among possible 836 haplotypes (4 x 209 genotypes) in %; Apparent allele frequency: Number of genotypes containing the haplotype in %. Simplex (s), duplex (d), triplex (t) and quadruplex (q): Allelic states of haplotypes per genotype

B.8 Continued list of table 20 (3.3.5) of calculated haplotypes of *Inv6*

No	Haplotype Sequence	True haplotype frequency % (no)	Apparent haplotype frequency % (no)	Allelic states n/s/d/t/q
19	GTGGCGTAT	2,9	10,0	191/18/3/0/0
16	GTGGCGAAT	2,6	9,1	192/16/3/0/0
15	GTGGCAAGT	1,9	7,2	195/14/1/0/0
3	AGGGCGTAT	1,8	6,7	196/13/1/0/0
4	AGGGCGTGA	1,4	5,7	198/12/0/0/0
17	GTGGCGAGA	0,8	3,3	203/7/0/0/0
8	GGAGTGTGT	0,5	1,9	206/4/0/0/0
10	GTAACATGT	0,5	1,4	207/2/1/0/0
14	GTGACGAGT	0,4	1,4	207/3/0/0/0
1	AGGACGTAT	0,1	0,5	209/1/0/0/0

Blue letter: associated SNP with positive effect. True allele frequency: Number of haplotypes among possible 840 haplotypes (4 x 210 genotypes) in %; Apparent allele frequency: Number of genotypes containing the haplotype in %. Simplex (s), duplex (d), triplex (t) and quadruplex (q): Allelic states of haplotypes per genotype

ACKNOWLEDGEMENTS

Whilst writing the last words of this thesis and thinking about the time passed, it becomes clear that this thesis is a result of three and half years full of support and help of many people.

First of all, I want to thank you, Christiane, for giving me the chance to do my PhD. I am very grateful that you offered me this position, probably already knowing that it will be a hard time, but definitely the most important and interesting one for me. Thank you for all your time, effort in discussions, critical reading and support in developing new ideas and of course for giving me plenty of rope to test many different approaches and hypotheses. Due to the manifold and very dynamic topic, it was very helpful that you always saw the value of every small result and that you helped me at the right moment to adjust my course.

Sincere thanks to Prof. Dr. Martin Hülkamp for kindly agreeing to be the second reviewer.

Also many thanks to Prof. Dr. Maarten Koornneef for the opportunity to be a part of his department.

My dear colleagues, I want to thank all of you for your support and guidance on all aspects of work and life. Thank you for sharing experiences, knowledge and chemicals. Special thanks to Astrid for all your effort and time and for the critical reading and detailed discussion of this work. Many, many thanks to Camila, Elske and Gabor for becoming very good friends and supporting me in all ups and downs. I think we can be very proud of having successfully answered many of the questions starting with: "As PhD, you must know...". Dear Birgit, thank you for being the good fairy. I am really impressed by your patience and honest care. Many thanks to Matthias and Markus for the tremendous support in all aspects of the protein work. Thanks to Claude, Sandra, Meki, Jude and Li Li for your support.

Finally, I want to thank my family for believing in me and supporting me in everything I do. Thanks for giving me security and a nice home, where I could always escape from all sorrows. Everything can be solved. My dear Burkhard, I do not know how to express how thankful I am for your enormous patience and taking me as I am. Without you, I would have never got so far.

ERKLÄRUNG

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

Ich versichere,

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

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

Köln, den 17.10.2011



LEBENS LAUF

Persönliche Daten

Name	Lena Katharina Schreiber
geboren am	20. Mai 1982
Geburtsort	Gütersloh
Familienstand	ledig, keine Kinder

Schulischer und akademischer Werdegang

Juni 2008 - heute	<p>Doktorandin am Max-Planck-Institut für Pflanzenzüchtungsforschung im Bereich Pflanzenzüchtung und Genetik (AG Gebhardt)</p> <ul style="list-style-type: none"> Thema der Dissertation: “Structural, functional and genetic analyses of novel candidate genes for tuber quality traits in potato (<i>Solanum tuberosum</i> L.)”
September 2007 - März 2008	Studium der Physik an der Universität Bielefeld
Oktober 2006 - August 2007	<p>Anfertigung der Diplomarbeit am Lehrstuhl für Genomforschung der Universität Bielefeld</p> <ul style="list-style-type: none"> Thema der Arbeit: Regulation der nukleo-cytoplasmatischen Partitionierung der KNOX- und BELL-Proteine in <i>Arabidopsis thaliana</i> durch Kernexport und Dimerisierung <p>Abschluss als Diplom-Biologin: Juni 2007</p>
Oktober 2003 - September 2006	Hauptstudium der Biologie an der Universität Bielefeld
Oktober 2001 - September 2003	<p>Grundstudium der Biologie an der Universität Bielefeld</p> <p>Abschluss des Vor-Diploms: September 2003</p>
August 1992 - Juli 2010	<p>Besuch des Städtischen Gymnasiums Harsewinkel</p> <p>Erhalt der Allgemeinen Hochschulreife: Juli 2001</p>
August 1988 – Juli 1992	Marienschule Marienfeld

