Analysis of Two Single Trait Loci Affecting Flavonol Glycoside Accumulations in *Arabidopsis thaliana* Natural Variations

Dissertation

to obtain the Academic Degree

Doctor of Natural Sciences

(Dr. rer. nat.)

at the Faculty of Biology

at Bielefeld University

presented by

Hirofumi Ishihara

from Osaka, Japan

Bielefeld, July 2007

Table of Contents

List of Abbreviations

°C	degrees Celsius
µ	micro
aa	amino acid
APS	ammonium persulphate
bp	base pairs
BAC	Bacterial Artificial Chromosome
cDNA	complementary DNA
cM	centi Morgan
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
DTT	dithiothreitol
DW	dry weight
EDTA	ethylene diamine tetraacetic Acid
ESI	electrospray ionisation
FW	fresh weight
g	gram
GC-MS	gas chromatography-mass spectrometry
GUS	ß-glucuronidase
h	hours
HPLC	high performance liquid chromatography
HPTLC	high performance thin layer chromatography
IPTG	isopropyl-beta-D-thiogalactopyranoside
Kb	kilo base
l	liter
LD	long days
m	milli
M	molar concentration (mol/liter)
Mb	mega base
m/z	Mass to charge ratio
min	minutes
MS	mass spectrometry
MSTFA	N-methyl-N-(trimethylsilyl)trifluoroacetamide

n	nano
nm	nanometer
OD	optical density
ORF	open reading frame
PCR	polymerase chain reaction
PEG	polyethylenglycol
PDA	photodiode array
pH	negative decadic logarithm of the proton concentration in mol per liter
RIL	recombinant inbred lines
RNA	ribonucleic acid
RNase A	ribonuclease A
RT	room temperature
RT-PCR	reverse transcriptase - PCR
s	seconds
SD	short days
SDS	sodium dodecyl sulphate
TEMED	N,N,N,N-tetramethyl ethylene diamine
TFs	transcription factors
Tris	2-amino-2-hydroxymethyl-1,3-propanediol
UDP	uridine 5'-diphosphate
UGT	UDP dependent glycosyl transferase
UTR	untranslated region
UV	ultraviolet
v/v	volume per volume
w/v	Weight per volume
w/w	Weight per weight
WT	Wild type
x g	Gravity force
X-gal	5-Bromo-4-chloro-3-indolyl-β-D- galactopyranoside
X-glc	5-Bromo-4-chloro-3-indolyl-β-D- glucuronide

Amino acid:

A	Ala	Alanine
R	Arg	Arginine
N	Asn	Asparagine
D	Asp	Aspartic acid
C	Cys	Cysteine
E	Glu	Glutamic acid
Q	Gln	Glutamine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
L	Leu	Leucine
K	Lys	Lysine
M	Met	Methionine
F	Phe	Phenylalanine
P	Pro	Proline
S	Ser	Serine
T	Thr	Threonine
W	Trp	Tryptophan
Y	Tyr	Tyrosine
V	Val	Valine

List of Figures

List of Tables

List of Appendices

Abstract

Various plant secondary metabolites, including flavonoids, are involved in plant adaptation to different environments. The needs of sessile lifestyle of plants may have increased the variation of enzymes which are required in the modification and/or accumulation of different flavonol derivatives. The probable mechanism for generating variants of the enzymes is by mutating the corresponding genes. Therefore, *Arabidopsis thaliana* wildtype accessions collected from different environments and locations may have differences in the accumulation of flavonoid derivatives. Using molecular markers, these differences can be investigated and used to identify genomic variations causing observed traits. The work presented in this thesis is focused on the identification of new genes involved in accumulation of flavonoid derivatives in *A. thaliana* wildtype accessions using both metabolomics and genomic approaches.

Three different analytical methods were employed to identify flavonoid derivatives in *A. thaliana* seedlings. First, high performance liquid chromatography photo diode array coupling with electro spray -ionization mass spectrometry (HPLC-PDA-ESI-MS) was employed to identify molecular mass of flavonoid aglycones and conjugate sugars. Second, high performance thin layer chromatography (HPTLC) was used to rapidly screen large amount of plant samples. HPTLC was also employed to purify flavonoid derivatives that were then used for the identification of sugar groups by gas chromatography mass spectrometry (GC-MS). Twelve different flavonoid derivatives were identified from seedlings, all were glycosylated in mono-, di-, or tri- forms with glucose and/or rhamnose.

Comparison of flavonol derivatives among twelve different *A. thaliana* wildtype accessions revealed quantitative differences in flavonol glycosides namely and quercetin 3-O-rhamnoside (quercitrin) and flavonol 3-O-gentiobioside 7-O-rhamnoside.

Quantitative trait mapping analysis of the quercitrin trait (Qr4) located a single trait locus on chromosome 5 at 99 cM. The previously characterized *TRANSPARENT TESTA 10* was identified to encode the Qr4 locus when comparing quercitrin accumulation between mutants and wildtypes (WTs) seed coats. Allelic difference in the *TT10* promoter sequence in L*er* was identified to

cause lower promoter activity than Col. The promoter difference was proposed to involve in the quercitrin accumulation.

From the metabolomic data, the candidate gene involved in qualitative accumulation difference in the flavonol 3-O-gentiobioside 7-O-rhamnoside (Km1 and Qr1) was assumed to be of the UDP glycosyltransferase family 1 protein (UGT). Linkage analysis was performed to identify the molecular markers linked to the locus. The trait was tightly linked to a molecular marker mi304 located on the chromosome 1 at 86.52 cM. The 1.8 Mb interval between flanking markers located both sides of mi304 did not contain any genes encoding UGT. Hence, the interval was narrowed down to a 90 Kb interval and sequenced. Again no UGT was found in the L*er* genomic sequence. However, comparison of genomic sequences of the interval between Col and L*er* showed substantial modification by large insertions and deletions. This genome region may be involved to the Km1 and Qr1 trait. Finally, based on the genome annotation, four candidate genes were proposed for further investigation.

1 Introduction

1.1 Flavonoids in *A. thaliana*

To adapt to a sessile lifestyle plants have evolved to produce a variety of secondary metabolites (reviewed in Gornall and Bohm, 1978). Secondary metabolites are important compounds which have diverse functions they act as phytoalexins and allelopathic agents to defend plants against pathogens and herbivores (camalexin and glycosynolates), function as signal molecules to auxin transport (flavonols), protect plants from UV-B radiation (flavonols and sinapates), and controlling seed dormancy and longevity (Tsuji *et al.*, 1992; Li *et al.*, 1993; Sheahan, 1996; Shirley, 1996; Graham, 1998; Olsson *et al.*, 1998; Murphy *et al.*, 2000; Brown *et al.*, 2001; Ryan *et al.*, 2001; Winkel-Shirley, 2001; Kliebenstein *et al.*, 2001a; Debeaujon *et al.*, 2003). More than 170 secondary metabolites from seven major classes have been identified in *A. thaliana*. These include indole and indole-sulfur compounds, phenylpropanoids, glucosinolates, benzenoids, terpenoids, flavonoids, and fatty acid derivatives (reviewed in D'Auria and Gershenzon, 2005). Major phenylpropanoids in *A. thaliana* are flavonoids such as anthocyanins, catechins, and sinapates.

Flavonoids are one of the most abundant secondary metabolites which are mostly found in glycosylated forms (Böhm *et al.*, 1998). They are divided into three classes depending on the attachment site of the aromatic B-ring at the heterocycle C-rings: 2-phenylbenzopyran, isoflavonoids, and neoflavonoids (Figure 1A). In *A. thaliana*, only 2-phenylbenzopyran derivatives are reported because of the lack of isoflavone synthase genes in the genome (Liu *et al.*, 2002). Oxidation and saturation of the heterocycle C-rings led to the division of the 2-phenylbenzopyran into sevral subgroups (Figure 1B); flavanones, flavones, hidydroflavonole, flavonols, flavan-3,4-diols, flavan-3-ols, and anthocyanidins (Grotewold, 2006). Because of the similarity and complexity among the flavonoid derivatives, the identification of the chemical structures is difficult (reviewed in D'Auria and Gershenzon, 2005). In spite of the chemical structures, the regulation of flavonoid biosynthesis in *A. thaliana* is well studied because of availability of mutants involved in flavonoid biosynthesis (Shirley *et al.*, 1995).

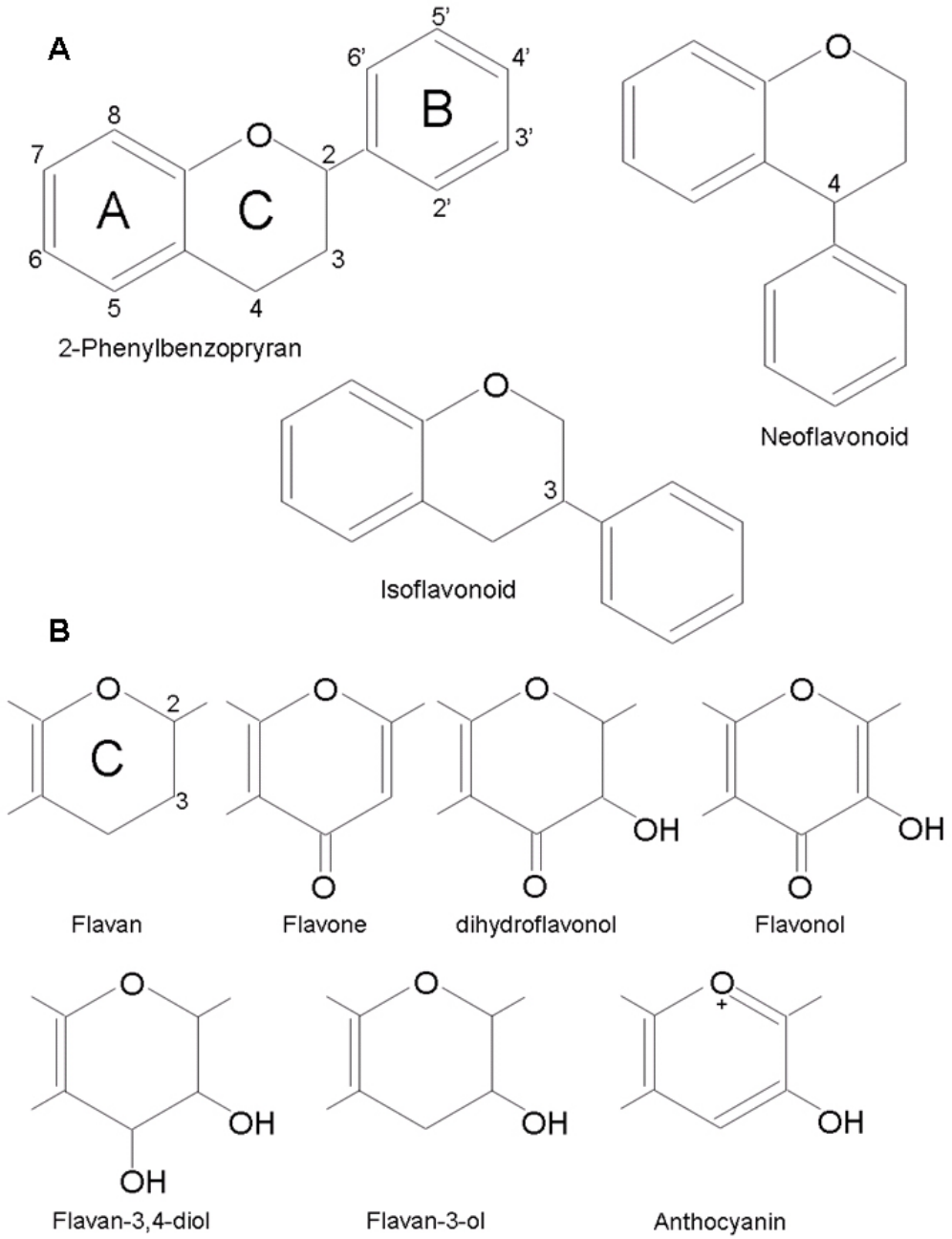


Figure 1. Flavonoid backbone structures and their modification.
(A) Flavonoid backbone structures 2-phenylbenzopyrans, isoflavonoids, and neoflavonoids. Flavonoids are comprised of two aromatic cycles A and B rings, and a heterocycle C-ring.
(B) 2-phenylbenzopyrans can be divided into several subgroups by oxidation of C-ring which are flavan, flavones, dihydroflavonols, flavonols, flavan 3,4-diols, flavan 3-ols, and anthocyanidins. Modification of the B-ring further enhances the diversity of flavonoids.

1.1.1 Flavonoid biosynthesis and its regulation in *A. thaliana*

Flavonoid biosynthesis is initiated from the phenylpropanoid pathway (Figure 2) (reviewed in Koes *et al.*, 1994). The flavonoid pathway starts with the condensation of one molecule of 4-coumaroyl-CoA and three molecules of malonyl-CoA to naringenin chalcone by the activity of the enzyme CHALCONE SYNTHASE (CHS). Isomerization of naringenin chalcone to naringenin flavonone is performed by the CHALCONE ISOMERASE (CHI). FLAVANONE 3 HYDROXYLASE (F3H) catalyzes the conversion of naringenin-flavanones to dihydrokaempferol which is either converted to kaempferol (flavonol) by the FLAVONOL SYNTHASE (FLS) activity or hydroxylated to dihydroquercetin by a FLAVONOID 3' HYDROXYLASE (F3'H). From this point, the flavonoid pathway diversifies in to small branches wherein dihydroquercetin is either hydroxylated to quercetin (flavonol) by FLS or reduced to leucocyanidin by the action of DIHYDROFLAVONOL REDUCTASE (DFR). Leucocyanidin, is converted to cyanidin (anthocyanidin) by LEUCOANTHOCYANIDIN DIOXYGENASE (LDOX). Cyanidin is either reduced to epicatechin by ANTHOCYANIDIN REDUCTASE (ANR) or glycosylated by GLYCOSYLTRANSFERASE (GT) to form anthocyanins.

The pathway is regulated at defined time points in development and organs by transcriptional control mediated by specific transcription factors (reviewed in Weisshaar, 1998; Marles *et al.*, 2003). Transcription factors of six different families have been described so far to be involved in the regulation of flavonoid biosynthesis: bHLH, MADS-box, R2R3-MYB, WD40, WIP and WRKY factors (reviewed in Weisshaar, 1998; Marles *et al.*, 2003). The transcription factors involved in the regulation of the flavonoid pathway are listed in Figure 2. The R2R3-MYB transcription factors are known to be the main regulator of flavonoid biosynthesis genes. TRANSPARENT TESTA 2 (TT2; MYB123) requires the co-factors TT8 (bHLH42) and TTG1 (WD40-repeat protein) to form a ternary complex to regulate *DFR*, *ANR* and *TT12*/*MATE*: thus influencing accumulation of proanthocyanidins (PAs) also known as condensed tannin, in seed coat (Nesi *et al.*, 2000; Nesi *et al.*, 2001; Baudry *et al.*, 2004). PRODUCTION OF ANTHOCYANIN PIGMENT1 (PAP1)/MYB75 and PAP2/MYB90 are responsible for bHLH co-factor dependent regulation of anthocyanin biosynthesis genes (Borevitz

et al., 2000). PAP1 and PAP2 activate the *PHENYLALANINE AMMONIA LYASE* (*PAL*), *CHS*, *DFR* and *GLUTATHIONE-S-TRANSFERASE* (*GST*). In contrast, R2R3-MYB4 is known as a negative regulator of *CINNAMATE 4-HYDROXYLASE* (*C4H*) which converts cinnamic acid to 4-coumaric acid (Jin *et al.*, 2000). Recently co-factor independent MYBs, namely, MYB11, MYB12, and MYB111 have been reported to be involved in the accumulation of flavonol glycosides and are therefore named PRODUCTION OF FLAVONOL GLYCOSIDE (PFG), PFG1/MYB12, PFG2/MYB11, and PFG3/MYB111 respectively (Mehrtens *et al.*, 2005; Stracke *et al.*, 2007). They individually regulate flavonol accumulation in different organs of developing seedlings (Stracke *et al.*, 2007). For example, PFG1 mainly controls flavonol biothynthesis in root, while PFG3 predominantly regulates the biosynthesis in cotyledon. Flavonol accumulation requires these three factors because the triple mutant *pfg1 pfg2 pfg3* seedlings can not produce flavonol glycosides, but the production of anthocyanins is not affected. Targets of PFGs are suggested to be *CHS*, *CHI, F3H, FLS* and UDP glycosyltransferases family 1 proteins (UGT) (Mehrtens *et al.*, 2005; Stracke *et al.*, 2007). PA biosynthesis is regulated in the seed coat after fertilization (Debeaujon *et al.*, 2003). Six transcription factors were shown to is involved in PA biosynthesis namely, TT1, TT2, TT8, TT16, TTG1, and TTG2 (Walker *et al.*, 1999; Nesi *et al.*, 2000; Nesi *et al.*, 2001; Johnson *et al.*, 2002; Nesi *et al.*, 2002; Sagasser *et al.*, 2002). *TT1* encodes a WIP type zinc finger protein which is expressed in the endothelial cell layer of the developing seeds (Sagasser *et al.*, 2002). *TT16*/*ARABIDOPSIS BSISTER* (*ABS*) encodes a MADS domain protein expressed in the endothelium cell during the early flower and seed developmental stages (Nesi *et al.*, 2002). Both transcription factors are necessary for PA accumulation in a core region of the endothelium but not in the micropyle and chalazal areas where a small amount of PA accumulation is still detected (Nesi *et al.*, 2002; Sagasser *et al.*, 2002). Other transcription factors mutants like the *tt2*, *tt8*, and *ttg1* lack PAs in the seeds. *TT2* is expressed specifically in the endothelium (Nesi *et al.*, 2001; Debeaujon *et al.*, 2003), while *TT8* and *TTG1* are found to not only be expressed in seeds but also in vegetative cells to regulate production of anthocyanins in seedlings (Baudry *et al.*, 2006).

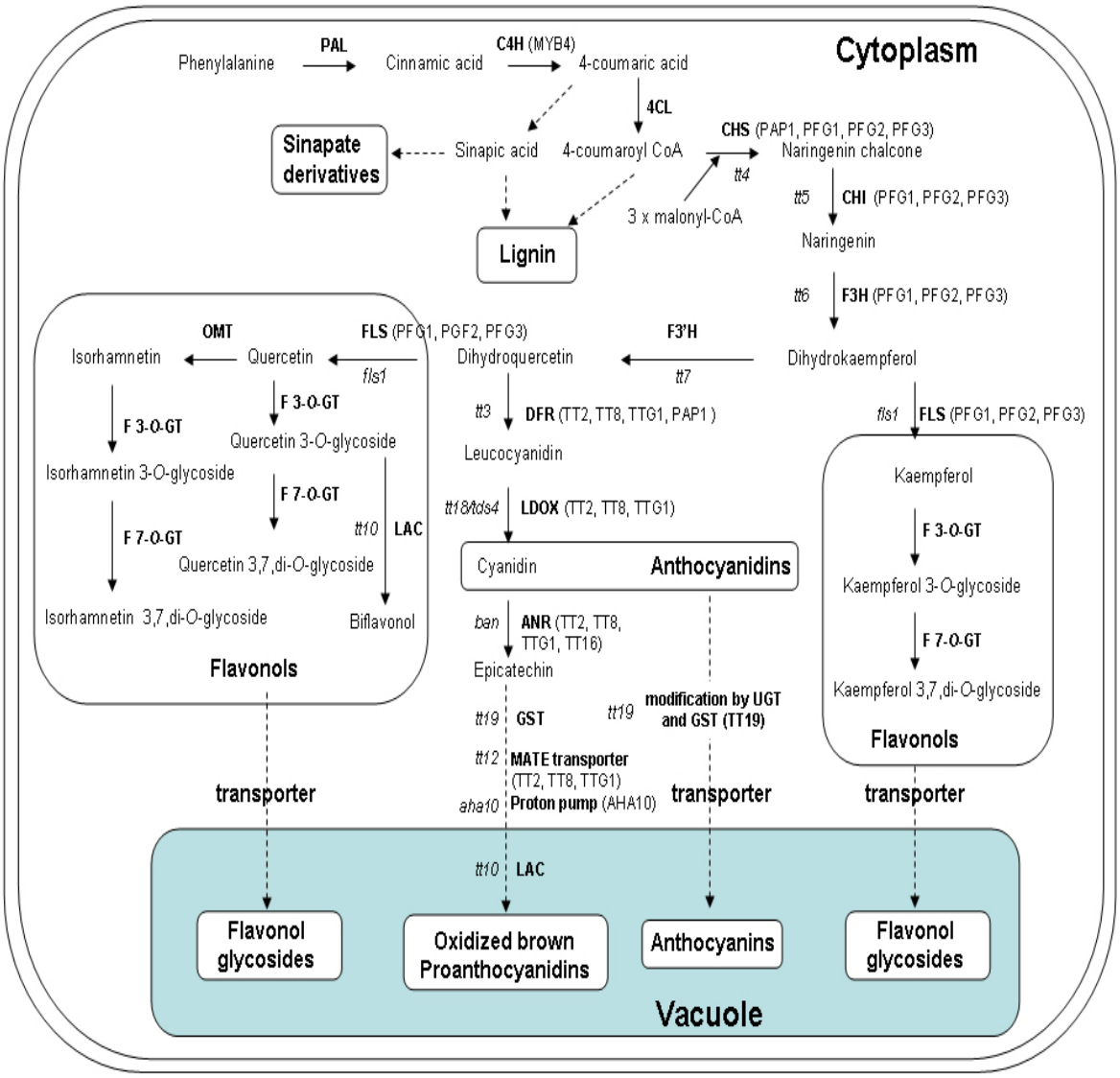


Figure 2. Scheme of the phenylpropanoid pathway in *A. thaliana.*
Solid lines indicate that the pathway supported by biochemical and genetic evidence, whereas black dotted arrows indicate multiple reactions. The major enzymes are listed with boldface capital letters. Transcription factors regulating the expression of the enzymes are given with capital letters between brackets. Mutants of the genes encoding the enzymes are listed with small italic letters. 4-coumaroyl-CoA ligase (4CL), P-type H^+-ATPase (AHA10), anthocyanidin reductase (ANR), cinnamate 4-hydroxylase (C4H), chalcone isomerase (CHI), chalcone synthase (CHS), dihydroflavonol reductase (DFR), flavanone 3 hydroxylase (F3H), flavonoid 3' hydroxylase (F3' H), flavonol 3-*O*-glycosyltransferase (F 3-*O*-GT), flavonol 7-*O*-glycosyltransferase (F 7-*O*-GT), flavonol synthase (FLS), glutathione S-transferase (GST), laccase (LAC), leucocyanidin dioxygenase (LDOX), multidrug and toxic compound extrusion-type transporter (MATE), *O*-methyltransferase (OMT), transparent testa (TT), transparent testa glabra (TTG). Modified from Chapple *et al.* (1992), Kitamura *et al.* (2004), Routaboul *et al* (2006), Liang *et al.* (2006), Sharma *et al.* (2005) and Stracke *et al.* (2007).

1.1.1.1 Accumulation mechanism of flavonoids in *A. thaliana*

Flavonoid derivatives are assumed to be stored in the vacuole as glycosylated forms because of their autotoxicity and reactivity (reviewed in Vogt and Jones, 2000). Glycosylation of flavonoid aglycones are suggested to decrease these effects due to reduced electron transfer to other compounds and result in blocking the reactive site (Osbourn, 1996). The enzymes involved in glycosylation of the flavonoids belongs to the UDP glycosyltransferase 1 family (UGT) (reviewed in Bowles *et al.*, 2006). In *A. thaliana*, the intercellular localization of UGT is not yet investigated however some study shows that all glycosylation takes place in the cytoplasm, whilst mono- and di- glycosylations occur in the cytoplasm but tri-glycosylation activity of luteolin diglycoside however is found in the vacuole from *Secale cereale* L (Anhalt and Weissenböck, 1992). In the *A. thaliana* genome, 112 open reading frames containing UGTs and eight pseudogenes with frame-shift mutations were identified (Paquette *et al.*, 2003). Recombinant enzyme studies indicate that many UGTs are region-specific rather than substrate specific and UGTs in *A. thaliana* have been demonstrated to preferentially transport sugars to the 3-OH and 7-OH position of flavonol backbones (Lim, 2004). Several UGTs involved in glycosylation of flavonol and anthocyanins have been identified and characterized in *A. thaliana* (Jones *et al.*, 2003; Tohge *et al.*, 2005b; Yonekura-Sakakibara *et al.*, 2007). UGT78D1 (At1g30530) has been functionally characterized to be involved in catalyzing the transfer of rhamnose from UDP-rhamnose to the 3-OH position of quercetin and kaempferol (Jones *et al.*, 2003). In contrast, UGT73C6 (At2g36760) was shown to catalyze the transfer of glucose from UDP-glucose to the 7-OH position of only flavonol glycosides namely, kaempferol-3-O-rhamnoside and quercetin-3-O-rhamnoside (Jones *et al.*, 2003). Most recently UGT89C1 (At1g06000) has been reported to convert kaempferol 3-*O*-glucoside to kaempferol 3-*O*-glucoside-7-*O*-rhamnoside (Figure 1) (Yonekura-Sakakibara *et al.*, 2007). Moreover UGT78D2 (At5g17050) was described as a flavonoid 3-*O*-glucosyltransferase which transfers glucose to the 3-OH sites of anthocyanins and flavonols, but UGT75D1 (At4g15550) transfer glucose to the 5-OH position of anthocyanins and thus is named anthocyanin 5-*O*-glucosyltransferase (Tohge *et al.*, 2005b). Regulators of the *UGTs* have also been identified: PAP1 was shown to be involved in activation of almost all genes

encoding anthocyanin biosynthetic enzymes including *UGT75D1* and *UGT78D2* (Tohge *et al.*, 2005b). There are some UGTs that remain to be discovered, like the UGTs involved in the production of flavonol 3-O-gentiobioside 7-O-rhamnoside (Veit and Pauli, 1999), kaempferol 3-*O*-rhamnoside 7-*O*-rhamnoside (Veit and Pauli, 1999), and kaempferol -3-O-rhamnosyl-glucosyl glucoside 7-O-rhamnoside (Stobiecki *et al.*, 2006). Recently, the R2R3-MYB transcription factors PFG1, PFG2 and PFG3 have been implicated in the accumulation of flavonol glycosides and the regulation of the UDP-glycosyltransferase genes *At5g17040*, *UGT84A1* (*At4g15480*), and *UGT91A1* (*At2g22590*) in seedlings (Stracke *et al.*, 2007). Glycosylated flavonoids such as anthocyanidins are reported to be transported to the vacuole by a multidrug and toxic compound extrusion-type transporter protein (MATE; TT12) which acts as a vacuolar flavonoid/H^+-antiporter at the seed coat endothelium (Debeaujon *et al.*, 2001). *TT19* which encodes a GST protein is reported to be required for not only the uptake of anthocyanidins in the vacuole but also the sequestration of anthocyanidins into the vacuoles in immature seed coats (Kitamura *et al.*, 2004). However, the exact biochemical functions of TT12 and TT19 are not very clear (Sharma and Dixon, 2005). A recently characterized proton pumping ATPase, AHA10, is also required for accumulation of proanthocyanidins (PA). It is predicted to be localized to the plasma membrane and could function in vesicular trafficking (Baxter *et al.*, 2005). Once anthocyanidins accumulate in the vacuole, they are possibly polymerized to PAs by a laccase (TT10) and stored in such a form (Pourcel *et al.*, 2005; Liang *et al.*, 2006).

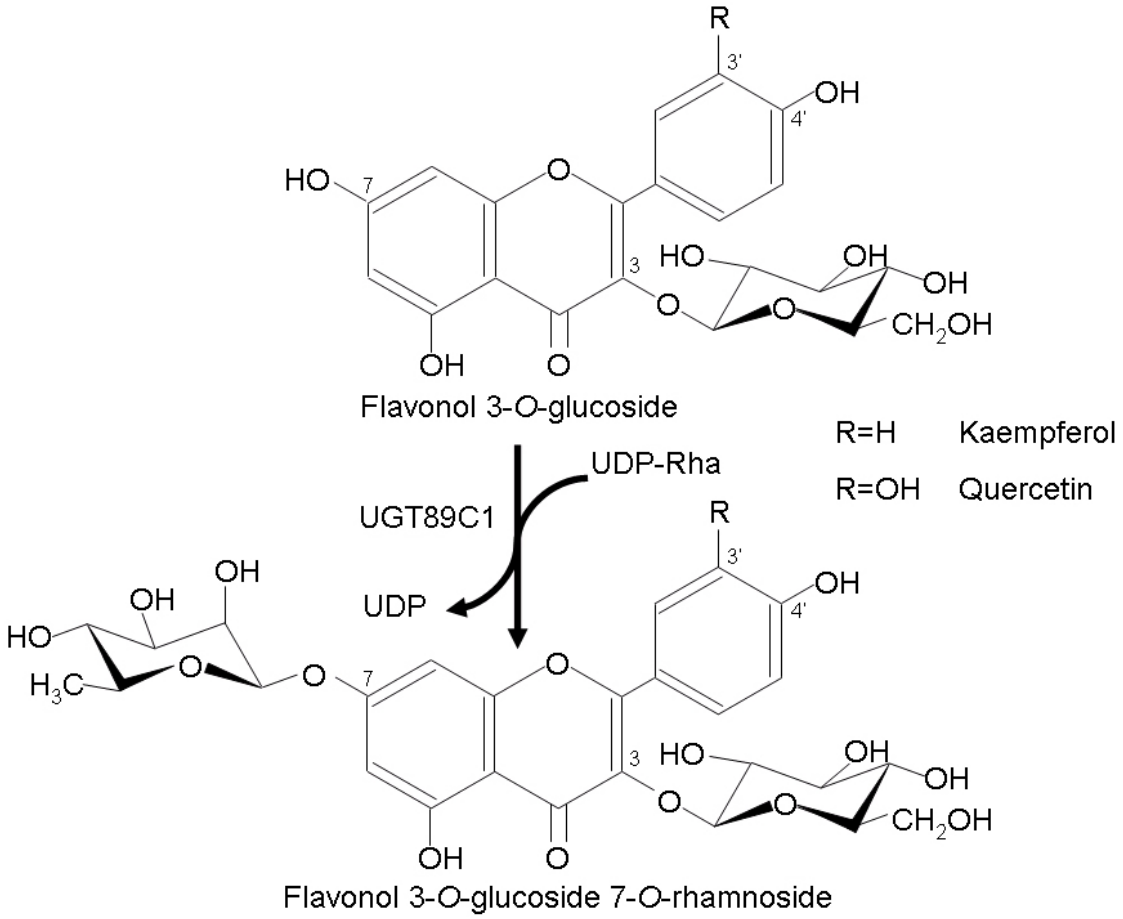


Figure 3. Glycosylation of flavonol 3-O-glucoside by flavonol 7-O-rhamnosyltransferase (UGT89C1).
The reaction catalyzed by UDP89C1 is addition of rhamnose from an UDP-rhamnose donor to the 7-OH position of flavonol 3-*O*-glucoside. Rhamnose (Rha). (Yonekura-Sakakibara *et al.*, 2007)

1.1.1.2 Flavonoids accumulation patterns in *A. thaliana* flavonoid-deficient mutants

Various *A. thaliana* mutants defective in the phenylpropanoid pathway have been identified (Table 1). In *A. thaliana*, the genes encoding the enzymes involved in flavonoid biosynthesis are known to be single copy genes (Shirley *et al.*, 1992; Pelletier and Shirley, 1996; Pelletier *et al.*, 1997). The exception is FLS which is known to be coded by a small family of genes (Pelletier *et al.*, 1997). Thus, the single mutants show phenotypic differences in terms of appearance and flavonoid accumulation patterns (see Figure 2). Wild-type *A. thaliana* produces seeds of brown color due to a flavonoid derived brown pigment. Mutants lacking this pigment have been named *transparent testa* (*tt*) (Koornneef, 1981). Due to the

transparency of the testa, the underlying yellow color of the cotyledon in the mature embryo becomes visible (Figure 4).

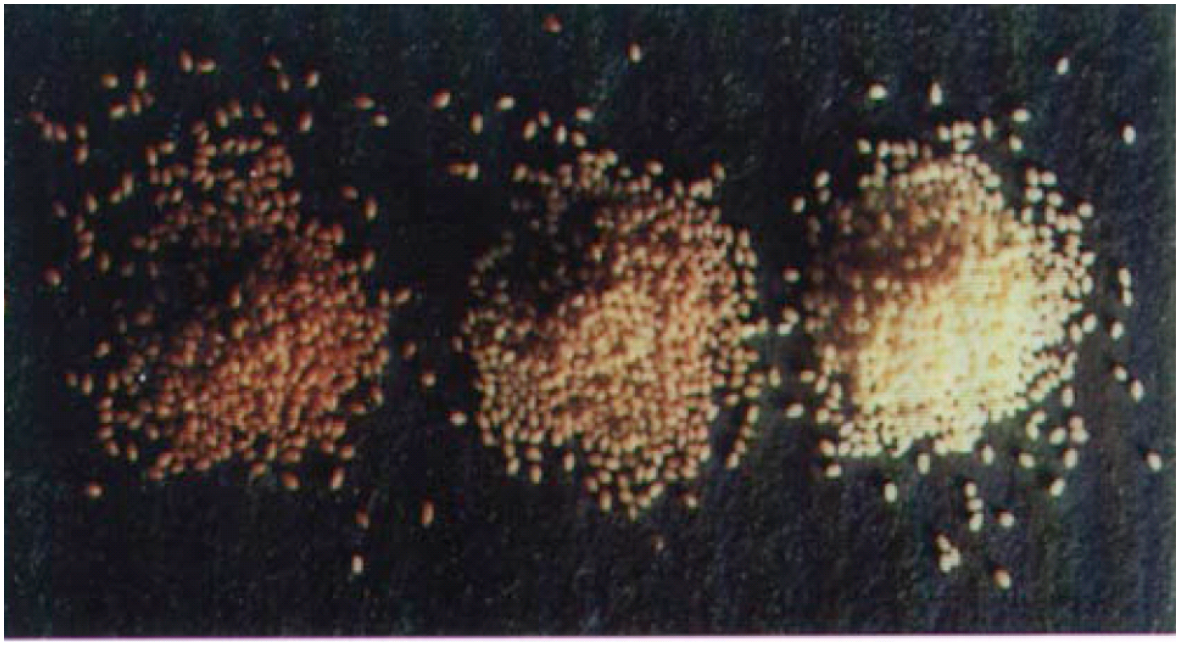


Figure 4. Comparison of seed color phenotype between *A. thaliana transparent testa* mutant and wildtype.
The yellow seed phenotype of *transparent testa* mutants are compared to the wildtype seeds. Col (left) displayed brown seed color. In contrast, *tt15* (center) and *ttg1* (right) mutants have pale blown and yellow *ttg1* (right) seeds, respectively. The picture was taken from Focks *et al.* (1999).

The production of brown pigments in the seed coats is due to the oxidative polymerization of flavonoids, mainly PAs. Chemically, the latter are polymers of flavan-3-ol subunits such as epicatechin and catechin (Pourcel *et al.*, 2005). The mutants *tt1*, *tt2*, *tt8*, *tt9*, *tt10*, *tt12*, *tt15*, *tt19* and *aha10* appear to be only altered in seed color (Debeaujon *et al.*, 2001), while others are also effected the accumulation of other flavonoids (Shirley *et al.*, 1995). CHS and CHI are key enzymes to initiate the flavonoid pathway. The corresponding mutants *chs* (*tt4*) and *chi* (*tt5*) are deficient in the production of the flavonoid precursors naringenin chalcone or naringenin flavanone, respectively (Shirley *et al.*, 1995). These mutants accumulate higher amounts of sinapate derivatives and do not contain any flavonoids (Shirley *et al.*, 1995). *F3'H*, which is also known as *TT7*, encodes a P450 monooxygenase of flavonoid biosynthesis catalyzing the 3'-hydroxylation of the flavonoid B-ring (dihydrokaempferol) to the 3',4'-hydroxylated state

(dihydroquercetin) (Schoenbohm *et al.*, 2000). The corresponding mutant *f3'h* (*tt7*) is deficient in the production of quercetin derivatives and accumulates a higher amount of kaempferol and pelargonidin (Abrahams *et al.*, 2002). Further analyses of these mutants demonstrated that flavonols, especially quercetin, are important for protection against UV light (Ryan *et al.*, 2001). The *FLAVONOL SYNTHASE* (*FLS)* gene has an important role in the production of flavonols, since it is required for the conversion of dihydrokaempferol and dihydroquercetin to kaempferol and quercetin, respectively (Wisman *et al.*, 1998). The *fls1* mutant accumulates dihydroflavonols (Stracke *et al.*, unpublished data). The mutants were particularly useful when analyzing the physiological role of flavonoid derivatives. Flavonols have been reported to be negative regulators of auxin transport (Murphy *et al.*, 2000; Brown *et al.*, 2001) and protect plants from UV-B radiation (Ryan *et al.*, 2001). The study of the *tt4* mutant revealed that flavonols are required for the fertilization process of petunia and maize but they are not necessary for the fertilization of *A. thaliana* (Burbulis *et al.*, 1996; Ylstra *et al.*, 1996). Because of differences in flavonoid accumulation, the mutants can be use to identify flavonoid derivatives.

Table 1. List of known mutants involved in flavonoid biosynthesis.
The table is adapted from Koornneef and Lepiniec *et al.* (Koornneef *et al.*, 1980; Jones *et al.*, 2003; Tohge *et al.*, 2005b; Lepiniec *et al.*, 2006; Stracke *et al.*, 2007; Yonekura-Sakakibara *et al.*, 2007). Abbreviations are listed in the legend of Figure 2. *Yellow seed color turn to brown after 6 to 12 months of storage (Pourcel *et al.*, 2005). Chromosome1 (chr 1), n/a- data not available

Locus code	Seed color	Gene product	ATG or Locus
Structural genes			
tt3	Yellow	DFR	*At5g42800*
tt4	Yellow	CHS	*At5g13930*
tt5	Yellow	CHI	*At3g55120*
tt6	Pale brown spotted	F3H	*At3g51240*
tt7	Pale brown spotted	F3'H	*At5g07990*
tt10	Yellow / Brown*	Laccase	*At5g48100*
tt12	Pale brown	MATE secondary transporter	*At3g59030*
tt18/ tds4/ tt11	Yellow	LDOX	*At4g22880*

tt19/tt14	Dark yellow	GST	*At5g17220*
aha10	Pale brown	AHA10	*At1g17260*
ban	Pale gray/gray	ANR	*At1g61720*
fls1	Brown	FLS	*At5g08640*
Regulatory genes			
egh3	Brown	bHLH002	*At1g63652*
gl3	Brown	bHLH001	*At5g41315*
pap1	Brown	MYB75	*At1g56650*
pap2	Brown	MYB90	*At1g66390*
pfg1	Brown	MYB12	*At2g47460*
pfg2	Brown	MYB11	*At3g62610*
pfg3	Brown	MYB111	*At5g49330*
tt1	Brown	WIP type Zinc Finger protein	*At1g34790*
tt2	Yellow	MYB123	*At5g35550*
tt8	Yellow	bHLH42	*At4g09820*
tt16 /abs	Yellow	MADS	*At5g23260*
ttg1	Yellow	WD-repeat protein	*At5g24520*
ttg2	Yellow	WRKY44	*At2g37260*
Glycosyltransferase			
ugt73c6	n/a	UGT73C6	*At2g36790*
ugt75d1	n/a	UGT75D1	*At1g15550*
ugt78d1	n/a	UGT78D1	*At1g30530*
ugt78d2	n/a	UGT78D2	*At5g17050*
ugt89c1	n/a	UGT89C1	*At1g06000*
Unknown loci			
tt9	Pale gray/dark	Unknown	unknown
tt13	Pale brown	Unknown	unknown
tt15	Pale brown	Unknown	Chr 1
tt17	Pale brown	Unknown	unknown
tds1,2, 3, 5, 6	Pale brown	Unknown	unknown

1.2 *Arabidopsis thaliana* natural variations as a tool for genome research

The genome of *A. thaliana*, a member of the mustard (*Brassicaceae)* family, is consists of five chromosomes mount up to a relatively small genome size estimated at 125 Mb, of which 114.5 Mb were sequenced in the year 2000 (The Arabidopsis Genome Initiative, 2000). Now, in 2007, the genome size has been

re-estimated to 135 Mb, of which 119 Mb have been sequenced (http://www.arabidopsis.org). *A. thaliana* is native to Europe and Central Asia and is naturalized to a broad range of environments in the Northern hemisphere between the latitude of 68°N to 0° and some in the Southern hemisphere such as South Africa and Australia (Figure 5) (reviewed in Koornneef *et al.*, 2004). *A. thaliana* wildtype accessions have been collected from around the world and at present, more than 750 different wildtype accessions are commercially available from the two major seed stock centers, the Arabidopsis Biological Resource Center (ABRC) and the Nottingham Arabidopsis Stock Center (NASC) (http://www.arabidopsis.org/). When the accessions are compared under same environments, the distinct differences in phenotypes among the accessions are observable. For instance, the quantitative differences in root system size (Fitz Gerald *et al.*, 2006), flowering time (Koornneef *et al.*, 1998), disease resistance (Deslandes *et al.*, 1998; Wilson *et al.*, 2001), secondary metabolite accumulation (Kliebenstein *et al.*, 2001c; Keurentjes *et al.*, 2006) and many other have been reported among the accessions to date. Because of the selfing nature of *A. thaliana*, most plants collected in the nature are homozygous due to inbreeding and so have high genetic variations between accessions (reviewed in Koornneef *et al.*, 2004). DNA marker analyses of 142 *A. thaliana* wildtype accessions with 79 amplified polymorphic (AFLP) markers have shown the high genetic isolation by distance between the different wildtype accessions from Central Asia and Southern Europe (Sharbel *et al.*, 2000). Moreover, expressed sequence tag (EST) sequences generated from twelve different *A. thaliana* wildtype accessions were exclusively compared and showed high genetic diversity (Schmid *et al.*, 2003; Torjek *et al.*, 2003). Comparison of redundant genomic sequences from *A. thaliana* wildtype accessions revealed that they are highly polymorphic throughout the genomes (Nordborg *et al.*, 2005). Correspondingly, wildtype accessions from distinct locations are highly genotypically and phenotypically polymorphic (reviewed in Koornneef *et al.*, 2004).

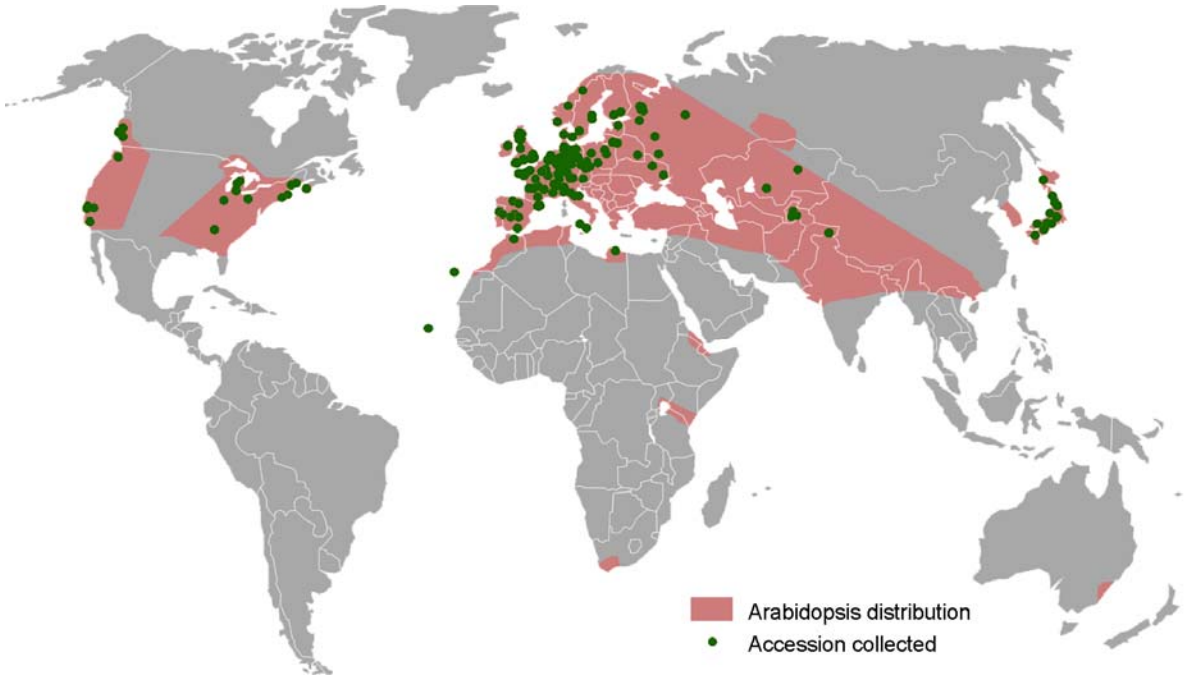


Figure 5. Geographical distribution of *A. thaliana* wildtype accessions.
The figure is adapted from the Natural EU project homepage (http://www.dpw.wau.nl/natural/).

A common approach to identify a gene function is via mutagenesis. The forward genetic approach starting from a mutant phenotype to find the corresponding genotype is usually rather slow and laborious. Most of the genes altered in *tt* mutants were identified by forward genetic approach because distinct phenotypes in seed color could be used easily as molecular markers to find the positions of the loci in the *A. thaliana* genome (Koornneef, 1990). When the locus of the trait was mapped on the chromosome, the sequence differences of candidate genes selected based on knowledge of flavonoid pathway were compared between the mutants and WTs. Final proofs of the candidate genes were tested by complementation of the mutant allele with the candidate genes to rescue WT phenotypes. However, attributing the functions of genes seems limited because many mutations are essentially undetectable in typical phenotype screenings (reviewed in Koornneef *et al.*, 2004). In contrast reverse genetics from genotype to phenotype needs a target gene to establish mutations for investigating the gene functions. In *A. thaliana,* this approach is more feasible because of the availability of the annotated genome sequence (The Arabidopsis Genome Initiative, 2000). For instance, most recently, the transcription factors MYB11/PFG2, MYB12/PFG1, and MYB111/PFG3 were identified to be involved in the

accumulation of flavonol glycoside after comparing the mutant phenotype to corresponding WTs (Mehrtens *et al.*, 2005; Stracke *et al.*, 2007). However, the high redundancy of same gene functions will make it difficult to analyze all gene functions (reviewed in Koornneef *et al.*, 2004). Since *A. thaliana* wildtype accessions are naturally occurring mutants, any phenotypic differences found among the accessions are corresponding to the difference in their genotypes. Any difference in the phenotype between two accessions can be associated to the genotype (gene) using a large number of a segregating population (mapping population) and molecular markers generated from the two accessions (reviewed in Tanksley, 1993). The quantitative trait loci (QTL) mapping approach is one of the methods which allows the detection of genes by associating the phenotype to genotype using molecular markers and mapping populations such as Recombinant Inbred Lines (RIL) (reviewed in Tanksley, 1993).

1.2.1 Molecular markers and Recombinant Inbred Line populations

Molecular markers are used as a tool to determine the genotype of loci and recombination events among the chromosomes by analyzing the mapping populations (Tanksley, 1983). The increasing number of polymorphic markers distributed within the *A. thaliana* genome increases the probability of finding new loci influencing qualitative and quantitative traits between any two accessions (Jansen, 1996). Therefore, large sets of markers which equally cover all chromosomes are necessary (Jansen, 1996). Up to date, more than 1000 markers and genotype data of markers from 100 lines from the L*er* x Col RIL population (Lister and Dean, 1993) are available from NASC (http://arabidopsis.info/new_ri_map.html). The markers and the genotype data from L*er* x Cvi (Alonso-Blanco *et al.*, 1998) and Col-*gl1* x Kas-1 (Wilson *et al.*, 2001) RIL populations are freely accessible from the Natural project homepage (http://www.dpw.wau.nl/natural/). Most of the markers deposited within the database are DNA-based markers.

Developing DNA-based molecular markers such as Simple Sequence Repeat polymorphism (SSR), Restriction Fragment Length Polymorphism (RFLP), and especially Amplified Fragment Length Polymorphism (AFLP) markers have increased the number of polymorphic markers among the chromosomes and

improved the resolution of the Arabidopsis genetic linkage map (Alonso-Blanco *et al.*, 1998; Sharbel *et al.*, 2000; Peters *et al.*, 2001). The two most abundant polymorphic markers are Single Nucleotide Polymorphism- (SNP) based marker Cleaved Amplified Polymorphic Sequence (CAPS), and Insertion-Deletions (InDels) (Figure 6) that have recently gained more attention. Jander *et al.* have sequenced partial L*er* genomic sequences and compared this to Col genomic sequence (Jander *et al.*, 2002). They have reported an average of one SNP every 3.3 Kb and an InDel every 6.6 Kb (2002) and donated this information to TAIR for the development of SNP and InDel markers (Jander *et al.*, 2002) (TAIR: Monsanto Arabidopsis Polymorphism and L*er* Sequence Collections). Moreover, expressed sequence tag (EST) sequences among 12 different *A. thaliana* wildtype accessions were compared exclusively and this showed high genetic diversity among the accessions (Schmid *et al.*, 2003; Torjek *et al.*, 2003). The SNP data generated from twelve different accessions can be queried in the SNP database available from the Max-Planck Arabidopsis SNP Consortium (MASC: http://www2.mpiz-koeln.mpg.de/masc/index.html) (Schmid *et al.*, 2003; Torjek *et al.*, 2003).

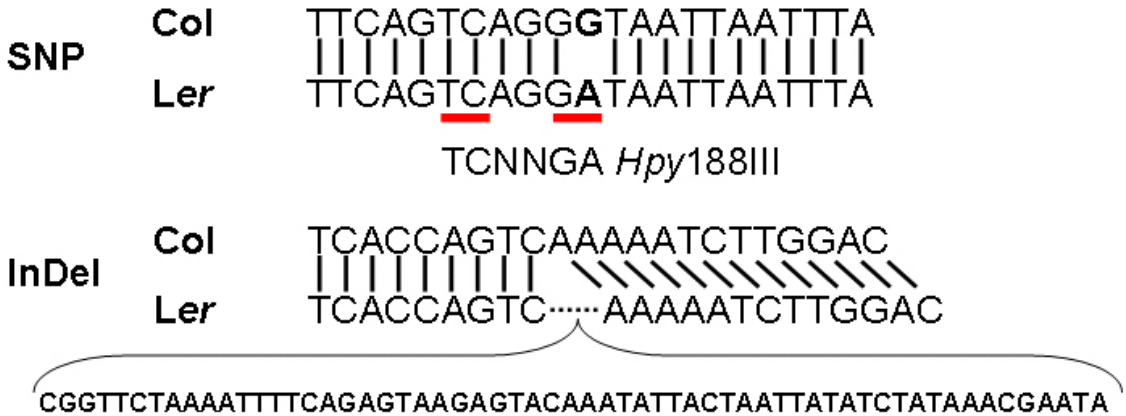


Figure 6. Example of a single nucleotide polymorphism (SNP) and insertion and deletion polymorphism (InDel) found between Col-0 and L*er*.
An example of a SNP (Col: **G** to L*er*: **A**) and an InDel (60 nucleotide insertion in L*er* versus Col) identified at the promoter region of *At5g48100*. The SNP can be used as a cleaved amplified polymorphism marker (CAPS) because the restriction site of *Hpy*188III is only present in L*er* sequence. The sequence recognition motif for *Hpy*188III is TCNNGA and listed with under lines.

In *A. thaliana*, the recombinant inbred line (RIL) populations are the most commonly used type of mapping populations because of several advantages compared to any F2 or F3 population previously reported (Lister and Dean, 1993). The main disadvantage of F2/F3 populations is the loss of genetic characteristics to the next generation. When the seed stocks are exhausted, new mapping populations with segregation of marker data has to be generated in the new population however, (Lister and Dean, 1993). In Figure 7, a RIL population is constructed by crossing two accessions with distinct phenotypic differences and then the selected F2 hybrids are self-crossed carried through eight to ten generations until the hybrid genome reaches a homozygous state (summarized in Wilson, 2000). Thus, a RIL is also a single descendent line. At the F8 generation, the majority of RI lines have homozygous alleles. So, the same genetic characters of each line can be maintained and transmitted to many generations (Lister and Dean, 1993). As a result, several RIL populations have been generated and are maintained at the stock centers (Figure 7). Koornneef Cvi x L*er* RIL (Alonso-Blanco *et al.*, 1998) and Lister and Dean L*er* x Col RIL (Lister and Dean, 1993) populations are the most widely used mapping populations. For both these RIL populations, the genetic maps have been linked by 49 common markers, mapping at similar positions (Alonso-Blanco *et al.*, 1998). The use of polymorphic markers allowed the genotyping of more than a thousand loci (summarized inWilson, 2000). The statistical methods like a Mendelian linkage analysis and a QTL mapping are applied to locate the loci by association of any observed phenotypes to regions of the genome (reviewed in Griffing and Scholl, 1991; reviewed in Alonso-Blanco and Koornneef, 2000).

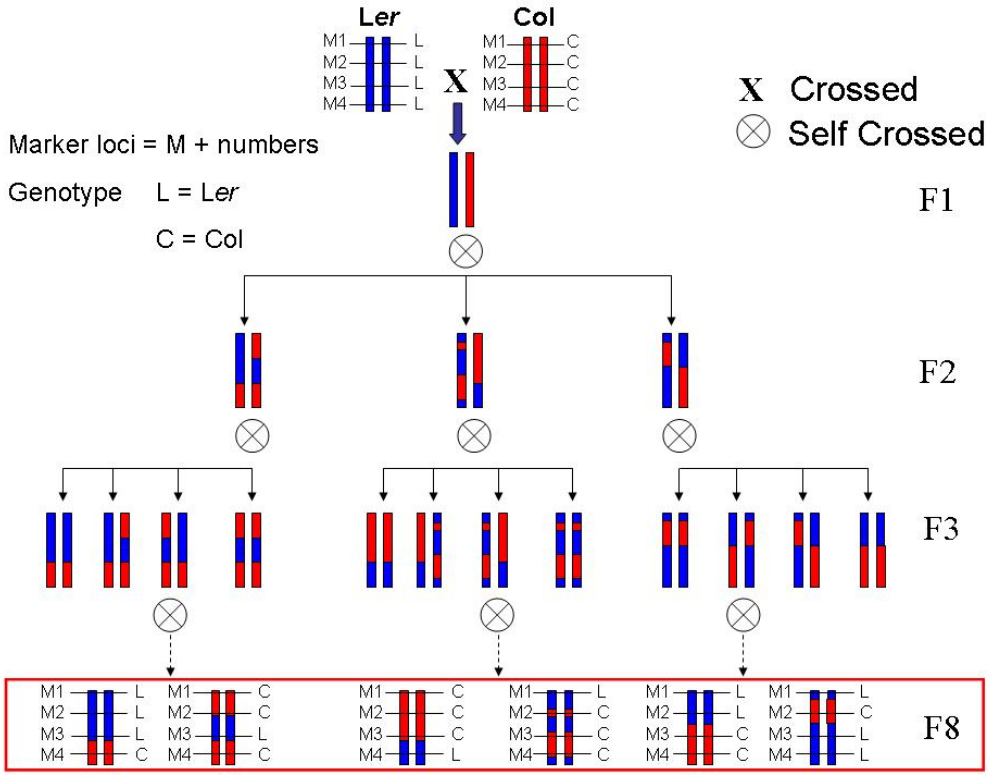


Figure 7. Recombinant Inbred line (RIL) construction strategy and schematic illustration of recombinant chromosome.
Construction of a RIL population from Landsberg *erecta* (L*er*) and Columbia (Col) accessions which have distinct genetic variations (L or C) at different loci (M1 through M4). The recombination events occurred in the F1 at meiosis I which results in recombinant chromosomes in the F2 population. Selected lines from the F2 population are further selfed to produce the F3 population. By repeating this process, the genome of the lines becomes stable at the F8 population. In genetic mapping, the distance between two loci (measured in centimorgans, cM) is measured using the recombination frequency, which is just the number of recombination events between two marker loci divided by the number of individuals of the mapping population. Modified after Wilson (2000).

1.2.2 Mapping of qualitative and quantitative traits loci

Most trait variations among wildtype accessions are quantitative traits but some traits were shown to segregate as Mendelian traits, such as pathogens resistance (Deslandes *et al.*, 1998) and glucosinolate (GS) contents and structures (Kliebenstein *et al.*, 2001b; Zhang *et al.*, 2006). The qualitative traits are mainly a result of allelic variations at one or two loci (Tanksley, 1993). A qualitative trait is defined as a phenotypic distribution that is discrete (not continuous) in a segregating population such as a RIL. A molecular marker can therefore be used

as a qualitative trait since its distribution is due to allelic variation of its locus in segregating population. Hence, the position of a qualitative trait on an *A. thaliana* chromosome can be obtained by linkage analysis. Basically, the linkage analysis is dependent on the likelihood of the recombination events occurring between the trait locus and the neighboring loci (marker loci) on a chromosome. When the distance between the loci is smaller, the likelihood of recombination frequency between them is lower and the trait locus is closely linked to the neighboring loci. Conversely, when the distance is larger, the number of recombination events will increase and the genetic linkage of the loci becomes weaker. The goal of the linkage analysis is to locate the trait locus by the identification of closely linked flanking markers (Figure 8). The recombination frequencies between the trait locus and random molecular marker loci with known locations in the genome are evaluated statistically using the maximum likelihood principle. MAPMAKER (Lander and Botstein, 1989) is the most commonly used non commercial software for calculating a large amount of segregation data for linkage analyses (Wilson, 2000).

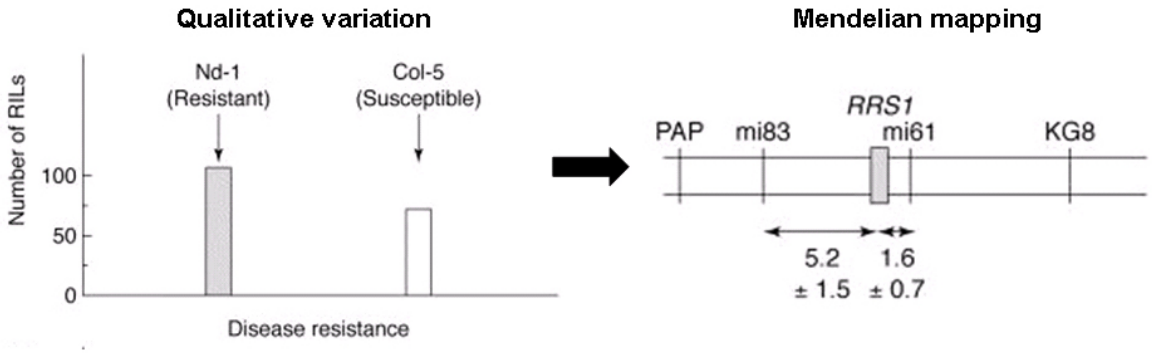


Figure 8. Qualitative trait variation and linkage map
Example of a qualitative phenotypic distribution of pathogen resistance to *Ralstonia solanacearum* in a Nd-1 x Col-5 RIL population. Map of the linkage analysis of *RRS1* (*RESISTANCE* to *RALSTONIA SOLANACEARUM* 1) locus. The recombination frequencies with standard errors are listed between the marker loci and *RRS1*. Adapted from Alonso-Blanco and Koornneef (2000)

In contrast, a quantitative trait can be attributed to several loci (and their dependence of environment) and shows continuous phenotypic distribution in a segregating population. Examples of quantitative traits are freezing tolerance, root length, seed size, leaf size and many more which are measurable by various units

(reviewed in Koornneef *et al.*, 2004) (Figure 9). The loci influencing a quantitative trait are called quantitative trait loci (QTLs). QTLs can be located by QTL analysis providing information to locate several loci involved in a quantitative phenotype in a segregating population (reviewed in Tanksley, 1993). There are different statistical methods to identify QTL by associating marker loci to a trait phenotype. These are analysis of variance (ANOVA), maximum likelihood (LOD score), and regression analysis. If the marker locus is associated to the trait, the genotypic means of the marker loci will be significantly different (reviewed in Tanksley, 1993; Jansen, 1996). One advantage of QTL mapping is that the analysis can show how many loci are affecting one phenotype that is not easily identifiable by mutagenesis approaches (summarized in Koornneef *et al.*, 2004). For example, a total of eleven QTLs including four main loci and seven minor loci affecting flowering time are found between Cvi x L*er* (Alonso-Blanco *et al.*, 1998). Among them, the Cvi alleles at *EARLY DAY-LENGTH INSENSITIVE* (*EDI*) and *FLOWERING H* (*FLH*) are found to be a main factor responsible for early flowering, while Cvi alleles at *FLOWERING F* (*FLF*) and *FLOWERING G* (*FLG*) were responsible for late flowering (Alonso-Blanco *et al.*, 1998). Interestingly the flowering time between Cvi and L*er* have little difference, although the segregating population shows distinct differences in flowering time, which indicates that phenotypes of the parental line are controlled by combination of four loci (Alonso-Blanco *et al.*, 1998). Later, *EDI* was identified to be encoded by *CRYPTOCHROME2* (*CRY2*) (El-Din El-Assal *et al.*, 2001). A single amino acid substitution of CRY2 increases protein stability of CRY2 in the Cvi allele which promotes early flowering (El-Din El-Assal *et al.*, 2001).

The QTL mapping procedure starts from finding any qualitative or quantitative differences among accessions such as metabolite accumulation differences. The identification of a quantitative locus depends on the accuracy of quantification, the size of the mapping population and the number of mapping markers. Especially, identification of a minor locus requires a large number of mapping populations to increase the effect of a QTL to reach statistical significance and marker size, which close up the distance and tighten the link between marker and QTL. Thus, it is a common approach to pursue major loci for the identification of the alleles (Jansen, 1993; Jansen, 1996). Once the QTL is

located on the chromosome by the QTL mapping software, identification of candidate genes in *A. thaliana* is more feasible because of the large number of recourses. Common non-commercial QTL softwares are MAPMAKER/QTL (Lincoln *et al.*, 1992), and QTL Cartographer (Wang, 2005).

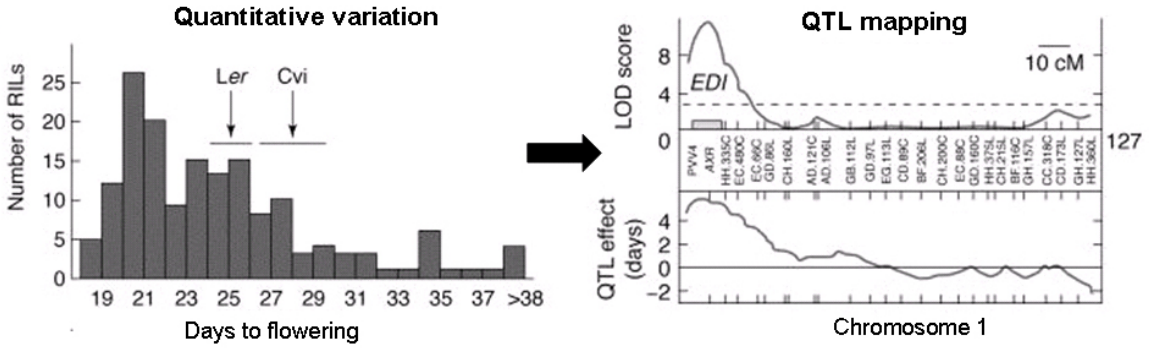


Figure 9. Quantitative trait variation and QTL map
Example of a quantitative phenotypic distribution of flowering time in Cvi x L*er* RIL population. Position of flowering time *EDI* (*EARLY DAY-LENGTH INSENSITIVE*) locus by QTL analysis. A presence of QTL is suggested when LOD score is over the threshold (dash line) on *A. thaliana* chromosome 1. QTL effect is estimated as the mean difference between the two RIL groups carrying the Cvi and L*er* alleles (Alonso-Blanco *et al.*, 1998). Adapted from Alonso-Blanco and Koornneef (2000)

The resources for the identification of gene functions in *A. thaliana* have been extensively developed. The availability of the complete *A. thaliana* genome sequence, T-DNA insertion lines, and easy gene transformation methods by *Agrobacterium tumefaciens* make positional cloning of QTLs possible. The database for the genetic variation is available through internet browsers from the Max-Planck Arabidopsis SNP Consortium (MASC: http://www2.mpiz-koeln.mpg.de/masc/index.html), The European Arabidopsis Stock Center (NASC: http://arabidopsis.info/), NATURAL project (http://www.dpw.wau.nl/natural/), Salk Institute Genomic Analysis Laboratory (SIGNAL: http://signal.salk.edu/) and the Arabidopsis Information Resource (TAIR: http://www.arabidopsis.org). A large number of mutants is available from the stock centers and can be used as final proof of the gene function. More importantly, the genome of *A. thaliana* accession Landsberg *erecta* has been partially sequenced and is available through TAIR (Jander *et al.*, 2002). Hence, this accelerates the discovery and assignment of new gene functions by QTL mapping.

1.3 Identification of novel flavonoid derivatives based QTLs

Flavonoid derivatives in *A. thaliana* accumulate in glycosylated forms (Veit and Pauli, 1999; Kerhoas *et al.*, 2006; Routaboul *et al.*, 2006; Stobiecki *et al.*, 2006). The glycosylation of the flavonoids increases polarity and thus decreases volatility of the compounds. The quantification and identification of flavonoid derivatives is key to identifying novel QTLs. Metabolite analysis based on gas chromatography-mass spectrometry (GC-MS) measurements can provide data comparing more than a thousand metabolites in one single analysis (Fiehn *et al.*, 2000). GC requires thermostable and less polar compounds. Those less volatile compounds are not easy to analyze using GC-MS since analytes need to be in the gas phase to be detected (Stobiecki *et al.*, 2006). The derivatization process converts polar compounds to nonpolar and thermostable ones. Polar metabolite functional groups such as –OH, -COOH, -NH2, or -SH groups are exchanged by single chemical reactions such as silylation, methylation and acetylation (Stobiecki and Kachlicki, 2006). However, it leads to structural rearrangements of the flavonoid –OH group and gives less structural information of flavonoid derivatives from MS (Stobiecki and Kachlicki, 2006).

Reversed phase high performance liquid chromatography (HPLC) coupled to electrospray ionization (ESI) and ion trap tandem mass spectrometry (MSn) are the most suitable analytical methods for the identification of flavonoid derivatives in *A. thaliana* (Tolstikov *et al.*, 2003; Lee *et al.*, 2005; Stobiecki *et al.*, 2006). Unlike GC-MS, HPLC analysis does not require special sample preparation; application of C_{18} monolithic capillary columns with acidified mobile phases allowed the separation of several hundreds of compound peaks derived from *A. thaliana* leaf extracts (Tolstikov *et al.*, 2003). For the detection of the flavonoids derivatives after HPLC separation, photo diode array (PDA) and mass spectrometry (MS) is commonly applied. MS is used to determine the molecular masses of flavonoid derivatives. Sequentially, the tandem MS analysis (MSn) is applied to fragment the ionized analytes at the collision cell where they interact with collision gas and are fragmented (Gaskell, 1997). MSn product ion spectra give the approximate structures of flavonoid aglycones, differentiation of C-6 and C-8 substituted glycosides and pattern of *O*-glycosylation (Stobiecki and Kachlicki, 2006).

High performance thin layer chromatography (HPTLC) is the simplest analytical method to analyze metabolites, is suitable for screening a large number of samples, and thereby reduces the analytical time. Basically, HPTLC analysis only requires a glass plate coated with silica gel or cellulose for sample separation. Separation of the sample can be improved by minimizing the size of sampling area and/or increasing the length of mobile phase. A retention factor (Rf) value is used for comparing the various metabolites. The Rf is the distance that the sample spot has traveled relative to the solvent front. Differentiating the flavonoid derivatives largely depends on the colors of the derivatives on the plates. The color of the derivatives can be enhanced by different staining methods (Wagner, 1996). Varieties of staining reagents are available for the detection of specific compounds (Wagner, 1996). A common staining reagent for flavonoid derivatives is diphenylboric acid-2-aminoethyl ester (DPBA) which visualizes kaempferol, quercetin, and sinapate derivatives to green, orange and light blue fluorescence under UV light. DPBA was used to characterize the flavonoids derivatives in *tt* mutants (Sheahan and Rechnitz, 1993). Abraham *et al.* adapted the dimethylaminocinnamaldehyde (DMACA) staining method which is used for detecting the presence of proanthocyanidin (PA) for TLC analysis (2005).

1.4 Aims and Objectives

The main focus of the thesis presented is the identification of new genes involved in *A. thaliana* secondary metabolite accumulation, especifically flavonoids derivatives, using combined metabolomic and genomic approaches. 1. First an analytical method had to be established to show qualitative but also quantitative differences in flavonoid accumulation in *A. thaliana* wildtype accessions. HPTLC, GC-MS, HPLC-ESI/MS_n techniques were used to achieve aims. The best qualitative and quantitative results could be obtained using HPLC-ESI/MS_n. 2. This method was applied to screen twelve different accessions for differences in their flavonoid composition. *A. thaliana* wildtype accession Col and L*er* showed the distinct differences in flavonoid accumulation. 3. *A. thaliana* wildtype accession Col and L*er* were chosen for a QTL mapping approach in order to identify genetic elements responsible for the differences in flavonoid composition.

2 Materials and Methods

2.1 Enzymes and chemicals

All restriction enzymes used were purchased from New England Biolabs/NEB (Frankfurt am Main). *Taq* DNA polymerase and buffers used in standard PCR analysis were obtained from Roche Applied Science (Mannheim). The PCR product purification kit, ExoSAP-IT was acquired from usb (Staufen). Superscript™II RNase H Reverse Transcriptase used for cDNA synthesis was purchased from Invitrogen (Paisley, UK). All chemicals and solutions had the highest degree of purity and were from the companies Duchefa (Haarlem, The Netherlands), Fluka (Buchs, Switzerland) Macherey-Nagel (Dueren), Merck (Darmstadt), Becton, BD (Sparks, MD/USA), Roth (Karlsruhe), Serva (Heidelberg) or Sigma-Aldrich (Steinheim).

2.2 Antibiotics

Final concentrations of antibiotics used for screening of transgenic bacteria and plants were as follows:

100 μg ml^{-1} ampicillin
50 μg ml^{-1} carbenicillin
12.5 μg ml^{-1} chloramphenicol
25 μg ml^{-1} gentamycine
50 μg ml^{-1} kanamycine
100 μg ml^{-1} rifampicillin
5 μg ml^{-1} tetracycline

2.3 Bioinformatics tools and databases

The bioinformatics tools and databases used are listed with uniform resource locator (URL) or producer as references.

A. thaliana databases

Genevestigator: https://www.genevestigator.ethz.ch/
Max-Planck Arabidopsis SNP Consortium (MASC):
http://www2.mpiz-koeln.mpg.de/masc/index.html

Munich Information center for Protein sequence (MIPS) Arabidopsis thaliana Database (MAtDB): http://mips.gsf.de/projects/plants
The European Arabidopsis Stock Centre (NASC): http://arabidopsis.info/
Salk Institute Genomic Analysis Laboratory (SIGNAL): http://signal.salk.edu/
The Arabidopsis Information Resource (TAIR): http://www.arabidopsis.org/
Monsanto Arabidopsis Polymorphism and L*er* Sequence Collections: http://www.arabidopsis.org/browse/Cereon/index.jsp
The TIGR Arabidopsis thaliana Database: http://www.tigr.org/tdb/e2k1/ath1/

Basic local alignment search tool (BLAST) servers

DNA data bank of Japan (DDBJ): http://blast.ddbj.nig.ac.jp/top-e.html
National center for biotechnology iInformation (NCBI) BLAST: http://www.ncbi.nlm.nih.gov/BLAST/
TAIR BLAST: http://www.arabidopsis.org/Blast/
TAIR Landsberg *erecta* BLAST: http://www.arabidopsis.org/Blast/cereon.jsp

***Cis*-regulatory element database**

Database of plant *cis*-acting regulatory DNA elements (PLACE): http://www.dna.affrc.go.jp/PLACE/index.html

Image analysis and processing tool

Image J: http://rsb.info.nih.gov/ij/

Mapping tools

MAPMAKER 3b: ftp://ftp-genome.wi.mit.edu/distribution/software/mapmaker3/
Windows QTL Cartographe: http://statgen.ncsu.edu/qtlcart/

Sequence alignment tools

ClustalW: http://clustalw.genome.jp/
GeneDoc: http://psc.edu/biomed/genedoc
Sequencer: Gene Codes Corporation (Ann Arbor, MI/USA)
T-Coffee: http://www.ch.embnet.org/software/TCoffee.html

Sequence annotation tool

Apollo: http://www.fruitfly.org/annot/apollo/install.html

PCR primer design tools

PerlPrimer: http://perlprimer.sourceforge.net

Proteomics Databases

InterPro: http://www.ebi.ac.uk/interpro/

SMART: http://smart.embl-heidelberg.de/smart/change_mode.pl

Sequence modification tool

Reverse and complement a sequence (RevSeq):
http://bioweb.pasteur.fr/seqanal/interfaces/revseq.html

2.4 Bacteria strains and DNA plasmid vectors

The bacteria strains and the DNA plasmid vectors used or made in this experiment are listed in the following tables (1-3)

Table 2. *Escherichia coli* (*E. coli*) strains.

E.coli strain	Genotype	Reference
GM2163	*F-arar-14, leuB6, thi-1, fhuA31, lacY1, tsx-78, galT22, supE44, galK2, hisG4, rps1134 (Strr), Xyl-5, mtl-1, dam13, Tn9 (Cam^r), dcm6 mcrB1, hsdR2 (rk- mk+), mcrA*	(Woodcock *et al.*, 1989) (NEB)
XL1-blue MRF'	Δ(*mcrA*)183, Δ(*mcrCB-hsd*SMR-*mrr*)173, *end*A1, *sup*E44, *thi*-1, *rec*A1, *gyr*A96, *rel*A1, *lac*, {F'*pro*AB, lacI^q, ZΔM15, Tn10(tet^r)]	Stratagene

Top10	F^-, *mcr*A, Δ(*mrr-hsd*RMS-*mcr*BC), ΦA*lac*ZΔM15, Δ *lac*X74, *rec*A1, *deo*R, *ara*D139, Δ(*ara-leu*)7697, *gal*U, *gal*K, *rps*L(Str^r), *end*A1, *nupI*G	Invitrogen

Table 3. *Agrobacterium tumefaciens* (*A. tumefaciens*) strains.

A. *tumefaciens* strains.	Characteristics	Reference
GV3101	pMP90RK-virulent plasmid, Gen^r, Kan^r, Rif^r	(Koncz and Schell, 1986)
COR308	pMP90-virulent plasmid, Gen^r, Rif^r, pCH20 (*vir*E, *vir*G, Tet^r),	(Hamilton, 1997)

Table 4. DNA plasmids used for cloning and transfection analysis.

Plasmid	Characteristics	Reference
pBT10-GUS	*col*E1*ori,* Amp^r , -46 to +8 *35S* promoter from CaMV, *uid*A-ORF, *nos*PA, MCS	(Sprenger-Haussels and Weisshaar, 2000)
pBT4-p35S:GUS	pBT4 with 35S promoter from CaMV	(Feldbrügge *et al.*, 1994)
pBT10-p*TT10*$_{Col}$:GUS (GH31)	-46 to +8 *35S* promoter of pBT10-GUS was replaced with 2 Kb *TT10* Col promoter	Huep unpublished
pBT10-p*TT10*$_{Ler}$:GUS (GH32)	-46 to +8 *35S* promoter of pBT10-GUS was replaced with 1.3 Kb *TT10* L*er* promoter	Huep unpublished
pBT2-ΔGUS	*col*E1*ori,* Amp^r, *uid*A-ORF, *nos*PA	(Sprenger-Haussels and Weisshaar, 2000)
pBT10-*UBI*-LUCm3	As described in pBT10-GUS, with replacement of 35S promoter and *uid*A-ORF with *PcUBI4-2* minimal promoter, and *Luc*-ORF with three mutations which eliminate *Eco*RI, *Cla*I and *Xba*I restriction sites	(Sprenger-Haussels and Weisshaar, 2000)
pBT4-ΔLUC	*col*E1*ori,* Amp^r, *Luc*-ORF, *nos*PA	(Sprenger-Haussels and Weisshaar, 2000)

pANGUS	Kan^r(*nptII*), MCS/TATA, MCS, *uid*A-ORF, *nos*PA, T-DNA LB and RB, *col*E1*ori,* Amp^r	Jakoby and Weisshaar, unpublished
pANGUS-p*TT10*$_{Ler}$-:GUS (GH33)	As described in pANGUS replacement of MCS/TATA with 2 Kb *TT10* Col promoter	Huep unpublished
pANGUS-p*TT10*$_{Col}$:GUS (GH34)	As described in pANGUS replacement of MCS/TATA with 1.3 Kb *TT10* L*er* promoter	Huep unpublished
pBiBAC2	Kan^r(*nptII*), *Bam*HI, *SacB*, HYG^r *nos*PA, T-DNA LB and RB, *col*E1*ori,* Kan^r	(Hamilton *et al.*, 1996)
pCR®4Blunt-TOPO	F1*ori, col*E1*ori, lac*Z, MCS, Amp^r , Kan^r	Invitrogen

2.5 Standard molecular biology material and methods

Standard molecular biology techniques were performed according to Sambrook and Russel (2001). The standard buffers and solutions were prepared as described (Sambrook and Russel, 2001).

2.5.1 Agarose gel electrophoresis

1 x TAE (Tris, Acetic acid, EDTA)

- 0.04 M Tris
- 0.02 M Acetic acid
- 0.001 M EDTA

10 x agarose gel loading dye

- 15 % (v/v) ficoll
- 0.25 % (w/v) xylene cyanol
- 0.25 % (w/v) bromophenol blue

The quality and size of DNA samples were analyzed by gel electrophoresis. 100 ml of 1 % agarose gel (1 % (w/v) agarose in TAE buffer) was melted using a microwave. 1 µL of ethidium bromide (10 mg ml^{-1}) was added to the agarose solution, cooled down to 60 °C and loaded into a mold with

combs. The solidified agarose gel without the combs was placed in an electrophoresis apparatus filled with TAE buffer. One-tenth volume of 10X agarose gel loading dye was added to DNA samples, mixed, and loaded into the wells. 5 µl of 1 Kb ladder (100 ng μl^{-1}) was loaded beside of sample for a size reference. The gel was electrophoresed at 150 mA until the required separation was done. The DNA fragments were visualized using 312 nm UV light and photographed using an Intas Gel Documentation System (INTAS, Goettingen).

2.5.2 Primer design

Oligonucleotide primers for Polymerase Chain Reaction (PCR) were designed using PerlPrimer. The annealing temperature and the amplicon size of the primers were set to between 60 to 64 °C and 1 to 1.2 Kb, respectively. PCR primers for genotyping were either designed by comparing the Col and L*er* sequences using either ClustalW, T-Coffee or were obtained from MASC (Schmid *et al.*, 2003; Torjek *et al.*, 2003). Alignment files (.aln) from those programs were imported to GeneDoc for fine tuning of the alignments. Finally, the primers were designed about 500 bp away from polymorphism sites. RevSeq was used to reverse and complement a DNA sequence for designing reverse primers. All designed primers were blasted against *A. thaliana* genomic sequence in NCBI or TAIR for their specificity. Oligonucleotides were purchased from either Metabion (Martinsried), or Invitrogen.

2.5.3 Polymerase Chain Reaction

Polymerase Chain Reaction (PCR) was employed to amplify DNA fragments for mapping, for genotyping of Arabidopsis plants, and for screening of transformed bacteria colonies. The general PCR reaction was performed according to the manual distributed with Roche *Taq* polymerase. The conditions for the PCR reaction are listed below:

PCR mixture per one sample

Template DNA	100 ng
PCR buffer	1 x (20 mM Tris-HCl (pH 8.4), 50 mM KCl, 1.5 mM $MgCl_2$)
Forward primer	10 pmole

Reverse primer	10 pmole
dNTP	10 mmole
Taq polymerase	2.5 units

Step	PCR cycle conditions	
1	94 °C	5 min
2	94 °C	20 sec
3	60 °C	20 sec
4	72 °C	1 min per 1 Kb amplicon
5	72 °C	5 min
6	20 °C	Hold forever

Step 2 through 4 are repeated 34 times. At cycle 34, the reaction goes to step 5 and 6.

2.5.4 Purification of PCR product

Unwanted dNTPs and primers from PCR products were removed using ExoSAP-IT. The ExoSAP-IT-treated PCR product was used for sequencing. QIAquick gel-extraction kit (Qiagen, Hilden) was used to isolate desired PCR fragments for sequencing or cloning purposes when multiple PCR products were synthesized by a PCR reaction.

2.5.5 DNA sequencing and assembling

DNA sequencing was performed at Bielefeld University Sequencing Core Facility (http://scf.cebitec.uni-bielefeld.de/SeqOrderGFDB/) using BigDye terminator version 3.1 on an ABI 3730XL sequencer (Applied Biosystems, Weiterstadt). DNA sequences were assembled using Sequencer software. Standard plasmid sequencing primers are listed in Table 5

Table 5. List of standard sequencing primers.

Plasmid	Sequencing Primer (5' – 3')	
BiBAC2	Sp6v4	GTAGGCCTAATTGGCCGCGGCCGC
	T7v3	CTTGACATTGTAGGACTATATTGC
pCR®4Blunt-TOPO	T3	AATTAACCCTCACTAAAGGG
	T7	GTAATACGACTCACTATAGGGC

2.6 Electro transformation of *E. coli* and *A. tumefaciens*

The protocol for the preparation of the electrocompetent cells and electroporation was adapted according to the Micro Pulser™ electroporation apparatus operation instructions and application guide (BioRAD, Munich).

2.6.1 Preparation of electrocompetent *E-coli* strain XL1-Blue

LB (Luria-Bertani) medium

1 % tryptone

0.05 % yeast extract

1 % NaCl

0.1 % glucose

A single *E. coli* strain XL1-Blue colony was inoculated into 5 ml of LB medium with 5 µg/ml tetracycline and incubated overnight at 37 °C at 200 rpm using a Multitron 2 incubator (INFORS HT, Bottingen). The preculture was further incubated in 500 ml LB medium with the antibiotic and grown to early saturation stage (overnight) at 37 °C with shaking at 200 rpm in the incubator. The cells were harvested by centrifugation at 4,761 x g at 4 °C. After discarding the supernatant, the pellet was washed 3 times with 500 ml ice-cold sterile Millipore water by centrifugation at 4,761 x g for 25 min at 4 °C. The cells were resuspended with 50 ml of ice-cold sterile 10 % (v/v) glycerol and transferred to a 50 ml Falcon tube. After the centrifugation of the tube at 4 °C at 4,761 x g for 15 min, the supernatant was discarded. Finally, the pellet was resuspended in 1 ml of 10 % (v/v) ice-cold glycerol. A 50 µl aliquot of cell suspension was tested for its performance. If the time constant was below 4 milliseconds, the cells were washed with 10 % (v/v) ice cold glycerol in the same procedure until the time

constant got above 4 milliseconds. Finally, the cell suspension was divided into aliquots of 50 µl and frozen using liquid nitrogen. The aliquots were stored at – 80 °C.

2.6.1.1 Transformation of *E. coli*

SOC medium

2 % (w/v) tryptone
0.5 % (w/v) yeast extract
0.05 % (w/v) NaCl
0.36 % (w/v) glucose)

LB agar plates

1 % (w/v) tryptone
0.5 % (w/v) yeast extract
1 % (w/v) NaCl
0.1 % (w/v) glucose
1.8 % (w/v) agar
0.0025 % (w/v) 5-bromo-4-chloro-3-inodlyl-β-D-galactopyranoside (X-gal)
1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG)

The transformation was performed with 500 ng plasmid DNA pipetted on top of 50 µl frozen electrocompetent *E. coli* XL-1 blue. The cell-plasmid mixture was transferred to a 2 mm gap electroporation cuvette (peqlab biotechnology, Erlangen). The transformation was made in a Micro Pulser™ electroporator (BioRAD, Munich) by using pre-programmed settings Ec2 (voltage 2.5 kV, number of pulse 1) for *E. coli*. The time constant was always close to 5 millisec. After transformation, 1 ml of SOC medium was added to the cells and the mixture was incubated for 1 h at 37 °C with shaking at 1,000 rpm using Eppendorf Thermomixer. Then the *E. coli* were spun down for 2 min at 2,500 *x g*, 900 µl of the medium was removed and the pellet was resuspended in the remaining medium. The cell suspension was plated on LB agar plates with ampicillin. For Blue-White color screening, X-gal and IPTG were added with the antibiotic. To

verify the presence of the insert, PCR was used to genotype the positive *E.coli* colonies.

2.6.2 Preparation of electrocompetent *A. tumefaciens*

YEP medium

2 % (w/v) peptone

1 % (w/v) yeast extract

0.015 % (w/v) tryprophan

A single *A. tumefaciens* colony was inoculated into 5 ml of YEP with appropriate antibiotics, and incubated overnight at 28 °C at 200 rpm in the Multitron 2 incubator. The preculture was further incubated in 500 ml YEP medium with the appropriate antibiotics and grown to early saturation stage (overnight) at 28 °C with shaking at 200 rpm in the incubator. The cells were harvested by centrifugation at 4,761 *x g* at 4 °C for 10 min. After discarding the supernatant, the pellet was washed 3 times with 500 ml ice-cold sterile Millipore water at 4,761 *x g* for 25 min at 4 °C. The cells were resuspended with 50 ml of ice-cold sterile 10 % glycerol and transferred to a 50 ml Falcon tube. After the centrifugation of the tube at 4 °C at 4,761 *x g* for 15 min, the supernatant was discarded. Finally the pellet was resuspended in 1ml of ice-cold 10 % glycerol. An aliquot of 50 µl cell suspension was tested for its performance. If the time constant was below 4 milliseconds, the cells were washed with 10 % (v/v) ice-cold glycerol in the same procedure until time constant got above 4 milliseconds. Finally, the cell suspension was divided into aliquots of 50 µl and frozen using liquid nitrogen. The aliquots were stored at - 80 °C.

2.6.2.1 Transformation of *A. tumefaciens*

YEP plate

2 % (w/v) peptone

1 % (w/v) yeast extract

0.015 % (w/v) tryprophan

1.8 % (w/v) agar

Electrotransformation was used to transfer binary plasmid DNA vectors containing a gene of interest in electrocompetent *A. tumefaciens* cells. The transformation was performed with 500 ng plasmid DNA pipetted on top of 50 µl frozen electrocompetent *A. tumefaciens*. The cell-plasmid mixture was transferred to a 1 mm gap electroporation cuvette. The transformation was made with the Micro Pulser electroporator with the following pre-programmed Agr (voltage 2.2 kV, number of pulse 1) settings for *A. tumefaciens*. The time constant was always close to 5 millisec. After transformation, 1 ml of SOC medium was added to the cells and incubated for 3 – 4 h at 4 °C with shaking at 1,000 rpm using the Thermomixer. Then the *A. tumefaciens* was spun for 2 min at 2,500 *x g*, 900 µl of the medium was removed and the pellet was resuspended in the remaining medium. The cell suspension was plated on YEP plates with appropriate antibiotics and incubated at 4 °C for 2 days. Colony PCR analysis was performed to verify presence of the plasmid in the *A. tumefaciens*.

2.6.3 Storage of bacterial cultures

Glycerol stocks of the bacterial liquid culture was prepared for longer storage purpose following the method from Sambrook and Russel (2001). A 2 ml screw caped tube containing 1.5 ml of bacterial culture was mixed with 500 µl of 60 % (v/v) sterile glycerol. The mixture was frozen by dipping into liquid nitrogen and stored at - 80 °C.

2.7 Plant materials and methods

2.7.1 Plant materials

Plant materials used for the identification of flavonoid diversity were obtained from NASC. *GK_146E10* was obtained from GABI-Kat.

Table 6. *A. thaliana* wildtype and mutant accessions used for detection of flavonoid accumulation differences. Adapted from Nordborg *et al.* (2005) and TAIR

Accession name	NASC number	Place of origin
Ak-1	N939	Achkarren/ Freiburg, Germany
Columbia-0 (Col-0)	N1092	Columbia, USA
Col-4	N933	Columbia, USA
Col-8	N60000	Columbia, USA
C24	N906	Coimbra, Portugal
Cape Verdi Islands-0 (Cvi-0)	N902	Cape Verdi Islands, Cape Verdi
CS22491	N22491	Konchezero, Russia
Eifel-2 (Ei-2)	N6689	Eifel, Germany
Landsberg *erecta*-0 (L*er*-0)	NW20	Landsberg, Poland
Gückingen -0 (Gü-0)	N1212	Gückingen, Germany
Lezoux-0 (Lz-0)	N22615	Lezoux (Puy-de-Dome), France
Niederzenz-0 (Nd-0)	N6803	Niederzenz, Germany
Weiningen-1 (Wei-1)	N1639	Weiningen, Switzerland
Wassilewskija-0 (Ws-0)	N1602	Wassilewskija, Ukraine
Lister and Dean Col x L*er* RIL population	N1899 (set 1) N4858 (set 2)	
tt4	NW85	
tt7	NW88	
tt10-1	NW110	
tt10-4	N502972	
GK_146E10	N395786	

2.7.2 *A. thaliana* seed sterilization

Seed sterilization buffer (SSB)

70 % (v/v) ethanol

0.05 % (v/v) Triton X-100

The desired amount of seeds to be surface-sterilized (maximum 200 mg) was soaked in 1 ml SSB. The sample was vortexed occasionally for 10 min, and then the washing solution was replaced with 1 ml of 100 % ethanol. After 5 min incubation with occasional vortexing, the ethanol was replaced with fresh 1 ml 100 % (v/v) ethanol. The seeds were dried on pre-sterilized filter paper (Ø = 70 mm, Macherey-Nagel) under the bench.

2.7.3 Plant growth conditions in green house

Seeds were sown on a moisturized filter paper placed in a petridish and incubated under complete darkness for 3 days at 4 °C. Then, the seed were transferred to a pot or a tray containing soil. The pots were incubated in a short day (SD) chamber for 2 weeks and then transferred to a long day chamber for seed production. The settings of SD chamber were 8 h of light, with temperature settings ranging between 20 to 22 °C during day, and 16 to 22 °C during the night. The conditions of LD chamber were 14 h of light and temperatures ranging between 20 to 22 °C during day and 16 to 22 °C during the night.

2.7.4 Plant growth conditions in light room

0.5 x MS (Murashige and Skoog) plate

0.5 x MS medium (Sigma-Aldrich)
2 % (w/v) sucrose
0.8 % (w/v) agar)
pH 5.7 adjusted with KOH

Surface sterilized seeds were sown on a 0.5 x MS plate and incubated under complete darkness for 3 days at 4 °C. Then, the plates were transferred to the room illuminated with three different light sources consist of Gro Lux F58W/GRO-T8, LumiLux de luxe L 78w/76 natural color and LumiLux de luxe L 58w/840 white fluorescent tubes (SYLVANIA-OSRAM, Munich). The plant seedlings were harvested at the seventh day after germination (7 DAG) for metabolite analysis. The settings of the light room were 16 h light per day and the room temperature at 21 °C.

2.7.5 *A. tumefaciens* mediated plant transformation

In order to transfer a gene to *A. thaliana* genome, *A. tumefaciens* mediated plant transformation was performed according to Clough and Bent (1998). Healthy *A. thaliana* plants grown in soil under LD conditions were used for the transformation (2.7.3). The first inflorescence of the *A. thaliana* plant was clipped to encourage proliferation of many secondary bolts. Roughly 4-6 days after clipping, the plants with many immature flower clusters were used for the transformation. *A. tumefaciens* carrying a binary plasmid DNA construct (2.6.2.1) were grown overnight at 28 °C in 500 ml YEP medium with appropriate antibiotics. A recently stationary culture was mixed with sucrose to make a sucrose concentration of 0 % (v/v) in 500 ml final volume. Silwet L-77 (100 µl/l) was added to the *A. tumefaciens* solution. The above ground parts of the plants were dipped in the solution for 30 sec, with gentle agitation. The dipped plants were covered with plastic bags and were kept in darkness overnight to maintain high humidity and cool temperature. After overnight storage, the plants were transferred to a green house and grown without plastic bags in LD conditions. The plants were watered and grown normally until the seeds matured. Upper parts of the matured plants with siliques were covered with paper bags for harvesting the seeds. Harvested seeds were sterilized (2.7.2) and grown on MS plates (2.7.4) with appropriate selective medium for isolating transgenic plants containing T-DNA. The transgenic plants were transferred to soil for seed production under LD conditions.

2.7.6 Histochemical analysis of β-glucuronidase (GUS) activity in plant tissues

Fixing solution

0.3 % (v/v) formaldehyde
10 mM pH 5.6 MES-KOH
300 mM mannitol

NaPi buffer

50 mM Na_2HPO
pH 7.0 adjusted with phosphoric acid

5-bromo-4-chloro-3-indolyl-β-D-glucuronic acid (X-Gluc) solution

0.5 mg ml^{-1} X-Gluc in NaPi buffer

Transgenic plants carrying the T-DNA containing target promoter fused with *uidA*-ORF were analyzed. A histochemical assay was used to identify the promoter specificity by observing β-glucuronidase activity according to Jefferson et al (1987). Parts of transgenic plants or whole seedlings were soaked in fixing solution at least for 45 min at RT. The fixing buffer was replaced with NaPi buffer and incubated at RT for at least 45 min. During the 45 min period the buffer was exchanged once to the fresh buffer. After fixing and washing, the vacuum infiltration method was use to infiltrate X-Gluc solution to plant tissues. Then, the infiltrated sample was incubated at 37 °C for overnight in the solution. The stained sample was washed with NaPi buffer and destained in 100 % (v/v) EtOH at RT on an Unimax 1010 orbital shaker (Heidolph Instruments, Schwabach) at 120 rpm. The samples were softened by storing in a 30 % (v/v) glycerol solution. The samples were analyzed using Zeiss Binocular Stemi 2000-C (Carl Zeiss Microimaging, Goettingen) mounted with a JVC digital camera KY-F750 3CCD (Victor Company of Japan, Japan) and two SCHOTT KL15000 LCD Cold Light sources (SCHOTT, Mainz). Digital images were taken using DISCUS software version 4.50.1638 (Hilgers, Koenigswinter)

2.7.7 Quick plant DNA extraction

Extraction buffer

200 mM Tris-HCl pH 7.5

250 mM NaCl

25 mM EDTA

0.5 % (w/v) SDS

TE (Tris, EDTA)

10 mM Tris-Cl pH 7.5

1 mM EDTA

Genotypes of transgenic plants were analyzed using the DNA extracted according to Edwards *et al.* (1991). A small amount of leaf material was homogenized with sterile sand and 400 µl of extraction using a pellet pestle. The sample was centrifuged at 16,100 *x g* for 2 min at RT in a table top centrifuge. Only the supernatant was mixed with one volume of isopropanol and incubated for 2 min at RT. The mixture was centrifuged at 16,100 *x g* for 5 min at RT for 5 min. The pellet was washed with 70% ethanol and again centrifuged at 16,100 *x g* for 2 min, dried and resuspended in 50 µl of TE. One µl of the extract was used for PCR reactions.

2.7.8 Large scale plant DNA extraction

CTAB1 buffer

2 % (w/v) CTAB
100 mM Tris-HCl pH 8.0
20 mM EDTA
1.4 M NaCl
0.5 % (w/v) Polyvinylpyrrolidone 40 (PVP40) (Sigma-Aldrich)
10 mM DTT

CTAB2 buffer

1 % (w/v) CTAB
50 mM Tris-HCl pH 8.0
10 mM EDTA
0.20 % (w/v) PVP40)

The CTAB plant DNA extraction method was used to obtain plant DNA which required to be stored for longer periods of time using the protocol described by Rosso *et al.* (2003). About 100 mg of plant material was harvested in a 96 well collection plate. The samples were crushed with two metal beads (Φ = 2 mm) with 150 µl CTAB1 using the Tissuelyzer (Qiagen). The plate was incubated at 65 °C for 20 min. Then it was cooled down for 10 min at RT. Dichloromethane (150 µl) was added to each well. The plate was firmly vortexed and centrifuged at 16,412 *x g*, 4 °C for 10 min using a SIGMA drum rotor 09865 and a SIGMA 4K15C

refrigerated bench top centrifuge (SIGMA Laboratory Centrifuge, Osterode am Harz). The supernatants were transferred to a fresh deep-well plate (Macherey-Nagel) with 300 µl CTAB2 buffer and vortexed. After centrifuging at 16,412 *x g*, 4 °C for 10 min with the SIGMA centrifuge, only the pellets were dried at RT. Subsequently, the pellets were resuspended in 1 M NaCl (100 µl) and mixed with 0.7(v/v) isopropanol. The mixture was centrifuged for 30 min at 16,412*x g* at 4 °C. The pellets were washed twice with 70 % (v/v) EtOH, dried, resuspended in 100 µl TE containing 7.5 µg/ml RNase A and incubated for 40 min at 37 °C. The DNA samples were stored at 4 °C.

2.8 Metabolite analyses materials and methods

2.8.1 Metabolite analyses materials

All metabolite standards were at least of HPLC grade and were purchased from either Sigma-Aldrich or Roth. All HPLC gradient running solutions were purchased from J.T. Baker (Phillipsburg, NJ, USA).

2.8.1.1 Metabolite analyses plant materials

Because *A. thaliana* growth conditions influence accumulation of flavonoid derivatives, more attention were taken in account for growth conditions and extraction methods. *A. thaliana* has shown to induce flavonoids by stress like; the roots of seedlings exposed to light has been shown to induce flavonoids 100 folds higher than the unexposed roots (Hemm *et al.*, 2004). Flavonoid compositions in the plants can be influenced by the stresses, for instances, quercetin derivatives accumulate in leaves where kaempferol derivatives usually accumulate after UV-light treatment (Ryan *et al.*, 2001) and anthocyanins can be induced by amount of sucrose and shown red pigmentation in seedlings (Teng *et al.*, 2005). To minimize variations in flavonoid accumulations at each experiments and samples, the seedlings which have advantage in minimum cultivation area and time (harvested 7 days after germination) were grown under same condition as possible (2.7.4). The seedling samples were harvested quickly in liquid nitrogen which was important to stop enzymatic reaction for metabolite analysis (Fiehn, 2002). Moreover frozen sample were freeze-dried to take water from samples to stop enzymatic activity (Fiehn, 2002).

2.8.2 Plant metabolite extraction

Extraction buffer

10 ng ml^{-1} Genistein (internal standerd)

80 % (v/v) Methanol

Aqueous mixture with methanol was used to extract flavonoid except anthocyanins which need acidified extraction buffer for accurate quantification of the derivatives and the acidified buffer might produce artifacts like acetylated derivatives (reviewed in Robards and Antolovich, 1997; Tohge *et al.*, 2005b; Stobiecki *et al.*, 2006). An internal standard genistein was added prior to extract metabolite that compensates for signal loss attributed to sample preparation and analyses (Boyd, 1993). Freeze dried plant material (5 – 10 mg) mixed with 80 % methanol (MeOH) containing 10 ng ml^{-1} genistein (500 – 1000 µl) was added to a 2 ml screw cap tube with 0.5 g of 1.0 mm silica beads (Roth). The cells were disrupted three times in a FastPrep Instrument (Qbiogene, Heidelberg) at 6.5 m/s for 45 s. The mixture was further incubated at 70 °C for 15 min at 1,400 rpm in a Thermomixer. After centrifugation at 16,100 *x g* for 20 min at RT using a table top centrifuge, the supernatant was evaporated to complete dryness in a SPD SpeedVac concentrator (Thermo Scientific, Dreieich). The final concentrations of plant material and genistein were adjusted in methanol to 100 µg/µl plant dry weight (DW) and 100 ng/µl genistein respectively.

2.8.3 DPBA staining of the seedlings

Diphenylboric acid-2-aminoethyl ester (DPBA) solution

0.20 % (w/v) DBPA

0.00375 % (v/v) Triton X-100

Prepared in sterile water

DPBA staining of *A. thaliana* seedlings was use to visualize flavonoid derivatives *in planta*. The protocol was adapted from Sheahan and Rechnitz (1992). 7 DAG *A. thaliana* seedlings grown on filter paper were stained with a saturated DPBA solution for 1.5 h at RT. Fluorescence was visualized on a Leica DM5500 B

epifluorescence minroscope (Leica Microsystems, Wetzlar) mounted with the JVC digital camera KY-F750 3CCD using Leica Filtercube A with an excitation wavelength of 340 – 360 nm and a 425 nm long pass filter. Pictures were taken using the DISKUS digital imaging software.

2.8.4 High performance thin layer chromatography (HPTLC) analysis

HPTLC running bluffer
10 ml ethyl acetate,
1.2 ml formic acid,
1.2 ml acetic acid,
2.6 ml water

DPBA solution
1 % (w/v) DBPA
Prepared in 100 % (v/v) MeOH

Polyethyleneglycol (PEG) 4000 solution
5 % (w/v) PEG
Prepared in 100 % (v/v) MeOH

HPTLC analysis was used to screen the flavonoid composition of a large number of plant samples and to identify aglycones of flavonoids. The protocol was adapted from Wagner and Bladt (1996). Two µl of the plant extract (2.8.2) was applied on a spot which was marked uniformly 1 cm away from bottom of 10 x 10 cm HPTLC silica gel 60 plate (Merck). The plate was placed in a HPTLC Flat Bottom chamber for 10 x 10cm plates (CAMAG, Muttenz, Switzerland) containing 15 ml running buffer made of ethyl acetate, formic acid, acetic acid, and water (10:2.6:1.2:1.2). The plate was removed when the solvent front reached the last 5 mm of the plate. When the plate dried completely, it was sprayed with 1 % (w/v) DPBA methanolic solution and left to dry for 20 min. Subsequently, 5 % (w/v) polyethyleneglycol (PEG) 4000 methanolic solution was sprayed to enhance the fluorescence of flavonol derivatives. After the plate was completely dried, the stained chromatograms were examined under UV light (312 nm) and

photographed. DPBA-stained kaempferol, quercetin and sinapate derivatives fluoresced green, orange and light blue under 312 nm UV light.

Retention factor (Rf) value is used to compare different flavonoid samples. Rf is the distance traveled by applied sample relative to solvent front. Degital image of HPTLC plate is used to measure the distances require for calculation of Rf using Image J software. Figure 10 explains calculation of Rf.

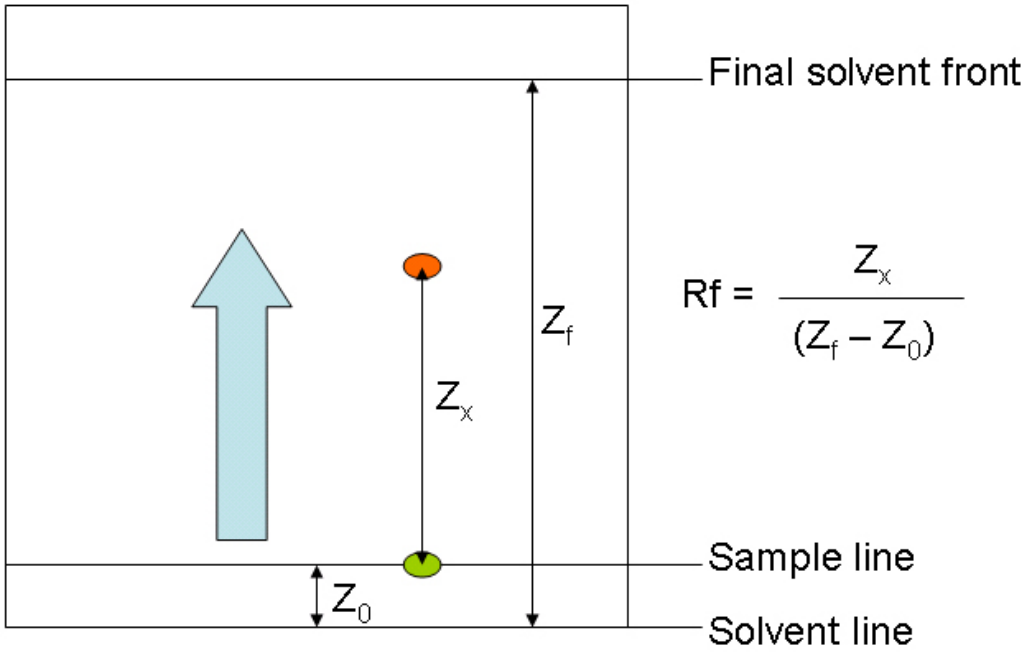


Figure 10. Calculation of Retention factor (Rf)
The retention factor of the sample separated by HPTLC can be calculated by the formula as listed. The blue arrow indicates the direction of mobile phase. Z_0 is the distance from solvent line to the line which sample is applied. Z_f is the distance of the sample traveled. Z_f is the distance the solvent traveled. Figure and the formula are adapted from a chromatography application note AN9 from Isco Inc. (http://www.isco.com/).

2.8.4.1 Flavonoid isolation using HPTLC silica gel plate

Flavonoid derivatives that appeared in HPTLC plates were extracted for identification of conjugated sugars by GC-MS. Figure 11 explains the metabolite extraction method developed for the identification of flavonol derivatives on HPTLC plate. Using a 0.5 µl glass capillary tube (Merck), 2 µl of methanolic extracts were applied on the HPTLC plate. The plate was placed in a HPTLC flat

bottom chamber with the running buffer as the mobile phase (indicated as arrow) in a closed glass tank. After separation and air drying, the center lanes(except for the first and the last), were covered with Parafilm® (Pechiney Plastic Packaging, Chicago, IL, USA) to avoid a chemical reaction between the flavonoids and the staining reagents. The covered plate was sprayed with the 1 % DPBA. After viewing the UV-fluorescence and marking of the specific band with a ruler, the unstained flavonol derivative bands were scratched out from the HPTLC plate accordingly to the parallel marks as indicators. The extracted band was dissolved in 500 µl 80 % MeOH. The mixture was vortexed and incubated at RT for at least 1 h. After precipitating the silica gel by centrifugation at 16,100 *x g* for 15 min, only the supernatant was collected and dried. The dried sample was resuspended in 80 % (v/v) MeOH of the desired volume. The isolate was directly used for structure identification. The plate was stained again as described in step 4 to check absence of targeted flavonol.

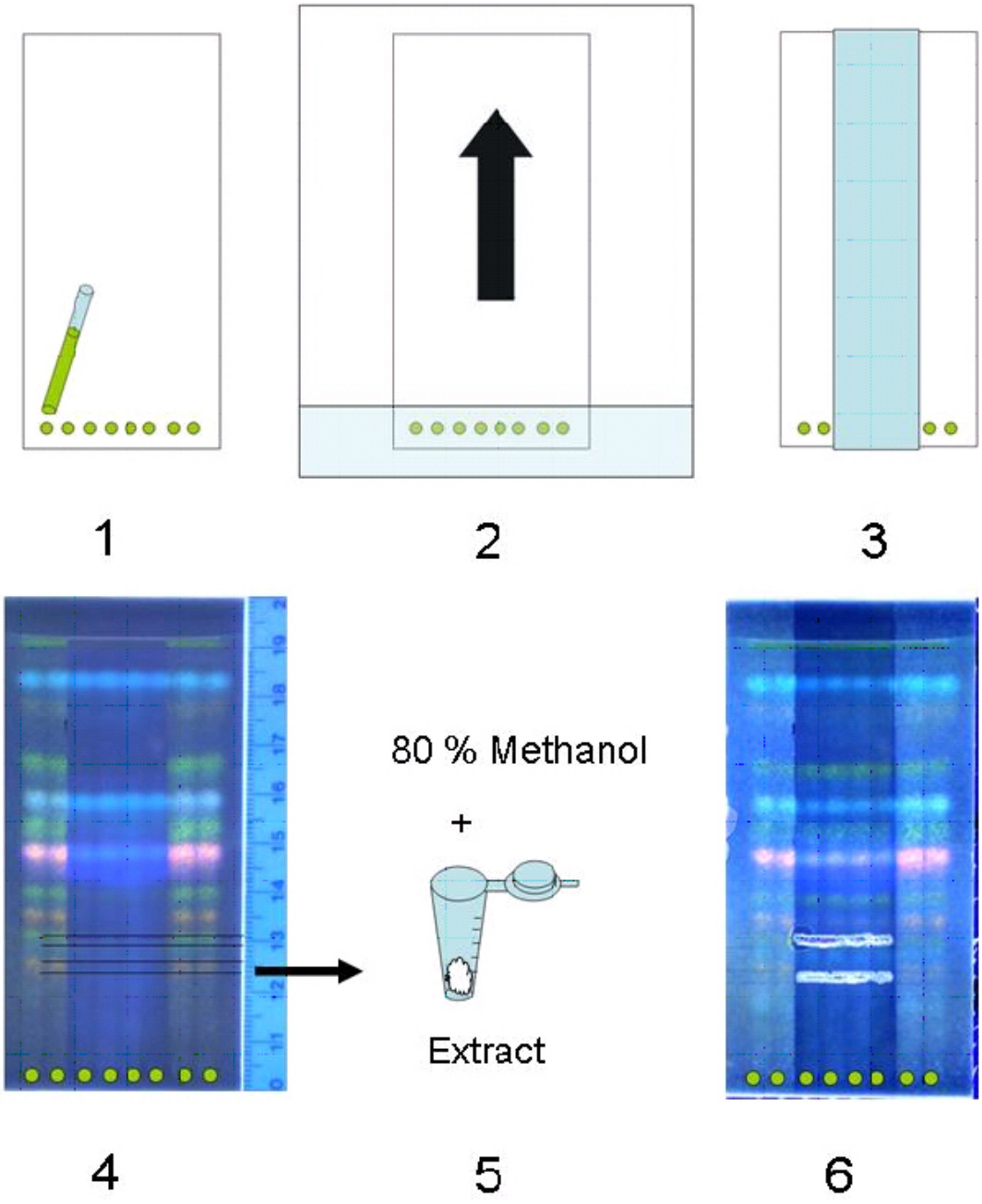


Figure 11. Scheme of flavonoid band extraction from HPTLC plate.
Stepwise flavonol purification method using a HPTLC silica gel 60 plate.
1. A sample is applied on the plate using a 1 µl glass capillary tube. 2. The direction of mobile phase is indicated with an arrow. 3. The plate is covered with a strip of parafilm. 4. Positioning of target flavonoids is performed by taking a picture with a ruler as an indicator. The targeted sites are marked with a carbon pencil. 5. The silica gel-containing fraction of the flavonoid band is dissolved in 80 % (v/v) MeOH for further analysis. 6. The plate is stained with DPBA to check for the absence of the target flavonoid bands.

2.8.5 High Performance Liquid Chromatography - photo diode array (HPLC - PDA) and electro spray ionization/mass spectroscopy (ESI/MS) analyses

Quantitative and structure analysis of flavonoid derivatives were performed by using a HPLC-PDA-ESI/MS system, composed of Finnigan Spectra System P4000 HPLC, UV6000photodiode array, Autosampler AS3000 and Finnigan LCQ DECA mass spectrometer (Thermo Scientific, San Jose, CA/USA). A custom spray head with spray angle of 30° was used to analyze 10 µL of 1 to 10 dilutions of the methanolic extracts (2.8.2) and HPTLC purified flavonol derivatives (2.8.4.1). HPLC was carried out on a LiChroCART C18.4 µm, 2- × 125 mm column (Merck) at a flow rate of 0.2 ml min^{-1}. Solvent A was acetonitrile with 0.1 % formic acid (J.T. Baker), and solvent B was 0.1 % formic acid in water (J.T. Baker). Elution gradient program was a linear gradient from solvent A to solvent B. The gradient program was started with 0 % A (5 min), linearly increased to 20 % A (30 min) and ended with 100 % A (35 min). Photo diode array (PDA) was used for detection of UV-visible absorption in the range of 200–700 nm. Nitrogen was used as sheath gas (6 bar at a flow rate of 0.8 l min^{-1}) for the positive-ion ESI-MS which was performed at capillary temperature and voltage of 320 °C and 5.0 kV, respectively. The tube lens offset was set at 10.0 V. Full scan mass spectra were acquired from 80 to 2,000 m/z at 2 scans/sec. Stepwise fragmentation analysis was carried out with different activation energies depending on the compound class.

2.8.6 Gas chromatography - mass spectroscopy (GC-MS) analysis

GC-MS was used to analyze sugars conjugated to HPTLC-purified flavonoid derivatives (2.8.4.1). 100 nM solutions of pure reference substances were placed into 1 ml tight reacti-vials (Supelco, Bellefonte, PA/USA). 0.5 ml methanolic HCl (at 1.25 M concentration) was added and the samples were incubated at 80 °C for 6 h. Subsequently, the samples were dried under a nitrogen stream and 50 µl of both pyridine and N-methyl-N-(trimethylsilyl) trifluoroacetamide (Macherey-Nagel) were added. The samples were stirred at 37 °C for 30 min. After hydrolysis and derivatization, the samples were analyzed on a Trace GC/PolarisQ bench-top GC/MS system equipped with an AS 2000 autosampler (Thermo Scientific). The

GC separation was performed on a 30 m Equity5 column with 0.5 µm film thickness (Supelco). The temperature of the injection port was adjusted to 200 °C and to 250 °C for the transfer line. Helium was used as carrier gas at a flow rate of 1 ml/min. Samples of 1 µl were injected in the splitless mode. The initial column temperature was 80 °C. This temperature was held for 3 min and was gradually increased at a rate of 5 °C/min to a final temperature of 325 °C. To equilibrate the system for the next sample injection, the temperature was lowered to 80 °C again and held constant for additional 5 min. Mass spectra were recorded with a scanning range of 50 to 650 m/z at 2 scans/sec.

2.9 QTL mapping and linkage analysis materials and methods

2.9.1 Mapping Materials

The published genotype data collected from the 101 Lister and Dean RIL were obtained from NASC (http://arabidopsis.info/new_ri_map.html). Windows QTL Cartographer version 2.5 (WinQTLCart) was obtained from the North Carolina State University Bioinformatics Research Center and Programs in Bioinformatics and Statistical Genetics (http://statgen.ncsu.edu/brcwebsite/home.php). MAPMAKER version 3.0b was obtained from Whitehead Institute for Biomedical research at Massachusetts Institute of Technology (ftp://ftp-genome.wi.mit.edu/distribution/software/mapmaker3/).

2.9.2 PCR based genotyping

PCR-based cleaved amplified polymorphisms (CAPS) (Konieczny and Ausubel, 1993) or Insertion Deletion (InDel) (Bhattramakki *et al.*, 2002) markers were used for genotyping of the Lister and Dean L*er* x Col recombinant inbred line (RIL) population (Lister and Dean, 1993). Table 7 lists all the PCR primer sets and restriction enzymes used for genotyping of the RIL. PCR primers were designed by either comparing Col and L*er* genomic sequence using TAIR L*er* BLAST search or were obtained from MASC (Schmid *et al.*, 2003; Torjek *et al.*, 2003). For CAPS analysis, 5 µl of the amplicons were digested with 0.5 units of the appropriate restriction enzyme and 1 x enzyme buffer filled with water to a total volume of 10 µl in each well from a 96 PCR plate. The PCR plate was incubated at suitable

temperatures either 37 or 65 °C depending on type of enzymes for at least 3 h. The sizes of fragments were checked with 1 % (w/v) agarose gel (2.5.1). Amplicons from InDeL markers were directly analyzed using 1 % (w/v) agarose gel (2.5.1). The genotypes of the RIL population were scored as A (Columbia), B (Landsberg *erecta*), - (Heterozygous or no data). The scored genotypes were collected and organized in an excel worksheet and saved as a text file for genetic mapping purpose.

Table 7. List of PCR based molecular markers.
[a]The marker is on *A. thaliana* chromosome 1. [b]The marker is on *A. thaliana* chromosome 5. *CAPS markers were indicated with the name of restriction enzymes used for cleaving the amplicon.

Marker	Name of primer and sequence (5'-3')	Position	*Type
[a]H078	H078 GCTGTTTGATCTTGGTCTCAAAC	22490401	*Hyp*188I
	H079 TAGCTTACTACAATCTCGAGAGC		
[a]H096 (MASC 02080)	H096 ACAGGTACAGTATTTCATAACATTC	22704745	*Hyp*188I
	H097 ACACACTCTTTGTTCTCGAGATG		
[a]H098 (MASC 02574)	H098 GGGTCTGAAGCTATATCGGG	22290039	*Bmg*I,
	H099 GTTATCTGCAGAAGTTTCATCC		*Taq*$^{\alpha}$I
[a]H100 (MASC 01983)	H100 TTAGGGATTGACTTTAGTTTGAGCG	20139508	*Hga*I
	H101 GCTGGACTTGGGTTAGGGTTAG		
[a]H102 (MASC 07418)	H102 TACAGGGTGCGTAGTTAAACATC	23104168	AciI
	H103 CAAAGAAGATAATCATCACTCTCC		
[a]H121	H121 ATCAATATTCGTGGACGACTAGG	23039704	*Hin*P1 I
	H122 AAGCTGGAGACCGTGAACGAC		
[a]H125	H125 TAGGAACAATAATGGGTCATCCG	23062461	*Aci*I
	H126 GGATTGCCCAAGACATTATCCAT		
[a]H127	H127 GGCTAGAACCGAAGTCTCAGG	22263044	*Hin*dIII
	H128 ATCACCATCAAACTTGACCCCC		

[a]H137	H137 CAATTGCATATTCATTAGGTACCC	22267814	*Eco*RV
	H138 TTGAAACTTCTCTGGAGTGAAAGG		
[a]H170	H170 CCCAAGAAACGGTTTACTGCGGC	22904835	*Taq*$^{\alpha}$I
	H171 CTCTGAACCGGTTCATAGTAGCTG		
[a]H174	H174 GCACAACAAGGTCGATTCTTTACC	22811330	*Taq*$^{\alpha}$I
	H175 TGTCTCGTTCCATCTCCATCCAAC		
[a]H180	H180 AACTTCTCCTGCCCGTATAATGCC	22573851	*Hyp*188I
	H181 CTCCCAAATGTCACCATCATCCTC		
[a]H184	H184 TCTCAATGGAGCAATTTGCACAAGC	22357338	*Dra*I
	H185 CCTTACATCATTCTTCTCGTTATGAG		
[a]H206	H206 ATCAACTCGTCTCTTGTTGCTGCC	21419862	*Hin*P1 I
	H207 GACACATCCCCACATAATCAGGAG		
[a]H216	H216 AGATCGCTGCCAAAGATACTCCTC	21840436	*Taq*$^{\alpha}$I
	H217 CAGTTGTTGGAGGAGTAAGGAGAG		
[a]H220	H220 CAAACTTCAAACTGAACCAGCGTAC	22025808	*Hae*III
	H221 GAGACGACGACGAGAGATGCCG		
[a]H238	H238 ATCCCCAGAGGCCAAGTCTAATAG	22184065	*Eco*RI
	H239 AAAGTGACATGTCCTTATTTTAACTCC		
[b]H322	H322 GTGACCAAAACCAAACCATGGGAAC	21938415	InDel
	H323 CCCTTAAGTTGTACGGTTTGTGATG		
[b]H379	H379 GAAAGAGGAGATCATACAAGTAGAATC	19511082	InDel
	H380 CCTAGCTACTCCTCCTCATTCAGG		

2.9.3 Construction of the quercitrin standards curve

The standard curve generated using HPLC-PDA assay is used to measure amount of quercitrin (ng) per 100 µg DW of seedling samples. The standard curve was obtained by plotting six different concentrations of commercially available quercitrin samples (0 µg ml^{-1} 0.5 µg ml^{-1}, 1 µg ml^{-1}, 2.5 µg ml^{-1}, and 5 µg ml^{-1}) on the x-axis and their assay measurement (µAU) on the Y-axis. 10 µl of the each standard samples were used for HPLC-PDA analysis which resulted in 0, 5, 10, 15, 25, and 50 ng of quercitrin peaks in the chromatograms respectively. The standard curve and its slope was extrapolated using Microsoft Excel (Figure

12). The slope was used to calculate the amount (ng per 100 µg DW) of quercitrin in the plant extracts.

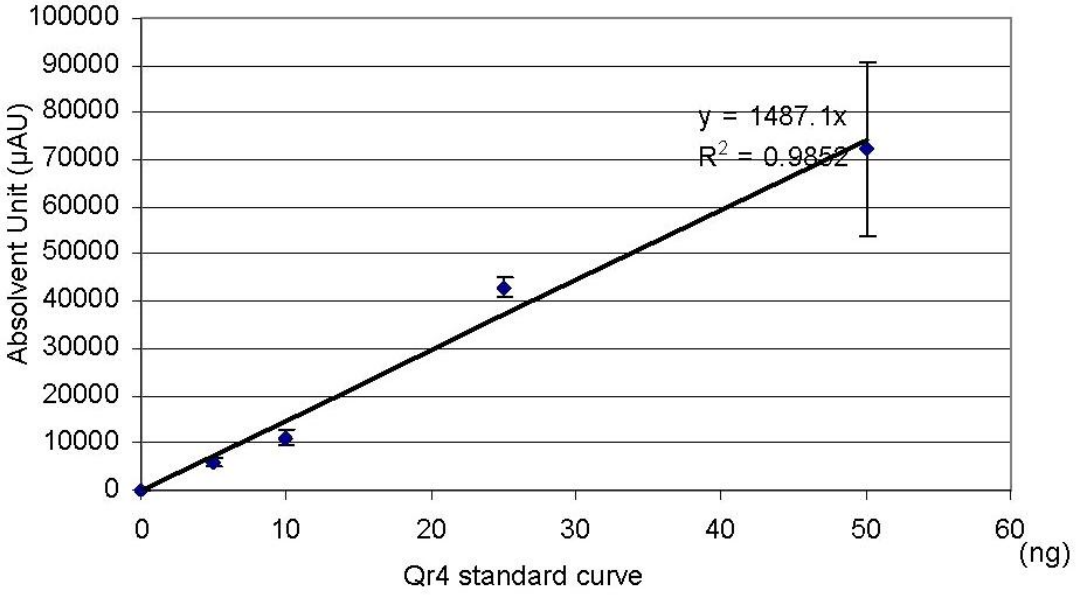


Figure 12. The standard curve used for calculating qr4 amount per 100 µg of DW sample.
Absolvent unit and amount of Qr4 are listed on the ordinate and abscissa. The formula of the standard curve obtained from quantification of Qr4 is y = 1487.1 X (Y = µAU, X = ng). The error bars indicate the standard deviation of the four independent measurements.

2.9.4 Normalization of quercitrin peaks for QTL analysis

For each QTL analysis, IS peaks from each RIL samples were divided by the mean of IS peak area (µAU) from 92 RIL samples. The obtained value was used for the normalization of the Qr4 peaks from each sample. The normalized Qr4 values (µAU) were used to calculate amount of quercitrin (ng per 100 µg DW) using the standard curb (2.9.3) and used for QTL analysis.

2.9.5 QTL mapping using Windows QTL Cartographer

Single-marker, interval-mapping (IM) and composite interval mapping (CIM) were used to detect quantitative trait loci (QTL) controlling quercitrin yield according to the manual from Windows QTL Cartographer version 2.5 (WinQTLCart) (Wang, 2005). The Qr4 quantitative value from the RIL (2.9.4) was used to create a trait data file for WinQTLCart. Marker genotype, chromosome number and marker number per chromosome, marker label, marker position, and cross information of

the Lister and Dean L*er* x Col RIL population were created to separate text files based on the information obtained from NASC. The genotype data from 279 markers with known positions, selected from more than 1000 genetic markers, were taken from NASC (2.9.1). Numbers of the markers per the chromosomes (in cM) were: chromosome 1 (Chr1)-67 markers/134 cM, Chr2-40 markers/109.9 cM, Chr3-30 markers/92.9 cM, Chr4-74 markers/124.7 cM, and Ch5-68 markers/142.02 cM. The data were combined to create a source data file in MCD format. The source file was used for QTL analysis. Single marker analysis was used to find association between a marker genotype and trait value. IM and CIM methods evaluate a possibility of a target QTL at multiple analysis points across each interlocus interval. Additionally CIM reduces residual error by analyzing the effect of background markers shown to be associated with the trait in elsewhere and minimizes the peak and noise ratio. The genome-wide significance threshold of 5 % was estimated with a permutation-test analysis, 1000 permutations of phenotypic data were analyzed using the composite interval mapping model for Qr4 data as suggested (Wang, 2005). The walk speed (the genome scan interval) chosen for the QTL analysis was the default set of 2 centi Morgan (cM).

2.9.6 Calculating map positions of new markers (Linkage analysis)

The genotype data from 542 marker were also selected from the published genotype data from NASC (2.9.1) for linkage analysis because it was not possible to use the genotype data containing more than 564 marker for linkage analysis using a computer equipped with Pentium 4 1.8 GHz CPU with 523.24 kilobytes random access memory (RAM). In order to integrate the new molecular (Table 7) and qualitative trait markers into the *A. thaliana* genetic map, the new marker positions were estimated using the genotype data from NASC and MAPMAKER (version 3.0b) (Lander, 1987). The nearest markers and the distance (cM) to the new markers were calculated according to the manual (Lincoln *et al.*, 1993). Detailed information about MAPMAKER can be obtained from Whitehead Institute for Biomedical Research. A simplified example of MAPMAKER analysis is provided in Figure 13. The nearest markers close to H379 were determined by using a command “near.” All markers which matched the criteria of more than LOD (logarithm of the odds) score of 2 and less than distance of 4 centi Morgans

(cM) were manually collected using a command “seq.” Curly brackets and without bracket specified that there were more than one sequence orders or only one order. Afterward marker orders were estimated using a command “compare.” By comparing to the estimated marker orders to RIL marker database at NASC, only one marker order was chosen. Finally, the command “map” was used to find the distances between the markers. The markers G2400 and Ve027 were located 100.6 and 104.4 cM on chromosome 5. The estimated H379 position was 101 cM on chromosome 5.

1> ***near*** 552
Linked Markers in Sequence at min LOD 3.00, max Distance 50.0

Marker-1	Marker-2	Theta	LOD	cM
552	418	0.08	9.44	9.12
552	419	0.03	15.93	3.45
552	420	0.05	13.27	5.72
552	421	0.03	15.66	3.50
552	422	0.00	6.61	0.04
552	423	0.00	6.31	0.04
552	424	0.02	18.49	1.99
552	425	0.03	17.15	2.70
552	426	0.06	6.13	6.07
552	427	0.07	10.36	7.45

2> ***seq*** {421-425 552}
sequence #1= {421-425 552}

3> compare
Best 10 orders:
1: 425 424 422 423 552 421 Like: 0.00
2: 425 424 423 422 552 421 Like: 0.00
3: 425 424 423 552 422 421 Like: -0.01
4: 425 424 422 552 423 421 Like: -0.01
5: 425 424 552 422 423 421 Like: -0.01
6: 425 424 552 423 422 421 Like: -0.01
7: 425 422 424 423 552 421 Like: -0.03
8: 425 423 424 422 552 421 Like: -0.03
9: 425 423 424 552 422 421 Like: -0.04
10: 425 422 424 552 423 421 Like: -0.04
order1 is set

4> seq 425 424 423 421 552 421
sequence #2= 425 424 423 421 552 421
warning: marker(s) listed in sequence multiple times...

5> ***map***
==========
Map:

Markers		Distance
425	mi2aa	0.5 cM
424	mi271	0.6 cM
423	g13698	0.6 cM
422	g2400	0.6 cM
552	H379	3.3 cM
421	ve027	----------

5.6 cM 6 markers log-likelihood= -47.55
==========

Figure 13. Scheme of linkage analysis using MAPMAKER program.
The commands used in MAPMAKER are marked with bold italics. Position of a marker H379 (552) was estimated using MAPMAKER. Logarithm of the odds (LOD), Centi Morgan (cM), theta-recombination fraction.

2.10 Bacterial Artificial Chromosome (BAC) related materials and methods

2.10.1 BAC related materials

Landsberg *erecta* genomic library (L*er* BIBAC) made with a binary vector BiBAC2 (Accession number CD4-35) (Chang *et al.*, 2003) was obtained from ABRC. The BAC library constructed with 11,520 clones had an average insert size of 162 Kb and represented the *A. thaliana* L*er* genome with coverage of 11.5 times.

2.10.1.1 BAC screening filter

The L*er* BIBAC library screening filter was prepared at Chair of Genome research at University Bielefeld. The Hybond-N+nylon filters (Amersham Pharmacia Biotech, Piscataway, MD/U.S.A.) were placed on the 22.2 x 22.2 cm Q-tray plates (Genetix, Munich) filled with LB agar (LB + 1.8 % (w/v) agar) with 50 μg ml^{-1} kanamycin. High-density colony filters were prepared using a Flexy Robotic Workstation (Genome Solutions, Ann Arbor, MI, U.S.A.) to spot the colonies on the Q-trays. Each plate position was spotted twice in a specific design noted on the figure (Figure 14). When the colonies grew to 1 – 2 mm in diameter, the filters were prepared for hybridizations.

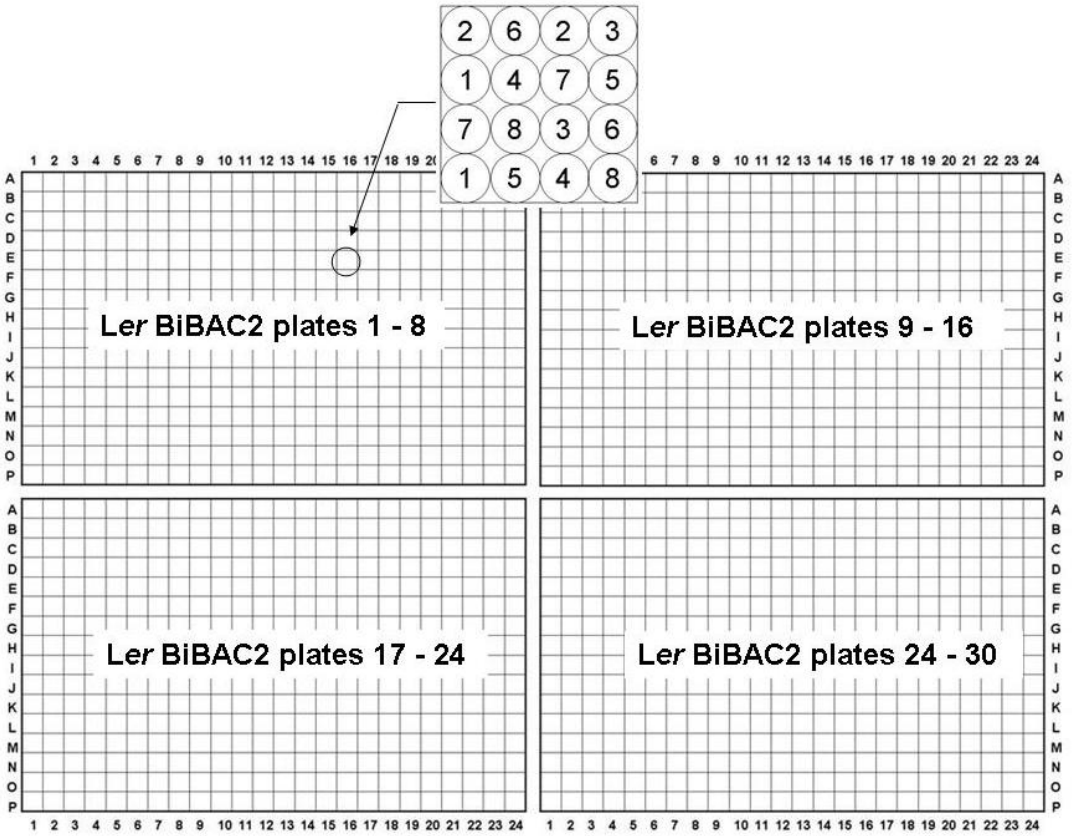


Figure 14. Layout of BAC screening filter prepared from the L*er* BiBAC library.
11520 L*er* BIBAC clones were spotted twice on each plate position as listed.

2.10.1.2 BIBAC custom sub-library

A custom sub-library of BAC clone 22K22 construction was ordered at GATC Biotech (Constance). The pCR®4Blunt-TOPO vector was used for construction of 22K22 shotgun library. After physical fractionation of the BAC DNA, the random sheared fragments were blunt-ended by using T4 DNA-Polymerase and then ligated into the pCR4Blunt-TOPO vector (Invitrogen). Approximately 6912 clones containing about 2 Kb insert length and 2304 clones with about 5 Kb insert length were produced from the BAC 22K22.

2.10.1.3 L*er* BIBAC22K22 clone consensus sequence

All clones from the sub-library were sequenced and assembled at the Sequence Core facility at Chair of Genome Research at Bielefeld University. The plasmids

from the shotgun library were sequenced and assembled at the sequence core facility. Approximately ten-fold redundant sequences were assembled using the PreGAP4 and Phrap from Staden software package (Medical Research Council Laboratory of Molecular Biology, Cambridge, UK). The sequence was submitted to the EMBL Nucleotide Sequence Database (accession number: AM748036).

2.10.1.4 L*er* genome annotation

Sequence annotation of BAC 22K22 was performed at the Plant Computational Biology at the Max Planck Institute for Plant Breeding Research Cologne. A piece of Col Chr1 spanning roughly the same region as the L*er* BAC for comparison (CHR1_22117000-22321000) was downloaded from MIPS. It contained the old MIPS annotation and the newer TAIR 6 annotation. Genes were annotated by comparing the homologous sequences in public databases using the Apollo Genome Annotation Curation Tool. The Col and L*er* targeted sequences were also compared using Sequencer alignment program.

2.10.2 Screening of the L*er* BiBAC library

20 x SSC stock solution

- 3.0 M NaCl
- 0.3 M sodium citrate
- 0.5 M EDTA pH 8.0

20 x SSPE stock solution

- 3.6 M NaCl
- 0.2 M Na_3PO_4
- 0.02 M EDTA
- Adjest pH to 7.7

100 x Denhardt's

- 2 % (w/v) bovine serum albumin (BSA)
- 2 % (w/v) ficoll 400
- 2 % (w/v) polyvinylpyrrolidone (PVP-40)

Hybridization buffer

5 x SSPE

0.5 M EDTA pH 8.0

5 x Denhard's

0.5 % (w/v) SDS

Washing solution 1

2 x SSC

0.1 % (w/v) SDS

Washing solution 2

0.1 x SSC

1 % (w/v) SDS

BAC screening was operated according to Meksem *et al.* (2005). The PCR product was radioactively labeled using the manual from the Ready-To-Go Labeling Beads tube (Amersham Pharmacia Biotech). The high density colony filters were placed in hybridization buffer for two h at 65 ˚C. Twenty ng of the PCR product and water were mixed to the total volume of 45 µl and placed on the heating block for five min at 95 ˚C. The mixture was cooled to RT. Five µl of 3,000 Ci mM^{-1} α-^{32}P dCTP (Hartmann Analytic, Braunschweig) was added to the mixture and transferred to a Ready-To-Go DNA Labeling Beads tube. The radioactive dCTP was incorporated to single stranded PCR at RT for 30 min and denatured at 95 ˚C for 5 min. Purification of labeled probe was performed according to the manual from MicroSpinTM S-200 HR Columns (Amersham Pharmacia Biotech). The mixture was passed through a MicroSpinTM S-200 HR Column by centrifugation at 400 *x g* for two min. The flowthrough radioactive probe was transferred to the hybridization tube containing the pre-hybridized filter and the buffer. After 16 h incubation at 65 ˚C, the filter was washed twice with 150 ml of washing solution 1 for 10 min each time, and further washed twice with 150 ml washing solution 2 for 10 min. The washed filter was sealed around with plastic, and exposed to a phosphoimager screen (Kodak, Rochester, NY, USA) for

3 h. The phosphoimager was scanned using a Typhoon 8600 imager (Amersham Pharmacia Biotech).

2.10.3 Large scale low copy number plasmid extraction method

2 x YT medium

1.6 % (w/v) tryptone

1 % (w/v) yeast extract

0.5 % (w/v) NaCl

pH 7.0 with NaOH

The BAC plasmid for sequencing was extracted using a Tepnel T1000 Automated DNA Purification System and Nucleoplex^{TM} BAC DNA kit (Tepnel Life Sciences PLC, Manchester/ UK). Protocol was followed according to the manual supplied from the company. Positive clones found from the BAC filters were inoculated in the wells of a 96 deep-well plate containing 1.3 ml 2 x YT medium with 50 µg ml^{-1} kanamycin. The plate was sealed with a gas permeable membrane and incubated for 24 h at 280 rpm in 37 ˚C incubator. The cell culture was pelleted by centrifugation at 1,500 *x g* for 15 min. The supernatant was completely removed using paper towel. The plate was placed in T1000 and alkaline lyses based extraction was performed automatically by the machine. The BAC plasmids from the 96 deep-well plate were extracted within 2.5 h.

2.10.4 Large scale high copy number plasmid extraction method

TB (Terrific broth) medium

1.2 % (w/v) bacto-tryptone

2.4 % (w/v) bacto-yeast extract

0.4 % (v/v) glycerol, 0.17M KH_2PO_4

0.72 M K_2HPO_4

Solution I

50 mM Glucose

10 mM EDTA pH 8.0

25 mM Tris-HCl pH 8.0

Solution II

0.2 M NaOH

1 % (w/v) SDS)

Solution III

3 M potassium acetate

11.0 % (v/v) glacial acetic acid

The extraction method was adapted from Sambrook and Russel (2001). The 96 well plates containing sub- clones of BAC 22K22 were inoculated in deep-well plates containing 1.5 ml TB medium with 50 µg ml^{-1} kanamycin. The plates were incubated on a shaker at 290 rpm at 37 ˚C overnight. The overnight cultures were spun for 15 min at 10,621 *x g* and supernatants were discarded. Two hundred µl of solution I was added to each well and the pellets were suspended by vortexing. Then 200 µl of solution II was mixed gently with the cell suspension and kept for 2 min at room temperature (RT). Next 200 µl of solution III was added to each of the wells, which were stored on ice for 10 min before centrifugation at 10,621 *x g*, at RT for 15 min using a SIGMA drum rotor 09865 and a SIGMA 4K15C refrigerated bench top centrifuge. After centrifugation, only supernatants were transferred to a fresh deep-well plate. For precipitation of the plasmid DNA, 400 µl of isopropanol was added to the wells and centrifuged for 30 min at 10,621 *x g*. Pellets were washed with 70 % (w/v) ethanol and dried for 1 h on the bench. The pellets were suspended with 100 µl of water with RNase A (10 µl ml^{-1}) and incubated at 37˚C for 1 h. Purification of the plasmid DNA was performed by adding 100 µl of 4 M ammonium acetate with 600 µl 100 % (v/v) EtOH. The mixture was centrifuged for 30 min at 10,621 *x g*. The pellets were dried and dissolved in 100 µl TE buffer

2.11 RNA materials and methods

2.11.1 RNA materials

Total RNA extracted from the protoplasts from At7 cell culture derived from *A. thaliana* accession Columbia was obtained and described by Hartmann *et al.* (Hartmann *et al.*, 1998).

2.11.2 RNA Isolation for transcript profiling

Total RNA from 7DAG *A. thaliana* seedling was extracted using TriReagent (Molecular Research Center, Cincinnati, OH, USA) according to the protocol provided by the manufacturer. Tri Reagent (1 ml) was added to 100 mg of plant tissue in 1.5 ml screw-cap microcentrifuge tube. The sample was homogenized using the Tissuelyser. After incubating the sample for 10 min at RT, chloroform (200µl) was added to the samples and vortexed. After 10 min at RT incubation, the sample was centrifuged at 16,100 *x g* at 4 °C for 10 min. The clear upper phase containing the total RNA was collected to a new microcentrifuge tube containing 0.7 x the total volume of isopropanol. The sample was incubated at RT for 5 min and centrifuged at 16,100 *x g* at 4 °C for 10 min. The supernatant was discarded and the pellet was washed with 70% ethanol. The pellet was air dried for 10 min at RT and dissolved in RNase free water. Total RNA was treated with Ambion DNase I (Applied biosystems) before reverse transcription.

2.11.3 cDNA synthesis

cDNA synthesis was carried out following a protocol from Mehrtens *et al.* (2005). Reactions were performed in a final volume of 20µl using 2µg of total RNA (2.11.2), 250µM oligo (dN-18T), 0.5 µM primer MJ389 (5'-ATTCTAGAGGCCGAG-GCGGCCGCCATG$(T)_{30}$VN-3'), 1 x First Strand Buffer (Invitrogen), 5 mM $MgCl_2$, 500 µM of each dNTP, 10 mM of dithiothreitol and 20 units of SuperScript II reverse transcriptase (Invitrogen). The reaction mix was incubated for 60 min at 42 °C, followed by inactivation of the enzyme by incubation at 90 °C for 10 min. One µl of cDNA was used for PCR.

2.11.4 Semi quantitative reverse transcriptase –PCR (RT-PCR)

One µl of cDNA (2.11.3) was used as a template in the PCR (2.5.3) with the following parameters: 28 cycles of 20 sec, 95 °C; 20 sec, 57 °C and 80 sec. 10 µL of the amplicons were analyzed by agarose gel electrophoresis and visualized with ethidium bromide (2.5.1). The amplicon from *actin-2 (ACT2)* transcripts were used for normalization of the result. The PCR primers used in RT-PCR are listed in the table (Table 8).

Table 8. List of the oligonucleotide primers used for semi quantitative RT-PCR analysis.

ATG code	RT-PCR primer (5' -3')
At3g18780 (ACT2)	RS469 TCCGCTCTTTCTTTCCAAGCTCAT
	RS470 TCCAGCACAATACCGGTTGTACG
At5g48100 (TT10)	H364 TTGTTCATACTCGCTGCTAATCTCG
	H371 ACCATGAAGTCCCCATAATTTTGGG

2.12 Transfection experiments materials and methods

Protoplast isolation and transfection experiments for the detection of transient expression analysis were performed at the cell culture facility at the Chair of Genome Research at Bielefeld University. The materials and methods were described by Mehrtens *et al.* (2005).

2.12.1 Growth condition of *A. thaliana*-suspension culture At7

MS-medium

1 x MS (Sigma-Aldrich)

1 *x g*amborg's vitamin (Sigma-Aldrich)

1 mg/l 2,4-dichlorophenoxylacetic acid (2,4D) (Sigma-Aldrich)

1.0 % (v/v) Diamant Raffinade zucker (Pfeifer & Langen, Cologne)

Cultivation of *A. thaliana* suspension culture At7 ecotype Columbia (1993) was used for transient expression analysis. The culture was cultivated under dark condition with MS-medium and continuously shaken with the condition at 105 rpm,

26 °C using the Multitron 2 incubator for 5 days. On the 5th day, the subcultivation of the culture was performed for the isolation of protoplasts.

2.12.2 Protoplast isolation

Cellulase solution

11.7 U/ml Cellulase Onozuka R10 (Serva, Heidelberg)

0.25 % (w/v) Mazerozyme R10 (Serva)

0.24 M $CaCl_2$

Floating medium

Gamborg's B5 medium (Sigma-Aldrich)

0.4 M sucrose

1 mg/l 2.4D pH 5.7

Cells from a 5 day-old 40 ml At7 culture (2.12.1) were collected by centrifugation (5 min, 500 *x g*, RT). The pellet was resuspended in 30 ml 0.24 M $CaCl_2$ and centrifuged again. After one repetition of the $CaCl_2$ washing step, the cells were resuspended in enzyme solution and incubated overnight at 26 °C on a shaker (20 rpm). Protoplasts sediment from a centrifugation at 250 *x g*, RT for 5min was washed once with 0.24 M $CaCl_2$, and resuspended in 15 ml floating medium. The suspension was centrifuged at 138 *x g*, RT for 8 min and the floating protoplasts were transferred in a new tube. The floating step was repeated once. Finally, the protoplasts were collected and used immediately for transfection experiments.

2.12.3 Plasmid DNA extraction for transfection analysis

Plasmid DNAs for transfection experiments were prepared from the dam- and dcm-methylation-deficient *E. coli* strain GM2163 (Table 2). The plasmid DNA was extracted by column chromatography using a JET star plasmid purification system (GENOMED, Loehne). The extraction method was according to the instruction manual.

2.12.4 Transfection of *A. thaliana*-protoplasts

In the transfection experiments, a total of 25 µg of plasmid DNA (2.12.3), consisting of 10 µg of a reporter plasmid (GH31 or GH32), 5 µg of a standardization plasmid (pBT10-*UBI*-LUCm3) and 10 µg of the substitution plasmid (pBT4-ΔLUC) (Table 4) was transfected into 200 µl protoplast suspension (2.12.2). Reporter plasmids pBT10-ΔGUS and pBT10-ΔLUC were used as negative controls for GUS and LUC analysis and pBT4-35SGUS were used for positive controls of the GUS experiments. Then, 200 µl of sterilized PEG solution (20 % PEG 6000, 100 mM $Ca(NO_3)_2$, 450 mM Mannitol, adjust final pH to 9 with KOH) was added to plasmid protoplast mixture and incubated at RT for 15 min. The reaction was stopped by adding 5 ml of 275 mM $Ca(NO_3)_2$ solution (pH 6). The mixture was centrifuged at 138 *x g*, RT for 8 min and the floating protoplasts were re-suspended in 7 ml of B5 medium. The B5 suspended protoplasts were incubated for 20 h at 26 °C in the dark. After the incubation, the transfected protoplasts were mixed with 20 ml of $CaCl_2$ and centrifuged at 1,349 *x g*, RT for 10 min. The supernatant was carefully removed by leaving about 1 ml solution. The protoplasts were re-suspended in the remaining solution, transferred to a 1.5 ml Eppendorf tube and again centrifuged at 1,349 *x g*, RT for 10 min. Only the supernatants were instantly frozen in liquid nitrogen and stored at – 80 °C

2.12.5 Protein extraction

Protein extraction buffer

100 mM KH_2PO_4, pH 7.5

1 mM DTT

For analysis of Luciferase (LUC) and β-Glucuronidase (GUS) enzyme activities, proteins were extracted from the transfected protoplasts (2.12.4). 800 µl of protein extraction buffer was added to the frozen protoplast (2.12.4). The mixture was vortexed for 30sec at maximum speed and centrifuged at 4 °C, 16,100 *x g* for 10 min. The supernatant was used for determination of GUS and LUC activity as well as for determination of protein concentrations by Bradford.

2.12.6 Bradford assay

Measurement of protein was performed according to the Bradford assay (Bradford, 1976). 10 µl of protein extract (2.12.5) was mixed with 1:5 diluted BioRad assay dye (BioRad) and the amount of protein was measured by the determination of the absorption at 595 nm after 20 min incubation using a FluoSTAR Optima plate reader. The OD_{595} of the Bovine serum albumin (BSA) standard solutions was used for making a standard curve used to calculate the amount of protein in the sample.

2.12.7 Measurement of luciferase activity

LUC-solution

20 mM Tricine
2.67 mM $MgSO_4$
0.1 mM EDTA
33.3 mM DTT
270 µM CoA
470 µM D-Luciferin
530 µM ATP

For determining the luciferase (LUC)-activity**,** 10µl of the protein extract (2.12.5) was mixed with 100 µl LUC-solution. For quantification of the LUC-activity, the photons emitted over an interval of 10 sec were immediately quantified by light emission at 562 nm with a MiniLum luminometer (Bioscan, Washington D.C., USA). Each well was read in triplicate and the mean values were used as the average Relative Light Units (RLU). The specific LUC activity was determined as RLU per µg protein per sec.

2.12.8 Measurement of β-glucuronidase activity

The activity of β-glucuronidase (GUS) was measured using 4-Methyl-Umbelliferyl β-D-Glucuronide (4-MUG) as a substrate. Upon hydrolysis by GUS, the fluorochrome 4-Methyl-Umbelliferone (4-MU) is produced. The velocity of the reaction was calculated by measuring the fluorescence emitted by 4-MU in every 20 min intervals in the time period of 60 min with an analysis temperature of 4 °C

and excitation at 355 nm and emission of 460 nm light using the FluoSTAR Optima plate reader (BMG LABTECH, Jena). The specific GUS activity is given in pMol MU per µg protein (2.12.6) per min. Standardized specific GUS activity (GUS') was calculated by multiplication of the specific GUS activity value with a correction factor calculated from the ratio between the mean value of specific LUC activity from the complete set of experiments and specific LUC activity of the given sample.

3 Results

3.1 Analysis of flavonoid derivatives in *A. thaliana*

This chapter focuses on the identification and characterization of flavonoid derivatives that accumulate in qualitative or quantitative manner in twelve different *A. thaliana* wildtype accessions that had been compared previously and show genetic variation throughout the genome (Schmid *et al.*, 2003; Torjek *et al.*, 2003). The data was used to determine candidate genes that control the level of flavonoid biosynthesis on the basis of the *A. thaliana* genome annotation and knowledge of the biochemistry of the respective metabolites identified in this section. So far no quantitative trait locus (QTL) involved in the accumulation of flavonoids in *A. thaliana* has been identified. QTL mapping requires a mapping population derived from parental lines differing in measurable phenotypes, for instance in the accumulation of secondary metabolites. Thus, finding flavonoid accumulation differences among the accessions is essential for starting the QTL analysis. First, the feasibility of metabolite assay methods was tested by comparing flavonoid profiles between wildtype accessions (WTs) and *transparent testa* (*tt*) mutants. In order to determine metabolite differences between the individual lines high performance liquid chromatography (HPLC) and high performance thin layer chromatography (HPTLC) were used. For HPLC experiments, genistein was used as an internal standard (IS) for normalization of the peaks on the chromatogram and for quantification of flavonoids among samples. It was chosen as an IS for the quantification of flavonoid derivatives because its occupancy is limited to leguminosae (Liu *et al.*, 2002). Quantitative and structural analyses of flavonoid derivatives were performed with a HPLC - photo diode array (HPLC - PDA) analysis coupled with an electrospray ionization/mass spectroscopy (ESI/MS). Extracts were separated by HPLC using a C_{18} column. Flow rate and composition of the gradient was optimized to obtain a good separation and maximum stability of the electrospray. A flow rate of 0.2 ml min^{-1} using a 2 mm column and a linear gradient starting from 5% (v/v) to 25% (v/v) acetonitrile in 30 min were found to be suitable for flavonoid analysis.

HPTLC analysis was chosen to screen large numbers of seedling extracts to find differences in flavonoid accumulation. A diphenylboric acid-2-aminoethyl ester (DPBA) known to visualize phenolic precursors (Sheahan and Rechnitz,

1992) was used to visualize different HPTLC-separated flavonoid derivatives in *A. thaliana* seedlings. The stain can be excited by ultra violet (UV) light at a wavelength of 312 nm, and the different flavonoids can be discriminated by the difference of their fluorescence and their retention factors (Rf). Flavonoid derivatives such as kaempferol, quercetin and sinapate derivatives are known to fluoresce in green, orange or light blue, respectively (summarized in Wagner, 1996). Therefore, metabolites that showed green, orange or light blue fluorescence were identified as kaempferol (Km), quercetin (Qr), or sinapate (Sn) derivatives, respectively.

In order to assess whether the analytical methods were sufficient for the detection of flavonoid derivatives in *A. thaliana*, the methods were tested using flavonoid deficient mutants for identification of the accumulation differences. The profiles of the flavonol derivatives between HPTLC and HPLC-PDA chromatograms were compared. *A. thaliana tt4* plants, which are deficient in the production of flavonoids due to a mutation in the *CHALCONE SYNTHASE* (*CHS*) gene, accumulated three sinapate derivatives with blue fluorescence which were separated on the HPTLC plate. The three major sinapate derivatives (Retention factor (Rf) 0.45, 0.58 and 0.89) might correspond to the peaks at retention times (RT) 24.5, 25.5, and 38.5 min in the HPLC chromatogram (Figure 15). *A. thaliana tt7* mutant, which is characterized by a deficiency in the production of quercetin derivatives due to a mutation in the *FLAVONOID 3'-HYDROXYLASE* (*F3'H*) gene, showed an accumulation of kaempferol and sinapate derivatives. The kaempferol derivatives accumulating on the HPTLC plate (Rf 0.23, 0.35, 0.51, 0.66) might correspond to the HPLC peaks at RT 29.7, 33.7, 36.5, and 40.5 min. By comparing the peaks between WT and *tt7* chromatograms, five quercetin derivative peaks at RT 27.5, 30.7, 33.0, 37.2, and 44.2 min were obvious in HPLC but only three quercetin bands were identified in HPTLC (Rf 0.18, 0.29, and 0.45; Figure 15). These results showed, that the methods applied were suitable to analyze the flavonoid accumulation. Thus the twelve *A. thaliana* wildtype accessions were analyzed in this way to find quantitative differences.

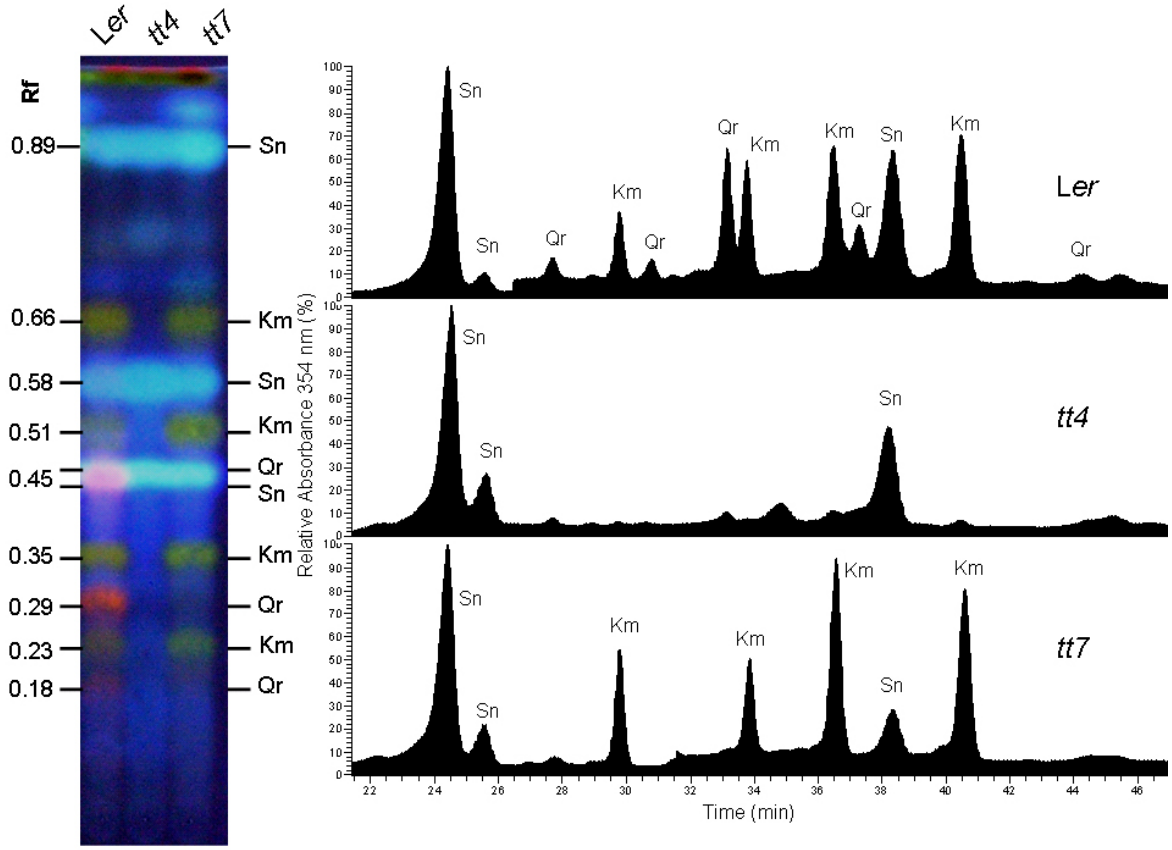


Figure 15. Comparison of flavonoid profiles in the seedlings of L*er* wildtype and *transparent testa* mutants.
Methanolic extracts from seedlings of L*er* and the flavonoid deficient mutants *tt4* and *tt7* were compared using high performance thin layer chromatography (HPTLC) and high performance liquid chromatography (HPLC)-photodiode array (PDA) chromatograms. The flavonoid peaks and bands are labeled as Km – kaempferol derivatives, Qr - quercetin derivatives, and Sn -sinapate derivatives. Color key: green – kaempferol derivative (Km), orange – quercetin derivative (Qr), faint blue – sinapate derivative (Sn), pink – overlapping quercetin and sinapate derivatives, dark red – chlorophyll. Retention factors (Rfs) are listed next to the compounds.

3.1.1 Identification of recombinant inbred line (RIL) parental lines showing differences in flavonoid accumulation

In order to identify a RIL population for QTL analysis, the differences in the flavonoid accumulation from seedling of twelve *A. thaliana* wildtype accessions were compared using HPTLC analysis (Figure 16). The HPTLC plates, loaded with defined sample volumes representing equal quantities of plant material, displayed qualitative and quantitative differences in kaempferol and quercetin derivatives among the accessions (Figure 16). The flavonoid bands were labeled

with numbers from the bottom to the top of the plate. The uppermost orange band was named quercetin derivative 4 (Qr4) and the lower most orange and green bands were named quercetin derivative 1 (Qr1) and kaempferol derivative 1 (Km1). The Km1 and Qr1 compounds displayed Mendelian inheritance patterns. They accumulated only in CS22491, C24, Cvi, Gue-0, L*er*, Lz-0, and Ws-0, but not in Ak-1, Col-0, Ei-2, Nd-0, and Wei-0. In contrast, the Qr4 trait showed quantitative difference among the accessions. The L*er* x Col RIL population (Lister and Dean, 1993) was chosen for the following analysis of the trait loci because the quantitative and qualitative differences could be found between the parental lines and a high density genetic map and a large number of mapping markers were available from NASC. The next step was the identification of the flavonoid derivatives for the determination of the genetic basis underlying the differences between two accessions found on HPLC chromatograms and HPTLC plates.

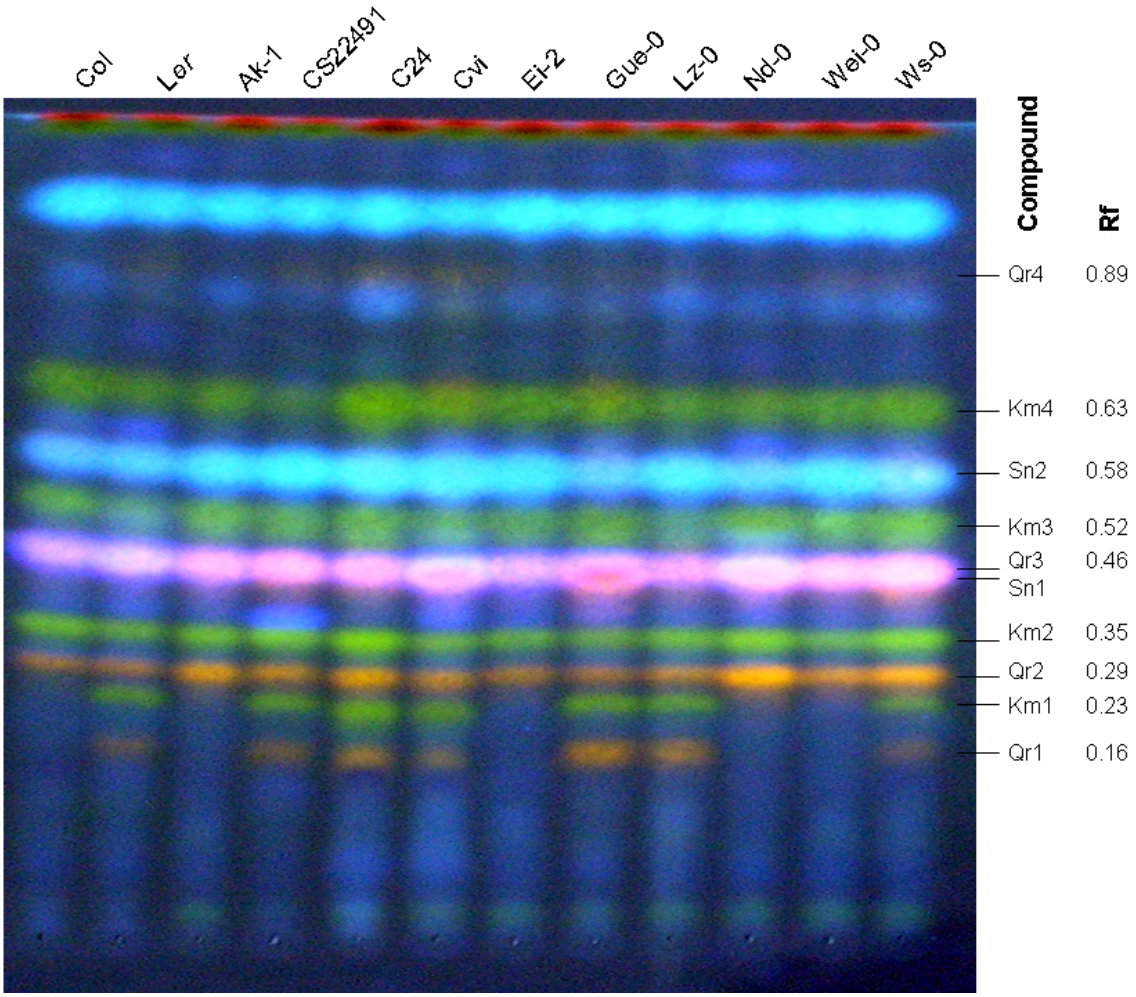


Figure 16. Comparison of flavonoid composition in seedlings of different *A. thaliana* wildtype accessions.
Methanolic seedling extracts from different *A. thaliana* wildtype accessions were analyzed by HPTLC. The flavonoids were detected by DPBA staining under UV illumination. Bands are named based on the color of their fluorescence and numbered from the bottom of the plate. Retention factors (Rfs) are listed next to the compounds.

3.1.2 Integration of HPTLC to HPLC chromatogram

While HPTLC allows a high throughput, HPLC offers the possibility to obtain structural information of the compounds by comparing the chromatography to mass spectroscopy. For this reason the identification of compounds by both methods need to be linked. The chromatogram profiles between HPTLC and HPLC were linked by a preparative HPTLC method. Individual flavonoids separated by thin layer chromatography were extracted from the HPTLC plates and subsequently analyzed by HPLC. A tandem MS approach was applied utilizing a HPLC-PDA-ESI/MS to identify aglycones and glycosylation

patterns. Finally, all flavonoid derivatives could be identified partially. A total of twelve compounds was detected and analyzed in more detail (Table 9 and Figure 17). The HPTLC purified band Km3 contained not only kaempferol di-glycoside (Km3) but also isorhamnetin glycoside (IR). The Qr3 band was identified to overlap with two sinapate glycosides called Sn1 and Sn1-2. By the linkage of the two applied methods, the flavonoid profile of *A. thaliana* seedlings could be identified. The L*er* seedling extract was used as a metabolite marker for further analyses to compare different samples.

Table 9. Analyzed flavonoid derivatives from *A. thaliana* seedlings of the accession L*er*.
List of flavonoid derivatives with retention factors (Rf) and retention times (RT) from HPTLC and HPLC analyses, respectively. *Overlapping quercetin and sinapate derivatives. Presence (+) and absence (-) of compounds in L*er* and Col accessions are listed.

	HPTLC	HPLC			
Compound	Rf	RT (min)	Color	L*er*	Col
Qr4	0.89	44.2	Orange	+	-
Km4	0.63	40.5	Green	+	+
Sn2	0.58	38.5	Light blue	+	+
Km3	0.52	36.5	Green	+	+
IR	0.52	37.2	Green	+	+
Qr3	0.46	33.3	*Pink	+	+
Sn1	0.46	24.5	*Pink	+	+
Sn1-2	0.46	25.5	*Pink	+	+
Km2	0.35	29.7	Green	+	+
Qr2	0.29	27.5	Orange	+	+
Km1	0.23	33.7	Green	+	-
Qr1	0.16	30.7	Orange	+	-

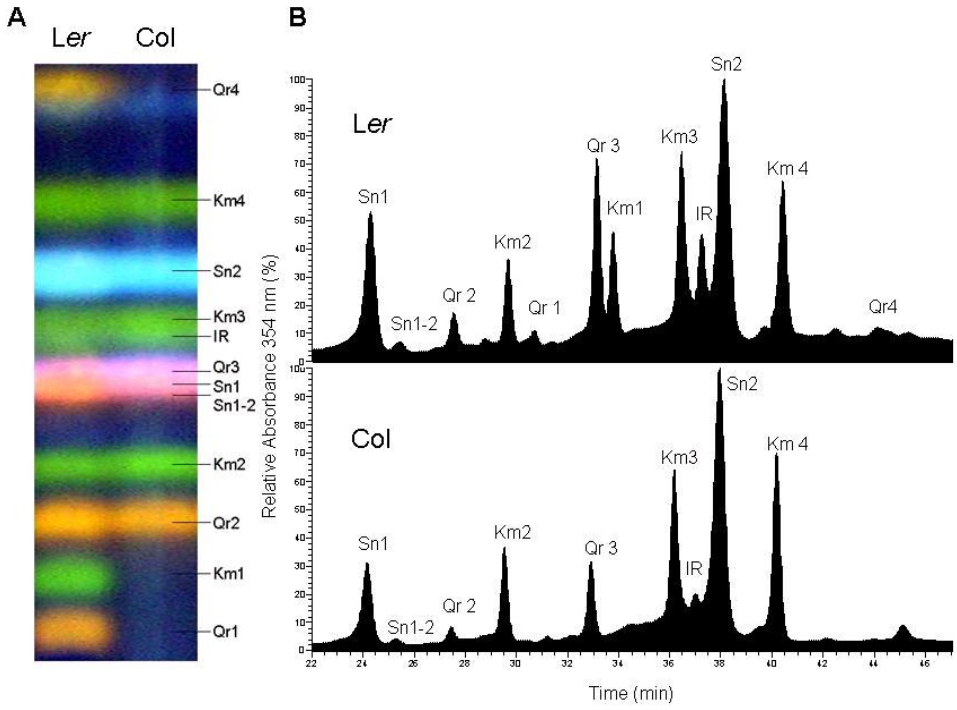


Figure 17. Flavonol derivatives identified in HPTLC and HPLC-PDA chromatograms.
(A) DPBA stained HPTLC plate with methanolic extracts from L*er* and Col seedlings. Isorhamnetin derivatives (IR) gave green fluorescence.
(B) HPLC-PDA relative observance chromatograms of methanolic extracts from L*er* and Col seedlings measured at UV 354 nm. Flavonoid derivatives purified from HPTLC plates were analyzed using HPLC. The corresponding peaks of the derivatives in the chromatograms were identified by co-chromatography analysis.

3.1.3 Structure analysis of flavonol derivatives

The HPTLC-purified flavonoid derivatives were used to identify the conjugated sugar groups. The identification of sugars bound to the flavonoid derivatives based only by their mass to charge (*m/z*) ratio was not possible, since this only gives the amount of carbon atoms present in the sugar. Therefore, the conjugated sugars of the flavonoid derivatives were released from the aglycone by hydrolysis. The hydrolyzed samples were further derivatized and analyzed by GC-MS. By comparing the *m/z* ratios of the fragmentation spectra and the retention times to the ones from the reference sugars, the sugar motifs could be identified as glucose and rhamnose (Figure 18).

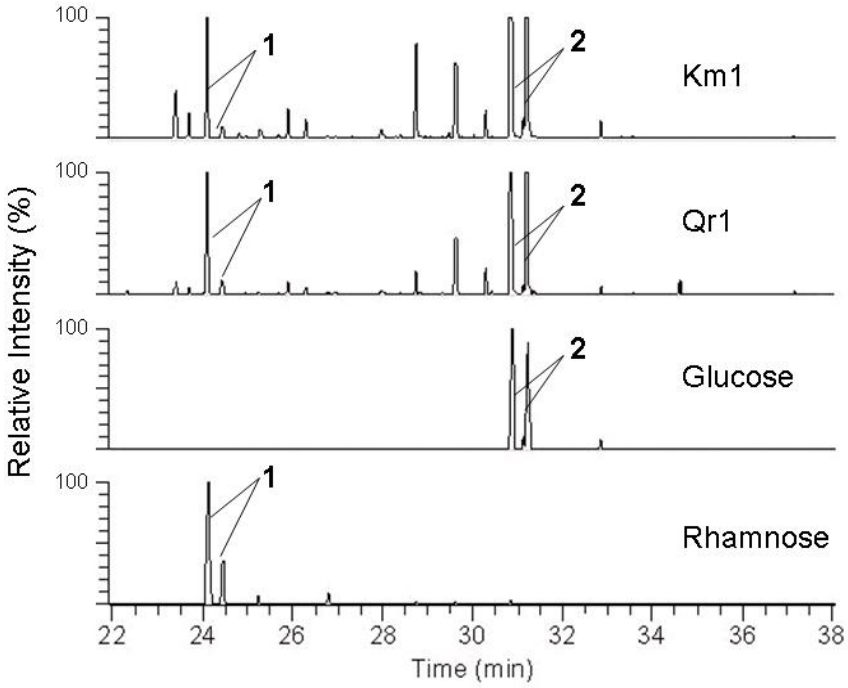


Figure 18. Identification of the conjugated sugars by comparison of gas chromatography – mass spectrometry (GC-MS) relative intensity chromatograms from hydrolyzed Km1 and Qr1 compounds and standard sugars.
Hydrolyzed HPTLC-purified Km1 and Qr1 compounds were fragmented and analyzed by GC-MS. Peak profiles between samples and the sugar standards were compared for identification of the conjugate sugars. Peak identification: 1 – rhamnose, 2 – glucose.

The GC/MS data revealed that the deoxyhexose at *m/z* 146 and the hexose at 162 *m/z* correspond to rhamnose (rha) and glucose (glc) respectively. Also the retention times were identical to the reference sugars. After that the identity of the attached sugars had been determined, tandem MS was used to confirm the GC/MS results and furthermore to gain information on the amount of each sugar attached to the flavonol. Figure 19 depicts the tandem MS analysis of the Km1 and Qr1 compounds. The full range MS scan revealed the Km1 (*m/z* = 779 $[Km1 + Na]^+$) and Qr1 (*m/z* = 795 $[Qr1 + Na]^+$) ions. These were independently extracted and dissociated by MS^2. The resulting fragment ions with the *m/z* ratios of 633 $[Km1 + Na – rha]^+$ and 649 $[Qr1 + Na – rha]^+$ indicate the dissociation of a rhamnose. The MS^2 ions were repeatedly extracted and fragmented to yield an MS^3 spectrum, which reveals the dissociation of glucose from the flavonols indicated by m/z ratios of 471 $[Km1 + Na – glc - rha]^+$ and 487 $[Qr1 + Na – glc - rha]^+$ as well as the mass of two glucose molecules bound to each other at

m/z 347. The fragmentation series clearly shows that one rhamnose and two glucoses bound to each other are attached to the flavonols. The aglycones stripped from the sugars should therefore have calculated *m/z* ratios of *m/z* = 287 and *m/z* = 303, representing the kaempferol and quercetin aglycones, respectively. However, these ions could not be detected with the settings of the MS instrument. Finally, Km1 and Qr1 were putatively identified as the known compounds kaempferol 3-*O*-[(glycosyl) glucoside]-7-*O*-rhamnoside (kaempferol 3-*O*-gentinobioside 7-*O*-rhamnoside) and quercetin 3-*O*-[(glucosyl) glucoside]-7-*O*-rhamnoside (quercetin 3-*O*-gentinobioside 7-*O*-rhamnoside). All other flavonol derivatives were also putatively identified from a comprehensive analysis of mass fragmentation profiles obtained from tandem MS analysis and comparison to the values from literature (references listed in Table 10).

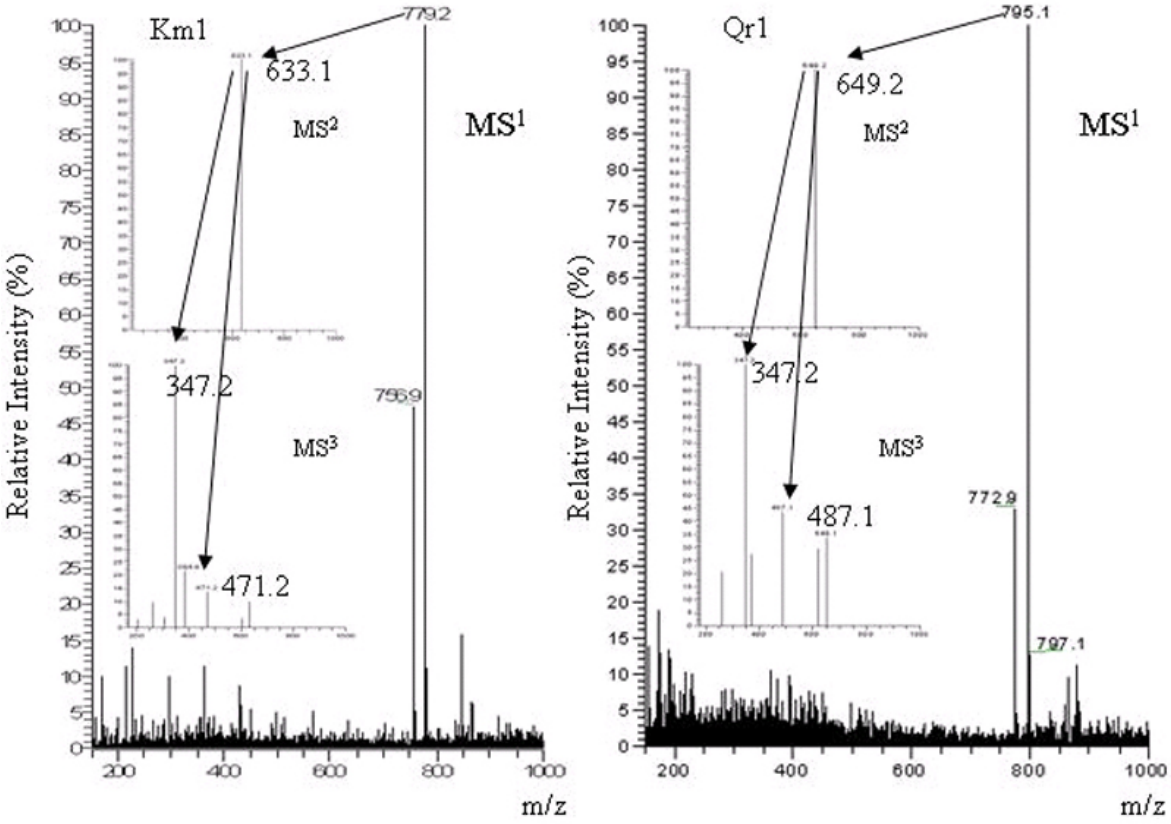


Figure 19. Tandem MS analysis of Km1 and Qr1 compounds.
Relative intensity MS spectra generated from the dissociation of Km1 and Qr1 by tandem MS analysis. Loss of *m/z* = 146 and 162 from main ions indicate dissociation of rhamnose and glucose moieties. Loss of *m/z* = 286 (633 – 347) and 302 (649 - 347) indicate the dissociation of kaempferol and quercetin aglycones from the MS^2 ions.

Table 10. Flavonoid derivatives identified in *A. thaliana* seedlings.
Flavonol derivatives were putatively identified by comparing tandem MS and GC/MS analysis data to values from literatures. *Detected in MS and/or tandem MS data. Peak names correspond to the ones listed in Figure 17. glucose – glc, rhamnose – rha.

Peak Name	ESI-MS (*m/z*)	*Fragment (*m/z*)	Compound Identified	Reference
Sn1	387 $[M+H]^+$ 409 $[M+Na]^+$	n.a.	sinapoyl glucose	(Le Gall *et al.*, 2005)
Sn1-2	773 $[M+H]^+$ 794 $[M+Na]^+$	n.a.	sinapate derivative (conjugate of peak 1 compound)	n/a
Qr2	757 $[M+H]^+$ 779 $[M+Na]^+$	303 $[qr+H]^+$ 449 $[qr+rha+H]^+$ 611 $[qr+rha+glc+H]^+$	quercetin [rhamnoside-glucoside] rhamnoside	(Stobiecki *et al.*, 2006)
Km2	741 $[M+H]^+$ 763 $[M+Na]^+$	287 $[km+H]^+$ 433 $[km+rha+H]^+$ 595 $[km+rha+glc+H]^+$	kaempferol [rhamnoside-glucoside] rhamnoside	(Bloor and Abrahams, 2002)
Qr1	772 $[M+H]^+$ 795 $[M+Na]^+$	303 $[qr+H]^+$ 347 $[glc+glc+Na]^+$ 487 $[qr+glc+Na]^+$ 649 $[qr+glc+glc+Na]^+$	quercetin [glucosyl-glucoside]-rhamnoside	(Stobiecki *et al.*, 2006)
Qr3	611 $[M+H]^+$ 633 $[M+Na]^+$	303 $[qr+H]^+$ 449 $[qr+rha+H]^+$	quercetin glucoside - rhamnoside	(Bloor and Abrahams, 2002)
Km1	756 $[M+H]^+$ 779 $[M+Na]^+$	287 $[km+H]^+$ 347 $[glc+glc+Na]^+$ 471 $[km+glc+Na]^+$ 633 $[km+glc+glc+Na]^+$	kaempferol [glycosyl-glucoside] rhamnoside	(Veit and Pauli, 1999)
Km3	595 $[M+H]^+$ 617 $[M+Na]^+$	287 $[km+H]^+$ 433 $[km+rha+H]^+$	kaempferol glucoside rhamnoside	(Bloor and Abrahams, 2002)
IR	625 $[M+H]^+$ 647 $[M+Na]^+$	339 $[IR+Na]^+$ 501 $[IR+glc+Na]^+$	isorhamnetin glucoside rhamnoside	(Routaboul *et al.*, 2006)
Sn2	681 $[M+H]^+$ 703 $[M+Na]^+$	n.a.	sinapate derivative	n/a
Km4	579 $[M+H]^+$ 601 $[M+Na]^+$	287 $[km+H]^+$ 433 $[km+rha+H]^+$	kaempferol rhamnoside rhamnoside	(Bloor and Abrahams, 2002)
Qr4	449 $[M+H]^+$ 471 $[M+Na]^+$	303 $[qr+H]^+$ 449 $[qr+rha+H]^+$	quercetin 3-*O*-rhamnoside	(Routaboul *et al.*, 2006)

3.1.4 Summary of metabolite analysis in *A. thaliana* seedlings

The flavonol derivatives accumulating in *A. thaliana* seedlings were identified and structurally characterized. Twelve flavonoids were found in mono-, di-, and tri-glycosylated forms. Glucose and rhamnose were found to be the only sugars attached to the flavonols. Flavonol accumulation traits among the twelve different *A. thaliana* wildtype accessions were found. The Km1 and Qr1 compounds (flavonols 3-O-gentinobioside 7-*O*-rhamnoside) accumulated in qualitative manner among different accessions, while quercitrin (Qr4) accumulated in quantitative manner among the accessions. The qualitative and quantitative trait loci and the corresponding alleles will be deal within the next two chapters.

3.2 Identification of loci influencing quercetin 3-*O*-rhamnoside accumulation in *Arabidopsis thaliana* wildtype accessions

A quantitative difference of quercitrin (Qr4) was found between the accessions Col and L*er* (Figure 20). L*er* seedlings extracts contained about seven times more Qr4 compound than extracts from Col seedlings. Therefore it was assumed that this phenotype is due to different levels of gene expression or enzymatic activity. To locate the loci influencing the Qr4 compound accumulation differences in the *A. thaliana* genome, quantitative Qr4 data from 92 lines of the Lister and Dean L*er* x Col RIL population (Lister and Dean, 1993) were taken for QTL mapping using the HPLC-PDA method. Actually the population consisted of 100 lines but eight lines failed to germinate and hence could not be analyzed.

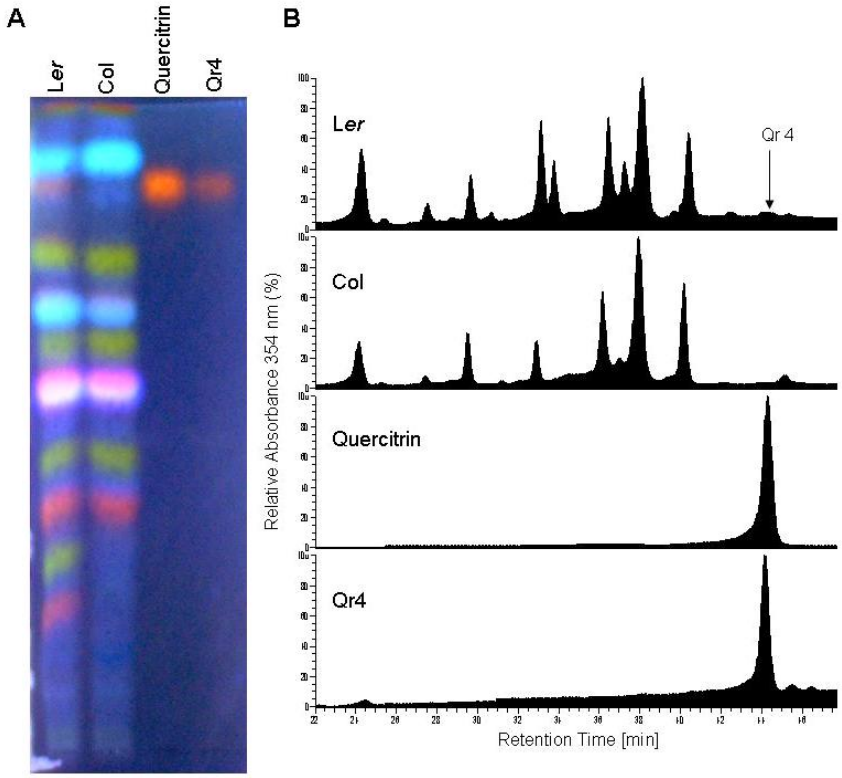


Figure 20. Quercitrin accumulation differences identified between the *A. thaliana* wildtype accessions Columbia and Landsberg *erecta*
(A) HPTLC chromatogram showing HPTLC-purified Qr4 compared to a quercitrin standard. Methanolic extracts from 7 DAG *A. thaliana* seedlings from Col and L*er* accessions were loaded as reference.
(B) The extracts used in HPTLC analysis were also characterized using HPLC-PDA. Relative absorbance chromatograms measured at UV 354 nm are given. The arrow indicates the position of the quercitrin peak.

3.2.1 Phenotyping of the Lister and Dean L*er* x Col RIL population

The L*er* x Col RIL population was used to determine quantitative data for the Qr4 trait. The amount of Qr4 per 100 µg of dried seedlings was measured from 92 lines from the L*er* x Col RIL population. The frequency distribution graph of Qr4 in the RIL population is shown in Figure 21. The mean value of the amount of quercitrin in the population was 24.3 ng per 100 µg of dry weight (DW). The average amounts of quercitrin in Col and L*er* accessions measured from at least three independent extracts were 4.9 (SD ± 6.33) and 30.3 ng (SD ± 2.93) per 100 µg DW. Four lines were found to contain more than 100 ng Qr4. 42 lines contained the Qr4 compounds in the range of 0 to 10 ng corresponding to the Col Qr4 mean value. 50 lines had more than 10 ng of Qr4 per 100 µg DW.

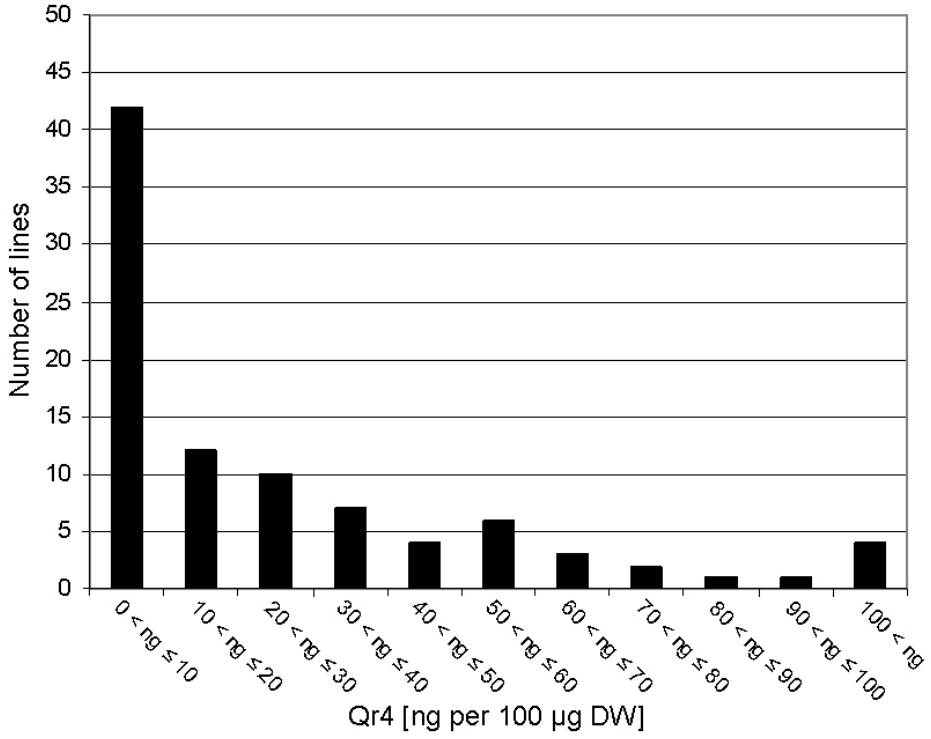


Figure 21. Frequency distribution graph of Qr4 in seedlings of the Lister and Dean L*er* x Col RIL population.
The ordinate and abscissa indicate the number of RILs and the amount of Qr4 [µg/100 µg DW]. The lines were grouped to different Qr4 weight categories depending on the amount of Qr4 compound measured. The mean of Qr4 in the population was 24.3 ng/100 µg DW.

3.2.2 QTL mapping of the Qr4 trait

QTL mapping of the Qr4 trait was performed by associating the Qr4 quantitative and genotype data from the L*er* x Col RIL using WinQTLCart software (Wang, 2005). Data from genotyping analyses of 278 polymorphic markers distributed among the five *A. thaliana* chromosomes was obtained from the European Arabidopsis Stock Centre NASC (http://arabidopsis.info/; 2.9.1). The data was used to generate genotype source data for QTL mapping. The average interval between flanking markers was 2.33 cM. The genome wide optimal threshold value for the trait was automatically calculated to 2.5 by permutation. By using the composite interval mapping analysis function of the software, the logarithm of the odds favoring linkage (LOD) curve above the calculated threshold of 2.5 was detected on the lower arm of chromosome 5, ≈ 99 cM with LOD_{max} of 5.11 between the two flanking markers SGCSNP219 at 96.83 cM (18,763,754 bp) and SGCSNP162 at 104.08 cM (19,881,822 bp) (Figure 22). Thus, the annotated genes in the interval were investigated.

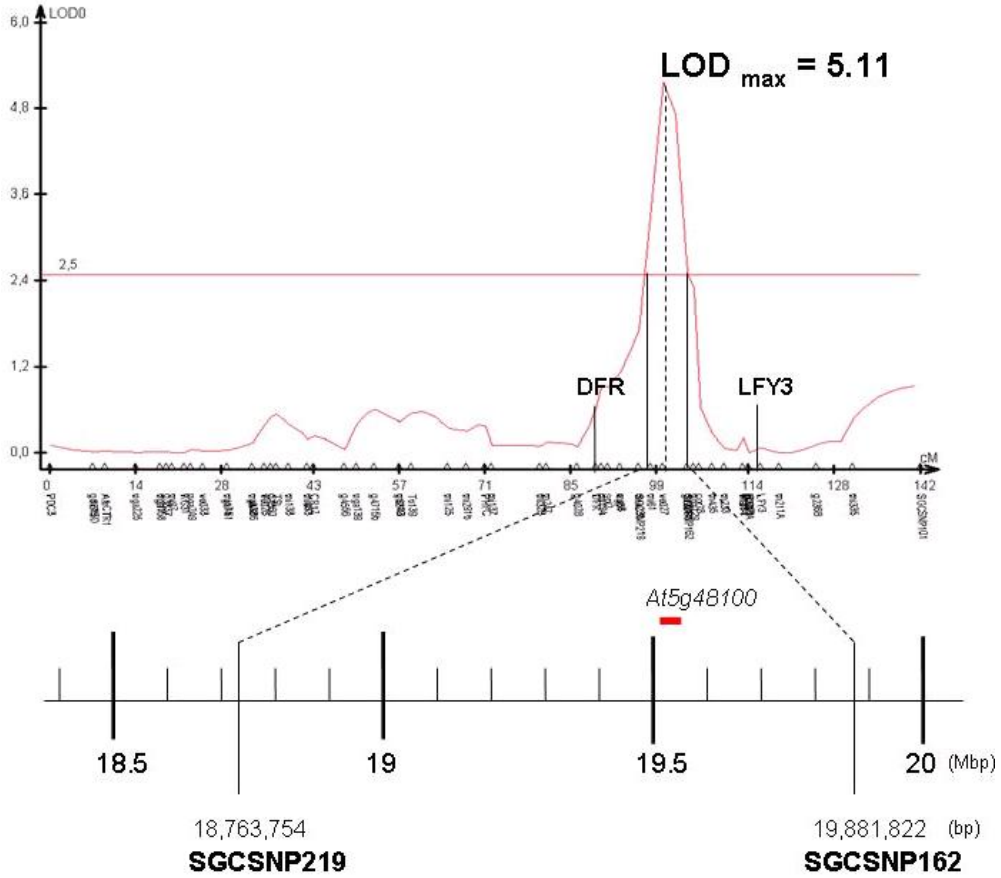


Figure 22. QTL likelihood map of the Qr4 trait located on chromosome 5.
The genetic map of the Arabidopsis linkage group is shown in the abscissa. The logarithm of the odds favoring linkage (LOD) score is listed on the ordinate. The trait was located between the flanking markers SGVSNP219 and SGCSNP162 which are positioned at 18,763,754 bp and 19,881,822 bp of Col chromosome 5 genomic sequence. LOD maximum score (LOD_{max}) of the trait was listed as 5.11 at 99 cM. The position of *At5g48100* is shown at 19.5 Mb.

3.2.3 Identification of the candidate gene

In the identified interval, 305 annotated genes (*At5g46240* to *At5g49000*) are located. The glycosyltransferase UGT78D1 (At1g30530) catalyses the transfer of rhamnose from uridinediphospho(UDP)-rhamnose to the 3-OH position of flavonols to produce quercitrin and kaempferol 3-O-rhamnoside (Jones *et al.*, 2003). Also, UGT73C6 (At1g36790) transfers glucose from UDP-glucose to the 7-OH position of flavonol 3-*O*-rhamnoside to form flavonol 3-*O*-rhamnoside 7-*O*-glucoside (Jones *et al.*, 2003). Muzac *et al.* (2000) reported the transfer of a methyl group to the 3'-OH position of quercetin by the flavonol 3'-*O*-methyltransferase AtOMT1 (At5g54160). Thus the enzyme having Col Qr4 allele

might converts quercitrin to isorhamnetin glucoside rhamnoside (IR), or quercetin glucoside rhamnoside (Qr3) more efficiently than the enzyme with L*er* allele. Neither a gene encoding an UDP-glycosyltransferase nor an *O*-methyltransferase was annotated in this region. However, the *TRANSPARENT TESTA 10* gene, encoding a laccase protein which has been described to produce quercitrin dimers from monomers (Pourcel *et al.*, 2007), was located within the interval. The dimerization of quercitrin led to reduced levels of the quercitrin monomers in the seed coat. The mutants show a *tt* phenotype and accumulate higher amounts of quercitrin in the seeds than WTs (Pourcel *et al.*, 2005). In order to measure differences in Qr4 accumulation between the mutants and wildtypes, *tt10-4* (Col-8 background) and *tt10-1* (L*er* background), the genotypes and phenotypes of which had been described by Pourcel *et al.* (2005), were obtained from NASC. Moreover, an insertion line, GK_146E10, which has a T-DNA insertion in the 5' UTR region of *At5g48100* in the Col-0 accession, was obtained from GABI-Kat. Sequence analysis of this mutant revealed a T-DNA insertion at – 101 bp (upstream) from the start codon (ATG) (Figure 24). The T3 seeds of GK_146E10 showed the typical yellow seed coat phenotype of *transparent testa* mutants at harvest (Figure 23). The yellow seed color is known to tern brown after six to twelve month of storage (Pourcel *et al.*, 2005), but the phenomenon has not been observed with GK_146E10 seeds. Qr4 was measured between the mutants and the wildtype using HPLC-PDA. The amount of Qr4 in the GK_146E10 seedling extracts contained a Qr4 amount that was 43-fold higher in comparison to Col-0. In contrast to Col *tt10* mutants, L*er tt10-1* accumulated only 1.5 times more Qr4 than the L*er* WT. These data indicated that *TT10* is the locus responsible for the difference in Qr4 accumulation between the Col and L*er* seedlings.

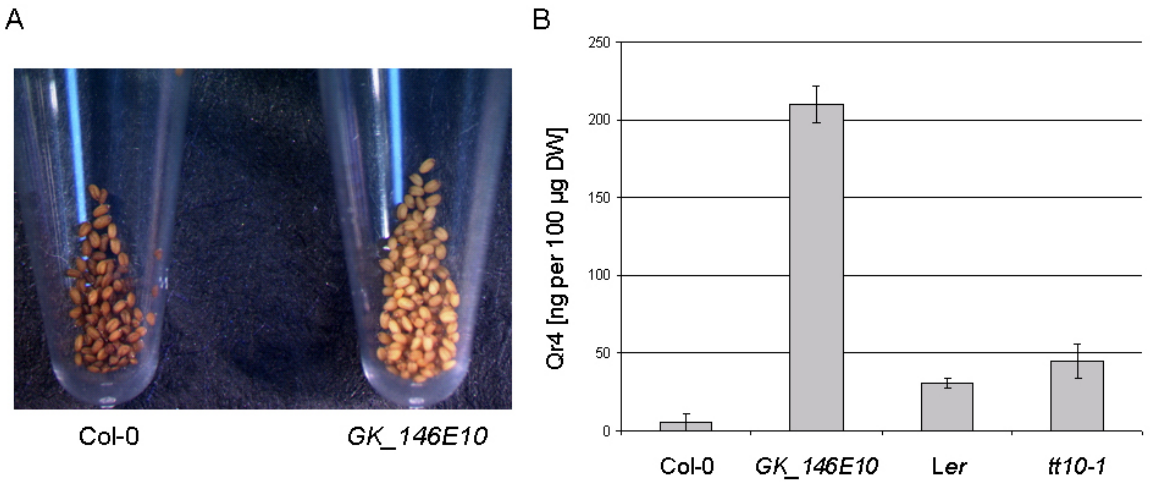


Figure 23. Phenotypical analysis of *GK_146E10* line.
(A) Comparison of the seed color. Homozygous T3 seeds from GK_146E10 showed yellow color.
(B) Measurement of Qr4 using HPLC-PDA. The amounts of Qr4 compound between WTs and corresponding mutant alleles were compared. The error bars indicate the standard deviation calculated from the values obtained from four independent measurements.

3.2.4 *TT10* sequence comparison

Since lower *TT10* expression level in L*er* as comparison to Col were reported by Pourcel *et al.* (2005), the *TT10* promoter sequences between Col and L*er* were compared. The L*er TT10* promoter sequence displayed an insertion of 61 bp and deletions of 16 and 752 bp compared to the Col *TT10* promoter sequence (Figure 24). The 2.2 Kb-long Col *TT10* (Appendix 1) and 1.5 Kb-long L*er TT10* promoter (Appendix 2) sequences, as well as the promoter sequence comparison results are listed in the appendices (Appendix 3). *In silico* analysis of *TT10* promoters were performed to identify differences in regulatory elements between Col and L*er*. PLACE, a database of plant *cis*-acting regulatory DNA elements, was used for the analysis the DNA sequences. Both promoter sequences contained *cis*-acting regulatory elements predicted to be involved in early responses to dehydration of the seed and in stress response in the root. ACGT containing *cis*-acting elements were also found in both promoter sequences. This element has been identified in plant genes regulated by diverse stresses and are known to be bonded by transcription factors and also found in the promoter of genes involved in the regulation of phenylpropanoid pathway (Foster *et al.*, 1994; Hartmann *et al.*, 1998). The Col 752 bp insertion contained a core GCC-box element, which was

also found in many pathogen-responsive genes. The L*er* promoter additionally contained a CT-rich motif also found in a 60-nucleotide region (S1) upstream of the transcription start site of the CaMV 35S RNA which is predicted to enhance gene expression. The presence of the insertion and deletion polymorphism (InDel) allowed investigating whether or not *TT10* is involved in the Qr4 accumulation differences. The InDel polymorphic marker (H379) based on the *TT10* promoter allelic difference was used to find a correlation between Qr4 accumulation and the *TT10* genotype in the RIL. After confirmation, *TT10* promoter activities between Col and L*er* were analyzed *in vitro* and in *A. thaliana* seedlings.

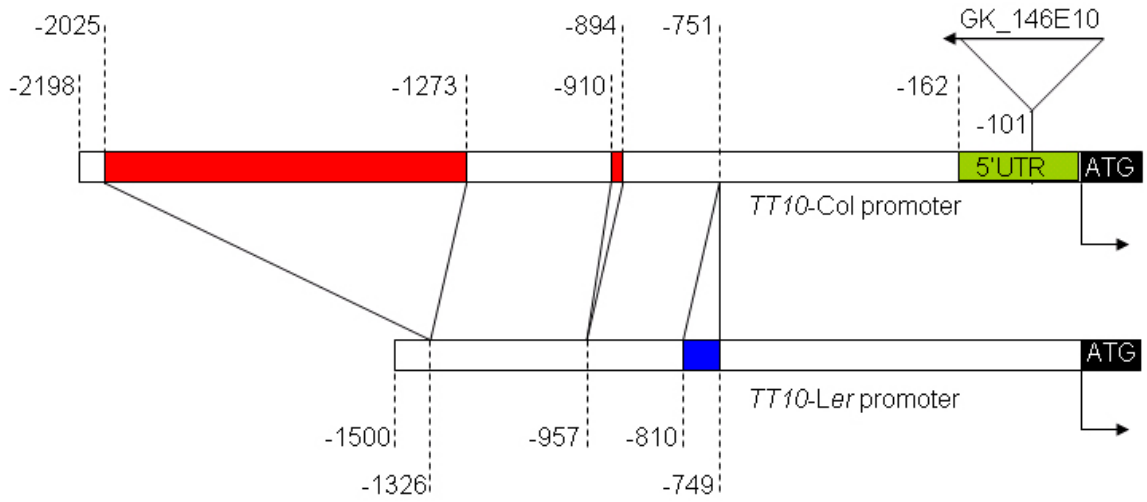


Figure 24. Comparison of *TT10* promoter sequences from Col and L*er*.
Numbers indicate positions of InDels and T-DNA insertion sites upstream from the start codons. The T-DNA insertion site in GK_146E10 is indicated with a triangle with an arrow. The arrow head indicates the orientation of the T-DNA right border. The 5'UTR identified by comparing cDNA to CDS is listed with a green box. The insertion sequence in Col compared to the L*er* promoter sequence are indicated by red boxes. The blue box indicates an additional sequence in L*er* compared to Col promoter sequence.

3.2.5 Association analysis of Qr4 accumulation with the *TT10* alleles

To investigate the allele effect on the accumulation of quercitrin, a two-sample *t*-test was performed. *TT10* genotypes from the RIL population (total of 92 out of 100 lines) were analyzed using the *TT10* promoter specific InDel marker H379 (Appendix 4). Seven lines out of 92 lines were excluded from the test because

they either did not have genotypic or phenotypic data. The genotyped lines were separated into two groups, either Col or L*er* background lines depending on their *TT10* promoter genotype. The allele effect was estimated as the mean difference between the two RIL groups carrying the Col and L*er* alleles. The average amount of Qr4 (in ng per 100µg DW) was calculated from both groups. The L*er* background lines contained 15 times more Qr4 than the lines with Col background. The *t*-score was capitulated to $t = 385$. The significance (*P*-value) was calculated from the *t*-score at a degree of freedom of 83 and had a *P*-value of less than 0.001. This implies a strong relationship between the genotype and the phenotype. The accumulation of Qr4 between the two groups were shown to be significantly different, because the *TT10* genotype strongly influences the amount of Qr4 in the population.

Table 11. Genotype performance in the L*er* x Col RIL population
Genotypes and phenotypes were collected in 85 L*er* x Col RILs consisting of 100 lines. Col (N) is the number of the lines having the Col *TT10* genotype. L*er* (N) is the number of the lines possessing the L*er* *TT10* genotype. A relative Qr4 accumulation difference was calculated between Col and L*er*. A two tailed *P*-value was calculated from *t*-test from the means of Col and L*er* Qr4 amount.
The degree of freedom (*df*) = Total N – 2. The values used for the analysis are listed in Appendix 4.

	Total (*N*)	Col (*N*)	L*er* (*N*)	Col Qr4 (ng/100µg DW)	L*er* (ng/100µg DW)	Relative Qr4 difference (%)	*t*-score *P*-value
L*er* x Col RIL	85	40	45	3.088	46.597	1508.889	$t = 385.329$ 0.001 > P

3.2.6 *In vivo TT10* promoter analysis by transfection experimets

Pourcel *et al.* reported that *TT10* in L*er* has a lower expression level than in Col (2005). To test if the promoter sequences are involved in *TT10* expression level differences, the 2 Kb and 1.37 Kb promoter fragments from Col and L*er* accessions were fused in front of the reporter gene *uidA* encoding β-glucuronidase (GUS) for the comparison of promoter activity levels (Figure 25A). The *A. thaliana* suspension culture At7 accession Columbia was used for transient expression analysis. Prior to the transfection analysis, *TT10* activity in

the cell culture was assayed by semi quantitative RT-PCR. Weak expression of *TT10* was found (Figure 25B). The promoter GUS constructs were transfected into the protoplasts derived from suspension cultures. The promoter activities were calculated by measuring hydrolyzation of 4-MUG by GUS activity. The result showed that even though the corrected specific GUS activity was low, a five times higher GUS activity was constantly observed from the Col promoter in comparison to the L*er TT10* promoter (Figure 25C).

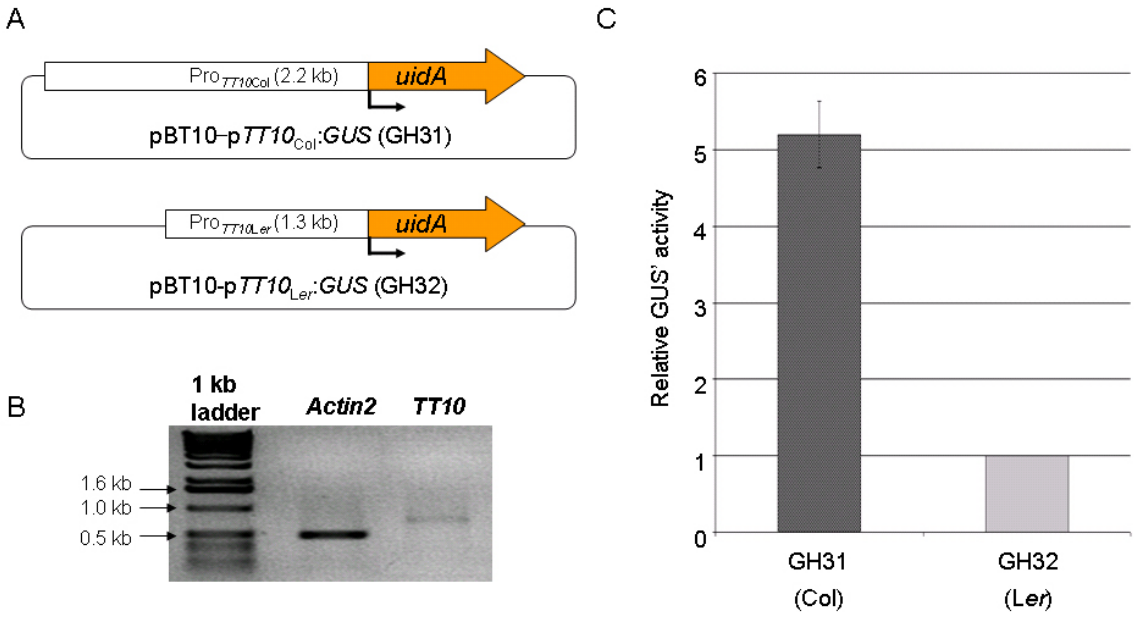


Figure 25. *TT10* promoter activity measured in At7 protoplasts.
(A) Schematic maps of promoter GUS constructs used for the transfection analysis in *A. thaliana* protoplasts.
(B) *TT10* transcript level were measured by semi quantitative RT-PCR (28 cycle) method using cDNA generated from total RNA extracted from At7 protoplasts. Actin2 was used for positive control.
(C) Specific corrected GUS' activity ratio between Col to L*er TT10* promoter activity is shown in the graph. The error bars indicate the standard error calculated from the averages from five independent experiments.

3.2.7 Analysis of *TT10* promoter activity in plant seedlings

To investigate where the *TT10* promoters is active in the seedling, they were studied using transgenic Col-0 plants transformed with promoter:GUS constructs. Two different GUS localization patterns were observed in the 7 DAG T2 transgenic seedlings grown on the selective MS agar plates. GUS activily (shown by blue staining) was found at the hypocotyl-root transition zone in all lines (Figure 26). The Col and L*er* p*TT10*:*GUS* line 1 had strong GUS activities at the root and hypocotyl-root transition zone. The Col p*TT10*:*GUS* Line3 and the L*er* p*TT10*:*GUS* Line2 seedlings also showed a very weak GUS activity. Only one seedling from the Col p*TT10*:*GUS* Line3 displayed GUS activity at the tips of the cotyledon. GUS activity was not found in the seed coats attached to the hypocotyl-root transition zone.

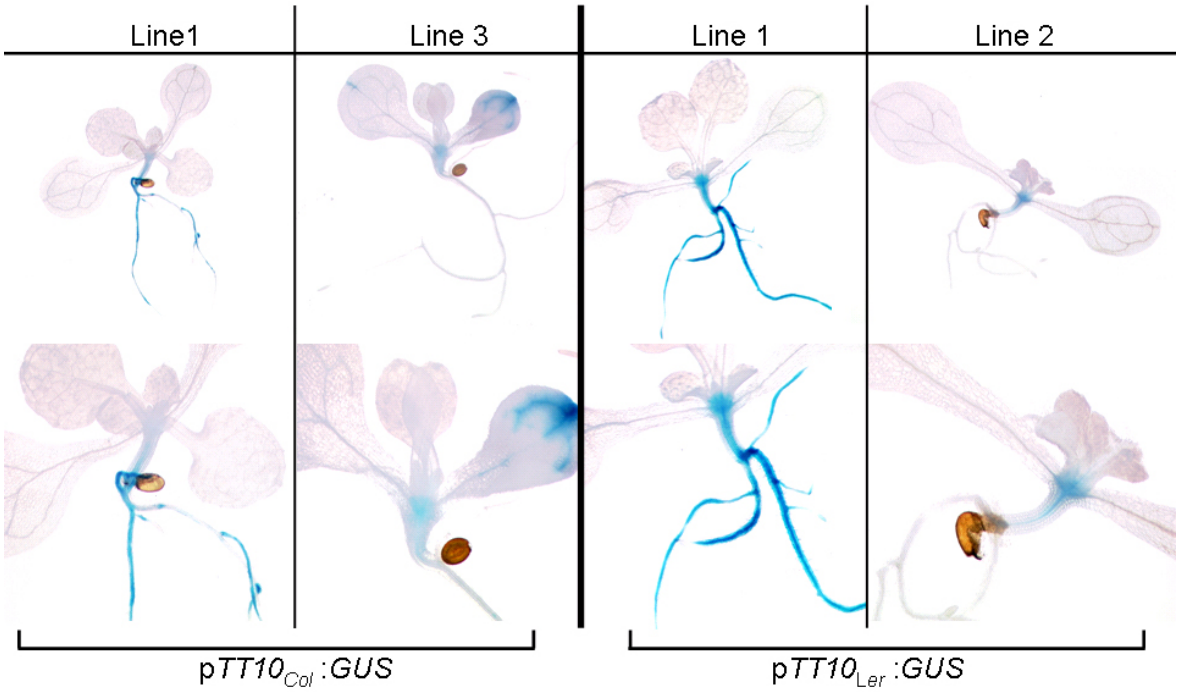


Figure 26. Expression of Col and L*er* *TT10* GUS in 7 DAG Col seedlings.
T2 transgenic 7 DAG seedlings containing the GUS reporter gene driven by *TT10* promoters isolated from Col and L*er* accessions. Col promoter GUS construct – p*TT10*$_{Col}$:*GUS*, L*er* promoter GUS construct - p*TT10*$_{Ler}$*:GUS*

3.2.8 Comparison of Qr4 accumulation between the seed coat and the seedling

Strong Qr4 accumulation in the seed coat was reported by Pourcel *et al.* (2005). Also, *cis*-acting regulatory elements involved in seed development were found in the *TT10* promoter sequences. The seedling samples used in this experiment were harvested with seed coats attached to the hypocotyl-root transition zone of the 7DAG seedlings. The phenotype observed from L*er* accession might be from the seed coat. Therefore the amounts of Qr4 in the seed coat and the seedling were analyzed. The seed coats attached to 7DAG seedlings from WTs were carefully removed. The separated seed coats (SC) and the seedlings (S) were used for extraction of Qr4. Moreover, the seedlings, which seed coats were removed at 2DAG and grew until 7DAG, were named seedling grown without seed coat (SW) and used for comparison of flavonol profiles among the samples. The extracts were compared using HPTLC and HPLC-PDA analyses. A quantitative Qr4 accumulation difference between Col-0 and L*er* was only found in the extracts from SCs (Figure 27). By comparing the seedling extracts between S and SW, no obvious difference in flavonol accumulation was seen. Additionally, only a small amount of Qr4 was detected in the L*er* seedlings from HPLC analysis. Therefore, the Qr4 trait was specific to the seed coats attached to the hypocotyls of the seedlings.

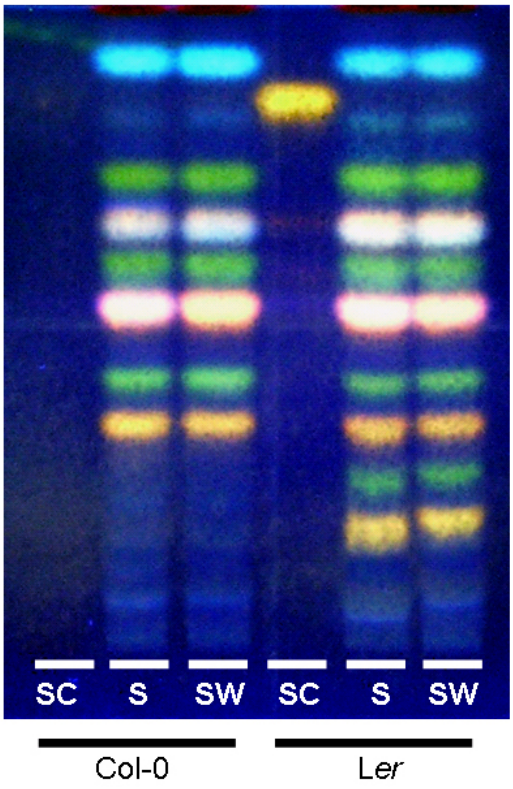


Figure 27. Comparison of Qr4 accumulation in seedlings and seed coats from 7 DAG *A. thaliana* seedlings.
HPTLC chromatograms of flavonoids extracted from seed coats and seedlings from Col and L*er*. SC-seed coat, S-seedling extracts whose seed coats were removed at 7 DAG, SW-seedling extract whose seed coats were removed at 2DAG and seedlings grew further until 7 DAG.

3.2.9 Summary

The Qr4 trait locus was mapped on chromosome 5 ≈99 cM. The candidate gene was identified as *TRANSPARENT TESTA 10* which encodes a laccase protein known to catalyze quercitrin dimer formation. Sequence comparison of the *TT10* promoter region displayed InDel polymorphisms. Association analysis of Qr4 accumulation with the *TT10* genotype in the used RIL population showed strong correlation that the Qr4 accumulation is depending on the *TT10* genotype. Transfection analyses of the promoters GUS constructs in Arabidopsis protoplasts displayed that the Col promoter has a higher activity than the L*er* promoter. Thus the allelic differences within the promoters might play a role in the Qr4 trait. The promoter activities were found in the root and the hypocotyl-root transition zone of the transgenic Col seedlings. Finally, the Qr4 trait was identified not to be the seedling-specific but the seed coat-specific trait.

3.3 Identification of qualitative loci influencing biosynthesis of flavonol glycosides

The Km1 and Qr1 traits identified by HPTLC and HPLC analyses of *A. thaliana* wildtype accessions appeared in a qualitative manner (Figure 28A and B). The loci involved in qualitative traits can be located in the *A. thaliana* genome by linkage analysis to find markers closely linked to the trait loci. The next step was to predict candidate genes for the trait loci, therefore, the flavonol derivatives found in Col and L*er* seedlings were compared to the *A. thaliana* flavonol glycosylation pathway at the *Arabidopsis thaliana* Biochemical Pathways database (http://www.arabidopsis.org). According to the database, the production of kaempferol 3-*O*-gentiobioside 7-*O*-rhamnoside (Km1) is derived from kaempferol 3-*O*-glu 7-*O*-rha (Km3) with UDP activated glucose by an UDP-dependent glycosyltransferase (UGT) family 1 protein whose gene has not been identified to date. Therefore, the candidate gene was predicted to be involved in the glucosylation of the flavonol glycosides (Figure 28 C). In order to prove that the phenotype was originating from the genotype of, the seedlings were dissected and analyzed for the content of flavonoids in the cotyledons, the hypocotyls, and the roots separately.

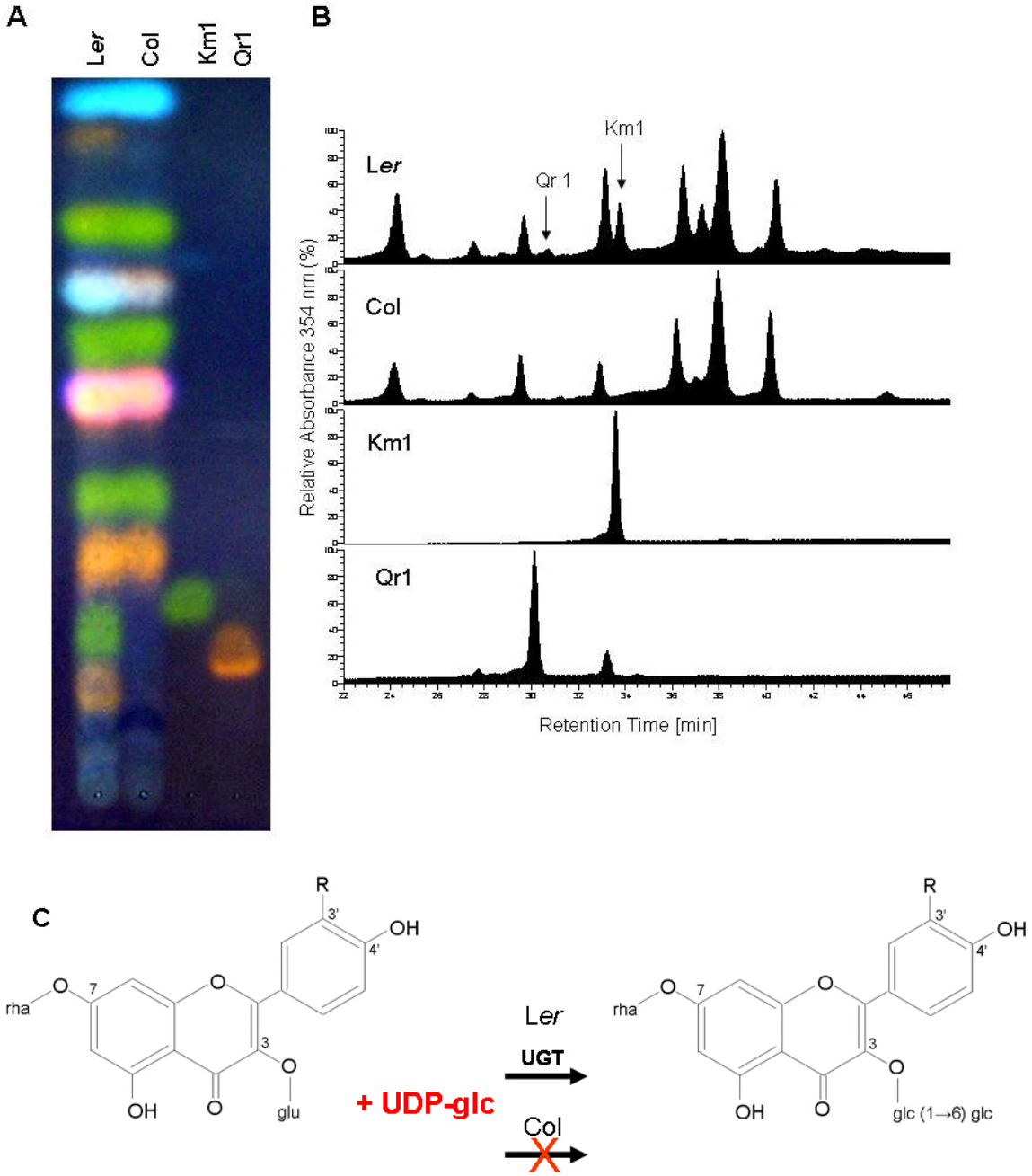


Figure 28. HPTLC and HPLC analysis of the Km1 and Qr1 traits showing qualitative differences between Col and L*er* *A. thaliana* wildtype accessions.
(A) HPTLC of the Km1 and Qr1 HPTLC purified extracts in comparison to total methanolic extracts from the accessions stated in the figure.
(B) HPLC chromatograms of the HPTLC purified Km1 and Qr1 compounds compared to the ones from total methanolic extracts from Col and L*er*.
(C) The working hypothesis of the candidate gene function involved in the Km1 and Qr1accumulations. L*er* might have an additional UDP dependent glycosyltransferase (UGT) family 1 protein which produces flavonol 3-*O*-gentiobioside-7-*O*-rhamnoside by transferring UDP activated glucose. glucose - glc, rhamnose - rha

3.3.1 Flavonoid accumulation profile in distinct parts of *A. thaliana* seedlings

Different flavonoid derivatives are known to accumulate in distinct parts of *A. thaliana* seedlings. The investigation of the Km1 and Qr1 accumulation profiles in the seedlings might provide more evidence for the candidate gene function. To investigate where the Km1 and Qr1 compounds accumulate in the seedling, the seedlings were dissected into three parts, namely, cotyledon, hypocotyl, and root. The flavonoid extracts from the dissected seedlings were analyzed using HPTLC (Figure 29A). The Qr1 compound accumulated exclusively in the root of L*er*. In contrast to the Qr1, the Km1 compound was mainly localized only in the cotyledon of L*er*. Additionally, kaempferol, quercitrin, sinapate derivatives were mainly found in the cotyledons, the roots, and the hypocotyls and the cotyledons from both accessions, respectively. Localization of the flavonoid derivatives were also observed by DPBA stained seedling. The DPBA staining of the seedling also showed that orange quercetin derivatives were accumulated in the root. Moreover, the kaempferol derivatives were exclusively localized in the apical meristem region of the seedling (Figure 29B). The Km1 and Qr1 compounds showed Mendelian inheritance patterns between Col and L*er*.

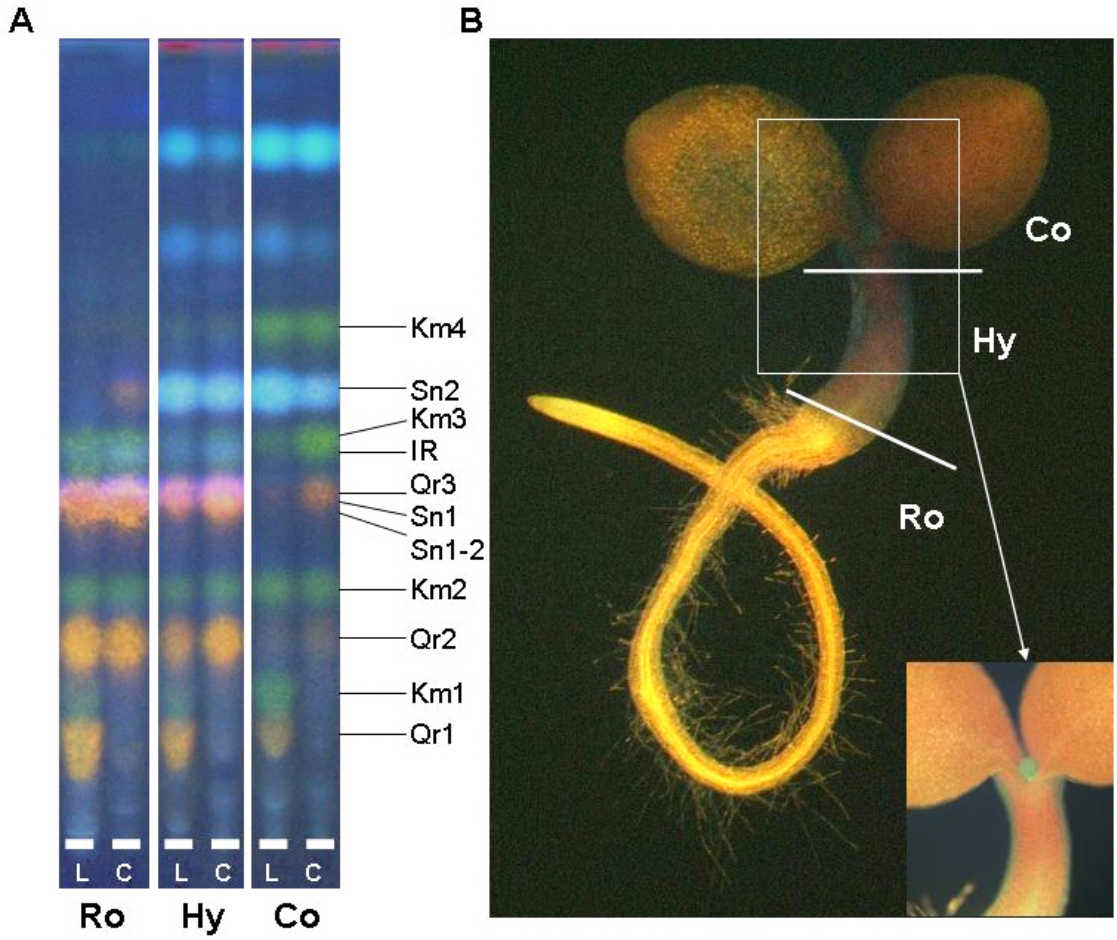


Figure 29. Flavonoid accumulation profile in distinct parts of *A. thaliana* seedlings.
(A) DPBA stained HPTLC plate with the methanolic extracts from distinct parts of 7 DAG seedlings showing different flavonoid compositions. *A. thaliana* seedlings. L – L*er*, C – Col, Ro - root, Hy - hypocotyl, Co – cotyledon.
(B) 7 DAG L*er* seedling stained with DPBA. The patterns of flavonoid accumulation in the seedling were observed under an epifluorescence microscope (M&M 1.8.4). Quercetin derivatives with orange fluorescence exclusively accumulate in the root and the root tip. Accumulation of kaempferol derivatives is found at the apical meristem region. The reddish fluorescence observed in cotyledon is from chlorophyll. The dissection sites are indicated with white lines.

3.3.2 Scoring of the Km1 and Qr1 traits in the L*er* x Col RIL population for a linkage analysis

The Km1 and Qr1 traits were scored for the linkage analysis by HPTLC. The linkage analysis compares the recombination frequencies between a trait locus and a marker locus from known positions in the chromosomes to find out whether or not the two loci are linked. To start the linkage analysis of the traits, the phenotypes of L*er* x Col RIL seedlings were evaluated using HPTLC

analysis. Depending on presence or absence of the Km1 and Qr1 traits, 95 lines from the L*er* x Col RIL population were classified into either Col or L*er* phenotypes. 31 lines containing the Km1 and Qr1 compounds were identified to have L*er* phenotypes and the other 64 lines were classified as Col phenotypes (Figure 30). The Km1 trait always segregated together with the Qr1 trait thereby indicating that both traits (from now on they are called the Km1 and Qr1 trait) were influenced by one locus. The phenotypic scores obtained from the RIL population are listed in Appendix 5. The data were used to determine the position of the Km1/Qr1 locus in the *A. thaliana* genome using the linkage analysis software MAPMAKER 3.0b (Lander, 1987).

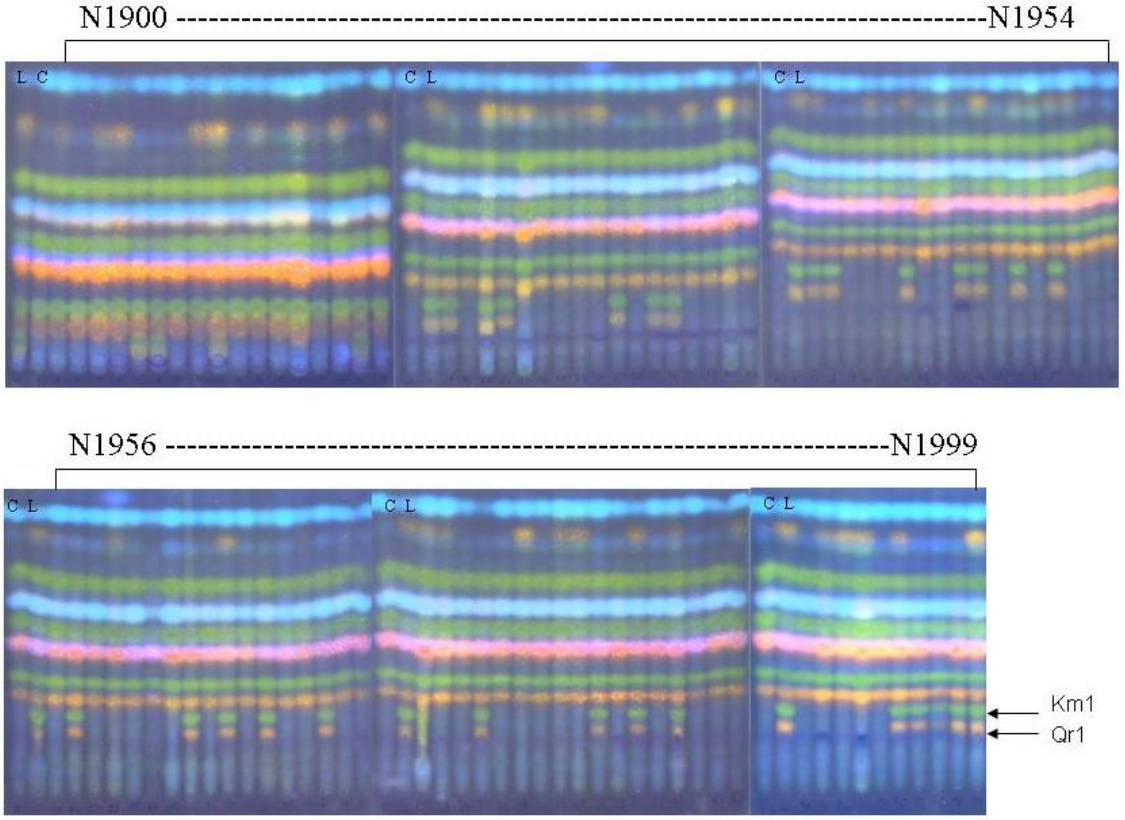


Figure 30. Scoring of the Km1 and Qr1 traits of L*er* x Col RIL populations for the linkage analysis by HPTLC analysis.
The data for the linkage analysis was collected by scoring the RILs by the qualitative appearances. The score data were used for identification of the locus location. The parental lines of RIL were always analyzed together with the siblings and listed as C - Col and L – L*er.* The data deduced from the HPTLC analyses were listed in Appendix 5.

3.3.3 Identification of the Km1/Qr1 locus in the *A. thaliana* genome by a linkage analysis

The location of the Km1/Qr1 locus on *A. thaliana* chromosome was identified by linkage analysis. The phenotype score (3.3.2) from the RIL was used for linkage analysis by comparing the score to the genotype data obtained from 542 polymorphic markers distributed among the five *A. thaliana* chromosomes (obtained from NASC). The linkage analysis between the Km1/Qr1 locus segregation data and the genotype data was carried out using MAPMAKER 3.0b. The Km1/Qr1 locus was tightly linked to mi304 (86.52 cM) with a distance of 0 cM. Physical position (bp) rather than statistically estimated position (cM) of mi304 was not known. Thus, the flanking markers with known physical position to the mi304 was searched and g4026 (87.02 cM- between 22,279,698 and 22,280,541 bp) and nga280 (83.83 cM at 20,877,364 bp) on chromosome 1 were identified (Figure 31). The flanking markers g4026 and nga280 were used as anchor markers to connect the genetic distance (cM) to the physical distance. No gene encoding an UGT family 1 protein was found in the interval. The closest UGT family 1 genes to the interval were *UGT79B5* (*At1g50580*) located at 18.73 Mb, *UGT89A1P* (*At1g51210*) at 18.99 Mb *UGT79B10* (*At1g64910*) at 24.12 Mb, and *UGT79B11* (*At1g64920*) at 24.12 Mb on chromosome 1. These genes are too far from the interval, thus, they were excluded from being candidates.

Jander and his colleagues (2002) reported the presence of large InDels between the Col and L*er* genomes. The InDel size varied between 1 bp and more than 38 Kb with a median and mean size of 4 and 175 bp, respectively. InDels can be found every 6.6 Kb in the Col sequence (Jander *et al.*, 2002). Therefore, it is possible that the L*er* genome has an extra UGT family 1 protein as an InDel within the mapped area. In order to prove this new hypothesis, it was essential to compare the L*er* genome to the Col sequence. Hence, we have tried to narrow down the interval for sequencing analysis.

To narrow down the area of the locus, CAPS markers between 22.02 Mb and 23.82 Mb on chromosome 1 were generated using the information gained from the Max-Planck Arabidopsis SNP Consortium (MASC) database (Schmid *et al.*, 2003; Torjek *et al.*, 2003). Finally, the Km1/Qr1 locus was located on the lower arm of chromosome 1 between the flanking markers nga280 and H102

(MASC07418), which are located at 20.87 and 23.10 Mb, respectively (Figure 31). The distance between the flanking markers was not shortened because of lacking recombination events in the interval. Thus, the number of the RIL population was increased to a total of 300 lines for identification of the lines containing the recombination events between the markers.

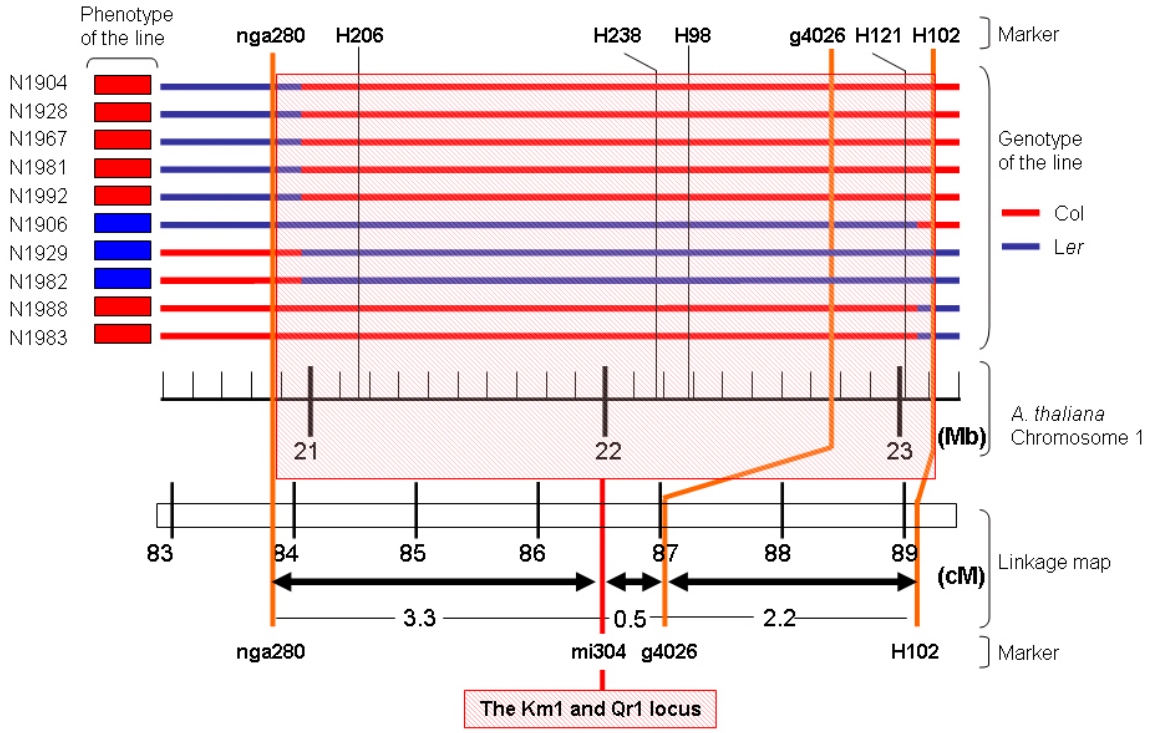


Figure 31. Positions of the Km1/Qr1 locus on chromosome 1.
The linkage map of the Km1/Qr1 locus area constructed using the MAPMAKER program was integrated to a physical map of *A. thaliana* chromosome 1 sequence by anchor marker H102, g4026, and nga280. Phenotypes of the lines were listed next to line names as Col with red and L*er* with blue boxes. The genotypes of the lines at the marker positions are indicated with red or blue solid lines with Col and L*er* genotypes accordingly. At the linkage map, the position of the Km1/Qr1 locus was listed with a red vertical line at 86.52 cM. The distance (in cM) between the markers and the trait locus at 86.52 cM are listed below the arrows.

3.3.4 Fine mapping of the Km1/Qr1 locus

The Lister and Dean L*er* x Col population from 200 additional RILs was analyzed using CAPS markers H100 and H102 to find the lines containing recombination events between the markers. The selected lines were further genotyped using CAPS markers located between H100 and H102 (2.9.2). Finally, the RILs N4662, N4674, and N4774 were identified to have recombination sites between the

flanking marker H98 (MASC02574) and H238 with position at 22,184,065 bp and 22,290,039 bp, respectively. The distance between the markers was calculated to be 106 Kb. Moreover, N4754 was identified to have a heterozygous locus of the targeted area. The phenotypes of the lines were verified using HPTLC and were compared to the genotypes (Figure 32). By combining genotypical and phenotypical analyses of the identified lines with the ones from the 95 lines, the Km1 and Qr1 locus was located between CAPS marker H238 and H98. By using sequence-based markers, the Km1 and Qr1 locus were further narrowed down between H238 and H139 with an estimated interval of 90 Kb between the markers.

Sequencing of the L*er* genomic sequence was performed using a binary-BAC (BIBAC) clone containing L*er* genomic sequence of the interval (2.10.1) The advantage of using the BIBAC vector is its readiness for use in *Agrobacterium tumefaciens*-mediated plant transformation with high-molecular-weight DNA (Hamilton, 1997). Thus, the BIBAC clones containing the Km1 and Qr1 targeted area were used for sequencing analysis. The complementation of the Km1 and Qr1 locus in Col plants failed although a large number of Agrobacterium tumefaciens flower dipping transformation experiments were performed

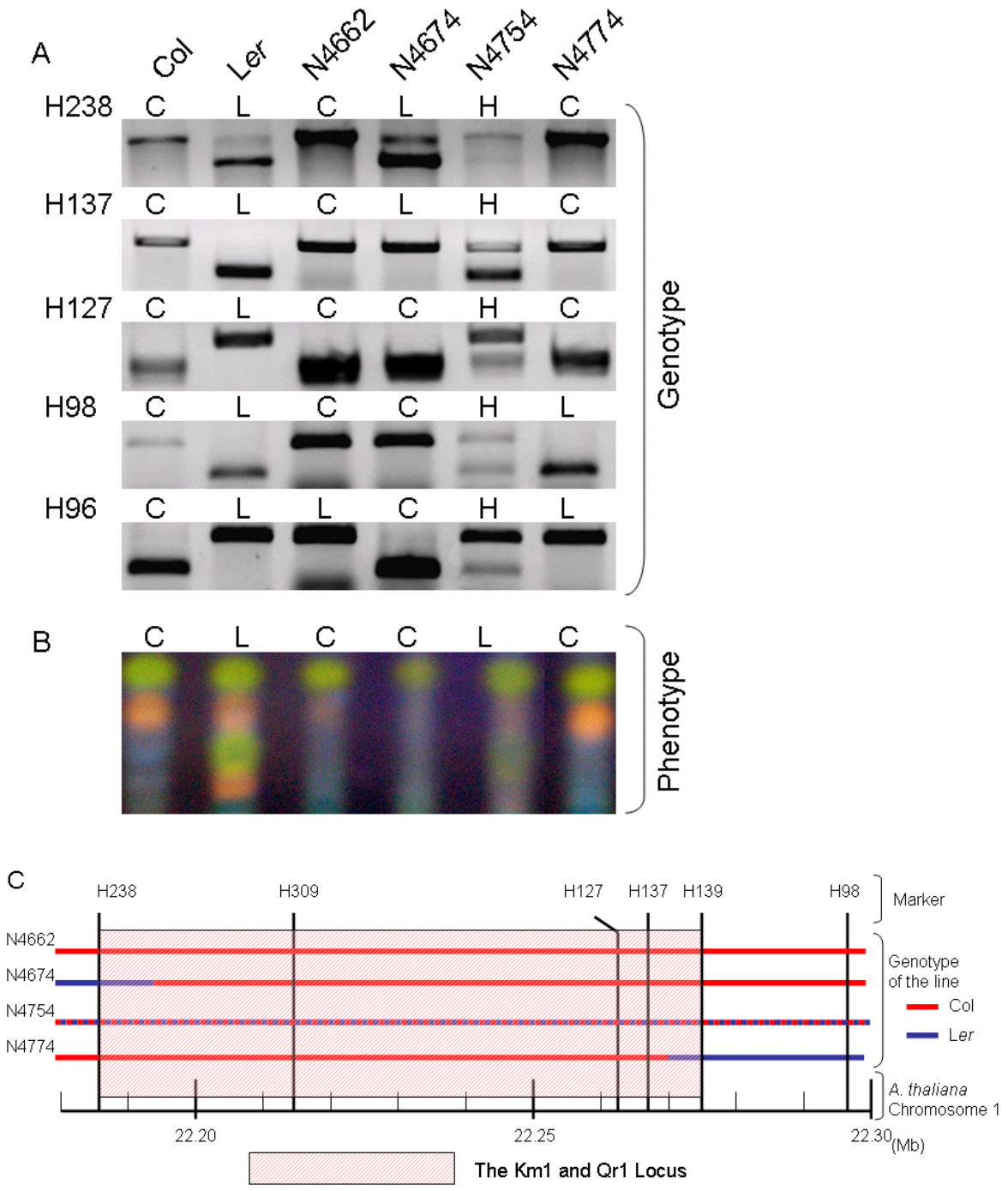


Figure 32. Genotypical and phenotypical analysis of the selected lines containing recombination events between flanking marker H96 and H238.
(A) Genotypes of selected RILsfrom the mapped area were genotyped using the CAPS markers. Genotype key: Col – C; L*er* – L; heterozygous - H.
(B) Phenotypes of the lines analyzed by HPTLC.
(C) Location of the Km1/Qr1 locus on *A. thaliana* chromosome 1 between 22.18 (H238) to 22.27 Mb (H139). H139 and H309 are sequence based markers and information of the sequences are listed in appendices 6 and 7. The location of the locus is indicated with a transparent red box.

3.3.5 Identification of L*er* BIBAC clones containing the target area

To determine the sequence of the 90 Kb region of L*er* genomic DNA, the BIBAC clones containing the interval were isolated by colony hybridization analysis from L*er* BIBAC library. The radiolabeled PCR product amplified from Col DNA using the pair of H127 primers, which is located between the flanking markers, was used to identify the BIBAC clones by screening the colony hybridization filters made from the L*er* BIBAC library. PCR analysis of positive clones using the H127 marker was used for the verification of the insert. Finally, seven positive clones were identified to contain the targeted area of the L*er* genomic sequences (Figure 33).

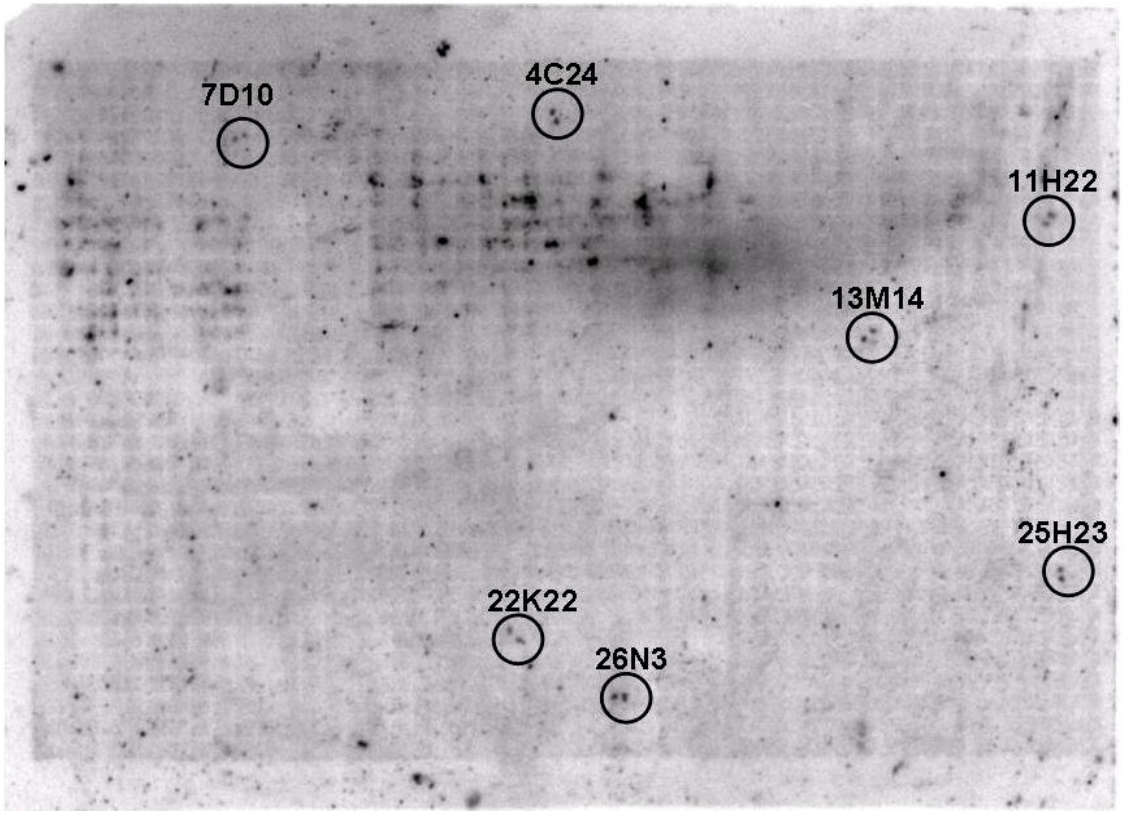


Figure 33. Colony hybridization screening of the L*er* BIBAC library using the radiolabeled H127 CAP marker.
The L*er* BIBAC library screening filter was hybridized with the H127 PCR fragment labeled with α-^{32}P dCTP. The seven positive clones are marked with circles. The BAC end sequences of the clones confirmed the positions of the positive clones in Col genomic sequence.

3.3.6 Integration of L*er* BAC inserts to Col Chromosome 1 genomic sequence

In order to verify the locations and relative sizes of the positive BIBAC inserts from the screening experiments, the right and left ends of the BIBAC insertions were sequenced. The BIBAC end sequences were aligned to the Col genomic sequence from *A. thaliana* chromosome 1 and the insert sizes and locations of the BAC inserts were estimated by subtracting the left from the right coordinate numbers (in bp; Table 12). The alignment results were used to select a BIBAC clone for analysis of L*er* genomic sequence at the Km1 and Qr1 loci. The BIBAC22K22 clone was identified to have an estimated insert size of 214 Kb that covered the entire targeted area between H238 and H139. Therefore, the L*er* BIBAC22K22 was used for the comparison of the genomic DNA sequences between Col and L*er* to identify the candidate gene.

Table 12. The list of BIBAC clones with estimated positions and sizes based on Col genomic sequence.
The BIBAC right end sequences are considered to be upstream of the BIBAC left end sequence.

Name of BAC clones	BAC left end position in Col (bp)	BAC right end position in Col (bp)	Estimated insert size in Col (bp)
04C24	22091457	22306280	214823
07D10	22146948	22283032	136084
11H22	22203732	22304036	100304
13M14	22080555	22275083	194528
22K22	22113456	22328096	214640
25H23	22093665	22272733	179068
26N03	22113456	22328009	214553

3.3.7 Comparison of the targeted area between Col and L*er* genomic sequences

For comparative analysis of the Col and L*er* genome, the L*er* genomic sequence of the targeted area was obtained using a shotgun sequencing method. The genomic DNA library consisting of 6912 clones with about insert length 2 Kb insertions and 2304 clones with 5 Kb insert length was constructed from BIBAC22K22. Approximately ten-fold redundant sequences were produced and

then assembled using PreGAP4 and Phrap programs from the Staden software package. The sequence was submitted to the EMBL Nucleotide Sequence Database (accession number: AM748036). The BIBAC22K22 insertion size was 211 Kb which is 3.5 Kb smaller than the estimated insert size of 214.6 Kb from Col genomic sequence. The difference was caused by the large InDels extensively found in the interval (Table 13). Surprisingly, a 9.66 Kb deletion in the mapped interval of the L*er* genome compared to the Col genomic sequence was found. The deleted sequence contains four complete genes which encode an apical meristem formation related protein (At1g60300), a Non-LTR retrotransposon family protein (LINE: At1g60310), a Toll-Interleukin-Resistance (TIR) domain containing protein (At1g60320), and a pseudogene similar to chalcone flavonone isomerase (At1g60330). A part of *At1g60340* (263 bp) encoding an apical meristem formation related protein was deleted in the L*er* genome but interestingly a 1040 bp insertion was found at 337 bp upstream of the predicted start codon of L*er At1g60340*. The L*er At1g60340* amino acid sequence had 15 amino acids less than the Col sequence. No UGT family 1 encoding gene was identified. The sequences from InDel were blasted to find any correlation to UGT family 1 protein. No UGT family 1 DNA sequence showed similarity to the insertion sequence identified from the L*er* genomic sequence.

Table 13. List of large InDels identified from the genome comparison analysis between Col and L*er* accessions.
Right and left coordinates of InDel in Col Chromosome 1 genomic sequence and InDel sizes were obtained from alignment of the end sequences to Col genomic sequence. Genes containing InDel are listed under the column "gene on the InDel." The insertion or deletion sizes in the Col genome comparing to the L*er* genome are indicated as Ins-insertion in Col comparing to L*er* genomic sequence. Del-deletion in Col comparing to L*er* gemomic sequence. Only Large InDels longer than 10 bp are listed in this table.

InDel	Left border position	InDel size (bp)	Right border position	Gene on the InDel
1	22224587	94^{Del}	22224588	*At1g60260*
2	22228223	265^{Ins}	22228488	
3	22232756	353^{Del}	22232757	
4	22233000	9660^{Ins}	22242662	*At1g60300* *At1g60310* *At1g60320* *At1g60330* *At1g60340*
5	22244407	1040^{Del}	22244408	*At1g60340*
6	22249425	683^{Del}	22249426	*At1g60370*
7	22250126	64^{Del}	22250127	*At1g60380*
8	22260274	311^{Ins}	22260587	

3.3.8 Comparison of annotated sequences in Col and L*er*

The nomenclature system reviewed by den Dunnen and Antonarakis (2001) was used to compare Col protein (p) and genomic (g) sequences with corresponding L*er* sequences. The L*er* genomic sequence obtained from BIBAC22k22 was manually annotated to search for UGT family 1 proteins at the Plant Computational Biology resource at the Max Planck Institute for Plant Breeding Research Cologne (http://bioinfo.mpiz-koeln.mpg.de/; Figure 34).

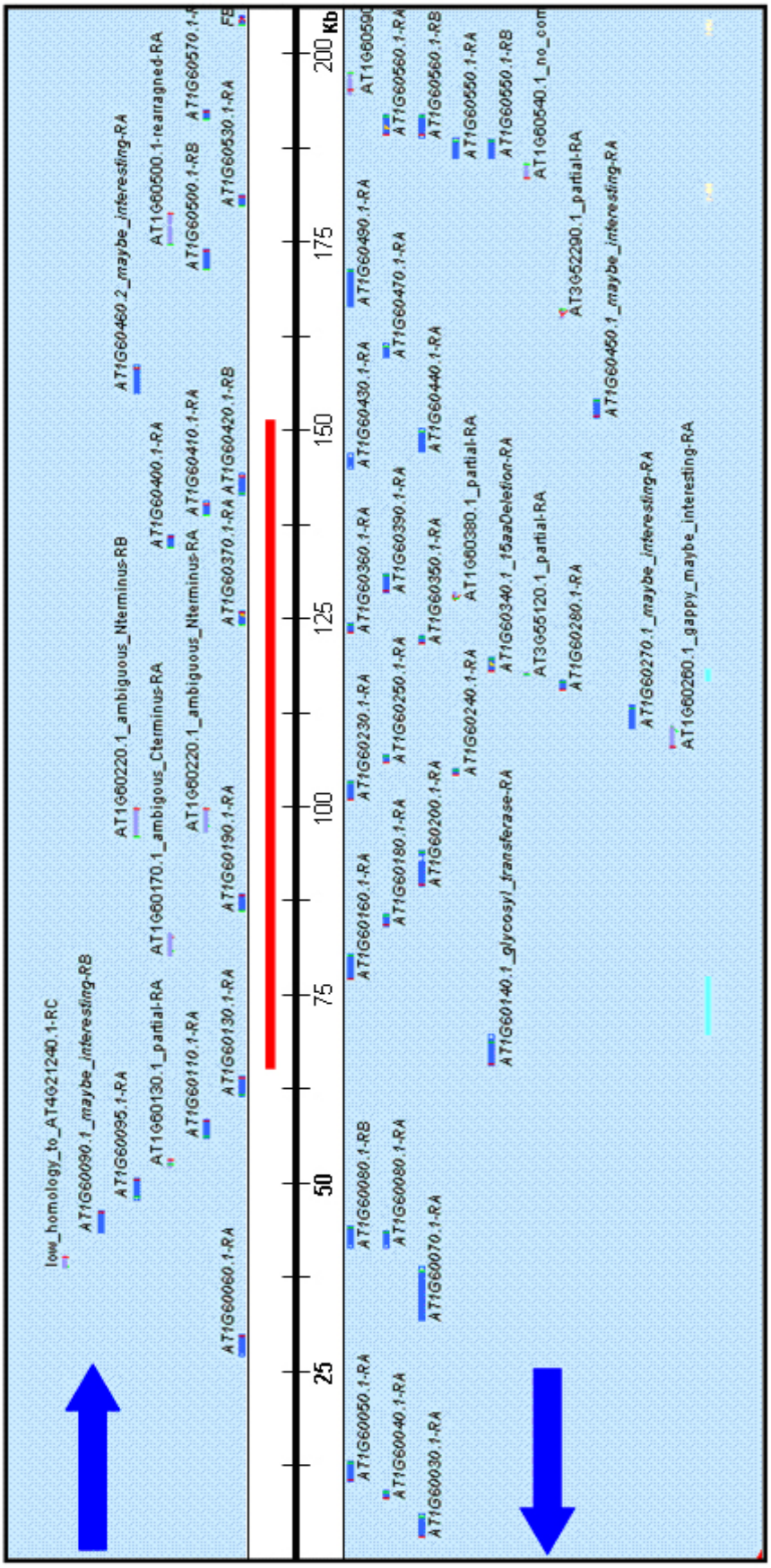


Figure 34. Annotation of the *Ler* genomic sequence from BIBAC22K22.
The 211 Kb long genomic sequence was annotated by comparison to Col homologous sequence on the chromosome 1 between 22117000 to 22321000 bp downloaded from MIPS (http://mips.gsf.de/proj/plant/jsf/athal/download/gamexmlForm.jsp). When the resulting ORF on the *Ler*-BIBAC deviated significantly from the best Col-Blast-Hit, the ORF are colored with lavender instead of blue. The blue arrow indicates the direction of the ORF. The position of the Km1and Qr1 locus between H238 and H139 is indicated with a red line over the guideline.

The annotation of the L*er* genomic sequence from BIBAC22K22 did not indicate the presence of a gene encoding an UGT family 1 protein. Thus, the candidate genes at the interval were selected for functional analyses based on the differences in the predicted OFRs between the accessions. Also annotated functions involved in sugar modifications and regulation of gene explessions. Twenty-nine annotated genes are located between the flanking markers in the Col genomic sequence (Table 14). Among those genes, four candidate genes for the trait were chosen based on the predicted amino acid sequence comparisons. Two *GLYCOSYL HYDROLASE FAMILY 1* (*At1g60260* and *At1g60270*) genes were compared. Both genes are annotated as pseudogenes in TAIR but in the MIPS annotation they are classified as genes. When Col *At1g60260* encoding a protein of 449 amino acids from (Munich Information Center for Protein Sequences (MIPS) annotation was compared to the corresponding L*er* sequence, seven amino acid substitutions were identified (p.P17S, p.S125R, p.V167A, p.Y176N, p.L260F, p.K331N, p.N361H, and p.L441S; Appendix 8). Comparison of Col *At1g60270* encoding a protein of 477 long amino acid to the L*er* revealed four amino acid substitutions (p.D2S, p.E27K, p.N130S, and p.P343T; Appendix 9).

At1g60140 encodes *TREHALOSE PHOSPHATASE SYNTHASE 10* (*ATTPS10*) which is an enzyme putatively involved in trehalose biosynthesis and also known as an UDP-glycosyltransferase family 20 protein. The protein has a trehalose synthase (TPS)-like domain and the function of the enzyme is unknown *in planta* (Leyman *et al.*, 2001). *ATTPS10* amino acid sequences between Col and L*er* exhibited four amino acid substitutions (p.V116L, p.V122I, p.I336V, and p.A840T; Appendix 10).

Last but not least, *At1g60360* encodes a C3HC4 zinc finger of the *REALLY INTERESTING NEW GENE* (*RING*) type transcription factor class. Some of the RING transcription factors are known to regulate genes negatively by degradation of target proteins (Saijo *et al.*, 2003). The comparison of the sequences showed that single nucleotide deletion at 602 bp from the ATG (g.602delA) in L*er* (Appendix 11) produced premature stop codons (p.200VfsX210, p.200TfsX212, p.200LfsX217, and p.200Ifs251; Appendix 12).

Table 14. List of annotated genes in Col chromosome 1 between the flanking marker H139 and H238.
The descriptions of the annotated sequences were from the Arabidopsis genome annotation (TAIR7) at TAIR.

ATG number	TAIR7 annotation description
At1g60140	*ATTP10* encodes an enzyme putatively involved in trehalose biosynthesis. The protein has a trehalose synthase (TPS)-like domain that may or may not be active as well as a trehalose phosphatase (TPP)-like domain
At1g60150	Pseudogene, similar to putative AP endonuclease/reverse transcriptase
At1g60160	Potassium transporter family protein, similar to potassium transporter HAK2p
At1g60170	Pre-mRNA processing ribonucleoprotein binding region-containing protein
At1g60180	Expressed protein
At1g60190	Armadillo/beta-catenin repeat family protein / U-box domain-containing protein
At1g60200	Splicing factor PWI domain-containing protein / RNA recognition motif (RRM)-containing protein
At1g60210	Proline-rich protein family,
At1g60220	Similar to Ulp1 (ubiquitin-like protein)-specific protease
At1g60230	Radical SHOOT APICAL MERISTEM (SAM) domain-containing protein
At1g60240	Apical meristem formation protein-related
At1g60250	Zinc finger (B-box type) family protein
At1g60260	Putative, glycosyl hydrolase family 1
At1g60270	Putative, glycosyl hydrolase family 1
At1g60280	ANAC023 (Arabidopsis NAC domain containing protein 23); transcription factor; similar to apical meristem formation protein-related
At1g60290	Pseudogene, similar to Chalcone--flavonone isomerase
At1g60300	Apical meristem formation protein-related
At1g60310	Non-LTR retrotransposon family (LINE)
At1g60320	Toll-Interleukin-Resistance (TIR) domain-containing protein
At1g60330	Pseudogene, similar to Chalcone--flavonone isomerase
At1g60340	Apical meristem formation protein-related
At1g60350	ANAC024 (Arabidopsis NAC domain containing protein 24); transcription factor; similar to apical meristem formation protein-related

At1g60360	Zinc finger (C3HC4-type RING finger) family protein
At1g60370	F-box family protein
At1g60380	Apical meristem formation protein-related; similar to ANAC024
At1g60390	BURP domain-containing protein / polygalacturonase, putative, similar to polygalacturonase isoenzyme 1 beta subunit
At1g60400	F-box family protein
At1g60410	F-box family protein
At1g60420	DC1 domain-containing protein

3.3.9 Summary of the Km1 and Qr1 locus

The Km1and Qr1 compounds were found to accumulate only in L*er* but not in Col seedlings. The trait was found to segregate always together in a RIL that indicates that the trait was controlled by a single locus. The candidate gene was assumed to encode an UDP-glycosyltransferase family 1 protein according to the data obtained from the metabolite analyses. The locus was fine mapped on the estimated interval of 90 Kb on chromosome 1 between the flanking markers H238 and H139 with an estimated interval of 90 Kb. In order to compare the genomic sequences of the interval between Col and L*er*, the L*er* BIBAC clone 22K22 was isolated and sequenced. Large InDels and many SNPs are located in this area. Because of no gene encoding UGT family 1 protein locating in the interval, the four candidate genes which were annotated in functions of sugar modifications and regulation of gene expression were selected for further investigation.

4 Discussion

The sessile lifestyle of plants may have increased the need for variations of enzymes involved in the accumulation of different flavonoid derivatives by mutation and selection of the corresponding genes (Gornall and Bohm, 1978). Flavonoids are known to have antimicrobial activity, which is involved in the plant microbial interection, and protection against UV radiation. Camalexin which is a major phytoalexin in *A. thaliana* is known to reduce growth of the wheat pathogen *Pseudomonas syringae* pv syringae (Tsuji *et al.*, 1992). In contrast, isoflavonol luteolin produced by *Medicago sativa* L act as signal molecules which induce *nod* genes in *Rhizobium meliloti* for nitrogen fixing symbiosis. Moreover flavonol and simapate derivatives act as protectants against UV radiations (Sheahan, 1996; Ryan *et al.*, 2001). *A. thaliana* accessions collected from different location and environments have been shown to differ in the accumulation of flavonol glycosides. The initial idea was to identify novel genes involved in accumulation of flavonoid derivatives. Prior experiments have shown that *TT10* is responsible for the natural variation in the accumulation of quercitrin by QTL mapping. The locus involved in the biosynthesis of flavonol 3-*O*-gentiobioside 7-*O*-rhamnoside (the Km1 and Qr1) trait was mapped on chromosome1 by linkage analysis. So far the gene involved in that trait is unknown. The L*er* genome is predicted to contain the Km1/Qr1 locus which was sequenced and compared to the corresponding Col genomic region. The predicted candidate gene encoding an UDP-glycosyltransferase family 1 protein was not found at the mapped interval, however, several genes which might be involved in the accumulation of the Km1 and Qr1 compounds were suggested based on genome annotation.

4.1 Establishment of analytical method for identification of qualitative and quantitative differences in flavonoid accumulations in *A. thaliana* wildtype accessions.

To identify new genes involved in the accumulation of flavonoid derivatives in *A. thaliana* wildtype accessions an approach using both metabolomics and genomics methods was established. To achieve this goal the analytical methods were chosen for the analysis of flavonoid derivatives in *A. thaliana* seedlings. High performance liquid chromatography coupled to electrospray ionization and ion trap

tandem mass spectrometry (HPLC-PDA-ESI/MSn) assay was used to quantify flavonoid derivatives and also utilized for the estimation of the flavonoid chemical structures. It has been applied especially for the identification of flavonoid derivatives (Tohge *et al.*, 2005b; Kerhoas *et al.*, 2006; Stobiecki *et al.*, 2006) because it is suitable for analysis of polar compound like flavonoid glycosides. The disadvantage was that this method is time consuming making it a bad choice for the screening of large numbers of samples and it is not able to identify the conjugated sugars of each flavonoid derivatives. Since a large numbers of plant samples were needed to be analyzed, an HPTLC assay was applied for identification of flavonoid accumulation traits between *A. thaliana* wildtype accessions. HPTLC is a suitable assay for the screening of large numbers of plant samples in a short time. Up to 19 samples can be applied on a 10 x 10 cm HPTLC plate and analyzed within one hour. The other advantages of the HPTLC analysis are that any qualitative differences displayed between the parents of segregating mapping population in HPTLC plate can be used as molecular markers to find the positions of the respective loci. Thiese techniques can distinguish flavonol and sinapate derivatives however, HPTLC analysis can not distinguish the chemical structures of these derivatives. All flavonoid and sinapate derivatives' spots which were present on the HPTLC performed with plant extracts were extracted by preparative a HPTLC method developed in this study. The retention times and chemical properties of the HPTLC purified compounds were analyzed using HPLC-ESI/MSn. Moreover, HPTLC-purified flavonoid derivatives were used for the identification of the conjugated sugar moieties by comparing the retention times and mass to charge ratio (*m/z*) of the standards sugars using GC-MS. The knowledge of the compounds associated to the spots on the HPTLC plates enabled a high throughput screening of plant samples. However, the disadvantage of HPTLC analysis is that the sample separation is not as good as HPLC and GC and a quantification of flavonoid derivatives is not possible. The resolving power of the chromatography, however, could be maximized by limiting the size of the sample application area. Compared to the spotwise application, applying samples as narrow bands at starting zones gives higher resolution and separation. By using commercially available sample application and analysis instruments for example from GAMAG

(http://www.camag.com/), higher resolution of the compound in the sample and quantitative analysis of the flavonol derivatives is possible.

4.1.1 Differences in flavonoid compositions in twelve different *A. thaliana* wildtype accessions.

Twelve different *A. thaliana* wildtype accessions were screened for differences in the accumulation of flavonoid derivatives using the established methods. Quantitative and qualitative differences found between the accessions were used for QTL and linkage analysis, respectively. It was possible to distinguish between nine flavonol glycosides and three sinapape derivatives from the accessions. All flavonols were glycosylated in mono-, di- and tri- forms with glucose and/or rhamnose. The data were in common with already published data (Veit and Pauli, 1999; Tohge *et al.*, 2005b; Kerhoas *et al.*, 2006; Routaboul *et al.*, 2006; Stobiecki *et al.*, 2006). The accessions collected from different locations displayed quantitative and qualitative differences in accumulation of flavonol glycosides. Flavonol 3-*O*-gentiobioside 7-*O*-rhamnosides were only accumulated in the accessions namely CS22491, C24, Cvi, Gue-0, L*er*, Lz-0, and Ws-0 but not in Ak-1, Col, Ei-2, Nd-0, and Wei-0. Quantitative differences in the accumulation of quercitrin was also found in these accessions. The difference in flavonol glycoside accumulation might be involved in the adaptation to different environmental demands since they are known to involved in protecting plant against abiotic and biotic stresses (Winkel-Shirley, 2002). Due to the availability of large numbers of published molecular markers, RILs were generated from L*er* and Col (Lister and Dean, 1993) and used to identify the position of the genes involved in the qualitative and quantitative differences in the flavonol glycoside accumulation. The number of markers is one of the most important aspects for the accuracy of the mapped position and for identification of the corresponding genes (Jansen, 1996).

4.2 *Transparent Testa 10* involved in the quercitrin accumulation difference in *A. thaliana* wildtype accessions

A. thaliana wildtype accessions Col and L*er* showed a quantitative difference in the accumulation of quercitrin in the empty seed coats attached to the hypocotyl-root transition zone 7 days after germination (DAG). The locus influencing this so-called Qr4 trait was mapped on the lower arm of chromosome 5 at around 99 cM. No other QTL peaks above the calculated threshold have been identified on the other chromosomes indicating that this trait is influenced by a single locus. The availability of the complete *A. thaliana* genome sequence and intensive studies on the phenylpropanoid pathway made it possible to identify the gene involved in the mapped area. The locus affecting quercitrin accumulation was identified and shown to be influences by the *TT10* gene. The identity of this *TT10* gene was confirmed by comparison of Qr4 accumulation between *tt10* mutants and the according wildtype. Furthermore, the association of the *TT10* genotype and phenotype in the RILs were examined. The significance of the genotypic means between Col and L*er TT10* alleles also confirmed that the RILs carrying the L*er TT10* allele accumulate more Qr4 compound than the lines with the Col *TT10* allele. When Qr4 accumulation between wildtype and mutant alleles were compared, less differences in the Qr4 accumulation between L*er* and *tt10-1* was found. The *tt10-1* mutant might still be functional for the dimerization of quercitrin because it was generated by ethylmethanesulfonate (EMS) mutagenesis, which caused two amino acid substitutions and therefore is still expressed in the plants (Pourcel *et al.*, 2005). A similar situation was described for the *tt4-W85* (*tt4-1*) allele that is responsible for the accumulation of small amounts of flavonol derivatives but never the less resulted in yellow seed coats (Booij-James *et al.*, 2000; Dana *et al.*, 2006).

TT10 was previously identified as a laccase involved in enzymatic oxidation of proanthocyanidins. Initially the yellow seed coat phenotype of *tt10* was mapped on chromosome 5 by Shirley *et al.* (1995) later Pourcel *et al.* (2005) used a candidate gene approach to search for the gene encoding the protein involved in enzymatic oxidation of proanthocyanidins which is necessary for early browning of seed coats at the mapped interval. The two approaches were similar except of starting materials and targeted phenotypes. They have used forward genetic

approaches to identify the gene involved in the *tt10* mutant phenotype. In contrast, this study used quercitrin accumulation differences to identify the corresponding gene using *A. thaliana* wildtype accessions. Interestingly, the Qr4 trait was linked to the seed phenotype, which indicates that quercitrin might be involved in browning of the seed coats. Pourcel *et al.* suggested that oxidized quercetin aglycone, which is highly reactive could explained the oxidation of other compounds necessary for the browning of seed coats (2007).

4.2.1 Natural variation of the *TT10* promoter play a role in the accumulation of Qr4

In this study, a higher activity of the Col *TT10* promoter compared to the L*er TT10* promoter indicated that allelic difference in the promoter sequences causes the Qr4 compound accumulation phenotype in seed coats. RT-PCR analysis performed by Pourcel *et al.* (2005) also showed higher *TT10* transcript level in Col than in L*er*. In *A. thaliana* wildtype accessions, modification of promoters to control expression of genes involved in secondary metabolite accumulation might be a common way to adapt to their environments. So far, all identified quantitative trait involved in glucosinolate accumulation in *A. thaliana* wildtype accessions shows allelic differences caused by modifications in the promoter region to activate or even completely knockout expressions of corresponding genes. QTL involved in accumulation of glucosinolate derivatives, which are important for plant defense against herbivores have been reported and the genes involved in the QTL were assigned to two *2-OXOGLUTARATE-DEPENDENT DIOXYGENASES* (*AOP2* and *AOP3*) (Kliebenstein *et al.*, 2001b). L*er* predominantly expresses *AOP2* and accumulates 3-hydroxypropyl glucosinolate as the predominant short chain alphatic glucosinolate; in contrast Cvi exclusively expresses *AOP3* and accumulates allyl glucosinolate and 3-butenyl glucosinolate. Furthermore the Col accession contains only 4-methylthiobutyl which is the precursor of 3-hydroxypropyl, allyl and 3-butenyl glucosinolate because Col does not have active *AOP2* and *AOP3* alleles. The quantitative differences in the glucosinolate accumulation are caused by the differences in promoter sequences between the accessions which lead to the expression difference of the genes (Kliebenstein *et al.*, 2001b). Lambix also reported an allelic difference in promoters of

EPOTHIOSPECIFIER PROTEIN (*ESP*) between Col and L*er* accessions. EPS is involved in hydrolysis of glucosinolate to nitrile and influences *Trichoplusia ni* (the cabbage looper) herbivory (2001). A 10 bp deletion found in the Col promoter sequence was enough to knockout the gene expression which resulted in the accumulation of unhydrolyzed glucosinolate compound (Lambrix *et al.*, 2001).

Even thought, a five times higher GUS activity was constantly observed from the Col promoter in comparison to the L*er TT10* promoter by transfection experiments, the same effect was not found in the comparison of the Col and L*er TT10* promoters in Col transgenic seedlings. Both seedlings showed strong activity in the root and the hypocotyl-root transition zone but did not show difference in activity level between Col and L*er* promoters (Figure 25). First, the expression differences between the lines are likely caused by T-DNA copy number in the genomes or differences in insertion sites (Liu *et al.*, 2005). MacCaig *et al.* showed high *TT10* activity in the roots and the cotyledons of two weeks old seedlings (2005). This contradiction might be explained that *cis*-acting regulatory DNA elements, predicted to be involved in early responses to dehydration of the seed and in stress response in the root, was found in both promoter sequences. Unknown stresses might activate both genes resulting in no difference in promoter activity Pourcel *et al.* did not find a strong expression of *TT10* in seedlings (2005) but it also might explain the activation of the gene by unknown stresses.

TT10 could be an enzyme controlling the amount of quercitrin in the seedlings to prevent further modifications of the compound. Small amounts of quercitrin were found in L*er* but none was detectable in Col. Perhaps quercitrin was converted to dimers by TT10 and no quercitrin was detectable because of the higher Col *TT10* promoter activity. The presence of quercitrin in seedlings was also reported by Peer *et al.* (2001). They showed a quercitrin accumulation at the hypocotyl-root transition zone and in roots of 7-day old L*er* seedlings. Moreover, quercitrin should be produced in the seedling because Km 3-*O*-rha 7-*O*-rha (Km4) was identified from the seedling which is synthesised from Km 3-*O*-rha. Km 3-*O*-rha which is produced by UGT78D1 which is also known to form quercitrin (Jones *et al.*, 2003). In contrast, Qr 3-*O*-rha 7-*O*-rha was not detected in the seedlings. Thus, one could assume that TT10 might control the extent of

quercitrin monomer by dimerization and prevents further glycosylation or even deglycosylation. Liang *et al.* reported that *tt10* mutant seedlings had significantly shorter primary roots than the seedlings from wildtypes (2006). That implys that TT10 might control the amount of quercitrin which can be deglycosylated by a glycosyl hydrolase to release quercetin aglycone which is known as a negative regulator of auxin transport (Jacobs and Rubery, 1988; Murphy *et al.*, 2000; Peer *et al.*, 2001; Woo *et al.*, 2007). Routaboul *et al.* reported on the accumulation difference of quercitrin dimer in developing seeds (2006). If quercitrin dimers can be identified and quantified in seedlings, the biological function of quercitrin dimerization by TT10 could be linked by QTL analysis.

4.3 The flavonoid derivatives accumulate in distinct parts of *A. thaliana*

Flavonoid composition in different organs of *A. thaliana* have been investigated by several groups. They have shown that flavonol compositions differ at different plant organs. Arabidopsis seeds mainly contain quercetin derivatives (Kerhoas *et al.*, 2006) and accumulate quercetin 3-O-rhamnoside 7-O-rhamnoside which is not found in seedlings (Stracke *et al.*, 2007) and leaves (Veit and Pauli, 1999; Stobiecki *et al.*, 2006). Tetra glycosylated kaempferol could only be isolated from rosette leaves of the accession C24 (Stobiecki *et al.*, 2006). In this works, the flavonoid accumulation differences among distinct parts of seedling also identified that kaempferol and sinapate derivatives were accumulated in cotyledons, and quercetin derivatives were exclusively found in roots (Figure 29). The data imply that analyzing distinct plant organs can increase the number of new flavonoid glycosides and also the number of flavonoid accumulation traits. From published work it becomes obvious that, metabolite analysis using laser micro-dissection coupled to laser pressure catapulting (LMPC) might be a useful tool for collecting organ and even tissue-specific samples or even single cells from plant tissue sections (Schad *et al.*, 2005). Schad and colleagues extracted metabolites from vascular bundles and analyzed these by gas chromatography time-of-flight mass spectrometry (GC-TOF MS). More than half of the identified metabolites were either enriched or depleted in vascular bundles as compared to the surrounding tissues (Schad *et al.*, 2005). Employing this technique would dramatically increase the number of traits, however, the LMPC is not suitable for QTL analysis

because internal standards cannot be used during comparison. Thus the comparison of metabolite profiles from large mapping population which are needed for QTL is a difficult task according to the published work. HPTLC analysis of distinct organs of *A. thaliana* seedlings has shown clear differences in flavonoid accumulation. Because of the limited time of this study, other plant organs were not analyzed, however, HPTLC analysis of flavonoid derivatives at a variety of plant organs has a great potential to increase number of traits and identified flavonol derivatives.

4.4 Identification of candidate genes influencing qualitative traits

From the data described above, the Lister and Dean L*er* x Col RIL were chosen because Km1 and Qr1 could not be detected in Col accession. The flavonol 3-O-gentiobioside 7-O-rhamnoside trait (the Km1 and Qr1 trait) was identified between Col and L*er* accessions. The Km1 and Qr1 trait from the RILs was used as a molecular marker to find the position of the locus since genotypes of the target gene in the lines should correspond to the phenotype of lines as it was shown while mapping the *transparent testa* mutations (Koornneef, 1990). In this study, the position of the locus is validated by the analysis of RILs: when the lines had Col genotype at the interval, the line also displayed Col phenotype as it was the case for the L*er* genotype (Figure 31). Linkage mapping analysis linked the Km1/Qr1 locus to the molecular marker mi304 at 86.52 cM on chromosome 1. Similar work was performed by Keurentjes *et al.* (2006), they mapped a trait influencing accumulation of flavonols deoxyhexosyl dihexoside using a L*er* x Cvi RIL population (Alonso-Blanco *et al.*, 1998) on the chromosome 1 at 88.6 cM. They could not identify the sugar moieties but one could assume that the compounds are corresponding to the Km1 and Qr1. They also did not identify the gene involved in the accumulation of the flavonol glycosides. They suggested that the candidate gene could encode an UGT family 1 protein which is only functional or existing in the L*er* accession.

Comparing the mapping result to Keurentjes *et al.* (2006), Keurentjes *et al.* (2006) which used 160 RILs which were previously genotyped by 99 markers which are distributed on 5 chromosomes with an average distance of 5 cM and the largest genetic distance of 12 cM. In this study, for the linkage mapping analysis

of the Km1/Qr1 locus, 100 RILs genotyped by 553 markers were used to locate the Km1/Qr1 locus. 170 makers are located on chromosome 1 compared to only 25 markers that were used by Keurentjes *et al.* (2006). Additionally, the position of the trait locus was further narrowed down using 200 additional RILs with markers developed around the trait locus. Thus, the linkage analysis of the Km1/Qr1 locus has higher accuracy in the mapped position than compare to the one from Keurentjes *et al.* (2006). The difference in the mapped positions 2.08 cM in the *A. thaliana* genome is estimated to approximately 500 Kb (250 Kb per 1 cM) (Jander *et al.*, 2002). No UGT family 1 protein encoding gene was found 500 Kb up- and downstream of the flanking marker H238 and H98. The closest UGT family 1 protein encoding genes are *UGT89A1P* and *UGT79B10* located at 18.9 and 24.1 Mb on the chromosome 1. This observed difference in the mapping position is most likely caused due to the use of different RIL populations and the number of mapping markers. The recombination frequency rate of the two loci in the L*er* x Col RIL population is not the same as in the L*er* x Cvi population. Even though the genetic maps from two RIL populations are connected with 49 common markers (Alonso-Blanco *et al.*, 1998), differences between the mapping positions cannot be avoided. Thus a more accurate position of the locus was obtained using the linkage analysis because the phenotype corresponded to the genotype of the trait locus and the usage of large number of molecular markers provided tightly linked markers.

4.4.1 Complementation analysis of the Km1/Qr1 locus

In Figure 32, the line N4754 harboring a heterozygous locus of the interval displays the L*er* phenotype that indicates the necessity of a candidate gene with L*er* allele to accumulate the Km1 and Qr1 compounds. Thus, complementation analysis of Col accession with the Km1/Qr1 locus from L*er* was attempted using the BIBAC, clones contain the mapped area isolated from the L*er* genomic BIBAC library (Chang *et al.*, 2003). Hamilton *et al.* (1997) demonstrated that a BIBAC vector carrying 150 Kb T-DNA has been transferred into tobacco calli generated from tobacco leaves using a *A. tumefaciens* strain carrying the helper- plasmid pCH32. pCH32 contains additional copies of *A. tumefaciens* virulence genes *virG*

and *virE* which cause constitutive expression of other *vir* genes independent of *virA* and enhance transformation efficiency of high molecular weight DNA into plant (Hamilton *et al.*, 1996; Hamilton, 1997). It has been described that without the hyper-virulent plasmid it was not possible to transfer the 150 Kb T-DNA into tobacco calli. The large T-DNA was shown to be stable in regenerated plant that maintained the 150 Kb of foreign DNA in the plant nuclear genome for generations (Hamilton, 1997). The floral dipping methods using the BIBAC binary vector to transform to *A. thaliana* genome has also been reported by Bi *et al* (2000). 40 Kb of T-DNA cloned in BIBAC vector carrying a *Brassica* DNA fragment containing an *S* locus glycoprotein (*SLG*) and a plasma membrane-spanning receptor kinase protein (*SRK*) was successfully transferred to the *A. thaliana* genome for the analysis of self incompatibility (Bi *et al.*, 2000).

A. tumefaciens mediated transformation using BIBAC clones containing the relevant L*er* genomic sequence harboring the Km1/Qr1 locus was attempted. However, no transgenic plants were identified. Several reasons can explain why the experiment was not successful. Clough *et al.* have reported a high efficiency *A. tumefaciens* mediated transformation by a floral a dipping method (1998). The author mentioned that the health of the plant is the most important factor for the high transformation efficiency. Therefore, the reason of the failure might be that the condition in greenhouse was not suitable for the transformation, e.g., stressed by drought, nematodes or herbivores. The other reason might be that *A. tumefaciens* could not transfer T-DNA more than 100 Kb in size to the plant genome using the floal dipping method (Clough and Bent, 1998). Fray *et al.* (2001) have reported a comparison of the transformation efficiency of BIBAC vectors containing 150 Kb and 40 Kb T-DNA into the tomato cell cultures using the same transfer system. They could determine transfer rates of 0.3 and 2.2 percent, respectively, indicating the difficulty of transferring T-DNA larger than 150 Kb.

4.4.2 Identification of the candidate gene involved in the Km1 and Qr1 trait

Sequence comparisons of the mapped interval between Col and L*er* indicated substantial modification of the genomes at this position (Table 13) especially when compared to other parts of the genomes as previously described by the

Arabidopsis Genome Initiative (2000) and Jander *et al.* (2002). Strikingly, the size of InDels in this area was higher than average. The Arabidopsis Genome Initiative (2000).reported that 95 percent of the InDels are smaller than 100 bp with a median of 4 bp, and an InDel was found every 6.6 Kb. Not counting the InDels less than 10 bp, eight large InDels sized between 9.66 Kb to 64 bp were found in 90 Kb intervals (Table 13). The traits found in *A. thaliana* wildtype accessions were caused by enzyme activities which have been described to be the results of differences in the promoter regions causing variation in transcripton levels, or in the activity of the enzyme by modification of the protein sequence or absence of the proteins due to deletion of respective DNA sequences (Koornneef *et al.*, 2004). Failure to isolate the candidate gene encoding the Km1/Qr1 locus could imply that it may not encode a *UGT* or is not present in the L*er* genome. Four possible candidate genes located in the interval were suggested based on the annotation of predicted ORFs and their differences between the accessions.

Two of the possible candidate genes could encode GLYCOSYL HYDROLASE (GH) family 1 proteins (*At1g60260* and *At1g60270*) which are listed as pseudogenes in TAIR annotation release 7, but are classified as genes in MIPS *Arabidopsis thaliana* Database (MAtDB) version 2.0 (http://mips.gsf.de/proj/plant/jsf/athal/index.jsp). SNPs causing amino acid substitutions found in *At1g60260* and *At1g60270* might alter the activities and specificities of enzymes (Appendix 8 andAppendix 9). GH family 1 proteins hydrolyze substrates of the type G-O-X, where G represents a β-linked glucosyl, galactosyl, mannosyl, fucosyl, 6-phosphoglucosyl, 6-phosphogalactosyl, or glucuronide residue, whereas X represents another glycosyl residue or a aglycone (reviewed in Xu *et al.*, 2004). Both the Km1 and Qr1 compounds were described as potential targets of the enzyme (reviewed in Xu *et al.*, 2004). If it is functional only in Col and not in L*er*, the enzyme might catalyze hydrolysis of the Km1 and Qr1 to less glycosylated forms resulting in the loss of the Km1 and Qr1 compounds in Col seedlings. Figure 35 suggests the hypothetical flavonol glycoside deglycosylation pathway described above.

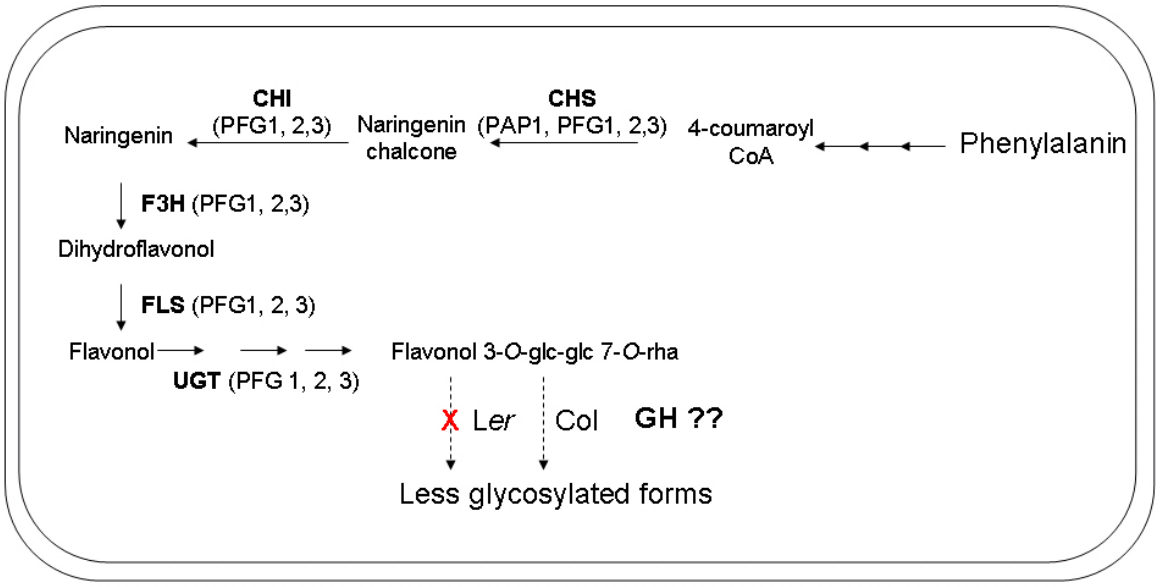


Figure 35. Hypothesis of the candidate gene involved in the Km1 and Qr1 compounds accumulation.
Solid lines indicate that the pathway is supported by biochemical and genetic evidence, whereas black dotted arrows indicate the hypothetical pathway. The major enzymes are listed with boldface capital letters. Transcription factors regulating the expression of the enzymes are given with capital letters brackets. Chalcone isomerase (CHI), chalcone synthase (CHS), flavanone 3 hydroxylase (F3H), glycosyl hydrolase (GH), glycosyltransferase (UGT).

Another possible candidate gene could be *At1g60140* encoding a trehalose synthase which is also known as UDP glycosyltransferase (UGT) family 20 protein. From the Geneinvestigator microarray database expression data, the gene is mainly expressed in seedlings and siliques, and is also regulated by abiotic and biotic stresses (https://www.genevestigator.ethz.ch/). Due to allelic differences in UGT family 20 proteins, the substrate specificity might change and result in the production of the Km1 and Qr1 compounds in L*er*. The UGT family 20 protein is suggested to participate in the production of trehalose by transferring UDP-glucose to a glucose moiety to form a 1-1 d linkage (Vogel *et al.*, 2001). Trehalose is known to accumulate in bacteria and fungi for storage purposes and stress protection, but its accumulation in plants is not known (reviewed in Leyman *et al.*, 2001). The enzyme is suggested to have a regulatory role in carbon metabolism but it is not restricted to housekeeping functions (reviewed in Leyman *et al.*, 2001). Even though glycosylation of flavonol glycosides by the enzyme is not known, the enzyme might transfer glucose to the

Km3 and Qr3 to produce the Km1 and Qr1 compounds as suggested in Figure 28.

Finally, the loss of gene function in *At1g60360* which encodes *Really Interesting New Gene* (*RING*) might cause a gain of the phenotype in L*er.* The gene is most likely not functional in L*er* because the L*er* allele contains stop codons introduced by a single nucleotide substitution (Appendix 11 andAppendix 12). The Col At1g60360 protein might be involved in the degradation of an enzyme which is involved in production of the Km1 and Qr1 compounds. The RING-type protein has E3 ubiquitin-protein ligase function. Ubiquitination leads to specific degradation of targeted proteins via 26S proteasome (summarized in Kosarev *et al.*, 2002). For instance, *CONSTITUTIVE PHOTOMORPHOGENIC 1* (*COP1*) encodes a E3 ubiquitin-protein ligase that contains a RING-finger domain, a coiled-coil domain and seven WD-40 repeats represses photomorphogenesis in the dark and negatively regulates *LONG HYPOCOTYL LOCUS 5* (*HY5*) by degradation of the protein (Osterlund *et al.*, 2000). Recently a gene (*GW2*) encodes a RING finger E3 ubiquitin-protein ligase which regulates grain width and weight was identified from *Oryza sativa* (Song *et al.*, 2007). Loss of the *GW2* function increased cell numbers and resulted in a wider spikelet hull, subsequently accelerating the milk filling rate of the grain and enhancing grain width, weight and yield (Song *et al.*, 2007). *GW2* was suggested to be involved in degradation of unknown target proteins required to suppress cell division in rice (Song *et al.*, 2007). Since *At1g60360* is not on the Affymetrix Arabidopsis AHT1 genome array, it is not possible to get direct expression data. But *At1g60360* could be co-expressed with genes involved in flavonoid biosynthesis. If *RING* is the gene responsible, the transcription factor involved in activation of *UGT* gene or UGT protein might be completely suppressed in Col thus the Km1 and Qr1 compounds can not accumulate.

4.5 Outlook

The results available so far are good bases for the identification of the gene involved in the accumulation of the Km1 and Qr1 compounds. Several experiments could be coordinated to identify the gene for the Km1/Qr1 phenotype in the L*er* accessions. The type of experiments depends on the molecular bases

underlying the observed phenotype. Two reasons could be responsible for the lack of Qr1 and Km1 in Col. 1). Missing gene upstream of biosynthesis leading to fast metabolisation of Km1 and Qr1 or 2). Presence of gene downstream of biosynthesis leading to production of Km1 andQr1.

If a loss of the gene function in L*er* is responsible for the accumulation of the Km1 and Qr1 compounds, the candidate genes can be analyzed using mutant lines, which certain a functional copy of the gene in question. Mutant plants should display the Km1 and Qr1 compounds on HPTLC.

Several approaches can be applied if the gain of gene function in L*er* is involved in the accumulation.

First, BIBAC clones located in the mapped interval could be used for functional complementation of the trait in Col-derived CK_UK cell culture (Berger *et al.*, 2007). A CK_UK cell culture which is transfected with the PFG1 transcription factor controlled by the 35S promoter was shown to accumulate large amount of flavonol glycosides which usually do not accumulate in the cell culture (Knepper, 2007). A L*er* BIBAC clone containing the up- and downstream of the suggested locus has been constructed using partial digestion with *Bam*HI restriction endonucleases (Chang *et al.*, 2003). Thus subclones with smaller inserts could be constructed with *Bam*HI restriction endonucleases to be used to transfect the CK_UK cell culture for the detection of the L*er* phenotype. This approach could minimize the number of candidate genes to be analyzed.

A second option would be to narrowing down candidate genes by increasing the number of recombination events in the mapped interval. Offsprings from the Lister and Dean L*er* x Col RIL N4774, which have a heterozygous locus of the mapped area, could be useful to initiate recombination events at the interval. Another approach to find more recombinations at the interval would be to use different RIL populations. The density of recombinations between different RIL population has been shown to be different (Alonso-Blanco *et al.*, 1998; Loudet *et al.*, 2002). Thus the Koornneef L*er* x Cvi RIL (Alonso-Blanco *et al.*, 1998) could be used to find recombination events at the interval. This approach has been successfully performed to isolate a QTL involved in glucosinolate accumulation (Kliebenstein *et al.*, 2001b).

Once the candidate gene is identified, it could then be analyzed using the CK_UK cell culture co-transfected with *PFG1*. The MYB12 transfected cell culture was shown to accumulate flavonol glycoside. The co-transfection of the candidate gene with MYB12 can speed up the identification of the candidate genes.

4.6 Concluding remarks

The aim of this study was to identify new genes involved in accumulation of flavonoid derivatives in *A. thaliana* wildtype accessions. Therefore, metabolite assay method was optimized for identification and quantification of flavonoid derivatives from large number of plant samples. The established HPTLC method has shown to be an effective and a simple method to analyse the flavonoid composition of a large numbers of the samples in short time. This method enabled the differentiation of the flavonoid profiles from distinct plant organs. It could also be applied to increase number of new flavonoid derivatives by comparing flavonoids profiles of distinct plant organs from different *A. thaliana* wildtype accessions.

The metabolite QTL mapping approach was shown to be an effective method to identify genes involved in flavonoid accumulation differences. Increasing number of flavonoid quantitative traits could lead to the identification of new genes. Mapping of the Km1/Qr1 locus indicated that there might be other factors affecting accumulation of flavonoid glycoside. The phenotype of Ler accumulating the Km1 and Qr1 compounds could be caused by loss or gain of function involved in the biosynthesis. It would be interesting to find the gene involved in the Km1 and Qr1 trait that might reveal a new enzyme controlling an accumulation of flavonol glycoside.

4.7 Summary

Twelve different *A. thaliana* accessions collected from different locations and environments have displayed differences in the accumulation of flavonol glycosides. The established HPTLC method has proved to be not only a simple but a powerful method to identify the differences in flavonoid accumulation from a large number of samples. This method could be applied to detect a number of

new flavonoid derivatives and the traits involved in flavonoid accumulation differences using distinct tissues of *A. thaliana* wildtype accessions.

The initial idea was to identify novel genes involved in the accumulation of flavonoid derivatives. However, a known gene was identified to be involved in natural variation in the accumulation of quercitrin. Allelic difference in the *TRANSPARENT TESTA 10* promoters found between the accession Col and L*er* have been identified to be involved in quercitrin accumulation differences in the seed coats. TT10 has been described to produce dimerized quercitrin (Pourcel *et al.*, 2005) which explains why the Qr4 compound accumulates more in the L*er* than in the Col accession. TT10 perhaps controls the amount of the Qr4 compound by preventing further modification by a glycosyl hydrolase and/or a glycosyltransferase. Further quantification of the quercitrin dimer in seedlings may answer the question of the biological function of the dimerization of quercitrin by TT10.

The qualitative trait involved in the accumulation of flavonol 3-*O*-gentiobioside 7-*O*-rhamnoside is identified among *A. thaliana* wildtype accessions. The Km1/Qr1 locus is located on chromosome 1 within the interval of 90 Kb. Sequence analysis of this region revealed that the candidate gene might not be UGT family 1 protein as expected. The sequence comparison of the interval of the mapped locus between Col and L*er* displayed significant differences resulting from modification of the genome by large InDels. Four possible candidate genes located in the interval are suggested to be involvement in the accumulation of the Km1 and Qr1 compounds.

References

Abrahams, S., Tanner, G.J., Larkin, P.J., and Ashton, A.R. (2002). Identification and Biochemical Characterization of Mutants in the Proanthocyanidin Pathway in Arabidopsis. Plant Physiol. **130,** 561-576.

Alonso-Blanco, C., and Koornneef, M. (2000). Naturally occurring variation in Arabidopsis: an underexploited resource for plant genetics. Trends in Plant Science **5,** 22-29.

Alonso-Blanco, C., El-Assal, S.E., Coupland, G., and Koornneef, M. (1998). Analysis of natural allelic variation at flowering time loci in the Landsberg *erecta* and Cape Verde Islands ecotypes of *Arabidopsis thaliana*. Genetics **149,** 749-764.

Alonso-Blanco, C., Peeters, A.J., Koornneef, M., Lister, C., Dean, C., van den Bosch, N., Pot, J., and Kuiper, M.T. (1998). Development of an AFLP based linkage map of L*er*, Col and Cvi *Arabidopsis thaliana* ecotypes and construction of a L*er*/Cvi recombinant inbred line population. Plant J **14,** 259-271.

Anhalt, S., and Weissenböck, G. (1992). Subcellular localization of luteolin glucuronides and related enzymes in rye mesophyll. Planta **187,** 83-88.

Baudry, A., Caboche, M., and Lepiniec, L. (2006). TT8 controls its own expression in a feedback regulation involving TTG1 and homologous MYB and bHLH factors, allowing a strong and cell-specific accumulation of flavonoids in *Arabidopsis thaliana*. Plant J **46,** 768-779.

Baudry, A., Heim, M.A., Dubreucq, B., Caboche, M., Weisshaar, B., and Lepiniec, L. (2004). TT2, TT8, and TTG1 synergistically specify the expression of *BANYULS* and proanthocyanidin biosynthesis in Arabidopsis thaliana. Plant J **39,** 366-380.

Baxter, I.R., Young, J.C., Armstrong, G., Foster, N., Bogenschutz, N., Cordova, T., Peer, W.A., Hazen, S.P., Murphy, A.S., and Harper, J.F. (2005). A plasma membrane H+-ATPase is required for the formation of proanthocyanidins in the seed coat endothelium of *Arabidopsis thaliana*. Proc Natl Acad Sci U S A **102,** 2649-2654.

Berger, B., Stracke, R., Yatusevich, R., Weisshaar, B., Flugge, U.I., and Gigolashvili, T. (2007). A simplified method for the analysis of transcription factor-promoter interactions that allows high-throughput data generation. Plant J **50,** 911-916.

Bhattramakki, D., Dolan, M., Hanafey, M., Wineland, R., Vaske, D., Register, J.C., 3rd, Tingey, S.V., and Rafalski, A. (2002). Insertion-deletion polymorphisms in 3' regions of maize genes occur frequently and can be used as highly informative genetic markers. Plant Mol Biol **48,** 539-547.

Bi, Y.M., Brugiere, N., Cui, Y., Goring, D.R., and Rothstein, S.J. (2000). Transformation of Arabidopsis with a Brassica *SLG/SRK* region and *ARC1* gene is not sufficient to transfer the self-incompatibility phenotype. Mol Gen Genet **263,** 648-654.

Bloor, S.J., and Abrahams, S. (2002). The structure of the major anthocyanin in *Arabidopsis thaliana*. Phytochemistry **59,** 343-346.

Böhm, H., Boeing, H., Hempel, J., Raab, B., and Kroke, A. (1998). Flavonols, flavone and anthocyanins as natural antioxidants of food and their possible role in the prevention of chronic diseases. Zeitschrift für Ernahrungswissenschaft **37,** 147-163.

Booij-James, I.S., Dube, S.K., Jansen, M.A., Edelman, M., and Mattoo, A.K. (2000). Ultraviolet-B radiation impacts light-mediated turnover of the photosystem II reaction center heterodimer in Arabidopsis mutants altered in phenolic metabolism. Plant Physiol **124,** 1275-1284.

Borevitz, J.O., Xia, Y., Blount, J., Dixon, R.A., and Lamb, C. (2000). Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. Plant Cell **12,** 2383-2394.

Bowles, D., Lim, E.K., Poppenberger, B., and Vaistij, F.E. (2006). Glycosyltransferases of lipophilic small molecules. Annu Rev Plant Biol **57,** 567-597.

Boyd, R.K. (1993). Quantitative trace analysis by combined chromatography and mass spectrometry using external and internal standards. Rapid Commun. Mass Spectrom. **7**.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical Biochemistry **72,** 248-254.

Brown, D.E., Rashotte, A.M., Murphy, A.S., Normanly, J., Tague, B.W., Peer, W.A., Taiz, L., and Muday, G.K. (2001). Flavonoids act as negative regulators of auxin transport in vivo in arabidopsis. Plant Physiol **126,** 524-535.

Burbulis, I.E., Iacobucci, M., and Shirley, B.W. (1996). A Null Mutation in the First Enzyme of Flavonoid Biosynthesis Does Not Affect Male Fertility in Arabidopsis. Plant Cell **8,** 1013-1025.

Chang, Y.-L., Henriquez, X., Preuss, D., Copenhaver, G., and Zhang, H.-B. (2003). A plant-transformation-competent BIBAC library from the *Arabidopsis thaliana* Landsberg ecotype for functional and comparative genomics. TAG Theoretical and Applied Genetics **106,** 269-276.

Chapple, C., Vogt, T., Ellis, B.E., and Somerville, C.R. (1992). An Arabidopsis Mutant Defective in the General Phenylpropanoid Pathway. Plant Cell **4,** 1413-1424.

Clarke, J.H., Mithen, R., Brown, J.K., and Dean, C. (1995). QTL analysis of flowering time in *Arabidopsis thaliana*. Mol Gen Genet **248,** 278-286.

Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. Plant J **16,** 735-743.

D'Auria, J.C., and Gershenzon, J. (2005). The secondary metabolism of *Arabidopsis thaliana*: growing like a weed. Curr Opin Plant Biol **8,** 308-316.

Dana, C.D., Bevan, D.R., and Winkel, B.S. (2006). Molecular modeling of the effects of mutant alleles on chalcone synthase protein structure. J Mol Model **12,** 905-914.

Debeaujon, I., Peeters, A.J.M., Léon-Kloosterziel, K.M., and Koornneef, M. (2001). The T*RANSPARENT TESTA12* Gene of Arabidopsis Encodes a Multidrug Secondary Transporter-like Protein Required for Flavonoid Sequestration in Vacuoles of the Seed Coat Endothelium. The Plant Cell **13,** 853–872.

Debeaujon, I., Nesi, N., Perez, P., Devic, M., Grandjean, O., Caboche, M., and Lepiniec, L. (2003). Proanthocyanidin-accumulating cells in Arabidopsis testa: regulation of differentiation and role in seed development. Plant Cell **15,** 2514-2531.

den Dunnen, J.T., and Antonarakis, S.E. (2001). Nomenclature for the description of human sequence variations. Hum Genet **109,** 121-124.

Deslandes, L., Pileur, F., Liaubet, L., Camut, S., Can, C., Williams, K., Holub, E., Beynon, J., Arlat, M., and Marco, Y. (1998). Genetic characterization of *RRS1*, a recessive locus in *Arabidopsis thaliana* that confers resistance to the bacterial soilborne pathogen *Ralstonia solanacearum*. Mol Plant Microbe Interact **11,** 659-667.

Edwards, K., Johnstone, C., and Thompson, C. (1991). A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic Acids Research **19,** 1349.

El-Din El-Assal, S., Alonso-Blanco, C., Peeters, A.J., Raz, V., and Koornneef, M. (2001). A QTL for flowering time in Arabidopsis reveals a novel allele of *CRY2*. Nat Genet **29,** 435-440.

Feldbrügge, M., Sprenger, M., Dinkelbach, M., Yazaki, K., Harter, K., and Weisshaar, B. (1994). Functional analysis of a light-responsive plant bZIP transcriptional regulator. The Plant Cell **6,** 1607-1621.

Fiehn, O. (2002). Metabolomics - the link between genotypes and phenotypes. Plant Molecular Biology **48,** 155-171.

Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R.N., and Willmitzer, L. (2000). Metabolite profiling for plant functional genomics. Nat Biotechnol **18,** 1157-1161.

Fitz Gerald, J.N., Lehti-Shiu, M.D., Ingram, P.A., Deak, K.I., Biesiada, T., and Malamy, J.E. (2006). Identification of quantitative trait loci that regulate Arabidopsis root system size and plasticity. Genetics **172,** 485-498.

Focks, N., Sagasser, M., Weisshaar, B., and Benning, C. (1999). Characterization of *tt15*, a novel *transparent testa* mutant of *Arabidopsis thaliana* (L.) Heynh. Planta **208,** 352-357.

Foster, R., Izawa, T., and Chua, N.-H. (1994). Plant bZIP proteins gather at ACGT elements. FASEB Journal **8,** 192-200.

Frary, A., and Hamilton, C.M. (2001). Efficiency and stability of high molecular weight DNA transformation: an analysis in tomato. Transgenic Res **10,** 121-132.

Gaskell, S.J. (1997). Electrospray: Principles and Practice. Journal of Mass Spectrometry **32,** 677-688.

Gornall, R.J., and Bohm, B.A. (1978). Angiosperm Flavonoid Evolution: A Reappraisal. Systematic Botany **3,** 353-368.

Graham, T.L. (1998). Flavonoid and flavonol glycoside metabolism in Arabidopsis. Plant Physiology and Biochemistry **36,** 135-144.

Griffing, B., and Scholl, R.L. (1991). Qualitative and Quantitative Genetic Studies of *Arabidopsis thaliana*. Genetics **129,** 605-609.

Grotewold, E. (2006). The Science of Flavonoids. (New York, NY, USA: Springer Science + Bussiness Media, Inc.).

Hamilton, C.M. (1997). A binary-BAC system for plant transformation with high-molecular-weight DNA. Gene **200,** 107-116.

Hamilton, C.M., Frary, A., Lewis, C., and Tanksley, S.D. (1996). Stable transfer of intact high molecular weight DNA into plant chromosomes. Proc Natl Acad Sci U S A **93,** 9975-9979.

Hartmann, U., Valentine, W.J., Christie, J.M., Hays, J., Jenkins, G.I., and Weisshaar, B. (1998). Identification of UV/blue light-response elements in

the *Arabidopsis thaliana* chalcone synthase promoter using a homologous protoplast transient expression system. Plant Molecular Biology **36,** 741-754.

Hemm, M.R., Rider, S.D., Ogas, J., Murry, D.J., and Chapple, C. (2004). Light induces phenylpropanoid metabolism in Arabidopsis roots. Plant J **38,** 765-778.

Jacobs, M., and Rubery, P.H. (1988). Naturally occuring auxin transport regulators. Science **241,** 346-349.

Jander, G., Norris, S.R., Rounsley, S.D., Bush, D.F., Levin, I.M., and Last, R.L. (2002). Arabidopsis map-based cloning in the post-genome era. Plant Physiol **129,** 440-450.

Jansen, R.C. (1993). Interval mapping of multiple quantitative trait loci. Genetics **135,** 205-211.

Jansen, R.C. (1996). Complex plant traits: Time for polygenic analysis. Trends in Plant Science **1,** 89-94.

Jefferson, R.A., Kavanagh, T.A., and Bevan, M.W. (1987). GUS fusions: β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. The EMBO Journal **6,** 3901-3908.

Jin, H., Cominelli, E., Bailey, P., Parr, A., Mehrtens, F., Jones, J., Tonelli, C., Weisshaar, B., and Martin, C. (2000). Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in Arabidopsis. EMBO J **19,** 6150-6161.

Johnson, C.S., Kolevski, B., and Smyth, D.R. (2002). *TRANSPARENT TESTA GLABRA2*, a Trichome and Seed Coat Development Gene of Arabidopsis, Encodes a WRKY Transcription Factor. The Plant Cell **14,** 1359-1375.

Jones, P., Messner, B., Nakajima, J.-I., Schaffner, A.R., and Saito, K. (2003). UGT73C6 and UGT78D1, Glycosyltransferases Involved in Flavonol Glycoside Biosynthesis in *Arabidopsis thaliana*. J. Biol. Chem. **278,** 43910-43918.

Kerhoas, L., Aouak, D., Cingoz, A., Routaboul, J.M., Lepiniec, L., Einhorn, J., and Birlirakis, N. (2006). Structural characterization of the major flavonoid glycosides from *Arabidopsis thaliana* seeds. J Agric Food Chem **54,** 6603-6612.

Keurentjes, J.J., Fu, J., de Vos, C.H., Lommen, A., Hall, R.D., Bino, R.J., van der Plas, L.H., Jansen, R.C., Vreugdenhil, D., and Koornneef, M. (2006). The genetics of plant metabolism. Nat Genet **38,** 842-849.

Kitamura, S., Shikazono, N., and Tanaka, A. (2004). *TRANSPARENT TESTA 19* is involved in the accumulation of both anthocyanins and proanthocyanidins in Arabidopsis. Plant J **37,** 104-114.

Kliebenstein, D.J., Gershenzon, J., and Mitchell-Olds, T. (2001a). Comparative Quantitative Trait Loci Mapping of Aliphatic, Indolic and Benzylic Glucosinolate Production in *Arabidopsis thaliana* Leaves and Seeds. Genetics **159,** 359-370.

Kliebenstein, D.J., Lambrix, V.M., Reichelt, M., Gershenzon, J., and Mitchell-Olds, T. (2001b). Gene duplication in the diversification of secondary metabolism: tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in Arabidopsis. Plant Cell **13,** 681-693.

Kliebenstein, D.J., Kroymann, J., Brown, P., Figuth, A., Pedersen, D., Gershenzon, J., and Mitchell-Olds, T. (2001c). Genetic Control of Natural

Variation in Arabidopsis Glucosinolate Accumulation. Plant Physiol. **126,** 811-825.

Knepper, J. (2007). Etablierung eines transienten Expressionssystems durch Agrobakterien-vermittelten Gentransfer und Anwendung auf Fragen der Flavonoidgenregulation. In Department of Biology (Bielefeld, Germany: Universitaet Bielefeld), pp. 69.

Koes, R.E., Quattrocchio, F., and Mol, J.N.M. (1994). The flavonoid biosynthetic pathway in plants: function and evolution. BioEssays **16,** 123-132.

Koncz, C., and Schell, J. (1986). The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of Agrobacterium binary vector. Molecular General Genetics **204,** 383-396.

Konieczny, A., and Ausubel, F.M. (1993). A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers. Plant J **4,** 403-410.

Koornneef, M. (1981). The complex syndrome of *ttg* mutants. Arabidopsis Information Services **18,** 45-51.

Koornneef, M. (1990). Mutations affecting the testa colour in Arabidopsis. Arabidopsis Information Services **27,** 1-4.

Koornneef, M., Rolff, E., and Spruit, C.J.P. (1980). Genetic control of light-inhibited hypocotyl elongation in *Arabidopsis thaliana* (L.) HEYNH. Zeitschrift für Pflanzenphysiologie **100,** 147-160.

Koornneef, M., Alonso-Blanco, C., and Vreugdenhil, D. (2004). NATURALLY OCCURRING GENETIC VARIATION IN ARABIDOPSIS THALIANA. Annual Review of Plant Biology **55,** 141-172.

Koornneef, M., Alonso-Blanco, C., Peeters, A.J.M., and Soppe, W. (1998). GENETIC CONTROL OF FLOWERING TIME IN ARABIDOPSIS. Annual Review of Plant Physiology and Plant Molecular Biology **49,** 345-370.

Kosarev, P., Mayer, K.F., and Hardtke, C.S. (2002). Evaluation and classification of RING-finger domains encoded by the Arabidopsis genome. Genome Biol **3,** RESEARCH0016.

Lambrix, V., Reichelt, M., Mitchell-Olds, T., Kliebenstein, D.J., and Gershenzon, J. (2001). The Arabidopsis epithiospecifier protein promotes the hydrolysis of glucosinolates to nitriles and influences Trichoplusia ni herbivory. Plant Cell **13,** 2793-2807.

Lander, E., Green, P., Abrahamson J., Barlow A., Daley, M., Lincoln S., Newburg L. (1987). MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics **1,** 174-181.

Lander, E.S., and Botstein, D. (1989). Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121,** 185-199.

Le Gall, G., Metzdorff, S.B., Pedersen, J., N., B.R., and Colquhoun, I.J. (2005). Metabolite profiling of *Arabidopsis thaliana* (L.) plants transformed with an antisense chalcone synthase gene. Metabolomics **1,** 181-198.

Lee, J.S., Kim, D.H., Liu, K.-H., Oh, T.K., and Lee, C.H. (2005). Identification of flavonoids using liquid chromatography with electrospray ionization and ion trap tandem mass spectrometry with an MS/MS libray. Rapid Communication in Mass Spectrometry **19,** 3539-3548.

Lepiniec, L., Debeaujon, I., Routaboul, J.M., Baudry, A., Pourcel, L., Nesi, N., and Caboche, M. (2006). Genetics and Biochemistry of Seed Flavonoids. Annu Rev Plant Biol.

Leyman, B., Van Dijck, P., and Thevelein, J.M. (2001). An unexpected plethora of trehalose biosynthesis genes in *Arabidopsis thaliana*. Trends Plant Sci **6,** 510-513.

Li, J., Ou-Lee, T.M., Raba, R., Amundson, R.G., and Last, R.L. (1993). Arabidopsis Flavonoid Mutants Are Hypersensitive to UV-B Irradiation. Plant Cell **5,** 171-179.

Liang, M., Davis, E., Gardner, D., Cai, X., and Wu, Y. (2006). Involvement of *AtLAC15* in lignin synthesis in seeds and in root elongation of Arabidopsis. Planta.

Lim, E.K.A., D.A. Hou, B. Jackson, R.G. Bowles, D.J. (2004). Arabidopsis glycosyltransferases as biocatalysts in fermentation for regioselective synthesis of diverse quercetin glucosides. Biotechnol Bioeng. **85,** 623-631.

Lincoln, S.E., Daly, M.J., and Lander, E.S. (1992). Mapping genes controlling quantitative traits with MAPMAKER/QTL 1.1. 2.ed. In A Whitehead Institute for Biomedical Research Technical Report Second Edition (Cambridge: Whitehead Institute for Biomedical Research).

Lincoln, S.E., Daly, M.J., and Lander, E.S. (1993). Constructing Genetic Linkage Maps with MAPMAKER/EXP Version 3.0: A Tutorial and Reference Manual. In A Whitehead Institute for Biomedical Research Technical Report Third Edition (Cambridge: Whitehead Institute for Biomedical Research).

Lister, C., and Dean, C. (1993). Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. The Plant Journal **4,** 745-750.

Liu, C.-J., Blount, J.W., Steele, C.L., and Dixon, R.A. (2002). Bottlenecks for metabolic engineering of isoflavone glycoconjugates in Arabidopsis. PNAS **99,** 14578-14583.

Liu, P.P., Koizuka, N., Homrichhausen, T.M., Hewitt, J.R., Martin, R.C., and Nonogaki, H. (2005). Large-scale screening of Arabidopsis enhancer-trap lines for seed germination-associated genes. Plant J **41,** 936-944.

Loudet, O., Chaillou, S., Camilleri, C., Bouchez, D., and Daniel-Vedele, F. (2002). Bay-0 x Shahdara recombinant inbred line population: a powerful tool for the genetic dissection of complex traits in Arabidopsis. Theor Appl Genet **104,** 1173-1184.

Marles, M.A., Ray, H., and Gruber, M.Y. (2003). New perspectives on proanthocyanidin biochemistry and molecular regulation. Phytochemistry **64,** 367-383.

McCaig, B.C., Meagher, R.B., and Dean, J.F. (2005). Gene structure and molecular analysis of the laccase-like multicopper oxidase (LMCO) gene family in *Arabidopsis thaliana*. Planta **221,** 619-636.

Mehrtens, F., Kranz, H., Bednarek, P., and Weisshaar, B. (2005). The Arabidopsis Transcription Factor MYB12 Is a Flavonol-Specific Regulator of Phenylpropanoid Biosynthesis. Plant Physiol. **138,** 1083-1096.

Meksem, K., Ishihara, H., and Jesse, T. (2005). Chapter 9. Integration of Physical and Genetic Maps. In The Handbook of Plant Genome Mapping: Genetic and Physical Mapping, K.M. (Editor) and G.K. (Editor), eds (Weiheim, Germany: WILLY-VCH Verlag GmbH & Co. KGaA), pp. 215 - 232.

Murphy, A., Peer, W.A., and Taiz, L. (2000). Regulation of auxin transport by aminopeptidases and endogenous flavonoids. Planta **211,** 315-324.

Muzac, I., Wang, J., Anzellotti, D., Zhang, H., and Ibrahim, R.K. (2000). Functional expression of an Arabidopsis cDNA clone encoding a flavonol 3'-O-methyltransferase and characterization of the gene product. Arch Biochem Biophys **375,** 385-388.

Nesi, N., Jond, C., Debeaujon, I., Caboche, M., and Lepiniec, L. (2001). The Arabidopsis *TT2* Gene Encodes an R2R3 MYB Domain Protein That Acts as a Key Determinant for Proanthocyanidin Accumulation in Developing Seed. Plant Cell **13,** 2099-2114.

Nesi, N., Debeaujon, I., Jond, C., Pelletier, G., Caboche, M., and Lepiniec, L. (2000). The *TT8* gene encodes a basic helix-loop-helix domain protein required for expression of *DFR* and *BAN* genes in arabidopsis siliques. The Plant Cell **12,** 1863-1878.

Nesi, N., Debeaujon, I., Jond, C., Stewart, A.J., Jenkins, G.I., Caboche, M., and Lepiniec, L. (2002). The *TRANSPARENT TESTA16* locus encodes the ARABIDOPSIS BSISTER MADS domain protein and is required for proper development and pigmentation of the seed coat. The Plant Cell **14,** 2463-2479.

Nordborg, M., Hu, T.T., Ishino, Y., Jhaveri, J., Toomajian, C., Zheng, H., Bakker, E., Calabrese, P., Gladstone, J., Goyal, R., Jakobsson, M., Kim, S., Morozov, Y., Padhukasahasram, B., Plagnol, V., Rosenberg, N.A., Shah, C., Wall, J.D., Wang, J., Zhao, K., Kalbfleisch, T., Schulz, V., Kreitman, M., and Bergelson, J. (2005). The pattern of polymorphism in *Arabidopsis thaliana*. PLoS Biol **3,** e196.

Olsson, L.C., Veit, M., Weissenböck, G., and Bornman, J.F. (1998). Differential flavonoid response to enhanced UV-B radiation in brassica napus. Phytochemistry **49,** 1021-1028.

Osbourn, A.E. (1996). Preformed Antimicrobial Compounds and Plant Defense against Fungal Attack. Plant Cell **8,** 1821-1831.

Osterlund, M.T., Hardtke, C.S., Wei, N., and Deng, X.W. (2000). Targeted destabilization of HY5 during light-regulated development of Arabidopsis. Nature **405,** 462-466.

Paquette, S., Moller, B.L., and Bak, S. (2003). On the origin of family 1 plant glycosyltransferases. Phytochemistry **62,** 399-413.

Peer, W.A., Brown, D.E., Tague, B.W., Muday, G.K., Taiz, L., and Murphy, A.S. (2001). Flavonoid Accumulation Patterns of *Transparent Testa* Mutants of Arabidopsis. Plant Physiol. **126,** 536-548.

Pelletier, M.K., and Shirley, B.W. (1996). Analysis of Flavanone 3-Hydroxylase in Arabidopsis Seedlings (Coordinate Regulation with Chalcone Synthase and Chalcone Isomerase). Plant Physiol. **111,** 339-345.

Pelletier, M.K., Murrell, J.R., and Shirley, B.W. (1997). Characterization of Flavonol Synthase and *Leucoanthocyanidin Dioxygenase* Genes in Arabidopsis (Further Evidence for Differential Regulation of "Early" and "Late" Genes). Plant Physiol. **113,** 1437-1445.

Peters, J.L., Constandt, H., Neyt, P., Cnops, G., Zethof, J., Zabeau, M., and Gerats, T. (2001). A Physical Amplified Fragment-Length Polymorphism Map of Arabidopsis. Plant Physiol. **127,** 1579-1589.

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

Pourcel, L., Routaboul, J.M., Kerhoas, L., Caboche, M., Lepiniec, L., and Debeaujon, I. (2005). *TRANSPARENT TESTA10* Encodes a Laccase-Like Enzyme Involved in Oxidative Polymerization of Flavonoids in Arabidopsis Seed Coat. Plant Cell **17,** 2966-2980.

Robards, K., and Antolovich, M. (1997). Analytical Chemistry of Fruit BioflavonoidsA Review. Analyst **122,** 11R-34R.

Rosso, M.G., Li, Y., Strizhov, N., Reiss, B., Dekker, K., and Weisshaar, B. (2003). An *Arabidopsis thaliana* T-DNA mutagenised population (GABI-Kat) for flanking sequence tag based reverse genetics. Plant Molecular Biology **53,** 247-259.

Routaboul, J.M., Kerhoas, L., Debeaujon, I., Pourcel, L., Caboche, M., Einhorn, J., and Lepiniec, L. (2006). Flavonoid diversity and biosynthesis in seed of *Arabidopsis thaliana*. Planta **224,** 96-107.

Ryan, K.G., Swinny, E.E., Winefield, C., and Markham, K.R. (2001). Flavonoids and UV photoprotection in Arabidopsis mutants. Z Naturforsch [C] **56,** 745-754.

Sagasser, M., Lu, G.-H., Hahlbrock, K., and Weisshaar, B. (2002). *A. thaliana TRANSPARENT TESTA 1* is involved in seed coat development and defines the WIP subfamily of plant zinc finger proteins. Genes & Development **16,** 138-149.

Saijo, Y., Sullivan, J.A., Wang, H., Yang, J., Shen, Y., Rubio, V., Ma, L., Hoecker, U., and Deng, X.W. (2003). The COP1-SPA1 interaction defines a critical step in phytochrome A-mediated regulation of HY5 activity. Genes Dev **17,** 2642-2647.

Sambrook, J., and Russel, D.W. (2001). Molecular Cloning: A Laboratory Manual Third Edition. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).

Schad, M., Mungur, R., Fiehn, O., and Kehr, J. (2005). Metabolic profiling of laser microdissected vascular bundles of *Arabidopsis thaliana*. Plant Methods **1,** 2.

Schmid, K.J., Sorensen, T.R., Stracke, R., Torjek, O., Altmann, T., Mitchell-Olds, T., and Weisshaar, B. (2003). Large-Scale Identification and Analysis of Genome-Wide Single-Nucleotide Polymorphisms for Mapping in *Arabidopsis thaliana*. Genome Res. **13,** 1250-1257.

Schoenbohm, C., Martens, S., Eder, C., Forkmann, G., and Weisshaar, B. (2000). Identification of the *Arabidopsis thaliana* Flavonoid 3'-Hydroxylase Gene and Functional Expression of the Encoded P450 Enzyme. Biol Chem **381,** 749-753.

Sharbel, T.F., Haubold, B., and Mitchell-Olds, T. (2000). Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe. Mol Ecol **9,** 2109-2118.

Sharma, S.B., and Dixon, R.A. (2005). Metabolic engineering of proanthocyanidins by ectopic expression of transcription factors in *Arabidopsis thaliana*. Plant J **44,** 62-75.

Sheahan, J.J. (1996). Sinapate esters provide greater UV-B attenuation than flavonoids in *Arabidopsis thaliana* (Brassicaceae). American Journal of Botany **83,** 679-686.

Sheahan, J.J., and Rechnitz, G.A. (1992). Flavonoid-Specific Staining Of *Arabidopsis thaliana*. Biotechniques **13,** 880-883.

Sheahan, J.J., and Rechnitz, G.A. (1993). Differential Visualization of Transparent Testa Mutants in *Arabidopsis thaliana*. Analytical Chemistry **65,** 961-963.

Shirley, B.W. (1996). Flavonoid biosynthesis: 'new' functions for an 'old' pathway. Trends in Plant Science **1,** 377-382.

Shirley, B.W., Hanley, S., and Goodman, H.M. (1992). Effects of ionizing radiation on a plant genome: analysis of two Arabidopsis *transparent testa* mutations. The Plant Cell **4,** 333-347.

Shirley, B.W., Kubasek, W.L., Storz, G., Bruggemann, E., Koornneef, M., Ausubel, F.M., and Goodman, H.M. (1995). Analysis of Arabidopsis mutants deficient in flavonoid biosynthesis. The Plant Journal **8,** 659-671.

Song, X.J., Huang, W., Shi, M., Zhu, M.Z., and Lin, H.X. (2007). A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. Nat Genet **39,** 623-630.

Sprenger-Haussels, M., and Weisshaar, B. (2000). Transactivation properties of parsley proline rich bZIP transcription factors. The Plant Journal **22,** 1-8.

Stobiecki, M., and Kachlicki, P. (2006). Isolaiton and Identification of Flavonoids. In The Science of Flavonoids, E. Grotewold, ed (Springer Science + Bussiness Media, Inc.), pp. 47-69.

Stobiecki, M., Skirycz, A., Kerhoas, L., Kachlicki, P., Muth, D., Einhorn, J., and Mueller-Roeber, B. (2006). Profiling of phenolic glycosidic conjugates in leaves of *Arabidopsis thaliana* using LC/MS. Metabolomics **2,** 197-219.

Stracke, R., Ishihara, H., Huep, G., Barsch, A., Mehrtens, F., Niehaus, K., and Weisshaar, B. (2007). Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. Plant J **50,** 660-677.

Tanksley, S.D. (1983). Molecular markers in plant breeding. Plant Mol. Biol. Rep. **1,** 3-8.

Tanksley, S.D. (1993). Mapping polygenes. Annu Rev Genet **27,** 205-233.

Teng, S., Keurentjes, J., Bentsink, L., Koornneef, M., and Smeekens, S. (2005). Sucrose-specific induction of anthocyanin biosynthesis in Arabidopsis requires the *MYB75/PAP1* gene. Plant Physiol **139,** 1840-1852.

The Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408,** 796-815.

Tohge, T., Nishiyama, Y., Hirai, M.Y., Yano, M., Nakajima, J.-i., Awazuhara, M., Inoue, E., Takahashi, H., Goodenowe, D.B., Kitayama, M., Noji, M., Yamazaki, M., and Saito, K. (2005b). Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. Plant J **42,** 218-235.

Tolstikov, V.V., Lommen, A., Nakanishi, K., Tanaka, N., and Fiehn, O. (2003). Monolithic silica-based capillary reversed-phase liquid chromatography/electrospray mass spectrometry for plant metabolomics. Anal Chem **75,** 6737-6740.

Torjek, O., Berger, D., Meyer, R.C., Mussig, C., Schmid, K.J., Rosleff Sorensen, T., Weisshaar, B., Mitchell-Olds, T., and Altmann, T. (2003). Establishment of a high-efficiency SNP-based framework marker set for Arabidopsis. Plant J **36,** 122-140.

Trezzini, G.F., Horrichs, A., and Somssich, I.E. (1993). Isolation of putative defense-related genes from *Arabidopsis thaliana* and expression in fungal elicitor-treated cells. Plant Molecular Biology **21,** 385-389.

Tsuji, J., Jackson, E.P., Gage, D.A., Hammerschmidt, R., and Somerville, S.C. (1992). Phytoalexin Accumulation in *Arabidopsis thaliana* during the Hypersensitive Reaction to *Pseudomonas syringae* pv *syringae*. Plant Physiol **98,** 1304-1309.

Veit, M., and Pauli, G.F. (1999). Major flavonoids from *Arabidopsis thaliana* leaves. J Nat Prod **62,** 1301-1303.

Vogel, G., Fiehn, O., Jean-Richard-dit-Bressel, L., Boller, T., Wiemken, A., Aeschbacher, R.A., and Wingler, A. (2001). Trehalose metabolism in Arabidopsis: occurrence of trehalose and molecular cloning and characterization of trehalose-6-phosphate synthase homologues. J Exp Bot **52,** 1817-1826.

Vogt, T., and Jones, P. (2000). Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. Trends in Plant Science **5,** 380-386.

Wagner, H., Bladt, Sabine. (1996). Plant Drug Analysis A Thin Layer Chromatography Atlas. (Berlin: Springer).

Walker, A.R., Davison, P.A., Bolognesi-Winfield, A.C., James, C.M., Srinivasan, N., Blundell, T.L., Esch, J.J., Marks, M.D., and Gray, J.C. (1999). The *TRANSPARENT TESTA GLABRA1* locus, which regulates trichome differentiation and anthocyanin biosynthesis in Arabidopsis, encodes a WD40 repeat protein. The Plant Cell **11,** 1337-1350.

Wang, S., Basten, CJ, Zeng, ZB. (2005). Windows qtl cartographer. (Raleigh, NC.: North Carolina State University).

Weisshaar, B., Jenkins, G.I. (1998). Phenylpropanoid biosynthesis and its regulation. Curr Opin Plant Biol **1,** 251-257.

Wilson, I.W., Schiff, C.L., Hughes, D.E., and Somerville, S.C. (2001). Quantitative trait loci analysis of powdery mildew disease resistance in the *Arabidopsis thaliana* accession kashmir-1. Genetics **158,** 1301-1309.

Wilson, Z.A. (2000). Arabidopsis: A Practical Approach. (New York: Oxford University Press).

Winkel-Shirley, B. (2001). Flavonoid Biosynthesis. A Colorful Model for Genetics, Biochemistry, Cell Biology, and Biotechnology. Plant Physiology **126,** 485-493.

Winkel-Shirley, B. (2002). Biosynthesis of flavonoids and effects of stress. Curr Opin Plant Biol **5,** 218-223.

Wisman, E., Hartmann, U., Sagasser, M., Baumann, E., Palme, K., Hahlbrock, K., Saedler, H., and Weisshaar, B. (1998). Knock-out mutants from an *En-1* mutagenized *Arabidopsis thaliana* population generate new phenylpropanoid biosynthesis phenotypes. Proceedings of the National Academy of Sciences of the United States of America **95,** 12432-12437.

Woo, H.H., Jeong, B.R., Koo, K.B., Choi, J.W., Hirsch, A.M., and Hawes, M.C. (2007). Modifying expression of closely related UDP-glycosyltransferases

from pea and Arabidopsis results in altered root development and function. Physiologia Plantarum **130,** 250-260.

Woodcock, D.M., Crowther, P.J., Doherty, J., Jefferson, S., DeCruz, E., Noyer-Weidner, M., Smith, S.S., Michael, M.Z., and Graham, M.W. (1989). Quantitative evaluation of *Escherichia coli* host strains for tolerance to cytosine methylation in plasmid and phage recombinants. Nucleic Acids Research **17,** 3469-3478.

Xu, Z., Escamilla-Trevino, L., Zeng, L., Lalgondar, M., Bevan, D., Winkel, B., Mohamed, A., Cheng, C.L., Shih, M.C., Poulton, J., and Esen, A. (2004). Functional genomic analysis of *Arabidopsis thaliana* glycoside hydrolase family 1. Plant Mol Biol **55,** 343-367.

Ylstra, B., Muskens, M., and Van Tunen, A.J. (1996). Flavonols are *not* essential for fertilization in *Arabidopsis thaliana*. Plant Molecular Biology **32,** 1155-1158.

Yonekura-Sakakibara, K., Tohge, T., Niida, R., and Saito, K. (2007). Identification of a Flavonol 7-O-Rhamnosyltransferase Gene Determining Flavonoid Pattern in Arabidopsis by Transcriptome Coexpression Analysis and Reverse Genetics. J Biol Chem.

Zhang, Z., Ober, J.A., and Kliebenstein, D.J. (2006). The gene controlling the quantitative trait locus *EPITHIOSPECIFIER MODIFIER1* alters glucosinolate hydrolysis and insect resistance in Arabidopsis. Plant Cell **18,** 1524-1536.

Appendices

TGAGGCTTTCAGTTAAATGCATAAAAAAGATTTACTAGGATATAATTTAGGATCATATTGTTTCGAAAGTATA
GAAAGAGGAGATCATACAAGTAGAATCTAATAAGGAGAAAGTAGAGCAATGGGAGAAAAGAATCCTCAAGCGA
CTTTAGGAACAATTATCTACGATCACCATAAAAAGGTATAGAGGATACACAACACACTGATTTTGCTTGGAAT
GCGCCAAGCAAACATTTTTCACTTATTCTTATTAATAATAAATTTTTTACAAGAATTCTATTAGCTCGGCAAT
TACTTGACATATTCTACCCGAATAGTCAACCTCTACTAAAAACAAATCAAAGCCTAATTCTTTCATCCGTTTC
TTATCACACACCTGGGAAGAGAAAGAGAATTACAAACACCACTTTTTAAGCTTTTTCTAGGCTTCGAAAGCTA
TACACGTATGGCAATATTCTAGAGTGTTTCGGCGGCTCAATGCCTATTTCTAGAAAATTTATACTATTCTCCA
AGAATCGATGAGCGGGATCTTTCTTGAACGTCAAACCTATCTCTTTTCTTGTTGGGATTTTCCTTGATTTCCA
AGCCTATCACTTTCCTTGTTTGAGATCCTTCTTGATATTTAAGTCTATCTCTTTCCTTATTTGTTGAGAAGAA
TATTCACCAAATCTTTTCAACCGACTTTCGTCTTTGATCACCACGCTAAGTCAGCAAAGCGTCTTGGCACGTC
ACATCAACCACTCCAGCTCCATCATGTCAGCTTACGCGTCCATGTCAGCAATTCAGACGTTTGATGTAGATTT
GATGCAGCTTCCTGATATCCAAGAGTTTCCTAATGTCCACCATCTCGTTGTCCTGATATCCATCAACTTGTTA
TTCTGATATCTAACAGCTCCTGATATCCATCAGCTCGTTGTTGTCAGAATCTGCATCTTCACACCACACCCAT
AATCCATTGCTAGTGAAATTTTTGAGTAGGCTTGATCATTGTCCTTCCTCCTTGGGTCCTTGAGAGGATTCAT
CACCCATCACATAACCATATACTCCCATGTCAAACACAGCAAATTGATTGCTTACTAAGACACTGTTCTGCAC
AGTCTCATTTGCCAATGAACCTGAATGAGGAGGAGTAGCTAGGATAATTACCAAGGCTGGAGATCACTATTTA
TTTTTTTATCTTCCATTTTGCTATAATCTATAGAAGATGACATGAAGAGGTTTGACAAACAAAAAAAAAAAAC
AATATCATGAAGAGAAGCCATTGGGCTAAAATCCAAAAAAAAAAAAAAAAAATTTGCTACCAAACAAATTAACA
ATGTTTATAGGGGAACATTTCGTCCGGTTATTAAAGTTTTCCAAATTAGAAAAAGAAAAAAAAAAGAAGATTC
TTTGAAGTTCACCAGTTTAAGTTTCTCGTAATGCATTAGAAATCGACATTTCACCAGTCAAAAATCTTGGACA
AATATTCGCAAAATTAGTTTTAAAAACATTCTTTCAGTAAAATGTAAACAATTAAAAAAGAAACATCTCTTTC
GTTTCTTTTATTGTTAGCGAAACATCTCTTTCGTAAAAAAAGAAAATTGTTTTCAGGAATCTTAAAATTCTAT
ACGTTAATACGTACGTTACGTACATGTAGCTTTGGCGTGGAATGGTAGCTCCCCGACATCTAATTAGATGATA
ACTCGTTTTCATTTAATTCAGTGAACAATGTAAACTATGCACGTAAGAAACTAAATTAACACAAAAACGTAAC
GAGTTTCGGAATATTTCTGGTGTTGGTTTATATAAAATAAAGGTAATGCTTGGCCTAACAAAACTCATAAGTG
ATTTTATTTTATCTTTAGTAAAGACTTTGAACTAGTTGATTGATGTCACGTCTCTTACTATATATGATATAAC
ATGTGAAAATATAATCTAGAGTTAGATAAATACGACTTTACTGTTTTCGTTAACTTGAAAGTTTCGTTTGACC
AAGAAATGTAGTTAAATATTTGTGGGAATGTGAACTCATTAGTCAAAACGGACCAAGAAGATGATGCTTACCG
AAAACATATTTTGTAAAATACAAAATTCTTATATCCGAATGTAACTAATTTGAGAAAGACGGTCTCTATACGT
ATATAAAATATATGTATATATGCCTCTCTATTCTTTGATGAACTTCAGTCAGGGTAATTAATTTACTAAAACT
CTTCCAAAATG

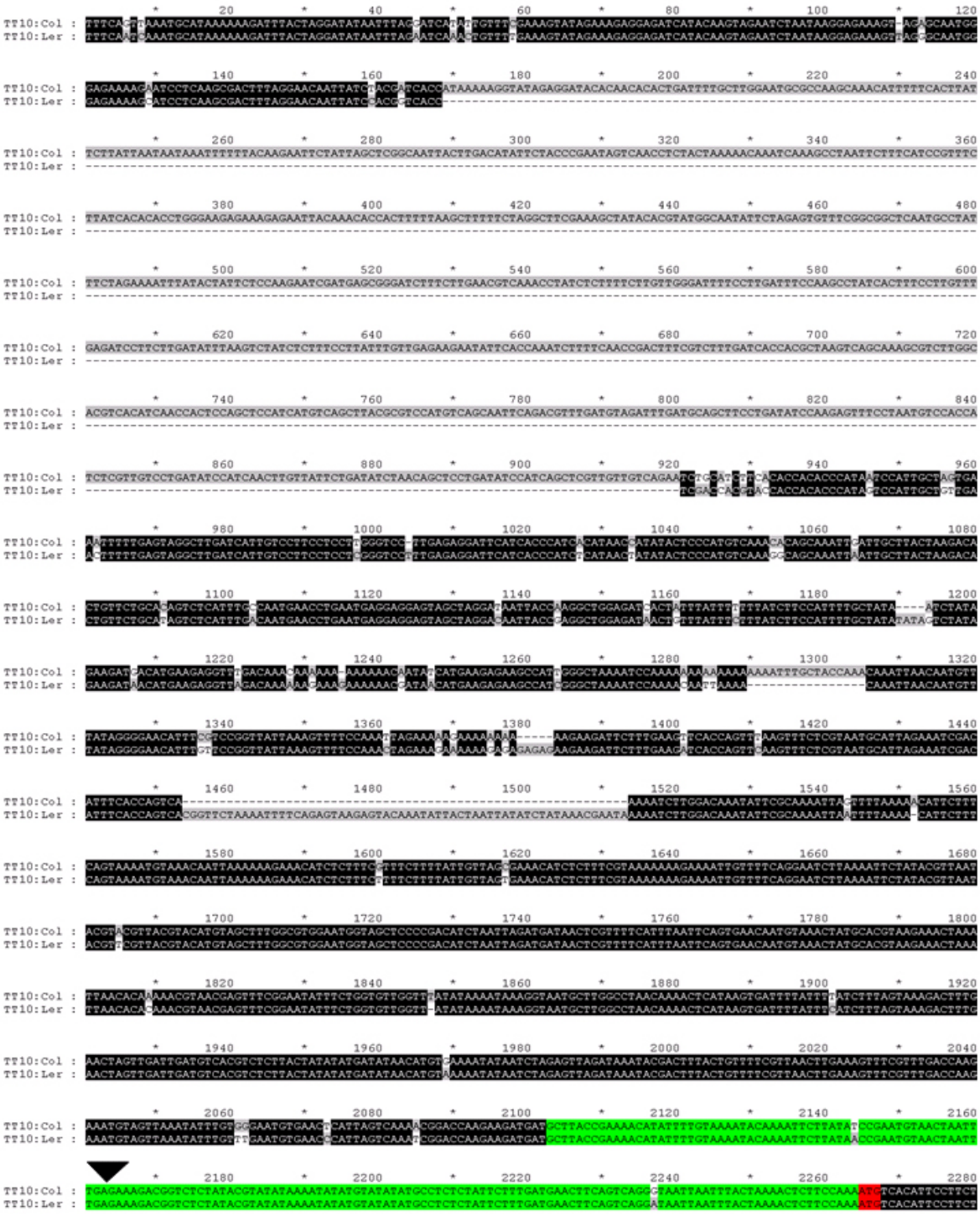
Appendix 1. The 2.2 Kb *TT10* promoter fragment from the accession Columbia.

Large Insertion (more than 5 bp) in the Pro*TT10*Col in comparison to Pro*TT10*L*er* are highlighted with a black background with white letters. The arrows with the gray highlights indicate large deletion sites (more than 5 bp). The start codon was marked with bold font and boxed. The yellow highlight with the gray triangle indicates the T-DNA insertion site at *GK_146E10*. Detailed information about the *TT10* 5' UTR sequence comparison between Col and L*er* is listed in Appendix 3.

AGAGGCTTTCAATCAAATGCATAAAAAAGATTTACTAGGATATAATTTAGAATCAAACTGTTTTGAAAGTATA
GAAAGAGGAGATCATACAAGTAGAATCTAATAAGGAGAAAGTTAGGGCAATGGGAGAAAAGCATCCTCAAGCG
ACTTTAGGAACAATTATCCACGGTCACCTCGACCACGTACCACCACACCCATAGTCCATTGCTGTTGAACTTT
TTGAGTAGGCTTGATCATTGTCCTTCCTCCTCGGGTCCTTTGAGAGGATTCATCACCCATCTCATAACTATAT
ACTCCCATGTCAAAGGCAGCAAATTAATTGCTTACTAAGACACTGTTCTGCATAGTCTCATTTGACAATGAAC
CTGAATGAGGAGGAGTAGCTAGGACAATTACCGAGGCTGGAGATAACTGTTTATTTCTTTATCTTCCATTTTG
CTATATATAGTCTATAGAAGATAACATGAAGAGGTTAGACAAAAAAGAAAGAAAAAACGATAACATGAAGAGA
AGCCATCGGGCTAAAATCCAAAACAATTAAAACAAATTAACAATGTTTATAGGGGAACATTTGTTCCGGTTAT
TAAAGTTTTCCAAACTAGAAAGAAAAAAGAGAGAGAGAAGAAGATTCTTTGAAGATCACCAGTTCAAGTTTCT
CGTAATGCATTAGAAATCGACATTTCACCAGTCACGGTTCTAAAATTTTCAGAGTAAGAGTACAAATATTACT
AATTATATCTATAAACGAATAAAAATCTTGGACAAATATTCGCAAAATTAATTTTAAAACATTCTTTCAGTAA
AATGTAAACAATTAAAAAAGAAACATCTCTTTCTTTTCTTTTATTGTTAGTGAAACATCTCTTTCGTAAAAAA
AGAAAATTGTTTTCAGGAATCTTAAAATTCTATACGTTAATACGTTCGTTACGTACATGTAGCTTTGGCGTGG
AATGGTAGCTCCCCGACATCTAATTAGATGATAACTCGTTTTCATTTAATTCAGTGAACAATGTAAACTATGC
ACGTAAGAAACTAAATTAACACACAAACGTAACGAGTTTCGGAATATTTCTGGTGTTGGTTATATAAAATAAA
GGTAATGCTTGGCCTAACAAAACTCATAAGTGATTTTATTTCATCTTTAGTAAAGACTTTGAACTAGTTGATT
GATGTCACGTCTCTTACTATATATGATATAACATGTAAAAATATAATCTAGAGTTAGATAAATACGACTTTAC
TGTTTTCGTTAACTTGAAAGTTTCGTTTGACCAAGAAATGTAGTTAAATATTTGTTTGAATGTGAACCCATTA
GTCAAATCGGACCAAGAAGATGATGCTTACCGAAAACATATTTTGTAAAATACAAAATTCTTATAACCGAATG
TAACTAATTTGAGAAAGACGGTCTCTATACGTATATAAAATATATGTATATATGCCTCTCTATTCTTTGATGA
ACTTCAGTCAGGATAATTAATTTACTAAAACTCTTCCAAA**ATG**

Appendix 2. 1.5 Kb *TT10* promoter fragment from the accession Landsberg *erecta*.

Large Insertion (more than 5 bp) in the Pro*TT10*L*er* in comparison to Pro*TT10*Col are highlighted with a black background with white letters. The arrows with the gray highlights indicate deletion sites (more than 5 bp). The start codon was marked with bold font and boxed. Detailed information about the*TT10* 5' UTR sequence comparison between Col and L*er* is listed in Appendix 3.



Appendix 3. Comparison of *TT10* promoter fragments between Col and L*er.*
2.2 Kb Col and 1.5 Kb L*er* promoter fragment were aligned using ClustalW and fine-tuned using GeneDoc. Start codons are highlighted in red. 5'UTR sequences based on Col full length cDNA sequence are highlighted in green. The back triangle indicates the T-DNA insertion site at *GK_146E10*.

NASC Nr.	Lister & Dean	Qr4 (in ng)	H379	NASC Nr.	Lister & Dean	Qr4 (in ng)	H379	NASC Nr.	Lister & Dean	Qr4 (in ng)	H379
1900	CL4	20.40	L	1934	CL161	0.00	C	1968	CL283	11.68	L
1901	CL5	3.59	C	1935	CL166	-	L	1969	CL284	0.00	C
1902	CL7	-	-	1936	CL167	34.17	L	1970	CL288	0.00	C
1903	CL13	22.61	L	1937	CL173	17.05	L	1971	CL295	0.00	C
1904	CL14	74.15	L	1938	CL175	0.00	-	1972	CL296	57.37	L
1905	CL17	0.00	-	1939	CL177	0.00	C	1973	CL297	21.55	L
1906	CL19B	0.00	C	1940	CL179	28.59	L	1974	CL302	5.58	C
1907	CL25	-	L	1941	CL180	0.00	C	1975	CL303	0.00	-
1908	CL29	0.00	C	1942	CL181	-	L	1976	CL311	0.00	C
1909	CL30	49.40	L	1943	CL182	4.23	-	1977	CL321	86.23	L
1910	CL32	107.06	L	1944	CL188	10.23	L	1978	CL332	0.00	C
1911	CL33	0.00	C	1945	CL190	0.00	C	1979	CL342	26.40	L
1912	CL34	21.29	L	1946	CL191	67.78	L	1980	CL345	49.50	L
1913	CL25	0.00	-	1947	CL193	0.00	C	1981	CL349	*	-
1914	CL36	37.00	L	1948	CL194	16.15	L	1982	CL350	16.45	C
1915	CL37	54.42	L	1949	CL199	12.72	L	1983	CL351	0.00	C
1916	CL46	16.97	L	1950	CL209	-	C	1984	CL356	67.74	L
1917	CL52	0.00	C	1951	CL217	35.76	L	1985	CL358	0.00	C
1918	CL54	52.68	L	1952	CL214	32.02	L	1986	CL359	34.13	L
1919	CL59	10.45	L	1953	CL231	0.00	C	1987	CL363	0.00	C
1920	CL62	-	C	1954	CL232	0.00	C	1988	CL367	0.00	L
1921	CL67	0.00	C	1955	CL235	14.03	L	1989	CL370	53.90	L
1922	CL68	-	L	1956	CL237	7.42	C	1990	CL377	0.00	C
1923	CL71	112.49	L	1957	CL238	0.00	C	1991	CL378	21.73	L
1924	CL79	166.17	L	1958	CL240	32.72	L	1992	CL386	72.93	L
1925	CL84	23.64	-	1959	CL242	0.00	C	1993	CL390	48.28	L
1926	CL90	33.47	L	1960	CL245	20.88	C	1994	CL394	16.20	C
1927	CL107	41.36	L	1961	CL253	8.10	C	1995	CL395	62.26	-
1928	CL113	92.06	L	1962	CL257	8.17	C	1996	CL397	11.43	-
1929	CL115	0.00	-	1963	CL263	0.00	-	1997	CL398	0.00	-
1930	CL123	0.00	C	1964	CL264	56.34	L	1998	CL400	9.24	C
1931	CL125	21.01	-	1965	CL266	0.00	C	1999	CL259	139.31	L
1932	CL131	0.00	-	1966	CL267	0.00	C	4686	CL53	-	-
1933	CL160	53.30	L	1967	CL279	12.26	C	Ave.		24.39	

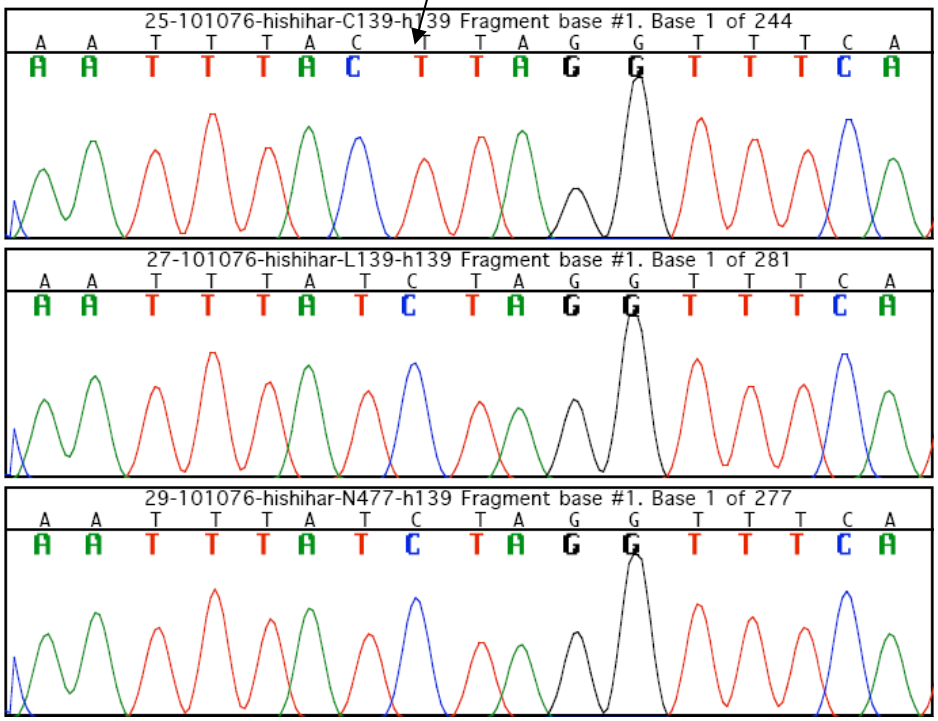
Appendix 4. Quercitrin quantitative and segregation data of a molecular marker H379 from the Lister and Dean L*er* x Col RIL.

The quantitative data (in ng per 100 µg DW) was obtained from HPLC-PDA. Ave.- Average amount of quercitrin found in the population. Genotypes: Col/C, L*er*/L and no data/-

NASC Nr.	Lister & Dean	Km1 & Qr1	NASC Nr.	Lister & Dean	Km1 & Qr1	NASC Nr.	Lister & Dean	Km1 & Qr1
1900	CL4	C	1934	CL161	C	1968	CL283	C
1901	CL5	C	1935	CL166	C	1969	CL284	L
1902	CL7	-	1936	CL167	C	1970	CL288	C
1903	CL13	C	1937	CL173	L	1971	CL295	C
1904	CL14	C	1938	CL175	L	1972	CL296	L
1905	CL17	L	1939	CL177	C	1973	CL297	C
1906	CL19B	L	1940	CL179	C	1974	CL302	C
1907	CL25	L	1941	CL180	C	1975	CL303	L
1908	CL29	C	1942	CL181	L	1976	CL311	C
1909	CL30	C	1943	CL182	C	1977	CL321	C
1910	CL32	L	1944	CL188	C	1978	CL332	C
1911	CL33	C	1945	CL190	L	1979	CL342	C
1912	CL34	C	1946	CL191	L	1980	CL345	C
1913	CL25	C	1947	CL193	C	1981	CL349	-
1914	CL36	C	1948	CL194	L	1982	CL350	L
1915	CL37	C	1949	CL199	C	1983	CL351	C
1916	CL46	C	1950	CL209	-	1984	CL356	L
1917	CL52	L	1951	CL217	L	1985	CL358	C
1918	CL54	C	1952	CL214	C	1986	CL359	L
1919	CL59	L	1953	CL231	C	1987	CL363	C
1920	CL62	-	1954	CL232	C	1988	CL367	C
1921	CL67	C	1955	CL235	C	1989	CL370	C
1922	CL68	L	1956	CL237	L	1990	CL377	C
1923	CL71	L	1957	CL238	C	1991	CL378	C
1924	CL79	C	1958	CL240	C	1992	CL386	C
1925	CL84	C	1959	CL242	C	1993	CL390	C
1926	CL90	C	1960	CL245	C	1994	CL394	C
1927	CL107	C	1961	CL253	C	1995	CL395	L
1928	CL113	C	1962	CL257	L	1996	CL397	L
1929	CL115	L	1963	CL263	C	1997	CL398	-
1930	CL123	C	1964	CL264	L	1998	CL400	L
1931	CL125	L	1965	CL266	C	1999	CL259	L
1932	CL131	L	1966	CL267	L	4686	CL53	-
1933	CL160	C	1967	CL279	C			

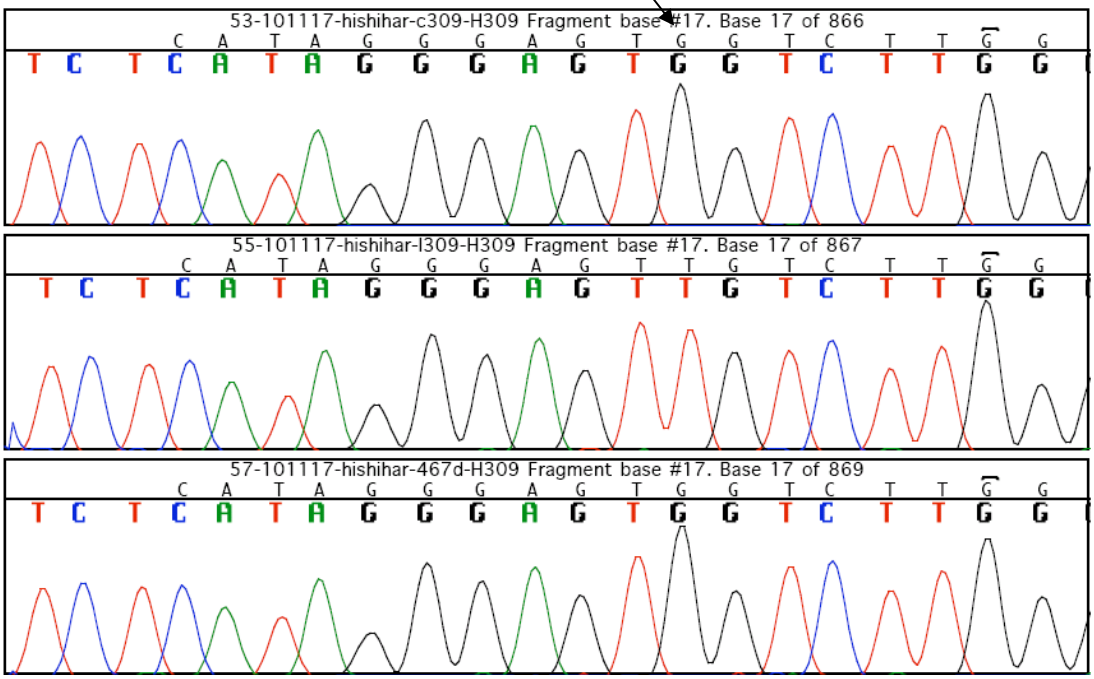
Appendix 5. Segregation data of the Km1 and Qr1 metabolite polymorphic marker in the Lister and Dean L*er* x Col RIL.
Genotypes: Col/C, L*er*/L and no data/-

```
H139_Col    AGATAAATATTATTTTTAATTATGGGTCTCTTAAATTTGGTGAAACTAAAACCAATTATC
H139_Ler    AGATAAATATTATTTTTAATTATGGGTTTCTTAAATTTGGTGAAACCAAAACCAATTATC
H139_N4774  AGATAAATATTATTTTTAATTATGGGTTTCTTAAATTTGGTGAAACCAAAACCAATTATC
            *************************** ****************** *************

H139_Col    TCTTAATTAAGAGACAATTTACTTAGGTTTCATCAAATTTAAGAGACTTATGTAATATTT
H139_Ler    TCTTAACTAAGAGACAATTTATCTAGGTTTCATCAAATTTAAGAGACTTATGTAATATTT
H139_N4774  TCTTAACTAAGAGACAATTTATCTAGGTTTCATCAAATTTAAGAGACTTATGTAATATTT
            ****** ************** ************************************

H139_Col    AAATCATATAAAATTTTTATTATTTGTTAAGAAACTTGTTTAGAAACCCCTATTGAGGGT
H139_Ler    AAATCATATAAATTTTTTATTATTTGTTAAGAAACTTGTTTAGAAACCCCTATTGAGGTT
H139_N4774  AAATCATATAAATTTTTTATTATTTGTTAAGAAACTTGTTTAGAAACCCCTATTGAGGTT
            ************ ********************************************* *

H139_Col    GCTCTTATTGTTTATTTTTTTCTTTCTTTGTATTCAATTTTTTATGATTACTATCTCTCC
H139_Ler    GCTCTTATTGTTTCTTTTTTTCTTTCTTTGTATTCAATTTTTTATGATTACTATCTCTCC
H139_N4774  GCTCTTATTGTTTCTTTTTTTCTTTCTTTGTATTCAATTTTTTATGATTACTATCTCTCC
            ************* ********************************************
```

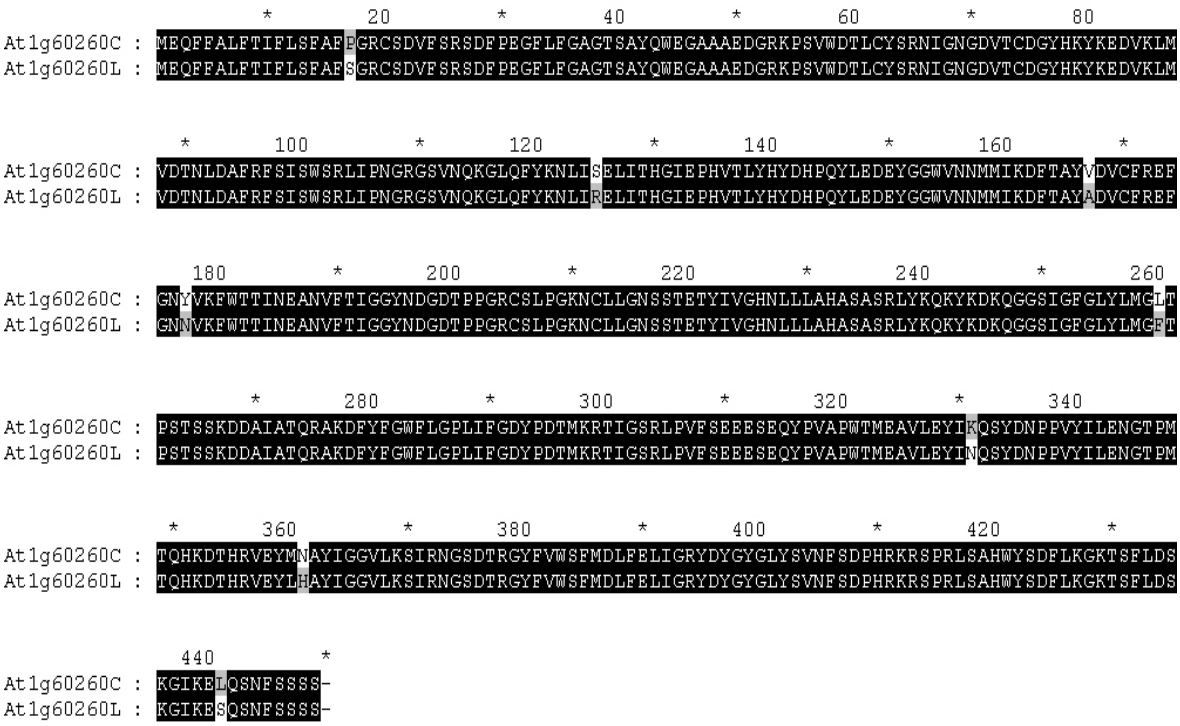


Appendix 6. Genotyping of N4774 using the marker H139 (At1g60420)
Genotypes of SNPs are highlighted with Columbia (red) and Landsberg *erecta* (blue). The allow points the position of the SNP at the sequence chromatogram.

```
H309_Col     GACATGACAATCTGGTACATAATAGAATGAGATGTATTGTAATTGCTCAACTTCTTTCTC
H309_Ler     GACATGACAATCTGGTACATAATAGAATGAGATGTATTGTAATTGCTCAACTTCTTTCTC
H309_N4674   GACATGACAATCTGGTACATAATAGAATGAGATGTATTGTAATTGCTCAACTTCTTTCTC
             ************************************************************

H309_Col     TCATAGGGAGTGGTCTTGGGAAAAATCTCCATCACAAAGTTCAAGGCGCCGTAAGGTATT
H309_Ler     TCATAGGGAGTTGTCTTGGGAAAAATCTCCATCACAAAGTTCAAGGCGCCGTAAGGTATT
H309_N4674   TCATAGGGAGTGGTCTTGGGAAAAATCTCCATCACAAAGTTCAAGGCGCCGTAAGGTATT
             ********** ***************************************************

H309_Col     CTTGCTTACTCCCGCTACTGTATATCTTGCAATTGCAGTTTTTACGTAGTCATTATAGTC
H309_Ler     CTTGCTTACTCCCGCTACTGTATATCTTGCAATTGCAGTTTTTACGTAGTCATTATAGTC
H309_N4674   CTTGCTTACTCCCGCTACTGTATATCTTGCAATTGCAGTTTTTACGTAGTCATTATAGTC
             ************************************************************

H309_Col     CTTAAGAAATTTACACCAGCAGAAGCATGACTCATTTTCTAAACCTTCTTGTTATCTCCC
H309_Ler     CTTAAGAAATTTACACCAGCAGAAGCATGACTCATTTTCTAAACCTTCTTGTTATCTCCC
H309_N4674   CTTAAGAAATTTACACCAGCAGAAGCATGACTCATTTTCTAAACCTTCTTGTTATCTCCC
             ************************************************************
```

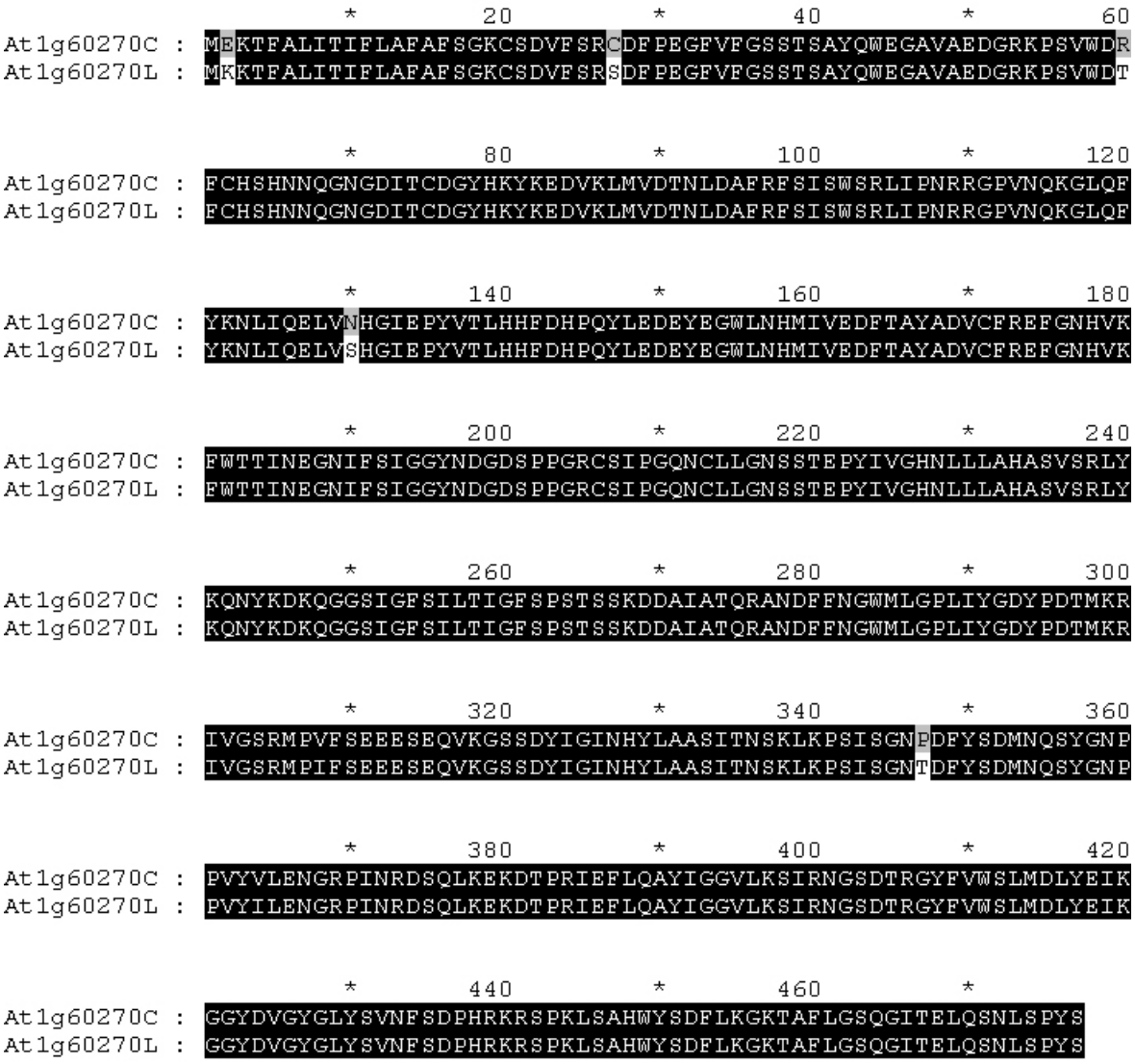


Appendix 7. Genotyping of N4674 using the marker H309 (At1g60220)
Genotypes of SNPs are highlighted with Col (red) and L*er* (blue). The allow points the position of the SNP at the sequence chromatogram.

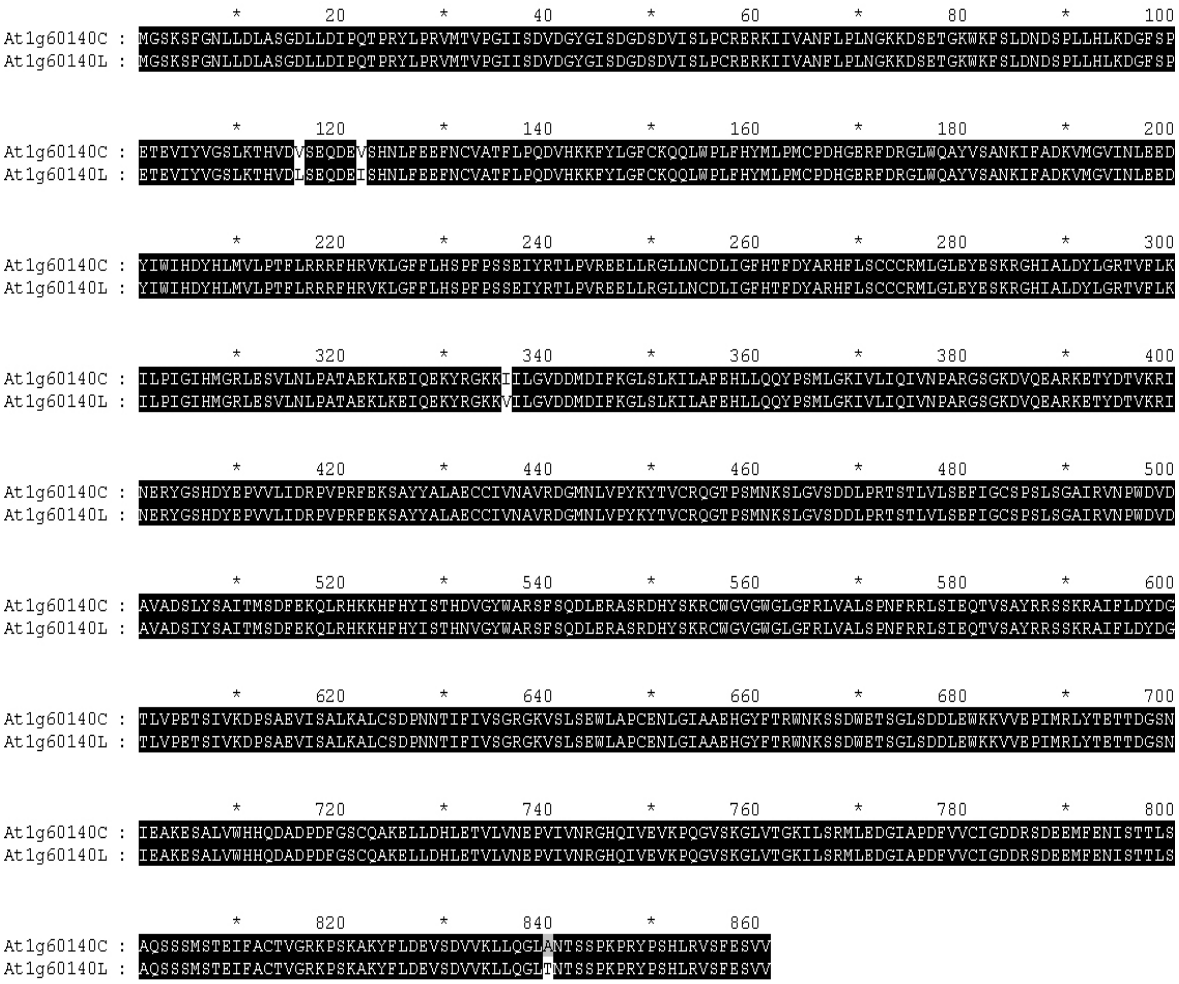


Appendix 8. Comparison of *At1g60260* predicted amino acid sequences between Col and L*er.*

Genotypes of the sequences are indicated with Col/C and L*er*/L.

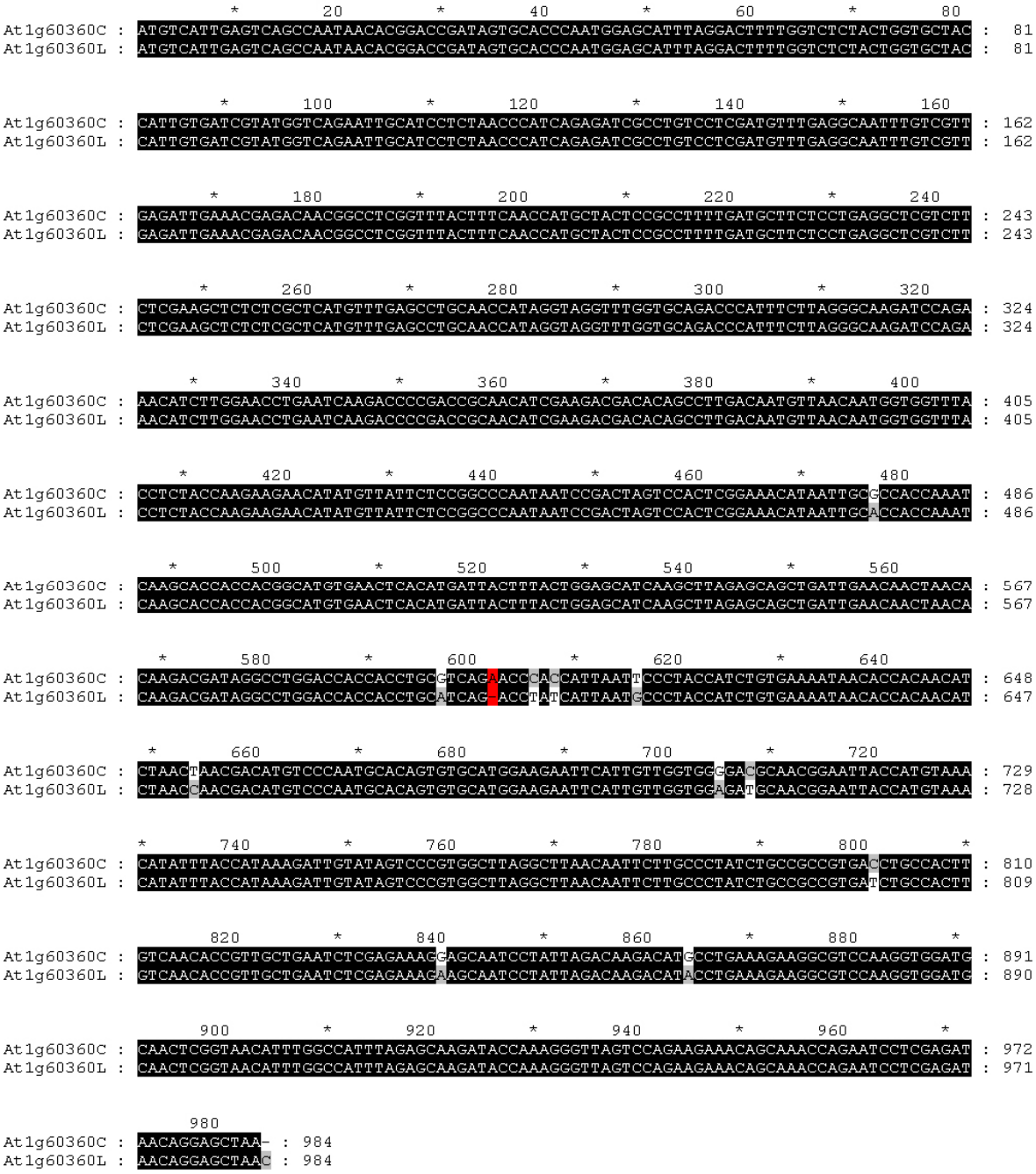


Appendix 9. Comparison of *At1g60270* predicted amino acid sequence between Col and L*er*.
Genotypes of the sequences are indicated with Col/C and L*er*/L.



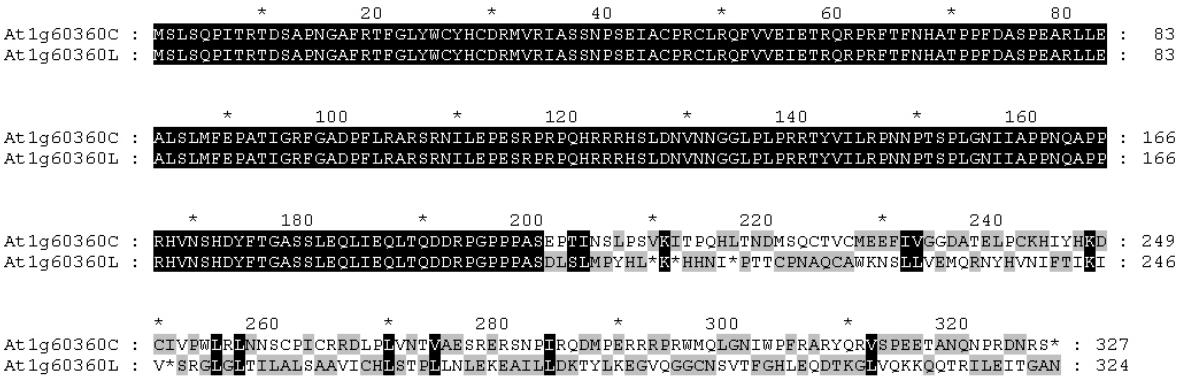
Appendix 10. Comparison of *At1g60140* predicted amino acid sequences between Col and L*er*.

Genotypes of the sequences are indicated with Col/C and L*er*/L.



Appendix 11. Comparison of *At1g60360* predicted ORF sequences between Col and L*er*.

A single nucleotide deletion is marked with red highlight. Genotypes of the sequences are indicated with Col/C and L*er*/L.



Appendix 12. Comparison of *At1g60360* predicted amino acid sequences between Col and L*er*.

Star signs within the sequence indicate stop codons. Genotypes of the sequences are indicated with Col/C and L*er*/L.

List of Publication

Stracke, R., Ishihara, H., Huep, G., Barsch, A., Mehrtens, F., Niehaus, K., and Weisshaar, B. (2007). Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. Plant J **50,** 660-677.

Curriculum Vitae

Name: Hirofumi Ishihara

Personal Information: Nationality: Japanese.
Date of Place: Osaka, Japan
Date of Birth: July 30, 1974.

Education:

Since 11/2003	PhD student at Bielefeld University, Germany International NRW Graduate School in Bioinformatics and Genome Research Center of Biotechnology
01/2001 – 05/2003	Southern Illinois University at Carbondale, USA Master of Science: May 2003 Genomics, Genetics and Biotechnology Lab Department of Plant, Soil and General Agriculture,
08/1997 - 12/2000	Southern Illinois University at Carbondale, USA Bachelor of Arts: December 2000 Major: Microbiology Minor: Chemistry
04/1994 – 03/1996	Kyoto Institute of Technology, Japan Major: Applied Biology (Transferred to Southern Illinois University at Carbondale after completing the sophomore year)

Acknowledgements

Finally, I would like to express my gratitude to people who help me to complete this work.

I would like to thank to Prof. Dr. Bernd Weisshaar for giving me an opportunity to work with him and introducing me to Germany. I want to thank to Prof. Dr. Karsten Niehaus for his kindness and giving me various interests toward not only in science but also in photography.

I want to thank to people in Chair of Genome Research. I thank to Dr. Martin Sagasser for correcting my Japanglish and giving me various advises in writing. My gratitude goes to Dr. Ralf Stracke for his help in scientific writing. I want to acknowledge Dr. Mario Rosso for his good and sincere advices. I also like to thank to PD. Dr. Thomas Merkle for offering his help. I would like to thank to Melli for providing teasty coffee every morning. Thank you also to Gunnar, Julian, SandraF, and Rashmi for providing their help in reading my writing.

I'd like to thank to Gunnar Huep for providing the transfection data. I also want to thank to Dr. Aiko Barsch for his excellent work in GC-MS analysis and providing good data. I'd like to thank to Dr. Maren Heese and Dr. Heiko Schoof for their work in annotation of the L*er* sequence. My special thanks go to Prisca Viehoever for her excellent work in the preparation of the BAC screening filters, and the sequencing and the assembling of the L*er* genomic sequence.

My heartfelt gratitude goes to two couples Aiko & Telli and Steven & Simone for their continuous supports, and sincere friendships. I could enjoy being here because of their presence. I also thank to Mr. and Mrs. Watt for their kindness and always inviting me to their family events. My thanks also go to my German mother Johanna Bode and her family, without their help I could have not survived in Germany. I want to thank to Dr. Rajsree Mungur for being my friend and her continuous help. Thank you also to Gregor Obernosterer for his friendship.

My special thanks go to my family for their continuous support.

And last but not least, I would like to thank to Prisca. Without her support and loving care, I could not reach this point.

Promotionsausschuss der Fakultät für Biologie

Herewith I affirm that I wrote the presented dissertation independently only using the cited references as an aid.

Furthermore I affirm that the presented dissertation was not submitted at any other faculty either partly or complete to obtain an academic title.

Herewith I apply to obtain the title Doctor of Natural Sciences (Dr. rer. Nat.) from the Faculty of Biology in the Bielefeld University for the first time. My supervisor is Pro. Dr. Bernd Weisshaar.

Bielefeld, ____/_____/_____ _____

Hirofumi Ishihara