Genetic bases of plant performance in different environmental scenarios using natural variation in *Arabidopsis thaliana*

PhD Thesis
Nadine Ilk

Genetic bases of plant performance in different environmental scenarios using natural variation in *Arabidopsis thaliana*

Inaugural-Dissertation
zur
Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln
vorgelegt von

Nadine Ilk
aus Herten

Jahr der Veröffentlichung 2012

Die vorliegende Arbeit wurde am MPI für Pflanzenzüchtungsforschung in Köln- Vogelsang in der Arbeitsgruppe von Dr. Matthieu Reymond und unter der Betreuung von Prof. Dr. Maarten Koornneef durchgeführt.

Berichterstatter (Gutachter): Prof. Dr. Maarten Koornneef
Prof. Dr. Ute Höcker

Tag der mündlichen Prüfung: 21.06.2011

Abstract

Plant performance is the result of many individual processes which are influenced by environmental factors. Flowering time and the production of secondary metabolites such as anthocyanins affect plant performance. These specific and other related traits were quantified on a new recombinant inbred line population of *Arabidopsis thaliana* derived from a cross between L*er* (Poland) and Eri-1 (Sweden). Altogether, 110 recombinant inbred lines were grown under two contrasting environmental scenarios: i) high light intensity and low temperature (HL4 for High Light 4°C) and ii) commonly used light intensity and temperature when growing *Arabidopsis thaliana* in growth chambers. This allowed us to detect QTL (Quantitative Trait Loci). In addition, Eri-1 accession was sequenced to obtain polymorphisms between the parental accessions.

Mapped QTL indicated that L*er*/Eri-1 variation in flowering time and/or anthocyanin accumulation is mainly caused by three interacting QTL. One of these QTL is located on chromosome III and is involved in flowering time in both tested conditions. Selected lines were phenotyped for validation and fine mapping of this QTL, which lead to the identification of two closely linked additive QTL. Candidate genes are proposed. Other QTL is located on chromosome V and is involved in both flowering time and anthocyanin accumulation under HL4 conditions. The validation and fine mapping lead to the identification of the gene *HUA2*, a pre-mRNA processing factor, as being responsible for flowering time variation observed for this QTL. *HUA2* has previously been shown to positively regulate a MADS box gene *FLC*, which is involved in the regulation of flowering time. In addition, expression analyses strongly suggest that *HUA2* is also responsible for the variation of anthocyanin accumulation. Finally, in accordance with epistatic interactions and sequence variation between the parental accessions, we propose a model that explains how *HUA2* could be involved in the regulation of late anthocyanin biosynthesis genes via the MYB-bHLH-WD40 transcriptional activation complex in HL4 conditions.

Sequencing accessions of *Arabidopsis thaliana* has become more accessible, facilitating the identification of candidate genes for detected QTL, as could be shown in this study. This project reveals a link between flowering time and secondary metabolism, highlighting the power of using natural variation to dissect the genetic architecture of responses of metabolic pathways to environmental scenarios.

Zusammenfassung

Pflanzenwachstum wird von verschiedensten Stoffwechselprozessen beeinflusst, welche unter den ständigen Auswirkungen von Umweltfaktoren stehen. Blühzeit und die Produktion von Sekundärmetaboliten (z.B. Anthocyane) sind zwei Fakoren, die Einfluß auf die Ausprägung des Pflanzenwachstums nehmen. Diese beiden und weitere verwandte Merkmale wurden mit Hilfe einer rekombinanten Inzuchtlinienpopulation von *Arabidopsis thaliana* quantitativ erfasst. Die Population stammt aus einer Kreuzung zwischen den Ökotypen L*er* (Polen) und Eri-1 (Schweden). Das Wachstum von insgesamt 110 rekombinanten Inzuchtlinien erfolgte in zwei unterschiedlichen Umweltbedingungen: (i) in Starklicht und niedriger Temperatur (4°C; „HL4“) und (ii) unter für *Arabidopsis thaliana* üblichen Standardbedingungen. Wir waren dadurch in der Lage, verschiedene QTL zu detektieren. Um Polymorphismen zwischen den elterlichen Linien bestimmen zu können, wurde eine vollständige Sequenzierung der Linie Eri-1 durchgeführt.

Durch die Kartierung der QTL fanden wir Hinweise, dass die beobachteten Unterschiede zwischen L*er*/Eri-1 in Blühzeit und/oder Anthocyangehalt vorwiegend durch drei interagierende QTL hervorgerufen werden. Ein auf Chromosom III liegendes QTL nimmt Einfluss auf die Blühzeit in beiden Umweltbedingungen. Um dieses QTL zu validieren und genauer zu kartieren, erfolgte eine Phänotypisierung ausgewählter Nachkommen. Dieses führte zur Identifizierung von zwei nah beieinander liegenden, additiven QTL. Es werden verschiedene Gene vorgestellt, die dem QTL Effekt zu Grunde liegen könnten. Ein weiteres, auf Chromosom V lokalisierter QTL, nimmt Einfluss auf die Blühzeit und den Anthocyangehalt unter HL4 Bedingungen. Innerhalb diesem QTL konnte im Verlauf des Validierungsprozesses und der genaueren Kartierung das Gen *HUA2* als genetische Grundlage für die unterschiedlichen Blühzeiten identifiziert werden.

Es ist bekannt, dass *HUA2*, ein prä-mRNA prozessierender Faktor, eine positive Regulierung des MADS-Box Gens *FLC* bewirkt, welches seinerseits Einfluss auf Blühzeiten nimmt. Zudem haben Expressionsanalysen gezeigt, dass *HUA2* in direktem Zusammenhang mit dem Anthocyangehalt steht. In Übereinstimmung mit epistatischen Interaktionen und Sequenzunterschieden zwischen den parentalen Linien stellen wir ein Modell vor, in dem *HUA2* die Regulation der späten Biosynthese von Anthocyanen, abhängig vom MYB-bHLH-WD40 Transkriptionsaktivierungskomplex unter HL4 Bedingungen, beeinflusst. Sowohl die ständig steigende Menge an Sequenzinformation

über *Arabidopsis thaliana*, als auch die vereinfachte Durchführbarkeit von neuen Sequenzierungen, sowie der kostengünstige Zugang zu Sequenzierungsplattformen, erleichtert, wie auch in dieser Arbeit, die Identifizierung von kausalen Genen innerhalb eines ermittelten QTL. Die innerhalb dieser Arbeit vorgestellte Verbindung zwischen Blühzeit und Sekundärstoffwechsel veranschaulicht in welcher Weise die natürliche Variation innerhalb einer Spezies genutzt werden kann, um genetische Strukturen zu erforschen. Die genetische Vielfalt ist vor allem vor dem Hintergrund der Plastizität von Stoffwechselwegen wichtig, die Einfluss auf das Pflanzenwachstum in unterschiedlichen Umweltbedingungen haben.

Contents

I.	Introduction	1
I.1.	Plant performance determinisms	1
I.2	Different factors influence plant performance traits	3
I.2.A.	Response of plant performance to environmental factors and the effects of oxidative stress	3
I.2.B.	Genetic influences on plant performance	5
I.2.C.	G x E interaction affecting plant performance	5
I.3.	*Arabidopsis thaliana*	6
I.3.A.	*Arabidopsis thaliana*: a model plant	6
I.3.B.	Mutant approach in *Arabidopsis thaliana*	7
I.3.C.	Natural variation in *Arabidopsis thaliana*	8
I.4.	Analysis of the genetic and molecular bases of natural variation	9
I.4.A.	QTL analysis and genome wide association mapping	9
I.4.B.	Elucidating the genetic basis of a QTL	11
I.4.BA.	Mapping populations commonly developed to detect QTL	11
I.4.BB.	Molecular markers for genotyping mapping populations	13
I.4.BC.	Toward the identification of the causal polymorphism(s) : QTL validation and fine mapping	14
I.4.BD.	Candidate genes in QTL region and complementation	16
I.5.	State of the art concerning genetic determinism of flowering time in *Arabidopsis thaliana*	17
I.6.	State of the art concerning genetic and molecular determinisms of anthocyanin accumulation in *Arabidopsis thaliana*	19
	Thesis Aim	24
II.	Materials and Methods	25
II.1.	Growth conditions and experimental design	25
II.2.	Plant Material	27

II.2.A.	Mapping population	27
II.2.B.	Other genetic materials	28
II.2.C.	Growing Eri-1 for DNA extraction in order to sequence its genome	28
II.3.	Molecular analyses	29
II.3.A.	Deoxyribonucleic acid extraction	29
II.3.B.	Revealing polymorphism between L*er*/Eri-1 using PCR based markers	29
II.3.BA.	PCR programs and gel-electrophoresis	29
II.3.BB.	Designing polymorphic markers between L*er* and Eri-1	31
II.3.BC.	Digestion of CAPS markers	32
II.3.BD.	Sequencing single regions/candidate genes	33
II.3.C.	Whole genome sequencing of Eri-1 with Illumina platform	33
II.3.D.	Genotyping of the L*er*/Eri-1 RIL population and building a genetic map	35
II.3.E.	Selection of Near Isogenic Lines	36
II.3.F.	Gene expression analysis	37
II.4.	Trait quantification for detection, validation and characterisation of QTL3, QTL5-1 and QTL5-2	38
II.4.A.	Quantification of flowering time related traits	38
II.4.B.	Quantification of biochemical – pigmentation related traits	39
II.5.	Statistical analyses	43
II.5.A.	Statistics	43
II.5.B.	QTL mapping	43
II.5.C.	Two-way marker interaction	44
III.	Results	45
III.1.	Genotyping of the RILs and construction of a L*er*/Eri-1 genetic map	45
III.2.	Phenotyping the RILs of the L*er*/Eri-1 mapping population	48
III.2.A.	Genetic variation in flowering time and number of leaves	48
III.2.B.	Genetic variation in pigmentation and photosynthetic related traits	49

III.2.C.	Phenotypic correlations between flowering-, pigmentation- and photosynthetic related traits over the two studied environments	51
III.3.	QTL mapping	54
III.3.A.	Detection of QTL involved in the variation of flowering, pigmentation and photosynthetic related traits in the CT and HL4 environments	54
III.3.B.	2-way QTL interactions mapping	59
III.3.C.	QTL x environment interactions	62
III.4.	Validation and characterization of the flowering QTL3	62
III.4.A.	Selection of NILs in L*er* or Eri-1 genetic backgrounds to validate the presence and effect of flowering QTL3	62
III.4.B.	Validation of flowering QTL3	65
III.4.C.	Characterization of the validated flowering QTL3	68
III.4.D.	Fine mapping of the flowering time QTL3	69
III.4.E.	Reducing the size of the QTL3 with additional markers	72
III.4.F.	Sequencing of Eri-1 accession	74
III.4.G.	Candidate genes for flowering time – genes in the region of QTL3	79
III.5.	Validation and characterization of QTL5	81
III.5.A.	Selection of NILs in L*er* or Eri-1 genetic backgrounds to validate the presence and effect of 2 linked QTL on chromosome V	81
III.5.AA.	Selection of NILs in Eri-1 background	81
III.5.AB.	Selection of NILs in L*er* background	84
III.6.A.	Validation of the flowering and pigmentation effects of QTL5	86
III.6.B.	Characterization of validated flowering QTL5	88
III.6.C.	Validation of separated flowering and anthocyanin QTL5-1 and flowering QTL5-2	91
III.6.CA.	Validation of flowering related trait QTL5-2 in HL4 condition	91
III.6.CB.	Validation of flowering and anthocyanin QTL5-1 in HL4 condition	92
III.6.D.	Fine mapping of QTL5-1	93

III.6.F.	Candidate genes in the QTL5-1 region for flowering related traits and anthocyanin content	96
III.6.G.	Allelism test with L*er-hua2-5*, L*er-HUA2* and L*er* NILs QTL5-1 to confirm *HUA2* as a candidate for QTL5-1	103
III.6.H.	Transcript level of genes involved in the regulation of anthocyanin production or biosynthesis of anthocyanins in the lines L*er-hua2-5*, L*er-HUA2* and Eri-1	108
IV.	Discussion	111
IV.1.	Building a genetic map for the newly developed L*er*/Eri-1 population	111
IV.2.	Effect of the environmental conditions on the correlations between the measured traits	112
IV.3.	Major QTL mapped in L*er*/Eri-1 population	115
IV.3.A.	Flowering time related QTL on chromosome V	115
IV.3.B.	Flowering time related QTL on chromosome III	120
IV.3.C.	Anthocyanin content QTL on chromosome I	121
IV.3.D.	Anthocyanin content, chlorophyll content and flowering related QTL on chromosome V	122
IV.4.	Fine mapping of QTL3 and fine mapping and allelism test of QTL5-1	124
IV.4.A.	Fine mapping of flowering QTL3 and sequence analysis of new sequenced Eri-1 accession revealed candidate genes for top QTL3-1 and bottom QTL3-2	124
IV.4.B.	Fine mapping of QTL5-1 and allelism test demonstrated that *HUA2* is the gene responsible for flowering time QTL5-1 and strongly supported that *HUA2* is also responsible for anthocyanin content variation in L*er*/Eri-1 population	128
IV.5.	Two-way QTL interactions mapping present genetic networks in flowering time and anthocyanin content in L*er*/Eri-1 population.	130
IV.5.A.	Interactions involved in flowering time variation in L*er*/Eri-1 population	130
IV.5.B.	Interactions involved in anthocyanin accumulation in L*er*/Eri-1 population	132
IV.6.	Proposed function of *HUA2* in the regulation of anthocyanin production via MBW complex	134

References 138

Appendix

Acknowledgements

Erklärung

List of Figures

Figure I.1	The production and use of recombinant inbred lines and near isogenic lines	13
Figure I.2	Flowering gene network; based on Ehrenreich et al. 2009	19
Figure I.3	Flavonoid biosynthetic and accumulation pathway; based on Grotewold 2006	22
Figure I.4	Regulatory network model of TTG1 dependent pathways MBW transcription factor complex; based on Zhang et al 2003	23
Figure II.1	Contrasted environmental conditions monitored in different growth cabinets	26
Figure II.2	HOBO data (temperature, light intensity and relative humidity)	27
Figure II.3	Image of different flowering RILs and Eri-1 as an example for scoring flowering (in leaf numbers)	38
Figure II.4	Image of a PS micro plate for quantification of anthocyanins by photometric tests and schematically display of the PS micro plate	42
Figure III.1	Genetic map of the L*er*/Eri-1 RIL population genotyped with 115 markers	46
Figure III.2	Images of fragment separation of PCR products from L*er*/Eri-1 RILs populations on a 3% agarose gel	47
Figure III.3	Principle component analysis plot of flowering, pigmentation and photosynthetic related traits in CT and HL4 environments	52
Figure III.4	L*er*/Eri-1 linkage map showing the locations of QTL for the analysed traits in control and high light and cold environment.	58
Figure III.5	Heat map of the log likelihood ratio from pairwise epistatic interactions detected by Epistat for all flowering related traits and chlorophyll and anthocyanin content detected under CT and HL4 conditions	61
Figure III.6	Selection strategy and genotype of Eri NIL QTL3 and L*er* NIL QTL3	65
Figure III.7	Rosette leaf numbers and total leaf numbers of L*er*, Eri-1, and the selected NILs L*er* NIL QTL3 (L*er* NIL97-3) and Eri NIL QTL3 (Eri NIL67-4-2) in CT and HL4 environment	67

Figure III.8	Flowering time of L*er*, Eri-1, and the selected NILs L*er* NIL QTL3 (L*er* NIL97-3) and Eri NIL QTL3 (Eri NIL67-4-2) in short day light condition.	69
Figure III.9	Fine mapping of QTL3 with progeny (BC1F5) of Eri NIL QTL3 (Eri NIL67-4-10)	72
Figure III.10	Frequency distribution for RLN in progeny of the line R7 (NIL67-4-10-36) grown in CT condition.	73
Figure III.11	Per base sequence quality of sequencing Eri-1 Illumina reads	76
Figure III.12	IGV browser image of Illumina sequence reads of Eri-1 aligned to the region 485494 Bp – 489804 Bp of chromosome III of the reference genome Col-0 based on TAIR9 information	77
Figure III.13	Zoom in of Figure III.12. IGV browser image of Illumina sequence reads of Eri-1 aligned to the region 487403Bp – 487469Bp of chromosome III of the reference genome Col-0 based on TAIR9 information	78
Figure III.14	Selection strategy and genotype of Eri-1 NIL QTL5, QTL5-1 and QTL5-2	83
Figure III.15	Selection strategy and genotype of L*er* NIL QTL5 and QTL5-1	85
Figure III.16	Rosette leaf numbers and anthocyanin content of L*er*, Eri-1, and the selected Eri NIL QTL5 (NIL53) in HL4 environment	88
Figure III.17	Comparison of rosette leaf numbers and total leaf numbers of L*er*, Eri-1, and the selected NILs L*er* NILs QTL5 (NIL41(3) and NIL37(1)) and Eri NIL QTL5 (NIL53-7) in CT and HL4 environment	90
Figure III.18	Rosette leaf numbers and anthocyanin content of L*er*, Eri-1, and the selected L*er* NILs QTL5 NIL37(1); L*er* NILQTL5-1 NIL18(2) and Eri NIL QTL5 (NIL53-7); Eri-NIL QTL5-1 (NIL8-118) and Eri NIL QTL5-2 (NIL60-22) in HL4 environment	93
Figure III.19	Fine mapping of QTL5-1 with the progeny (BC2F3) of Eri-1 NIL53-8	96
Figure III.20	Validation of *HUA2* as a candidate gene for QTL5-1 for flowering and anthocyanin content	99
Figure III.21	Comparison of the predicted protein sequences of HUA2 from Col-0, Eri-1, L*er* and L*er*-*HUA2*.	101

Figure III.22 Allelism test of F_2 crosses containing *hua2-5*^{Ler} alleles and *HUA2*^{Ler} and *HUA2*$^{Eri-1}$ alleles for flowering time 105

Figure III.23 Allelism test of F_2 crosses containing *hua2-5*^{Ler} alleles and *HUA2*^{Ler} and *HUA2*$^{Eri-1}$ alleles for anthocyanin content 107

Figure III.24 Anthocyanin content of L*er-hua2-5*, L*er-HUA2*, L*er* NILs QTL5, F_2 of the cross L*er-HUA2* x L*er* NIL18(2) and F2 of L*er* NIL18(2) x L*er* NIL84(1). 108

Figure III.25 Gene expression analysis of regulatory and biosynthesis anthocyanin genes tested in L*er-hua2-5*, L*er-HUA2* and Eri-1 110

Figure IV.1 Photo of branching phenotype of Eri NIL QTL3 (Eri NIL67-4-10-26) in HL4 condition. 128

Figure IV.2 Working model depicting the possible role of HUA2 together with PAP1/TT8/TTG1 complex in transcription activation of a structural anthocyanin biosynthesis gene (*DFR*) 137

List of Tables

Table I.1 Different types of marker systems 14

Table II.1 Primer and primer combinations used in the AFLP analysis performed at Wageningen University 35

Table III.1 Parental values and RILs, averages and ranges of FT etc. and heritability in the control and in high light and low temperature condition 50

Table III.2 Pearson correlation between the traits in the analysed environments 53

Table III.3 Characteristics of the detected QTL explaining flowering time related traits, photosynthesis related traits and anthocyanin content in L*er*/Eri-1 RIL population in CT and HL4 environment 57

Table A.1 Marker used for building genetic map of L*er*/Eri-1 and validation and fine mapping of the detected QTL

Table A.2 Marker used for sequencing candidate regions and qRT-PCR primer for expression analyses

Abbreviations

A. thaliana	*Arabidopsis thaliana*
ACS	automatic cofactor selection
AFLP	amplified fragment length polymorphism
AGI	Arabidopsis Genome Initiative
AI-RIL	advanced intercross recombinant inbred line
BC	backcross
bHLH	basic helix loop helix
bp	base pairs
Bur-0	Burren (Ireland)
BWA	Burrows Wheeler Aligner
°C	degree Celsius
CAPS	cleaved amplified polymorphic sequences
CCI	chlorophyll content index
cDNA	complementary deoxyribonucleic acid
Chla	chlorophyll a
Chlb	chlorophyll b
Chr	chromosome
CLN	cauline leaf number
cM	centi Morgan
Col-0	Columbia (USA)
ct	cycle treshold
CT	Control
Cvi	Cape Verde islands
DAS	days after sowing
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
EBGs	early biosynthetic genes
EDTA	Diaminoethanetetraacetic acid
Eri-1	Eringsboda (Sweden)
Est-1	Estland (Russia)
FT	flowering time
GC	guanine-cytosine
GWA	genome wide association mapping
GxE	gene x environment
h^2	heritability
HCL	hydrochloric acid
HIF	heterogeneous inbred family
HL4	high light at four degree Celsius
HR	hypersensitive response
Hz	hertz
IL	introgression line
IM	interval mapping
indel	insertions and deletions
kbp	kilo base pairs
Kro-0	Krotzenburg (Germany)
LBGs	late biosynthetic genes
LD	long day

L*er*	Landsberg *erecta* (Polen)
LLR	log likelihood ratio
Mbp	mega base pairs
MBW	MYB-bHLH-WD40
MetOH	methanol
min	minute
MQM	multiple QTL mapping
mRNA	messenger ribonucleic acid
NIL	near isogenic line
nm	nanometre
PCA	principle component analysis
PCR	polymerase chain reaction
PCR	polymerase chain reaction
QTGs	quantitative trait genes
QTL	quantitative trait locus
QTL x E	quantitative trait locus x environment
R^2	correlation coefficient
RFLP	restriction fragment length polymorphism
RIL	recombinant inbred line
RLF	rate of leaf production
RLN	rosette leaf number
RNA	ribonucleic acid
ROS	reactive oxygen species
SD	short day
sec	second
Sha	Shakdara (Pamiro-Alay) Tadjikistan
SNK	Students-Newman-Keuls
SNP	single nucleotide polymorphism
SSLP	single sequence length polymorphism
SSR	simple sequence repeat
Suc	sucrose
TLN	total leaf number
UV	ultraviolet
V	volt
W	watt
Ws	Wassilevkajia

I. Introduction

I.1. Plant performance determinisms

Plant performance is not the result of a single parameter, but of the co-operation of many components. In plants, the performance can be defined as the result of primary production like carbon fixation, mineral uptake and photosynthesis followed by allocation, storage and use of the produced assimilates. An important output of plant performance is its fitness, usually measured as its capacity to produce numerous and viable seeds. To optimize all the above mentioned processes and to increase fitness of a plant, the transition from the vegetative growth phase to the reproductive phase has to occur at the most suitable time (Koornneef et al. 2004, Ehrenreich and Purrugganan 2006). Indeed, one way of regulating and controlling fitness of a plant is the timing of flowering. Appraising the right moment to flower at the time when plants developed the optimal quantity of sources, can increase the number of seeds. Due to the connection between source and sink organs during flowering time, this trait can be quantified as one of the key traits for the estimation of plant performance.

The capability of plants to reproduce depends on many other different processes occurring during their life cycle. One important component that significantly affects plant performance is growth during the vegetative phase. Plant growth can be affected mainly at three levels: Firstly, at cellular level where growth can be quantified by cell proliferation and/or cell expansion (Qi and John 2007). Secondly, plant growth can be measured at organ level (Reymond et al. 2003). Plant organs like roots, leaves and flowers grow to reach specific sizes and shapes that are dictated by the plants genotype and the identity of the organ. All organs contribute to plant growth and can also be estimated at the whole plant level. This includes plant size and architecture of the plant. The architecture of a plant for example can be described by the number, size and shape of the leaves (Pérez-Pérez et al. 2010). Absolute growth at the organ level is the result of direct relation between source and sink organs. During the life cycle source-sink transitions of organs occurs, as well as changes regarding to the sink strength of individual organs. This source to sink relationship involves many biochemical processes such as photosynthesis. Photosynthetic efficiency is characterized by building up photosynthates which are mainly transported as sucrose. Sucrose is relocated to sink organs such as the roots, flowers, fruits or seeds. Outcome of higher production of carbohydrates is also shown in increased growth

rate and greater biomass production due to the induce of storage metabolism genes, such as those of starch biosynthesis (Rook et al. 2006). An increased growth rate is linked to the development of chlorophylls, a higher density of chlorophylls and light harvesting proteins in the leaves that support the photosynthetic activity (Labate et al. 2004). Other important traits of plant performance are time of maturation and time of germination of seeds, as well as production of secondary metabolites like flavonoids (Willmann et al. 2011, Bentsink et al. 2010, D´Auria and Gershenzon 2005).

Plant performance can be measured at different levels of organization and by using different methods. Beside the measurement of growth by estimating the leaf weight/area etc., biochemical traits related to the chlorophyll content, anthocyanin content, sugar cycles, enzymatic activities, metabolic profiles and redistribution of carbons into seeds can also be analyzed. Finally, the measurement of timing of germination, flowering and maturation can be scored also, in days after sowing (DAS). At the same time point as scoring flowering time (FT), other related traits can be scored like the number of leaves. In addition, in the context of growth quantification, FT can also be considered as a growth related trait because it is significantly correlated with the duration of the vegetative phase (Koornneef et al. 1991, Pieper et al, 2009).

The values of the above mentioned traits can be influenced by three factors:

- environmental factors
- genetic factors
- the interaction between environmental and genetic factors, usually mentioned as G x E

The influences of these factors are introduced in the next paragraphs.

I.2. Different factors influence plant performance traits

I.2.A. Response of plant performance to environmental factors and the effects of oxidative stress

Plants are growing under variable environments and all the traits described above are modulated under the effect of particular environmental parameters, affecting plant performance. Environmental factors can be identified as biotic or abiotic factors.

An example of biotic factors influencing performance and fitness of a plant is pathogen infection. Additionally, herbivores, especially insects, can reduce seed production and other plant performance related traits. It has been demonstrated that initial herbivore damage increased levels of chemical, physical and biotic defences. Glucosinolates play an important role in these plant/biotic interactions as natural pesticides (Burow et al. 2010). Plants have developed many defending mechanisms against fungal and bacterial infections which are causing diseases and plant death. An important mechanism preventing bacterial infection is the hypersensitive response (HR). The response of resistant plants to microbial pathogens is programmed cell death in which the plant sacrifices a few cells under infection, restricting pathogen growth into adjacent healthy tissues. The hypersensitive cell death, the induced programmed cell death is one of the most efficient plant innate immunity (Greenberg 1996). Several of the features diagnostic for programmed cell death, such as nuclear condensation, DNA fragmentation and cytoplast shrinkage, have been observed in plant cells undergoing HR (Lam 2004). Programmed cell death can also be induced by oxidative stress (He et al. 2011), which is mainly a response to abiotic factors like strong light intensities and excessive energy levels.

Abiotic factors also play a major role in influencing plant performance. These factors include many different environmental variables like above mentioned light quantity and quality, temperature fluctuations, changes in air humidity, drought, CO_2 availability and nutrient deficiency. As an example the effect of light on plant development programs is striking and illustrated by extreme morphological differences between plants grown in the light (and exhibiting photomorphogenesis) and plants grown in the dark (and exhibiting skotomorphogenesis). Proper regulation of these two developmental stages is important for plants to optimize their growth (Casal et al. 2004). Photoperiod, light quality and temperature are also influencing flowering time (Li et al. 2006, Balasubramanian et al. 2006). Variations in light intensity and quality can decrease the photosynthetic efficiency

and consequently result in sugar limited conditions that down regulate biosynthetic activity and thus influence leaf expansion and flowering time. Under high light conditions the photosynthetic electron flow becomes oversaturated and can adversely affect the photosynthetic system components like antenna pigments, reaction centres and more (Adir et al. 2003). Due to the high amount of light energy plants have to deal with photo inhibition and oxidative stress (Apel and Hirt 2004). As a consequence the accumulation of reactive oxygen species (ROS) can be in increased. In plants, the link between ROS production and photosynthetic metabolism are particularly important (Rossel et al. 2002). Oxidative stress arises from an unbalance in generation and removal of ROS (superoxide, hydrogen peroxide etc.). It is well known that plants increase the amount of ascorbate and other antioxidants in order to remove the hydrogen peroxide and avoid cell damage by ROS (Asada 1999). To build up ascorbate levels, accumulated carbohydrates, particularly glucose, being the main carbon initial precursor for ascorbate synthesis, is converted to ascorbate (Foyer and Noctor 2011). The protection process mentioned above is restricted in cold temperatures. It was suggested by Lokhande et al. (2003) that the ascorbate peroxidase level is decreased in cold environment even though the light intensity is high. Additional antioxidants like glutathione and anthocyanin are produced. Anthocyanins play a key role as `light filters´ against high light stress. It has been shown that they absorb more light in the green and yellow wavebands (500-600nm) than green leaves (Neill and Gould 2002). The light filtering aspect has been indicated many times to reduce photoinhibition and to accelerate photosynthetic recovery (Hatier and Gould 2009). Another function has been shown in chilling resistance of the plants (Winkel-Shirley 2001, Christie et al. 1994, Nozzolillo et al. 2002). The accumulation of anthocyanins in leaves, shoots and roots is stimulated by various environmental factors. Beside low temperatures and light stress conditions also biotic stressors lead to an accumulation of anthocyanins (Chalker-Scott 1999).

Temperature is another environmental factor that plays an important role in plant plasticity (Alonso Blanco et al. 2005, Alcàzar et al. 2009). Low temperature is a major environmental factor that limits the productivity of a plant. Cold response, the acclimation of the plant to recent temperature can be also accounted as a trait related to plant performance. Zhu et al. (2007) provide a snapshot of the complex transcriptional network that operates under cold stress and changes the expression of hundreds of genes resulting in the accumulation of protective proteins. Variation of plant performance and its response to

environmental factors is much more subtly in the wild, where plants carry functional allelic variants.

I.2.B. Genetic influences on plant performance

As mentioned before, one genotype can respond to a set of environmental factors. But variation of plant performance can also be observed in the same environmental scenario when studying a wide range of genotypes. Hence, genetic factors are also involved in the performance of a plant. Indeed, mutations in genes which are involved in important pathways can lead to a decrease of fitness and in some cases to extreme phenotypes like dwarfism. An example for dwarfism in sunflower is dwarf2 (Fambrini et al. 2011), which is generated by a deletion in the ent-kaurenoic acid oxidase1 gene sequence. Mutations can again react at different levels of plant performance. Mutations at the cellular level mostly influence the whole plant growth for example the *bui1* mutant (Yang et al. 2011) of rice (Oryza sativa) displayed pleiotropic phenotypes, including bent uppermost internode, dwarfism, wavy panicle rachis, and enhanced gravitropic response. Cytological observation with this mutant indicated that the growth defects of *bui1* are caused mainly by inhibition of cell expansion.

1.2.C. Genetic x Environment interaction affecting plant performance

Plant performance can be affected by a combination of genetic and environmental factors as explained above. Individual plants do not perform equally in one environment and they might respond differently to environmental factors. This term is called GxE interaction. In Prinzenberg et al. (2010) for example, two accessions of *Arabidopsis thaliana* have been shown to respond differently to diverse potassium and phosphate levels resulting in a variation of plant growth.

I.3. *Arabidopsis thaliana*

I.3. A. *Arabidopsis thaliana*: a model plant

Arabidopsis thaliana (later *A. thaliana*) is a small crucifer with a vegetative growth period that produces a leafy rosette followed by the bolting of an indeterminate reproductive shoot. Since approximately two decades *A. thaliana* is used as a model plant to analyze different plant processes on the genetically and molecular basis. The choice of utilizing this plant is based on its biological characteristics: Small size of the plant (20-25 cm height) makes cultivation on a large scale in small space and growing many plants at the same time possible. The short generation time that takes about six weeks from germination to mature seed (www.arabidopsis.org) and being a self-fertilizer, which can be easily outcrossed are two more advantages of using *A. thaliana* for genetic analysis. Another advantage is the existence of a complete genome sequence (Kaul et al. 2000; Arabidopsis Genome Initiative (AGI)) and the small genome size (five chromosomes of about 157 mega base pairs (Mbp) DNA and approximately 30.000 genes (Bennett 2003)) and diploidy which makes it very suitable for genetic studies. Resources of *A. thaliana* have been made available (Somerville and Koornneef 2002) and a huge collection of gene mutations (TAIR - www.arabidopsis.org) provide possibilities that are rare for higher plants. Natural variation in this species is also well documented: Accessions of *A. thaliana* have been collected in the north hemisphere and are available through stock centres (TAIR – www.arabidopsis.org). New collections from other regions have recently become available (Jorgensen and Mauricio 2004, Stenoien et al. 2005, Schmid et al. 2006, Beck et al. 2008, He et al. 2007, Picó et al. 2008). The complete sequencing of hundreds and soon thousands of genomes from *A. thaliana* accessions (Weigel and Mott 2009) makes the analysis of genome wide annotations of potential functional polymorphisms possible. Also the availability of new molecular tools like ultra-high resolution microarrays and next generation sequencing gives more informative input to dissect the genetic bases of traits in *A. thaliana*. All molecular tools and biological resources and the two decades of experiences in working with this plant summarize the attractiveness of *A. thaliana* as a model system (Alonso and Ecker 2006).

I.3.B. Mutant approach in *Arabidopsis thaliana*

Continued progress in the area of genomics, new molecular technologies and database management of the model plant *A. thaliana* makes it possible to use a mutant approach for almost every gene of interest. By comparing phenotypes of wild type and mutants in which a particular gene is disrupted one can get information on the function of the genes. Among the whole range of currently usable strategies, forward and reverse genetics are of the most importance (Richmond and Somerville 2000). In the forward genetics approach, a distinguished phenotype is observed when screening a population of mutants. The strategy is to start from this peculiar phenotype to find the gene which encodes for this phenotype. In the reverse genetic approach, the gene under study is already known and the strategy is to reveal a variation at the phenotypic level encoded by a mutation of this gene. Reverse genetics already provided mutants in nearly every gene in *A. thaliana* (Alonso et al. 2003). The understanding of plant biology has benefited tremendously from work done with artificially induced mutants and it still continues to be useful in the identification of gene function in *A. thaliana* (Page and Grossniklaus 2002).

Photosynthesis research was mainly based on the isolation of *A. thaliana* mutants via forward genetics (Stitt et al. 2010). Its use was supported by the development of techniques for chloroplast isolation and gas exchange analysis. Identification of photosynthetic mutants could be performed by isolating mutants with alteration in photosynthetic performance. The most common mutation is based on alteration in pigmentation and chlorophyll fluorescence parameters (Meurer et al. 1998). Nevertheless, it is known that mutations affecting photosynthesis do not necessarily have to show strong and obvious phenotypes.

Seed development has been extensively studied using mutants deficient in various aspects of this process (McElver et al. 2001). These studies showed that gene functions involved in plant survival can be identified only partially by induced mutant analyses, where mutants with reduced fitness are easily selected (Alonso-Blanco et al. 2009).

I.3.C. Natural variation in *Arabidopsis thaliana*

Induced mutagenesis has proven to be a powerful tool for discovering the functions of a gene (see paragraph above). However, there are some limitations in this approach. Genes with low effect on trait variation or genes with lethal mutations would rarely be detected in a mutant screen (see above). Furthermore, current mutant collections have typically been derived from mutagenesis of a limited number of genotypes like laboratory strains such Col-0 (Columbia) and L*er* (Landsberg *erecta*) or Ws (Wassilevkajia) and C24 accessions (Koornneef et al. 2011).

Genes with non-functional alleles already present in these accessions could also go undetected in a mutant screen. Interestingly, even within species, the gene content between accessions can vary substantially. Clark et al. (2007) showed that approximately 9.4% of *A . thaliana* protein coding genes are naturally absent or lost in wild accessions, limiting the mutant spectra that can be obtained from each accession. One way to solve those problems is to make use of natural variation as a resource for genetic variation. *A. thaliana* has emerged as a powerful platform to study the genetic bases of naturally occurring variation for different traits and this approach can be exploited to gain insight into the control of important processes in plants (Koornneef et al. 2004).

The traits and genes for which natural variation is present differ from induced mutants by the fact that they have been arisen spontaneously and have been selected by environmental specificities of local habitats. Adaptation to different environments and thus natural genetic variation has been found in *A. thaliana* for many different traits. The study of *A. thaliana* natural variation lead to the identification of hundreds of loci that are responsible for the variation of an abundance of traits and more than 30 of the underlying genes (Alonso-Blanco et al. 2009).

A. thaliana is naturalized at many contrasting habitats and shows a broad geographic distribution among the Northern hemisphere (Al-Shebaz and OKane 2002, Hoffmann 2002). It has a wider climatic amplitude than other well-investigated species of Brassicaceae and has an impressive latidunal range from 68°N (North Scandinavia) to 0° (mountains of Tanzania and Kenya) (Koornneef et al. 2004). A habitat from which a particular plant originates has its own specific environmental conditions. *A. thaliana* plants growing in different environments all over the world have adapted to those individual habitats. This leads to accessions with different alleles at many loci in the genome.

More than 750 natural accessions of *A. thaliana* collected all over the world are publically available at seed stock centres. Because these accessions are variable in terms of form, development and physiology they are useful for analyzing plant performance related traits and investigating the genetic basis of response to contrasted environments. Taken together, natural variation provides a relevant complementary resource to discover also novel gene functions as well as those allelic variants that specifically interact with the genetic background and/or the environment or show small phenotypic effects particular related to adaptation (Benfey and Mitchell-Olds 2008).

I.4. Analysis of the genetic and molecular bases of natural variation

I.4.A. QTL analysis and genome wide association mapping

Two sources of genetic variation are used to study the genetic variation of a trait. Observed genetic variation within species is the result of mutations. In the wild, these mutations spontaneously occurred and are targets for natural selection. Laboratory induced mutations result mostly in the disruption of a single allele (monogenic variation). The resulting phenotype of the obtained mutant is usually easily distinguishable from the wild type. Distinct phenotypic classes characterise qualitative traits. However, in the wild in most cases more than one gene is influencing the variation of a specific trait. This results in a continuous distribution of the trait among the population. In natural variation most of the variation is quantitative and determined by molecular polymorphisms at multiply loci and genes (multigenic), which are referred to as quantitative trait loci (QTL) and quantitative trait genes (QTGs) (Alonso-Blanco et al. 2009). The analysis of the genetic basis of quantitative traits in populations derived from experimental crosses is known as QTL analysis, QTL detection is a case of linkage mapping. QTL detection is a powerful approach being used to characterize the genomic loci involved in the variation of a quantitative trait. This procedure involves the evaluation of statistical and significant associations between phenotypic variation and specific alleles at and in between marker loci. QTL detection is of interest especially for plant performance related traits which are usually caused by several genes. QTL detection bears, like any statistical test, the risk of detecting false positives. This is less likely for major QTL than for those of small effects (Kroymann and Mitchell-Olds 2005). Knowing the number of QTL that explain variation

in the phenotypic trait reflects the genetic architecture of a trait. The contribution of genetic factors to the total observed variation between different genotypes is often expressed as the heritability of a trait. QTL analysis gives also information on the contribution of each QTL in the observed variance of the studied trait among the studied population.

The ultimate goal of QTL detection is the identification of the genes harbouring the DNA sequence polymorphism(s) that cause the detected QTL effects. Quantitative natural variation controls adaptive strategies to cope with biotic and abiotic factors. The understanding of quantitative natural variation can provide insights in ecological mechanisms and the evolutionary history of *A. thaliana* (Tonsor et al. 2005, Mitchell-Olds and Schmitt 2006). In the past five years there have been more than 300 articles published on the analysis of natural variation of *A. thaliana*, including more than 30 reviews covering this source of variation (Alonso-Blanco et al. 2009, Koornneef et al. 2011). Beside QTL detection these articles deal with genome wide association mapping (GWA), which involves searching for phenotype-genotype associations in a general population of individuals whose relatedness is unknown. Thus, it is not restricted like in traditional QTL mapping to the variation of segregation in a traditional bi-parental population (see below) and allows searching without *a priori* for genomic regions that are associated with a trait of interest in a set of several accessions. Association studies take advantage of the *A. thaliana* genome sequence information and recombination events which accumulated in the long history of *A. thaliana*. Moreover, it has been shown to be a powerful approach because of the increasing availability of natural accessions and high marker density (for instance 250.000 SNP Tilling arrays) to genotype these lines (Kim et al. 2007). Recently, Atwell et al. (2010) revealed common variants responsible for the variation in 107 phenotypes in a set of *A. thaliana* natural accessions. It was the first genome wide association mapping performed in plants. A disadvantage of GWA is the detection of false positive associations due to population structure. Statistical methods have been developed to control for population structure and thus reduce false positives (Atwell et al. 2010, Yu et al. 2006, Zhao et al. 2007). However, Brachi et al. (2010) introduced the complementary use of traditional QTL (linkage) mapping in controlled crosses and NILs additional to GWA mapping, as an alternative to the false positives reducing methods. They show that detected associations in GWA mapping validated by QTL mapping results enhance the ability to distinguish true from false associations. An advantage of traditional QTL mapping

compared to GWA mapping is the identification of rare alleles. Rare alleles would be diluted in the huge amount of accessions which are required for GWA. For QTL detection the possibility would be only true, if the accession carrying the rare allele is selected as a parent for the QTL detection.

I.4.B. Elucidating the genetic basis of a QTL

The genetic analysis of natural variation aims to determine the genetic architecture of quantitative traits, including estimation of the number of QTL, their position, their relative and absolute effects, their genetic interactions (QTL x QTL), and their interaction with environment (QTL x E) (Alonso-Blanco et al. 2006). All this information is given by QTL analysis. Mainly two things are important to perform QTL mapping: A mapping population and genotyping of lines of the mapping population. Once a QTL is detected, further analysis is required to underline the gene(s) which is/are responsible for the effect of the detected QTL. All this points are developed in the following paragraphs.

I.4.BA. Mapping populations commonly developed to detect QTL

In order to observe segregation of a trait in a population and to define the chromosomal regions involved in the variation of this trait, mapping populations are developed from crosses between lines which show either contrasting genotypes or contrasting phenotypes. Different approaches are commonly used to develop mapping populations. The presence of genetic recombinant events is the basis of all the mapping populations. These recombinations commonly occur during meiosis as chromosomal crossovers between paired chromosomes. This process leads to an offspring having different combinations of alleles from the parental lines and can produce new chimeric genotypes.

Only two generations are necessary to obtain such a mapping population. Two contrasting genotypes are crossed to produce F_1 hybrid plants which are genetically identical and heterozygous for all the loci of the genome. These F_1 plants are then selfed in the case of F_2 populations or backcrossed with one of the parental line in the case of the backcrossed (BC) populations. Most of the genetic background remains heterozygous but because of

the presence of recombination events, the genotype of all the lines obtained differ. This can lead to variation of the phenotype of the lines for a studied trait. F_2 and BC populations have been used for QTL analysis of different traits (Kowalski et al. 1994, Kuittinen et al. 1997, Werner et al. 2005a, reviewed Koornneef 2011). A disadvantage of these early mapping populations is the remaining heterozygosity, because the lines need to be genotyped again in every study and the same set of lines cannot be tested in several environments. This is a big disadvantage for studying GxE interaction in different environments.

In contrast to the F_2 and BC populations, recombinant inbred lines (RILs) are homozygous at almost all the loci in the genome. The main advantage of this homozygosis is that many identical seeds can be obtained and the phenotype of the lines can be estimated in many replicates and/or in contrasting environments. This makes these populations the most commonly used for QTL detection. RILs derive from the F_2 generation by single seed descendant and repeated selfing until F_7-F_{10} generation (see Figure I.1). During this process, the level of heterozygosity per locus is halved in each generation, which is resulting in nearly homozygous lines in the F_9/F_{10} generation. Due to virtual homozygosity of these lines, the genotype is not changing anymore and the lines can be referred to as immortal populations. This is a great advantage as it allows the exact same genetic material to be analyzed multiple times for different traits. Currently more than 60 RILs populations have been produced in different laboratories, of those 23 *A. thaliana* RIL populations are published until now (reviewed by Koornneef 2011). Since beginning 2011, the used L*er*/Eri-1 RIL population is also publicly available. Recently, also advanced intercross RILs (AI-RILs) are in use (Balasubramanian et al. 2009). The strategy increases the power to detect small effect QTL by using an intercross approach before inbreeding.

Another kind of immortal mapping population are Introgression Lines (ILs) which are obtained through repeated backcrossing and extensive genotyping. They are also named as Near-Isogenic Lines (NILs) (Monforte and Tanksley 2000). An allele of one parental line is introgressed into the other parental background (see Figure I.1). In general, recombination frequency in RIL populations is higher than the one observed in equally sized NIL populations, that allows using reduced numbers of individuals in the analyses (Keurentjes et al. 2007). In contrast to RILs, NILs contain only a single introgression from one parent into the genetic background of another per line, which increases the power to detect small effect QTL during QTL mapping. So far in *A. thaliana*, there are three

genome-wide IL-populations published involving the most common laboratory strains (Keurentjes et al. 2007a, Törjek et al. 2008).

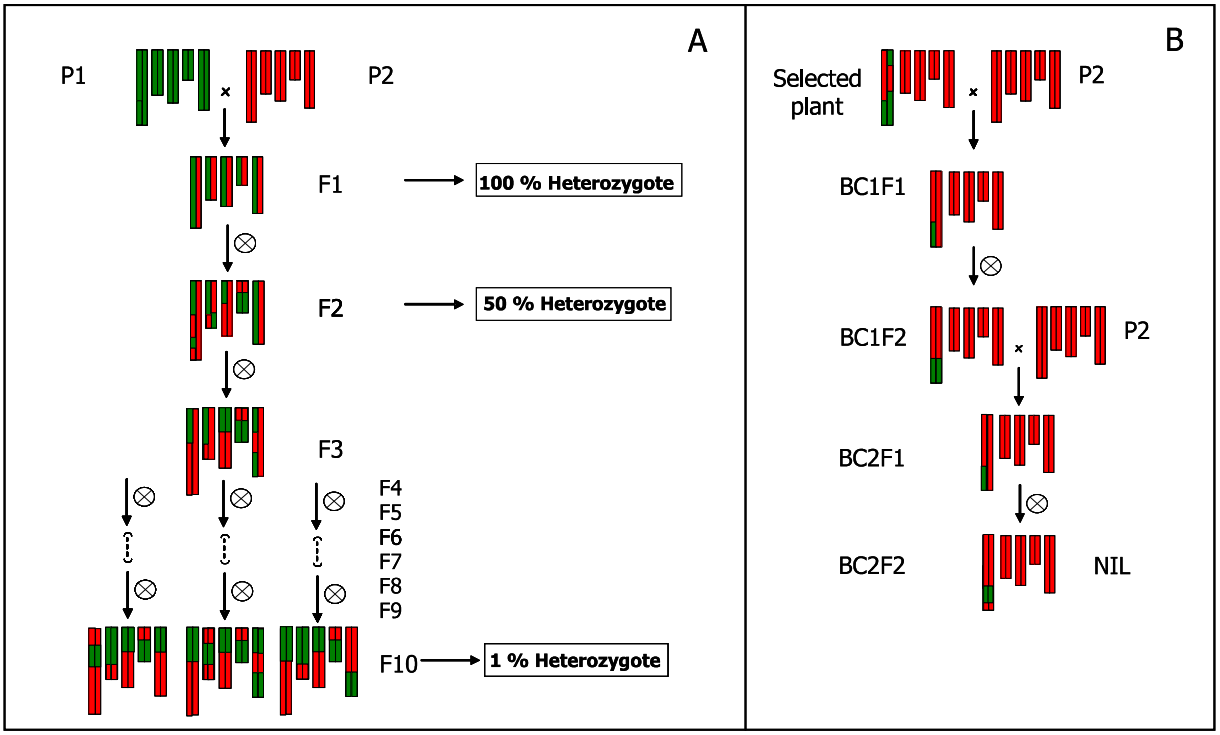


Figure I.1: The production and use of recombinant inbred lines (RILs) (A) and near isogenic lines (NILs) (B). The graphical genotypes of individual plants are depicted for all five chromosomes of *A. thaliana* in coloured bars. RILs build by single seed descendant from the second generation until the F_{10}. NILs are obtained by recurrent backcrossing and selection, Idealy only at one locus the NIL shows an introgression of a different allele.

I.4.BB. Molecular markers for genotyping mapping populations

To dissect the complexity in genetic involvement of quantitative traits, genotyping the line of a mapping population by using molecular markers is a crucial step. Once all the lines are genotyped, a genetic map for the studied population can be build. To detect the allelic value at particular loci on the genome a lot of different markers are available. Molecular markers are DNA markers. Several molecular marker techniques are used for genotyping, such as microsatellites, short sequence length polymorphisms (SSLP), restriction fragment length polymorphisms (RFLP), cleaved amplified polymorphic sequences (CAPS), amplified fragment polymorphisms (AFLP): All these techniques are PCR-based and make use of the polymorphisms occurring between the accessions of *A. thaliana*. Single nucleotide polymorphism (SNP) markers experience an increasing popularity for mapping

because of their frequency appearing in the genome (Delseny et al. 2010).To date, nearly unlimited numbers of SNP marker can be obtained from resequencing of 876 fragments in 96 accessions (Nordborg et al. 2005) or from almost full genome resequencing of 20 accessions by high-density arrays (Clark et al. 2007, polymorph.weigelworld.org). Additionally, many others were detected with next generation short read sequencing (Ossowski et al. 2008, Weigel and Mott 2009, summarized by Koornneef et al 2011). The accuracy of QTL mapping benefits from high resolution genetic maps of the analysed populations, which is mainly a function of the number of evenly distributed markers and the quality of the genotyping (Darvasi and Soller 1994, Charmet 2000).

Table I.1: Different types of marker systems. The name and designation of the marker is displayed as well as the appropriate method which is used. The dominance explains if it is possible to identify heterozygous parts in the mapping population.

Name	Designation	Method	Dominance
RFLP	Restriction Fragment Length Polymorphisms	PCR + Southern Blotting	CD
RAPD	Randomly Amplified polymorphic DNA	PCR (random primers)	D
AFLP	Amplified Fragment Length Polymorphism	PCR of adaptor ligated restriction fragments	D
CAPS	Cleaved Amplified Polymorphic Sequences	PCR + restriction digest	CD
SSR	Single Sequence Repeat	PCR (specific primers)	CD
	= Microsatellite		CD
SSLP	Single Sequence Length Polymorphism		CD
INDEL	Insertion – Deletion Polymorphisms		CD
SNP	Single Nucleotide Polymorphisms	Various methods	CD

I.4.BC. Toward the identification of the causal polymorphism(s): QTL validation and fine mapping

The accuracy of QTL mapping depends on statistical factors such as size of the mapping population and marker density of the genetic map (Price 2006). However, there is usually a need for QTL validation and further fine mapping (Paran and Zamir 2003, Weigel and Nordborg 2005). One common strategy makes use of near isogenic lines (NILs – presented in I.4.BA. Figure I.1), which carry an introgression of one genotype in the genetic

background of the alternative genotype, in the region of the QTL. The trait difference between those lines should only be due to the QTL effect, as it is the only genetic differences between them. If the trait difference is significant, the QTL is then validated. A validated QTL can be further fine mapped to unravel its genetic basis. A QTL region includes several genes (hundreds) and fine mapping consists of narrowing down the QTL region. It involves searching for lines with recombinant events within the QTL region. Co-segregation of the phenotype and genotype variations of a set of distinct recombinant lines (experimental population) allows the reduction of the size of the studied QTL. Thus, fine mapping enables to identify the causing genes and nucleotide polymorphisms underlying the QTL. Detecting polymorphisms and developing markers to perform in the experimental population is essential for a successful fine mapping. When examined in natural populations at high density, polymorphisms enable high resolution mapping through linkage disequilibrium (Kim et al. 2006, Atwell et al. 2010, Brachi et al. 2010, Li et al. 2010). The goal however, is to find the polymorphism which is responsible for the observed phenotypic variation. GWA studies point sometimes directly to the gene and thus fine mapping, which is a long lasting procedure, is not required. Among accessions, polymorphisms of sequences appear at an abundance of about one every 350 bp (Schmid et al. 2003). Not all these polymorphisms are leading to modifications and most of polymorphisms present between different accessions are neutral. However, some of them cause changes in gene function or expression. In detail, null or loss of function mutations derive from structural nonsense or insertions and deletions (indel) (Alonso-Blanco et al. 2005). Complete gene deletions caused by retro transposition can lead to a loss of function allele. Missense mutations in coding sequences might produce a change of function allele which alters protein structure, function or stability and might introduce phenotypic variation. Polymorphisms in regulatory sequences like indels nearby or in promoter regions of a gene might on the other hand result in differences in transcriptional efficiency (Keurentjes et al. 2007, West et al. 2006). Intronic mutations on the other hand, including insertions of transposons or differential expansion of microsatellite repeated sequences can also affect the expression of a gene (Sureshkumar et al. 2009). One expects expression differences or variation in mRNA stability caused by coding sequence polymorphisms, giving high input to natural variation in *A. thaliana* (Chen et al. 2005, Keurentjes et al. 2007).

An alternative approach to the marker based fine mapping is the use of microarray expression data or sequence predictions for highlighting inactivated genes within the QTL intervals (Roosens et al. 2008, Verbruggen et al. 2009). First demonstrated in induced mutations in *A. thaliana* is the use of direct sequencing instead of fine mapping of segregating populations to identify causative mutations (Schneeberger et al. 2009, Laitinen et al. 2010, summarized by Bergelson and Roux 2010).

I.4.BD. Candidate genes in QTL region and complementation

Fine mapping of QTL can allow narrowing down the region ideally to a single gene that is underlying the trait variation. This is not the case in many studies due to a lack of described polymorphisms and/or available recombinant lines. As a result, fine mapped QTL regions still contain multiple candidate genes due to the presence of DNA polymorphisms in genes of the region. As mentioned above, polymorphisms are rather frequent between natural accessions. In this case, candidate genes are chosen based on their functional annotation (Alonso-Blanco et al. 2005). However, one final proof to identify a gene as causal to the QTL effect is required and can be provided by complementation or allelism tests (Yano 2001, Salvi and Tuberosa 2005). Alleles from candidate genes in the QTL region can be cloned and transferred into the respective mutant line of the candidate gene or a line with a different allelic value at the QTL position. If the gene is responsible for the QTL effect, we will see a change in comparison to the mutant phenotype that is depended on the dominant nature of the allele. The trait value of the transformed line will change accordingly in comparison to the mutant phenotype. Complementation by plant transformation provides a convincing prove for the genetic basis of the QTL. Another test is the allelism test. A mutant line will be crossed with different parents to test the effect of the individual allele of each parent.

Once the genes responsible for the phenotypic variation have been described, the challenge is to explain that their nucleotide polymorphisms are involved in adaptation to particular environments and might not originate from spontaneous mutations that occurred during laboratory multiplication (Koornneef et al. 2011). By combining population genetics, evolutionary genetics with phenotypic studies, the composition of nucleotide variation

analysed in accession collections has shown potential signatures of natural selection (Koornneef et al. 2011).

I.5. State of the art concerning genetic determinism of flowering time in *A. thaliana*

The transition from the vegetative growth phase to the reproductive phase has to occur at the most suitable time for the plant in order to increase its fitness. When flowering, resources which accumulated in storage tissues during the vegetative growth phase are reallocated to the production of seeds. Flowering time is well studied in *A. thaliana* and nearly 70 genes involved in the variation of this trait have been identified via analyses of mutants and natural variation (Koornneef et al. 1998, Simpson and Dean 2002, Boss et al. 2004, Turck et al. 2008). In connection with natural variation flowering time has been analysed in nearly all *A. thaliana* mapping populations (Koornneef et al. 2011). *A. thaliana* flowering time is regulated by a complex genetic network composed of four main converging pathways (Roux et al. 2006): Vernalization pathway, photoperiod pathway, autonomous pathway and gibberellin pathway. These pathways integrate environmental and physiological factors in order to trigger the transition to flowering at an appropriate time (Brock et al. 2009).

Two extreme flowering behaviours have been described among *A. thaliana* accessions grown in laboratory conditions. These are winter (late) and spring (early) plants which show strong and weak vernalization responses, respectively (Alonso-Blanco et al. 2009). Most winter annual genotypes carry active alleles of *FRI* (Johanson et al. 2000) and *FLC* (Michaels and Amasino 1999) interacting genetically to delay flowering time. *FRI* and *FLC* were the first flowering time genes which were identified because of their natural variation. Natural polymorphisms altering flowering time have been functionally validated in greenhouse studies for nine flowering time genes: *CRY2, FRI, FLC, FLM, HUA2, PHYA, PHYB, PHYC* and *PHYD* (see Figure I.2). Interestingly, Weining et al (2002) detected different QTL in a RIL population grown in natural field conditions than the ones detected in controlled conditions. In studies of RILs grown in climatic chambers simulating the natural conditions in Sweden and Spain significant QTL x environment interactions were found. It gives way to the assumption that the genetic basis of flowering time is very diverse in laboratory and field conditions (Li et al. 2006, 2010, Brachi et al. 2010).

It has been shown that GWA mapping is a powerful tool for detection of candidate genes associated with a quantitative trait (Atwell et al. 2010). For 107 phenotypes, the highest number of significant associations was found for flowering time related phenotypes, whereas it most likely include a high number of false positives due to a clear geographical structure in these phenotypes. It has been indicated that flowering time in *A. thaliana* exhibit a latitudinal cline, suggesting that natural selection has shaped flowering time along a continental range to local climatic conditions (Caicedo et al. 2004, Lempe et al. 2005, Stinchcombe et al. 2004). This cline most probably interferes with population structure, which can be corrected by statistical models. Several of the most significant associations found for flowering time matched with a priori candidates such as *FRI* and *FLC*. But also some candidates like *CRY2*, *FLM*, *HUA2* or *MAF2* were not associated with flowering time in association studies. All three genes carry rare alleles which are not found in many accessions. Genome wide association studies fail when they include too few accessions with functional variant alleles, or if too many of the functional variant alleles are distinct from each other (Salomé et al. 2011). Increasing the sample size solves the problem like shown in the study of Li et al. (2010). A new candidate which is emerging from GWA is *DOG1* (Bentsink et al. 2006), which was not originally detected to be involved with flowering, but is highly associated with flowering phenotypes. Variation in environmental signals reveal new genetic loci associated with flowering time natural variation. For example, in Brachi et al. (2010) most candidates were related to the circadian clock, thus revealing a connection to the photoperiod pathway. They also found nine association peaks (supported by similar QTL results) which were far from candidate genes associated with natural variation in flowering, suggesting that GWA is beneficial in identifying genes that have not been previously described as candidates. All studies showing associations with flowering time and genes (previously not described as flowering related gene), highlight the importance of analysing flowering behaviours of *A. thaliana* in different environments to get an overview about flowering time natural variation and adaptation.

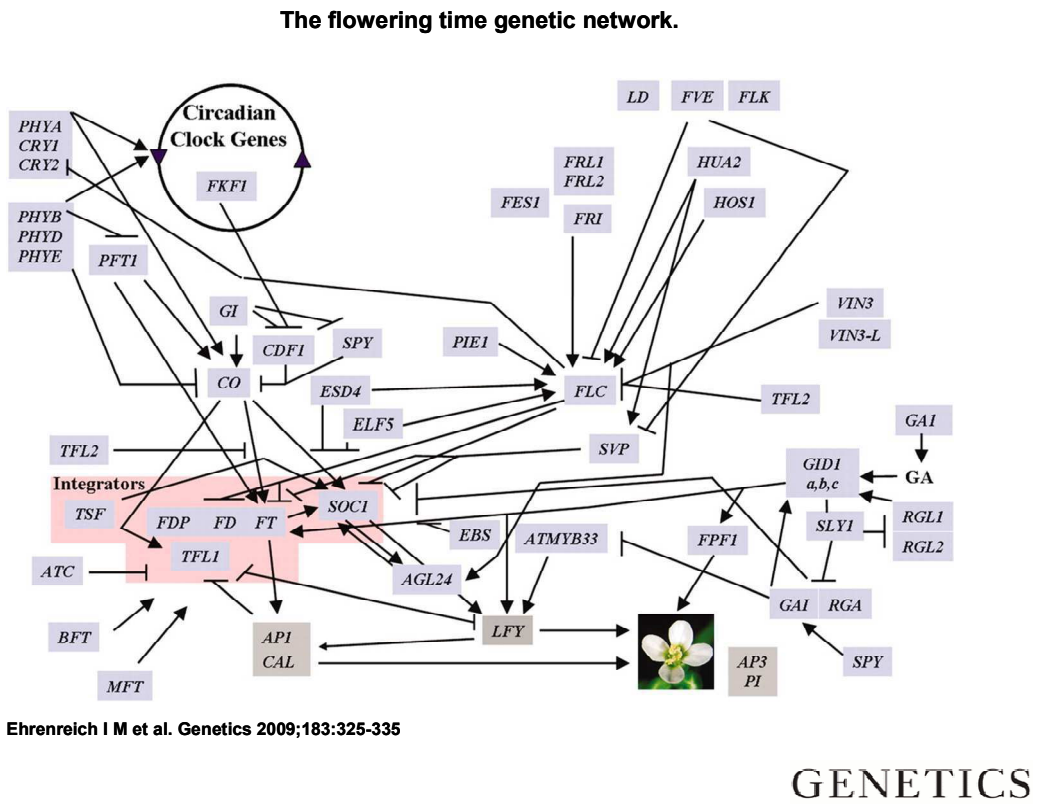


Figure I.2. Flowering time genetic network based on Ehrenreich et al. 2009

I.6. State of the art concerning genetic and molecular determinisms of anthocyanin accumulation in *A. thaliana*

Anthocyanins belong to the flavonoid family, consisting of a minimum of 6000 molecules. Flavonoids are secondary metabolites involved in an abundance of mechanisms in plant development, stress release and defence (Winkel-Shirley 2001, Lepiniec et al. 2006). They play a protective role against several abiotic stress factors in *A. thaliana*. Light (quantity as well as quality) is the key factor for anthocyanin accumulation. Anthocyanins accumulate in leaf epidermal cells, where they act as UV-B filters or complex with DNA as protection from oxidative damage (Sarma and Sharma 1999, Harborne and Williams 2000, Dixon 2005, Dixon et al. 2005, Aron and Kennedy 2008, Albert et al. 2009). In case of high light they protect the photosynthetic apparatus against excess of light and thus prevent photo inhibition. Other environmental cues like cold temperature, nutrient deficiency, water status, wounding and pathogen attacks also lead to an accumulation of anthocyanins (Christie et al. 1994, Leyva at al. 1995, Zhang et al. 2011, Bhattacharya et al. 2010). As plants are not able to avoid the stress environments, the production of anthocyanins

contributes to plant adaptation to these environmental conditions. To date the biosynthesis of flavonoids and anthocyanins is well studied and described by many research groups (Holton and Cornish 1995, Winkel-Shirley 2001, Koes et al. 2005). Flavonoids are synthesized in the cytosol and derive from phenylalanine via the general phenylpropanoid pathway in multiple enzymatic steps. The biosynthesis genes are divided into two subgroups, early biosynthetic genes (EBGs) and late biosynthetic genes (LBGs) (Cominelli et al. 2008; see Figure I.3). The EBGs are *CHS, CHI, F3H, F3´H,* which are common to different flavonoid sub-pathways and the LBGs: *DFR, LDOX, ANR* and *UF3GT*. In *A. thaliana,* regulation of structural gene expression appears tightly organized in a spatial and temporal way during plant development. It is organized by a tertiary transcription factor complex involving a R2R3 MYB factor, a bHLH protein and a WD40 protein (MBW complex) (Zhang et al. 2003, Cominelli et al. 2008, Gonzalez et al. 2008). The MBW complex regulates only the LBGs and not EBGs in binding the promoter and leading to the biosynthesis of anthocyanins. Each subunit of the MBW complex fulfils a specific function (reviewed in Hichri et al. 2011). The MYB factor binds via its MYB domain the *cis*-regulatory elements of the LBGs. The bHLH protein is responsible for the activation of the target gene expression. Connection between bHLH protein and the MYB protein and WD40 protein occurs at the MIR motif and WD40 motif, respectively. The WD40 protein in the MBW complex is responsible for the stabilization of the factor complex. It is assumed that WD40 proteins involved in anthocyanin content regulation do not have any catalytic activity like DNA binding or regulation of expression. In *A. thaliana* several genes encoding for members from the three transcription factor families are known to be involved in the regulation of the anthocyanin accumulation. The WD40 protein stabilizing the MBW complex in *A. thaliana* is TTG1 (Walker et al. 1999), which is acting as a physical link between MYB and bHLH domain (Baudry et al. 2004, Zhao et al. 2008). The bHLH proteins are represented by TT8, GL3 and EGL3. It has been shown that they assess redundant roles between each other and all of them bind to TTG1 (verified in yeast 2 hybrid tests – Baudry et al. 2004). The target gene specificity of the complex is conferred by the MYB factor (Zhang et al. 2003). PAP1/PAP2, TT2 and GL1 are MYB proteins initiating the anthocyanin or trichome production. There is evidence that the MYB factors play a major role in the light induction of anthocyanin biosynthesis (Core et al. 1993, Petroni et al. 2000, Piazza et al. 2002). It is worth noting that until now; only two studies are dealing with natural variation in anthocyanin production. Both publications (Teng et al.

2005 and Diaz et al. 2007) detected *PAP1* as the candidate gene for anthocyanin production in two different RIL populations. In Teng et al. (2005) sucrose (Suc) induces PAP1 to build the MBW complex and start the biosynthesis. Sugar is a common regulator of a number of genes involved in photosynthesis, pathogenesis and anthocyanin biosynthesis (Rolland et al. 2006, Teng et al. 2005, Solfanelli et al. 2006). It has been indicated by Loreti et al. (2008) that plant hormones act in concert with sugar in the presence of light to regulate anthocyanin accumulation. Ethylene has been verified as a negative regulator of anthocyanin production (Jeong et al. 2010). It acts through the regulation of MYBL2 and CPC, suppressing anthocyanin accumulation via MBW complex blocking. Both proteins are R3 MYB factors, which are not able to bind DNA, but still bind the bHLH transcription factor of the MBW complex. The anthocyanin accumulation will break up because the MBW is not binding to the promoter of LBGs anymore. Plants have evolved this positive and negative regulation system, because the accumulation of anthocyanins involves an investment of energy that may reduce light capture and carbon assimilation (Das et al. 2011). Therefore a down regulation of the anthocyanin biosynthesis is necessary in some environmental scenario. Even if the biosynthesis pathway of anthocyanin production itself is quite well understood, its regulation is still poorly understood, and the identification of new transcription factors involved in anthocyanin biosynthesis should be conducted together with investigation of the parameters controlling their expression (Hichri et al. 2011).

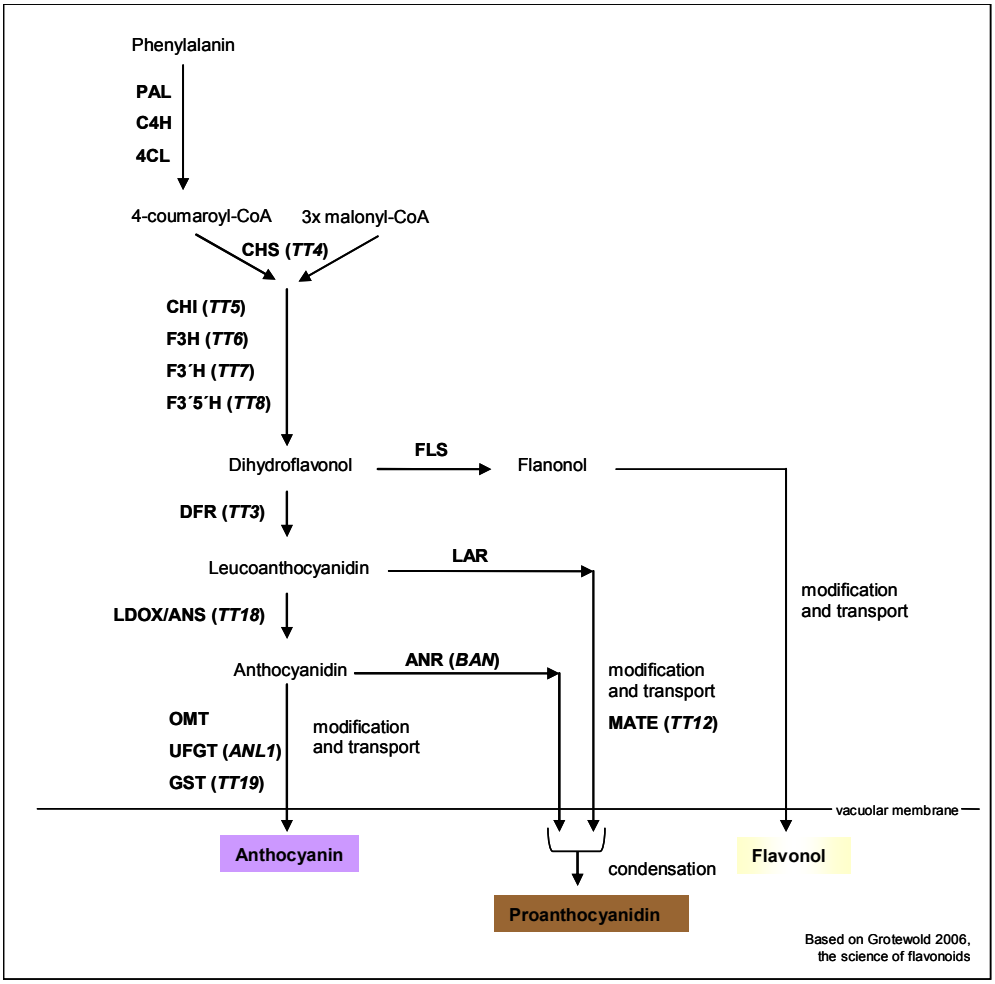


Figure I.3: Based on Grotewold 2006. Flavonoid biosynthetic and accumulation pathway. Enzymatic steps leading to the major classes of end products, anthocyanins, proanthocyanidins and flavonols, are presented and identified with colour boxes. Enzymes are indicated with standard abbreviations in bold; the encoding genes are indicated in italic and brackets. Abbreviations - Enzymes: ANR, anthocyanidin reductase; C4H, cinnamate-4-hydroxylase; CHI, chalcone isomerase; CHS, chalcone synthase; 4CL, 4-coumaryl:CoA-ligase; DFR, dihydroflavonol 4-reductase; F3H, flavanone 3-hydroxylase; F3´H and F3´5´H, flavanoid 3´and 3´5´hydroxylase; FLS, flavonol synthase; GST, glutathione transferase; LAR, leucoanthocyanidin reductase; LDOX, leucoanthocyanidin dioxynase (also known as ANS – anthocyanidin synthase); OMT, O-methyltransferase; PAL, phenylalanine ammonia-lyase; UFGT, UDP flavonoid glucosyl transferase. Genes: *ANL1, ANTHOCYANINLESS1; BAN, BANYLUS; TT, TRANSPARENT TESTA.*

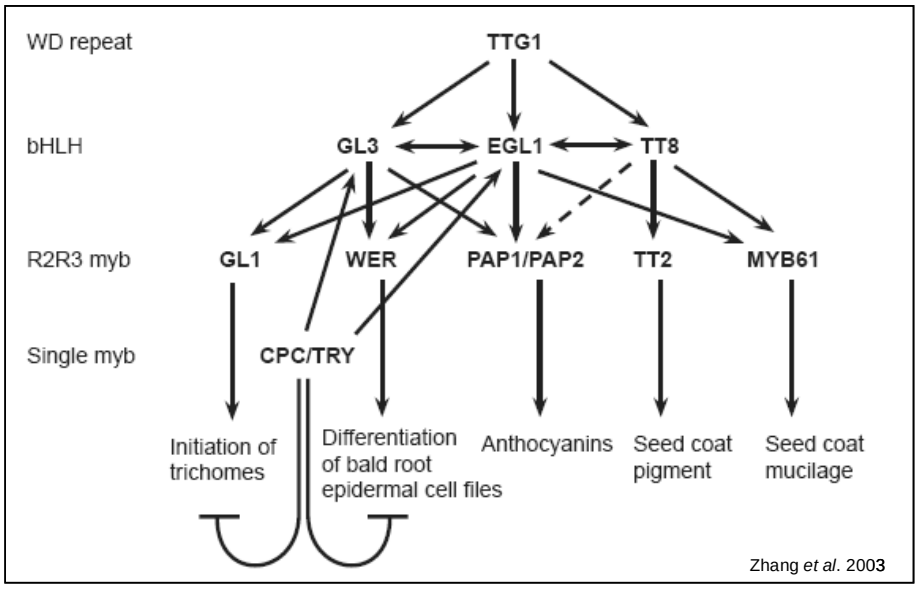


Figure I.4: Based on Zhang et al 2003. Regulatory network model of TTG1 dependent pathways. This reticulated model shows bHLH and MYB transcriptional regulators that function the TTG1-dependent pathways. The model illustrates the interactions among the proteins and genes.

Thesis aim

Accessions of *A. thaliana* from different habitats are publically available, offering a pool of biological samples exhibiting a wide range of responses to environmental factors. The accessions L*er* (Landsberg *erecta*, from Poland) and Eri-1 (Eringsboda-1, from Sweden) show variation in their responses of flowering time and anthocyanin accumulation when grown in high light intensity and low temperature.

The aim of the present study is to decipher the genetic and molecular determinism of the response of flowering time and anthocyanin accumulation to high light intensity and low temperature using natural variation in *A. thaliana*.

For this purpose, a new mapping population derived from the cross between L*er* and Eri-1 was used. The 110 RILs (recombinant inbred lines) were genotyped using evenly distributed markers along the genome of *Arabidopis thaliana* and a saturated genetic map was built. In addition, quantification of flowering time and anthocyanin accumulation in both conditions was carried out. Variation of phenotypic responses among this new population allowed detection of QTLs (quantitative trait loci) involved in flowering time and anthocyanin accumulation in each environment.

I developed and selected suitable lines in order to validate the presence of major detected QTL. After validation, the QTL was fine mapped in order to reveal the gene(s) which is (are) responsible for the effect of the detected QTL. Whole genome sequencing of Eri-1 was a crucial step to identify causal polymorphisms and to propose candidate genes. A candidate gene was identified for one QTL responsible for flowering time under high light intensity and cold environment. This QTL co-locates with another QTL responsible for anthocyanin accumulation. Analysis of a loss-of-function mutant of this gene allowed assessing whether the gene is also responsible for anthocyanin accumulation under this environmental condition.

Finally, gathering all the results generated in this study in addition to published literature, I propose a model explaining how anthocyanin accumulation could be regulated at molecular levels under high light and cold environment in the L*er*/Eri-1 mapping population.

II. Materials and Methods

II.1. Growth conditions and experimental design

Seeds were stratified for 3-5 days on filter paper moistened with demineralised water at 4°C in the dark before being transferred to soil. Stratification breaks dormancy and synchronizes germination. All plants were grown on a mixture of weak to strong decomposed bog peat supplemented with perlite and macro nutrients (PRO start, Gebr. Brill Substrates GmbH & Co.). To investigate responses to environment, two different growth conditions were set up. In the environment tested first (Figure II.1.A), referred to as "Control-CT" environment, the plants were subjected to long-day photoperiod conditions (16 hours light and 8 hours darkness) provided by fluorescent tubes (Philips F25T8/TL741) and incandescent light bulbs (regular 25W) with overhead lighting of the growth cabinet (BBC York Brown Boveri international No: BBC30). Light intensity during the day showed an average of 100 µmol m^{-2} s^{-1}. Growth cabinet temperature was maintained at 20°C during the day and 18°C during the night (Figure II.2), and relative air humidity was maintained at 70%. The second environment is referred as "HL4" (for high light at 4°C). In this environment, plants were exposed to two environmental constraints: cold temperature (4°C) and high light stress. In addition to the fluorescent tubes (Philips F25T8/TL741) of the cold growth cabinet (Rivacold, Montecchio, Italy), four highlight lamps (workshop MPIPZ) mounted above the trays (see Figure II.1.B) were added. All the lamps in the cabinet were set to long-day photoperiod conditions (16 hours light/high light and 8 hours darkness). Light intensity during the day showed an average of 430 µmol m^{-2} s^{-1}. Due to heat generated by the high light lamps, the temperature at the level of the plants was maintained at 11-12°C during the day and at 6-7°C during the night. Relative humidity ranged between 60 and 70% during the day and between 45 and 50% during the night (see Figure II.2). Throughout all experiments positioning of the pots in the growth cabinet were changed every second day in order to provide equal light distribution. The light intensity in µmol s^{-1} m^{-2} was determined from averaged values, measured with a light meter (LI-250A, LI-COR Biosciences, Bad Homburg). Additionally, temperature, light intensity and relative humidity were measured using a HOBO Data logger (onset® computer corporation, Pocasset, MA), which was positioned between the plants. The recorded data was checked every second day (see Figure II.2). Before being transferred to the HL4, all plants were subjected to CT condition for 10 days after sowing.

Figure II.1: Contrasted environmental conditions monitored in different growth cabinets.
A: Control condition; Ten trays with 35 plants each (112 RILs * 3 repetitions + L*er* * 7 repetitions + Eri-1 * 7 repetitions) were growing in long-day photoperiod conditions (16h/8h day/night) at light intensity of 100 µmol m^{-2} s^{-1} and a cabinet temperature of 20℃ day/18 ℃ night, 70% relative air humidity **B:** HL4 condition; Ten trays with 35 plants each (112RILs * 3 repetitions + L*er* * 7 repetitions + Eri-1 * 7 repetitions) were grown in long-day photoperiod conditions (16h/8h day/night) at a light intensity of 430 µmol m^{-2} s^{-1} and a cabinet temperature at 12℃ day/7 ℃ night, 70% relative air humidity during day/ 50% during night.

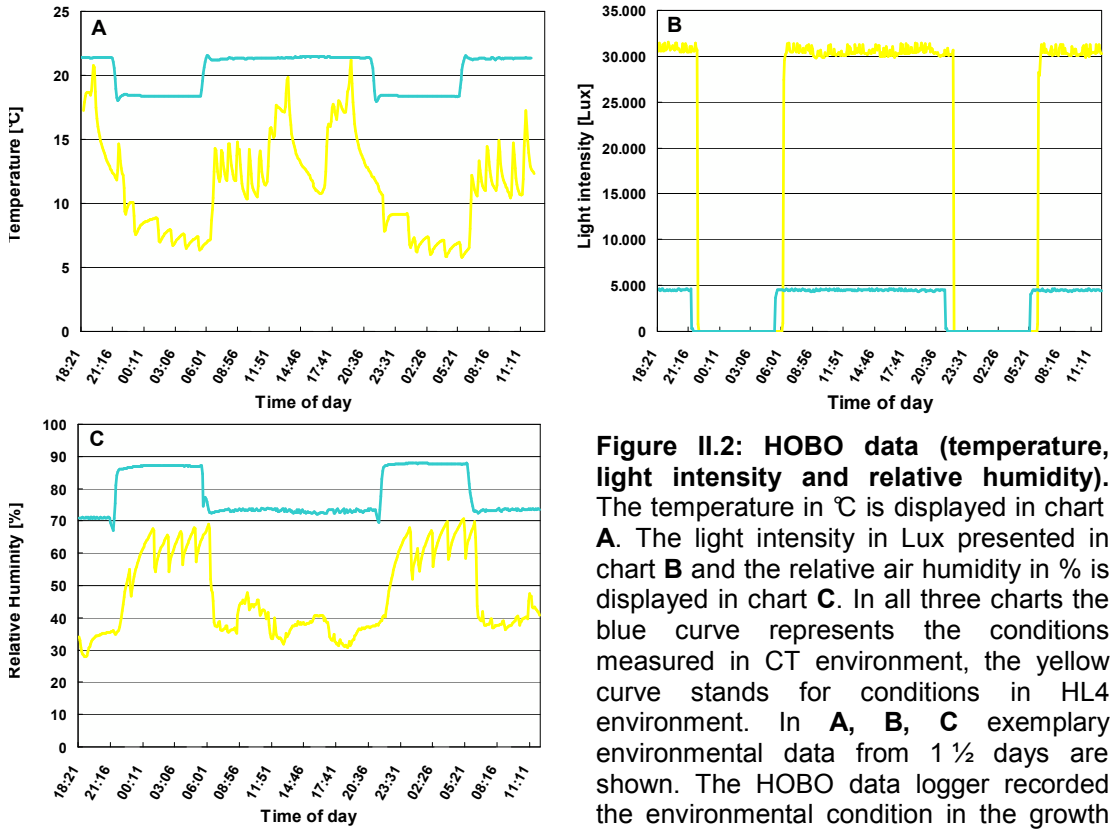


Figure II.2: HOBO data (temperature, light intensity and relative humidity). The temperature in °C is displayed in chart **A**. The light intensity in Lux presented in chart **B** and the relative air humidity in % is displayed in chart **C**. In all three charts the blue curve represents the conditions measured in CT environment, the yellow curve stands for conditions in HL4 environment. In **A, B, C** exemplary environmental data from 1 ½ days are shown. The HOBO data logger recorded the environmental condition in the growth cabinet every four minutes.

II.2. Plant Material

II.2.A. Mapping population

For QTL detection, a population of 112 recombinant inbred lines (RILs) was used. These lines have been generated at Wageningen University (Plant Genetics, Wageningen, the Netherlands) from a cross between two accessions originating from different geographic habitats. The Landsberg accession has been collected in western Poland in the Warta river region (15.23° longitude; 52.74° latitude). Landsberg *erecta* (L*er*) is a selected laboratory strain (spontaneous mutation) which harbors a mutation in the *erecta* gene in the Landsberg genetic background (Torii et al. 1996, van Zanten et al. 2009). This mutant shows a phenotype with round leaves, short petioles, short pedicels, is 20-22 cm high and the flowers are clustered at the top of the inflorescence. The Eri-1 accession originates from Eringsboda in South Sweden (15.35° longitude; 56.43° latitude). Eri-1 has a medium size rosette and is 42-50 cm in height. The progeny, derived from a cross between L*er* and

Eri-1, was grown through eight generations via self-pollinating, single seed descent producing nearly homozygous lines. Seeds from the F_9 progeny (F_{10} plants) were grown for QTL experiments.

II.2.B. Other genetic materials

A L*er* line with the wild type *HUA2* allele (Doyle et al. 2005) was provided by Dr. Rick Amasino, University of Wisconsin, USA.

II.2.C. Growing Eri-1 for DNA extraction in order to sequence its genome

Eri-1 seeds were sterilized in 1 ml Ethanol 96% for 5 min and with 1 ml sodium hypochlorite solution 6% supplemented with 0.05% Triton X for 8 minutes followed by three washes with sterile water. The accession was grown for 4 weeks under sterile conditions in MS media (Murashige and Skoog 1962) supplemented with sucrose 0.3%. The recipe is written below.

<u>1 l MS medium</u>
4.3 g MS salts including vitamins (Duchefa Biochemie, Haarlem, the Netherlands)
0.5 g MES (Sigma-Aldrich, Munich,Germany)
3 g sucrose (Duchefa Biochemie, Haarlem, the Netherlands)
Adjust pH to 5.6-5.8 with KOH 1 N
8 g/L phytoagar (Duchefa Biochemie, Haarlem, the Netherlands)
Autoclave (121°C, 20 min)

II.3. Molecular analyses

II.3.A. Deoxyribonucleic acid extraction

Large scale DNA extraction (plant material in boxes for 96 samples) was performed using BioSprint 96 workstation (QIAGEN, Hilden, Germany) and BioSprint 96 DNA Plant Kit according to the manufacturer instructions.

If only a few samples were analysed, the plant material was collected in Eppendorf reaction tubes and DNA extraction was performed, using a protocol adapted from Dellaporta et al. 1983. The samples were ground and 300 µl of extraction buffer (see recipe below) was added. After homogenizing and centrifugation, 200 µl of the supernatant was transferred to a new Eppendorf tube and precipitated with 20µl sodium acetate 3 M pH 4.8 and 400 µl ethanol 96% in -20°C between 1 h to overnight. After two washing steps with 70% ethanol, the DNA pellet was resuspended in 200 µl AE buffer.

The DNA extraction of the sterile grown plant material of Eri-1 accession for sequencing experiments was performed with DNeasy plant Maxi KIT (QIAGEN, Hilden, Germany) according to the manufacturer protocol.

<u>DNA extraction buffer</u>

200 mM Tris-HCl pH 8.0

250 mM NaCl

25 mM EDTA

0.5 % SDS

II.3.B. Revealing polymorphism between L*er*/Eri-1 using PCR based markers

II.3.BA. PCR programs and gel-electrophoresis

For revealing polymorphisms between L*er* and Eri-1, DNA of both accessions was amplified in several PCR with different marker combinations. The PCR protocol (mix and cycles) is detailed below. To amplify DNA fragments, the PCR thermo cycler (BioRad DNA Engine Tetrad, Biozym Peltier Thermo Cycler 96 Block, Munich, Germany) was set up with a standard PCR program. The general settings are shown in the table below. Specific annealing temperatures of all primers used in different experiments can be found in table A.1 in the appendix. To accomplish successful PCR, Taq-polymerase (5 u/µl) was

used. Subsequent to the PCR polymorphism between PCR products were visualized by gel electrophoresis. To identify polymorphism between L*er* and Eri-1, DNA products were loaded on a 3% agarose gel (my-budget Universal Agarose, Bio-Budget Technologies GmbH, Krefeld, Germany) using gel-electrophoresis method (PEQLAB Biotechnologies GmbH, Erlangen, Germany). When preparing the gel, electrophoresis-grade agarose and compatible electrophoresis buffer were mixed. 1x Tris-Acetate-EDTA (TAE, see recipe below) buffer was used as a running buffer. DNA samples were mixed with 1 µl gel loading dye buffer (see recipe below) before loading on the prepared gel. To identify polymorphism between L*er* and Eri-1 DNA PCR products were loaded on 3% agarose gels. By using high percentage gels, it is possible to separate small fragments of amplified DNA (100 – 700 base pairs (bp)). Depending on the fragment size a 1 kb marker (Invitrogen, Darmstadt, Germany) or a low range DNA ladder (Gene Ruler, Fermentas, St. Leon-Rot, Germany) was used to assess the PCR product size. For photo documentation of the samples the gel was stained during the run with ethidium bromide (2 µl stock (10 mg/ml) in 100 ml gel; Roth, Karlsruhe, Germany). Pictures of the gel were taken on an UV-screen with a digital camera (Gel Photodokumentationsanlage, INTAS, Göttingen, Germany) and the software INTAS GDS Windows.

PCR Mix (20 µl sample)

13.5 µl Aqua dest.

2 µl PCR reaction buffer 10x (Roche Diagnostics GmbH, Mannheim, Germany)

2 µl dNTPs (stock 10 µmol/100 µl – ready to use 1:100 = 1 mM; Roche Diagnostics)

1 µl primer forward (10 µM), (Invitrogen, Karlsruhe, Germany)

1 µl primer reverse (10 µM)

0.1 µl Taq polymerase

PCR standard program

1) Reaction start: 3 min at 93°C

2) Denaturation of the DNA: 15 sec at 93°C

3) Annealing of the primers: 45 sec at 45-60°C

4) Elongation of the primer sequence: 1.30 min at 72°C

5) Reaction stop: 5 min at 72°C

Step 2-4 repeated for 34 cycles

1% agarose gel

2.5 g agarose

250 ml 1x TAE running buffer

5 µl ethidium bromide

50x TAE-running buffer-stock solution (pH 8.5)

242 g Tris-Base

57.1 g Glacial acetic acid

100 ml 0.5 M EDTA (pH 8.0)

Distilled water to 1 l final volume

10x Orange D loading buffer

0.5% Orange D

40% sucrose

10 mM Tris-HCl pH 8.5

Distilled water to 10 ml final volume

II.3.BB. Designing polymorphic markers between L*er* and Eri-1

Databases like TAIR (www.arabidopsis.org) provide potential PCR-based marker information with known physical positions. From these databases, primer sequences for microsatellite markers were obtained. Another source of SSLP markers is described in the INRA MSAT database (http://www.inra.fr/internet/Produits/vast/msat.php). Because the accession Eri-1 was not sequenced at the beginning of marker development, already commonly available markers based on polymorphisms between L*er* and other accessions were tested for differences between L*er* and Eri-1.

Table A.1 (in appendix) shows additional polymorphic markers used for validating and fine mapping the QTL. Additionally to these already known markers, new molecular markers were designed based on published sequencing data (www.1001genomes.org, http://polymorph.weigelworld.org/cgi-bin/msqt-sbe.cgi). Taking newly released single nucleotide polymorphism (SNP) data of L*er* and Bur-0, Kro-0, C24 and already published

data of Est-1 into account; the SNP data was scanned for rare L*er* SNPs compared to the other four accessions and the reference genome of Col-0. With SNPs only found in the L*er* accession we assumed to have also different SNPs in Eri-1 compared to L*er*.
The SNPs were used for designing cleaved amplified polymorphic sequences (CAPS) markers. CAPS are based on the digestion at the position of the SNP between the genotypes. For designing marker primers the software Primer3 (Rozen and Skaletsky 1999) was used. Primers were designed to be between 18 bp and 25 bp in length with an GC-content between 40% and 60%, and no hair pin or dimer formation that would affect the 3'-end of the sequence. Due to the fact that most of the selected SNPs were carried from both accessions L*er* and Eri-1 and most of the developed markers were not polymorphic between them, sequencing PCR products from L*er* and Eri-1 respectively was performed as a next step to find SNPs between L*er* and Eri-1. Mainly intergenic regions and introns at the locus of interest were sequenced, since a higher density of allelic polymorphism is expected in those regions compared to exons. As a third step, the Eri-1 accession was sequenced and SNPs between L*er* and Eri-1 were called (see paragraph II.4.F). The new developed marker primers are listed in Table A.1 in appendix. All the marker primers were provided by Invitrogen (Karlsruhe, Germany).

II.3.BC. Digestion of CAPS markers

To distinguish SNP polymorphisms in PCR products, amplified fragments had to be digested with an enzyme with differential cutting sites at the position of the polymorphism. The digestion was done with 10 µl PCR-product. 5 µl of a digestion mix were added to the PCR-product, vortexed and incubated at the enzyme specific temperature overnight.

<u>Digestion Mix (10 reactions)</u>
35 µl Aqua dest. (LiChrosolv, Merck, Darmstadt, Germany)
15 µl buffer (10x buffer specific for the enzyme from New Englands Biolabs, Frankfurt am Main, Germany)
0.5 µl enzyme (New England Biolabs)

II.3.BD. Sequencing single regions/candidate genes

To verify that only a single amplicon was produced PCR products were checked by gel-electrophoresis before sequencing. 80 µl of the PCR products were cleaned according to the protocol of the High Pure PCR Cleaning Micro Kit (Roche, Basel, Switzerland) or Nucleo Spin Extract II (Macherey-Nagel, Düren, Germany). Concentration of these samples was determined (Nanodrop, ND-1000 Spectrophotometer, peqlab Biotechnologies GmbH, Erlangen, Germany) and adjusted to the requirements for sequencing by the MPIPZ DNA core facility on Applied Biosystems (Weiterstadt, Germany) Abi Prism 377, 3100 and 3730 sequencers using BigDye-terminator v3.1 chemistry. Premixed reagents were acquired from Applied Biosystems. For DNA sequence alignments and analyses MegAlign, Seqmen Pro and Seqbuilder moduls of the Lasergene package (DNASTAR Inc., Madison, WI) were used. Sequence analyses of different candidate genes in the QTL3, QTL5-1 and QTL5-2 region were performed in the L*er* and Eri-1 accession by sequencing the genomic loci of the genes. Overlapping primer combinations for sequencing single genes are listed in Table A.2 in appendix. Each PCR product was sequenced with forward and reverse primer. Predicted coding regions of genes in L*er* and Eri-1 as well as in the Col-0 reference sequence from TAIR were translated and aligned with the software MegAlign.

II.3.C. Whole genome sequencing of Eri-1 with Illumina platform

Genomic DNA of Eri-1 was sonified in a Bioruptor (Diagenode) to obtain fragments of 300-600 bp. Library preparation for Illumina sequencing was performed using the NEBnext DNA Sample Prep Reagent Set 1 (NEB) according to manufacturer's protocol with the following modification: Full length adaptors were used in the adaptor ligation step and the PCR step to amplify adaptor ligated DNA was omitted. During the Illumina sequencing procedure, adapter ligated single-stranded DNA was hybridized with a universal sequencing primer and extended with modified nucleotides. Each modified nucleotide acts as a terminator: it is labelled with a distinct fluorochrome that is chemically cleavable, so that DNA synthesis is reversible stopped at each position. An image is

recorded at each cycle in four channels so that the nucleotide incorporated can be determined. Then the cleavable attachment of the nucleotide is removed and the cycle repeated, thus allowing the incorporation of the next nucleotide (summarised by Delseny et al. 2010).

Sequencing was performed on two lanes of a Genome Analyser IIx (Illumina, San Diego, CA) using SBS v4 Cluster Generation and TruSeq Sequencing Kits (Illumina) at the Max-Planck Genome Center, Cologne, with an output of 32.987.304 total reads.

To call SNPs from these reads between the accessions L*er* and Eri-1 different software packages were used (performed by Jia Ding, bioinformatician of Koornneef Department MPIPZ, Cologne). The illumina sequence files are presented in a FASTQ format. The quality of the sequencing was controlled by the software FASTQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/) which provides a modular set of analyses to check the raw sequencing data. It controls the sequence quality per base as well as the sequence content per base and the GC content per sequence, respectively. The mapping was performed with the software Burrows-Wheeler Aligner (BWA) (http://bio-bwa.sourceforge.net). BWA uses an algorithm to align short (< 200 bp) reads to the reference genome of *A. thaliana* Col-0 (TAIR9).

BWA outputs alignments in SAM format which is supported by the SNP caller software SAMTOOLS (http://samtools.sourceforge.net, Li et al. 2009). SAMTOOLS provides various utilities for manipulating the alignments including sorting, merging, indexing and generating the alignments in a per-position format. For the Eri-1 sequence alignment a filter for mapping quality was used (q20 – low root mean square mapping quality) which only uses alignments above a mapping quality of 20 bp. With the software PICARD (http://www.picard.sourceforge.net/index.shtml) alignment duplicates were removed. After manipulating the alignments, SAMTOOLS converted the SAM file into a pile up format like BAM file which output can be visualized in the IGV browser (http://www.broadinstitute.org/software/igv) to view the mapping result. To call SNPs between the reference genome Col-0 and Eri-1 various filters were used additionally in SAMTOOLS. The maximum read depth was set to 100 bp, the minimum read depth to 10 bp. The SNP quality was set > 50 reads for SNPs and > 20 reads for Indels. After filtering, 397.836 different SNPs between Col-0 and Eri-1 were called. Adding published SNP data of L*er* (1001genomes.org), identical SNPs between L*er* and Eri-1 but different to Col-0 were removed (163042 SNPs). 184.924 different SNP between L*er* and Eri-1 could

be called whereas L*er* carried the same SNPs than Col-0. 212.912 were highlighted between L*er* and Eri-1 where Eri-1 carried similar polymorphisms than Col-0.

II.3.D. Genotyping of the L*er*/Eri-1 RIL population and building a genetic map

The 112 RILs generated from the cross between L*er* and Eri-1 were previously genotyped with molecular markers at the Wageningen University (Plant Genetics, Wageningen, the Netherlands). In this analysis, amplified fragment length polymorphism markers (AFLPs) were used to create a linkage map of these loci (Vos 1995). Using the AFLP-PCR technique, it was possible to detect polymorphisms between the parental DNA of L*er* and Eri-1 and to assign to each of the 112 RILs alleles present at each locus. Ninety AFLP marker fragments were amplified with these AFLP primer combinations (see Table II.1) and used to build a genetic map. Using the computer software Joinmap® 4 (Van Ooijen J.W. 2006), five *A. thaliana* linkage groups with 90 AFLP markers were built. Since *A. thaliana* has five chromosomes, each of the linkage groups generated represents one chromosome. The constructed genetic map can then be considered as saturated.

Table II.1: Primer and primer combinations used in the AFLP analysis performed at Wageningen University.

Primer	Selective nucleotides	M50 CAT	M60 CTC	M62 CTT
E12	AC		X	X
E13	AG			X
E14	AT		X	
E25	TG	X		X
E26	TT			X

E, *Eco*RI primer GACTGCGTACCAATTC
M, *Mse*I primer GATGAGTCCTGAGTAA

However, the AFLP markers are not anchored to the physical map of *A. thaliana* and each linkage group cannot be assigned to a specific chromosome. In addition to this AFLP genotyping, a new genetic map of the 110 RILs population (2 RILs were excluded, see results III.1.) was constructed including PCR-based markers which reveal a length-polymorphism (SSLP = single sequence length polymorphism or microsatellite) with known physical position. Several markers were designed by previous members of the

research group in Cologne, and tested to reveal polymorphism between accessions other than L*er* and Eri-1. Not all makers could be used and finally, 39 polymorphic microsatellite markers were chosen to genotype the 110 L*er*/Eri-1 RIL population. The segregation of the parental alleles for each marker locus was studied in detail and then all genotyping data including the AFLP information was combined. The AFLP genotyping was done in the F_8 of L*er*/Eri-1 population. By calculating the recombination fractions between the markers, grouping trees could be build and due to known positions of the SSLP markers, each chromosome could be identified. Grouping trees were converted into map order and genetic distances (cM), using Joinmap® 4 (Van Ooijen J.W. 2006) to construct a new genetic map of the L*er*/Eri-1 RIL population (Ghandilyan et al. 2009).
Table A.1 in appendix presents the information concerning the 39 microsatellite markers. These 39 microsatellite markers were also used to genotype the NILs and recombinant plants during validation and fine mapping of the detected QTL. In addition to these 39 markers new markers were developed (see paragraphs II.3BB.) and included for fine mapping of the recombinant plants.

II.3.E. Selection of Near Isogenic Lines

For validation of the detected QTL near isogenic lines (NILs) with introgressions at QTL3, QTL5-1 and QTL5-2 region in a L*er* or Eri-1 allelic background were generated by backcrossing useful RILs either to one of the parental lines L*er* and Eri-1 or to another RIL. For the crosses flowering plants with only a few yet developed siliques were chosen. Closed flower buds were taken as mother of the cross. The flower buds were opened with a pair of tweezers (No.5; A. Dumont & Fils, Montignez, Switzerland) and the anthers of the mother were removed, if possible, without removing the sepals and petals. Thereafter, mature anthers were taken from the pollen donor to pollinate the stigma of the mother plant. The NILs were then developed by marker assisted selection in the progeny of the crosses. The validation of the QTL with the selected NILs was performed in both environmental conditions to mimic the experimental design during QTL detection. The selection of recombinant NILs was performed in a greenhouse under long day conditions. The greenhouse was climate controlled for 20°C during the day and 18°C during the night and for 60% relative air humidity.

II.3.F. Gene expression analysis

RNA was isolated from adult leaf material (after DAS 43) of L*er*, Eri-1 and L*er HUA2* grown in HL4 condition (RNeasy, QIAGEN, Hilden, Germany). The RNA concentration of these samples was determined by Nanodrop (ND-1000 Spectrophotometer, peqlab Biotechnologies GmbH, Erlangen, Germany). For cDNA synthesis around 500 ng of isolated RNA was used as a template. First strand cDNA was synthesized with oligo dT primers using the SuperscriptII Kit (Invitrogen, Carlsbad, CA) according to the manufacturers instructions. Quantitative real time PCR reactions were conducted in a Mastercycler (ep realplex², Eppendorf, Hamburg, Germany). A total reaction volume of 20 µl reaction was used for real-time PCR containing 1 µl cDNA, 0.5 µl of each (forward and reverse) gene specific primer (10 µM), 8 µl IQ^{TM} SYBR Green Super Mix (BioRad, Hercules, CA) and 10 µl of water (Roth, Karlsruhe, Germany). The gene specific primers were designed with ProbeFinder version 2.45 (www.roche-applied-science.com/sis/rtpcr/upl/index.jsp), based on Primer3 software (Rozen and Skaletsky 2000). The primer combinations of the targeted genes were as listed in Table A.2 in appendix. The qRT-PCR reactions were carried out following the thermal profile: 95°C for 2 min, followed by 40 cycles of 95°C for 15 s, 60°C for 30 s and 68°C for 20 s. After 40 cycles absence of primer dimers and amplicon specificity was tested in a melting curve. Data analysis was performed using the realplex software version 2.2. All amplification curves were analysed with a normalized reporter (the ratio of the fluorescence emission intensity of SYBR Green to the fluorescence signal of the background) threshold of 0.2 to obtain the Ct-values (Cycle threshold). Quantification of expression differences of the targeted genes were calculated using the standard curve method (Alcázar et al. 2009). Standard curves for targeted genes and reference consisted of 1:2 serial dilutions of template cDNA. All samples and standard curves were analysed in triplicates. The expression data were normalized by subtracting the mean copy number of the reference gene *Actin2* from the mean copy number of the targeted gene.

II.4. Trait quantification for detection, validation and characterization of QTL3, QTL5-1 and QTL5-2

II.4.A. Quantification of flowering time related traits

During the QTL experiments the F_{10} of the 110 RILs were phenotyped in a randomized complete block display in triplicate (three plants observed per line). Flowering time was scored by daily inspection in each of the two environments as the number of days from sowing till opening of the first flower bud. The total leaf number (TLN) i.e., rosette leaf number (RLN, excluding the cotyledons) plus cauline leaf number (CLN) were also scored. Approximately 10 to 14 days after flowering, between three to four (dependent on plant size) rosette leaves were harvested from each plant for biochemical tests such as determination of the chlorophyll a (Chla), chlorophyll b (Chlb) and anthocyanin content. Figure II.3 shows images of different flowering plants growing in the HL4 environment. At the date of harvesting, RILs showed variation in accumulation of anthocyanins, noted by the different coloration of the leaves. After harvesting, leaves were directly frozen in liquid nitrogen and stored at -80°C until the anthocyanin and chlorophyll content measurement.

Figure II.3: Image of different flowering RILs and Eri-1 as an example for scoring FT in TLN. The five displayed plants are of the same age, the photos were taken with a digital camera (Sony Cybershot DSC F-828) on the same day (41 days after sowing). On the picture are RIL 48 (first plant repetition), RIL 65 (first plant repetition), RIL 69 (second plant repetition), RIL 108 (third plant repetition) and the parental line Eri-1 (left to right, up to down). Flowering started at different time points. RIL 65 started to flower one day before the photo was taken. Eri-1 flowered at the same day the photo was taken.

During validation experiments flowering time was also quantified in DAS when flowering occurred. The number of rosette and cauline leaves were counted after flowering had commenced. During experiments aimed at the selection of recombinant NILs all plants were phenotyped for flowering time related traits. The effect of the introgression in the NILs on the different traits was tested in comparison to the parental accessions L*er* and Eri-1.

II.4.B. Quantification of biochemical – pigmentation related traits

To quantify the chlorophyll and the anthocyanin content from the same sample at the same time, a protocol was used which enabled the subsequent extraction. The protocol is based on purification of plant pigments with acetone and subsequent isolation of anthocyanins with chloroform. After extraction of chlorophylls and anthocyanins the content was quantified by a photometric test. The harvested rosette leaves of the CT, HL4 environments were freeze-dried (Christ lyophilizer Alpha 1-4 LSC, Osterode, Germany) for three days. Depending on the harvested samples, three to four lyophilized leaves were ground with a mixer mill (Retsch MM 300, 220-240 V, 50/60 Hz laboratory mixer mill). Two steel beads (5 mm) were added to the leaves in 2 ml tubes (Eppendorf, Hamburg, Germany) before the tube rack was placed between adapter plates of the TissueLyser Adapter Set 2x96. The samples were homogenized two times for 1 min at 30 Hz. Because of the instability of Chla and Chlb, the samples were kept on ice and in the dark during the entire extraction process. Before this experiment was started, preliminary tests were performed with only the parental accessions L*er* and Eri-1, in order to determine the correct amounts of sample weight (2-10 mg) that should be used for quantifications and photometric analysis. During this pilot test, it was determined that plants grown in the CT condition; 10 mg of material was required to reach the level of detection (for anthocyanins) in the photometric test. From the plants grown in the HL4 environment, only 2 mg of leaf sample were needed. We showed that with an amount exceeding 2 mg, it is necessary to do extra purification steps to extract all anthocyanins from the weighted sample. Samples in the HL4 environment were collected in two times 2 mg and were weighed, extracted and quantified (technical replicate; two tests of the tissue material of the same plant). For calculating the chlorophyll and anthocyanin content the exact weight of the plant samples

was recorded. The powdered samples were purified with 100% acetone (Sigma-Aldrich Laborchemikalien, Seelze, Germany), vortexed (Heidolph RFAX 2000, Germany) and mixed (HLC Thermomixer HTMR 2-133, Bovenden, Germany) for ten minutes at the maximum speed level (9000g). After a centrifugation step for 5 min at 9000 g (Biofuge fresco, Heraeus instruments, Osterode, Germany), the green supernatant was kept and transferred to fresh 2 ml tubes. The remaining pellet was washed twice with 70% acetone and all the supernatants were mixed. To separate the anthocyanins from the chlorophylls chloroform 1:1 (Merck KGaA, Darmstadt, Germany) was added to the supernatant. The centrifugation step separated the chloroform phase and the water containing phase. Anthocyanins were dissolved in the aqueous phase and the organic phase comprised the chlorophylls. The organic phase was suitable to quantify the chlorophylls directly in these samples. For the quantification, the absorption at different wavelength of the samples was measured in a multiscan spectrophotometer (Thermo Electron Corporation, Dreieich, Germany). Because the chlorophylls were dissolved in chloroform, special photometer plates were required (PP microplate 96 well flat bottom, Greiner bio-one GmbH, Frickenhausen, Germany). In order to have a measurement replicate, each sample was added twice to the plate. For Chla, the absorbance at the wavelength of 663 nanometre (nm), for Chlb, the absorbance at 646 nm was measured. The additive measurement of the absorbance of 750 nm allowed the subtraction of the disturbance background. Porra (2002) reported the following equations to calculate the Chlorophyll a and Chlorophyll b content in 80% acetone.

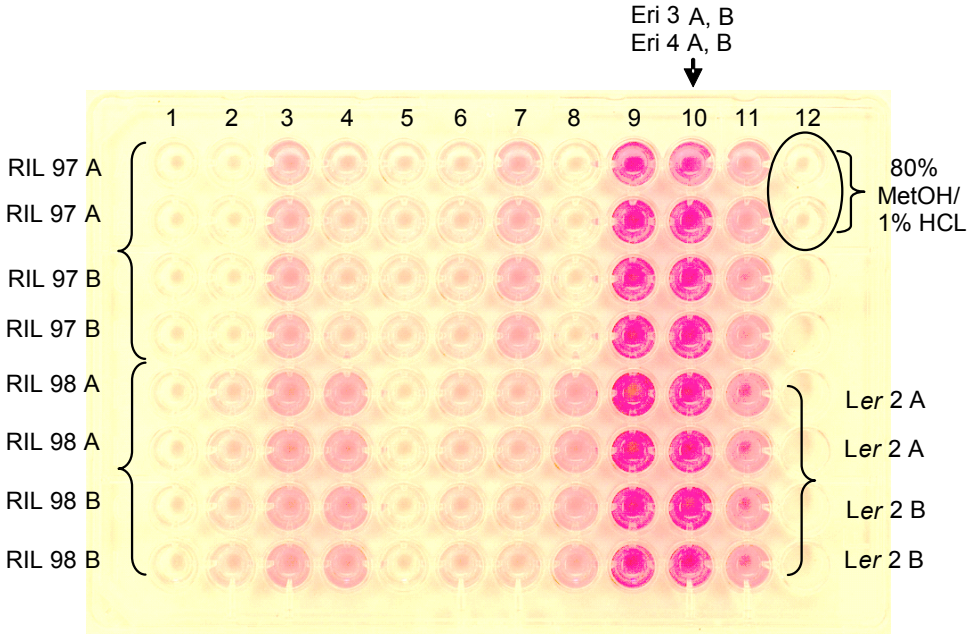
Chlorophyll a (µg/ml) = $12.25\ (A_{663} – A_{750}) – 2.55\ (A_{646} - A_{750})$
Chlorophyll b (µg/ml) = $20.31\ (A_{646} – A_{750}) – 4.91\ (A_{663} - A_{750})$

With A = absorbance at the subscript number in nm.

In order to quantify the anthocyanins samples were dissolved in 80% methanol / 1% hydrochloric acid (Merck KGaA, Darmstadt, Germany) and measured in PS microplates (96-well flat bottom, greiner bio-one GmbH). An example of a PS microplate filled with samples for anthocyanin quantification is displayed in Figure II.4. Owing to the extraction the anthocyanins were dissolved in water. With the use of a vacuum centrifuge, water was evaporated after at least twelve hours. The measurement has been carried out in the same photometer, measuring the absorbance of 530 nm for anthocyanins. Sims and Gamon (2002) reported equations for determining anthocyanin in methanol/HCL/water.

Anthocyanin (µg/ml) = (A_{530} / weight of used sample)* 1000

A blank with 80% methanol / 1% hydrochloric acid was measured at the same time point. The value of the measured blank was subtracted from the absorbance value at 530 nm to negate the error of background.



	1	2	3	4	5	6	7	8	9	10	11	12
A	RIL 97A	RIL 99A	RIL 101A	RIL 103A	RIL 105A	RIL 107A	RIL 109A	RIL 111A	Eri 1A	Eri 3A	Ler 1A	80%MetOH/1% HCL
B	RIL 97A	RIL 99A	RIL 101A	RIL 103A	RIL 105A	RIL 107A	RIL 109A	RIL 111A	Eri 1A	Eri 3A	Ler 1A	80%MetOH/1% HCL
C	RIL 97B	RIL 99B	RIL 101B	RIL 103B	RIL 105B	RIL 107B	RIL 109B	RIL 111B	Eri 1B	Eri 3B	Ler 1B	
D	RIL 97B	RIL 99B	RIL 101B	RIL 103B	RIL 105B	RIL 107B	RIL 109B	RIL 111B	Eri 1B	Eri 3B	Ler 1B	
E	RIL 98A	RIL 100A	RIL 102A	RIL 104A	RIL 106A	RIL 108A	RIL 110A	RIL 112A	Eri 2A	Eri 4A	Ler 2A	
F	RIL 98A	RIL 100A	RIL 102A	RIL 104A	RIL 106A	RIL 108A	RIL 110A	RIL 112A	Eri 2A	Eri 4A	Ler 2A	
G	RIL 98B	RIL 100B	RIL 102B	RIL 104B	RIL 106B	RIL 108B	RIL 110B	RIL 112B	Eri 2B	Eri 4B	Ler 2B	
H	RIL 98B	RIL 100B	RIL 102B	RIL 104B	RIL 106B	RIL 108B	RIL 110B	RIL 112B	Eri 2B	Eri 4B	Ler 2B	

Anthocyanin quantification

Figure II.4: **Image of a PS microplate for quantification of anthocyanins by photometric tests (top) and schematical display of the PS microplate (bottom)**. **Top:** The extracted anthocyanins from RIL leaves and the parental lines are dissolved in 80% MetOH/1% HCL. This plate shows the samples of the RILs 97-112 plus Eri-1 1-4 and L*er* 1, 2 dissolved in 80% MetOH/1% HCL. For each RIL a technical replicate was desired (two analysis (A, B) of one plant) and a measurement replicate (two photometric measurements of the same sample). Beside the photometric test of absorbance at 530 nm for anthocyanins the different level of the anthocyanin concentration in the RILs is obvious by the coloration of the samples. **Bottom**: The schematic display shows the filling order of the anthocyanin samples of RIL 97-112, Eri-1 1-4, L*er* 1, 2 in the wells of the PS microplate.

II.5. Statistical analyses

II.5.A. Statistics

All statistical analyses were performed using SPSS 13.0 for Windows (SPSS inc., Chicago, IL). Broad sense heritability (h^2) of the different FT related and pigmentation related traits were estimated as the proportion of variance explained by between- RILs differences based on measurements of three plants per genotype. The correlation between traits was obtained using principle component analyses (PCA) in the statistical package of SPSS. To test significant differences in the trait values of the parental lines in comparison to the NILs and recombinant lines a Student-Newman-Keuls test (SNK) or the univariate linear model of SPSS was used.

II.5.B. QTL mapping

For each RIL the mean value was taken for each measured trait in the two different environments for QTL analysis with the software package MapQTL® 5.0 (van Ooijen, 2004). MapQTL® 5.0 was used to detect QTL on the genetic map by using interval mapping (IM) and multiple-QTL model (MQM) mapping methods. For each environment a separate QTL detection was performed. In a first step, putative QTL involved in the variation of the traits were identified using interval mapping (Lander and Botstein 1989, also described by van Ooijen 1992). In IM the information of the data supporting presence of a QTL in a defined region of the genome is given by a likelihood ratio statistic, the LOD score (Alonso-Blanco et al. 1998). LOD = log (likelihood for the presence of a segregating QTL / likelihood for no segregating QTL). The LOD significance threshold was calculated by a permutation-test analysis (Churchhill, G.A. and Doerge, R.W. 1994) implemented in MapQTL®5.0 using at least 1000 permutations of the original data set for each trait, resulting in a 95% LOD threshold at 2.1. In IM the entire linkage map was scanned for QTL effects and the LOD score was calculated at every marker position and between markers (e.g. every cM). At the position where a QTL was expected to be located (with a LOD $\geq$ 2.1) the LOD profile showed a peak at this position and low LOD values outside the QTL regions. In a MQM the search for a single QTL is enhanced by taking into account all the previously detected QTL in the statistical model. This is done by using

markers closely linked to the previously detected QTL as cofactors. If a QTL explains a large proportion of the total variance, then the use of a linked marker as a cofactor will considerably enhance the power in the search for other segregating QTL. With MapQTL an automatic cofactor selection (ACS) is possible. The software uses a backward selection procedure to eliminate the markers with low association with the trait until only the markers involved significantly in the variation of the trait remain. The ACS procedure was applied chromosome per chromosome where QTL where obvious after the IM test to select markers to be used as cofactors for the MQM mapping method. Starting with the chromosome where the most significant QTL was detected during IM, the ACS was performed. Then the ACS was used at the next chromosome where also significant QTL was/were found, while keeping the resulting final selection of the previous chromosome. Markers that were selected during ACS but did not result in QTL with significant LOD scores were discarded. The final set of cofactors was used for MQM mapping. The results of MQM mapping also provided the estimates of total variance explained by each QTL as well as the additive allelic effect of each QTL.

II.5.C. Two-way marker interaction

Detection of significant pair-wise interactions between all the markers used to genotype the RILs were performed using the software Epistat (Chase et al. 1997). Both interacting and main effect QTL were then statistically tested using the general linear model module of SPSS 13.0. QTL models were composed of all statistically significant ($P<0.05$) main and interacting QTL.

III. Results

III.1. Genotyping of the RILs and construction of a L*er*/Eri-1 genetic map

The 112 RILs derived from the cross between L*er* and Eringsboda (Eri-1, Sweden) which were analysed in this study have been previously genotyped with 89 AFLP markers at the University of Wageningen (Wageningen, the Netherlands) in order to build a genetic map for this population. Most of the AFLP markers were not anchored on the reference physical map of *A. thaliana* (www.arabidopsis.org) leading to some difficulties in assigning different linkage groups obtained to the corresponding chromosomes. In addition, this set of AFLP markers was not evenly spread along the genetic map, resulting in chromosomal regions without genotyping information. In order to encompass these problems, a set of physically anchored SSLP markers have been selected. The selection of the SSLP markers was done according to (i) their physical position, in such way they are equally distributed across the *A. thaliana* genome; (ii) the high quality and specificity of their PCR products (*e.g.* good amplification; no unspecific amplifications); (iii) their level of polymorphism (difference in size of PCR product between the allele of L*er* and Eri-1). According to these criteria 39 SSLP markers (Table A.1 in appendix) were selected and used to genotype the 112 RILs in addition to the 89 AFLP markers. Closely linked markers (absence or low level of recombination observed among the population of RIL) were underlined and a set of thirteen AFLP markers was removed for further analysis. In total 115 markers were used to build the genetic map. The first genetic map created for the cross between L*er* and Eri-1 (Ghandilyan et al. 2009) resulted in a genome of 365 centi-Morgan (cM) (Figure III.1). This is in accordance with the genetic maps of *A. thaliana* obtained from different crosses (Alonso-Blanco et al. 1998, Loudet et al. 2002, El Lithy et al. 2006). The selected markers are evenly distributed along the five chromosomes of *A. thaliana* with an average genetic distance between two successive markers of four cM. Nevertheless, it is worth noting chromosomal regions of more than 12 cM without markers (chromosome II between T2N18 and F17A22; chromosome III between DF.76L and BH.120L-Col; chromosome IV between markers M4-36 and G3883; Figure III.1). No significant allelic distortion was observed for this new mapping population.

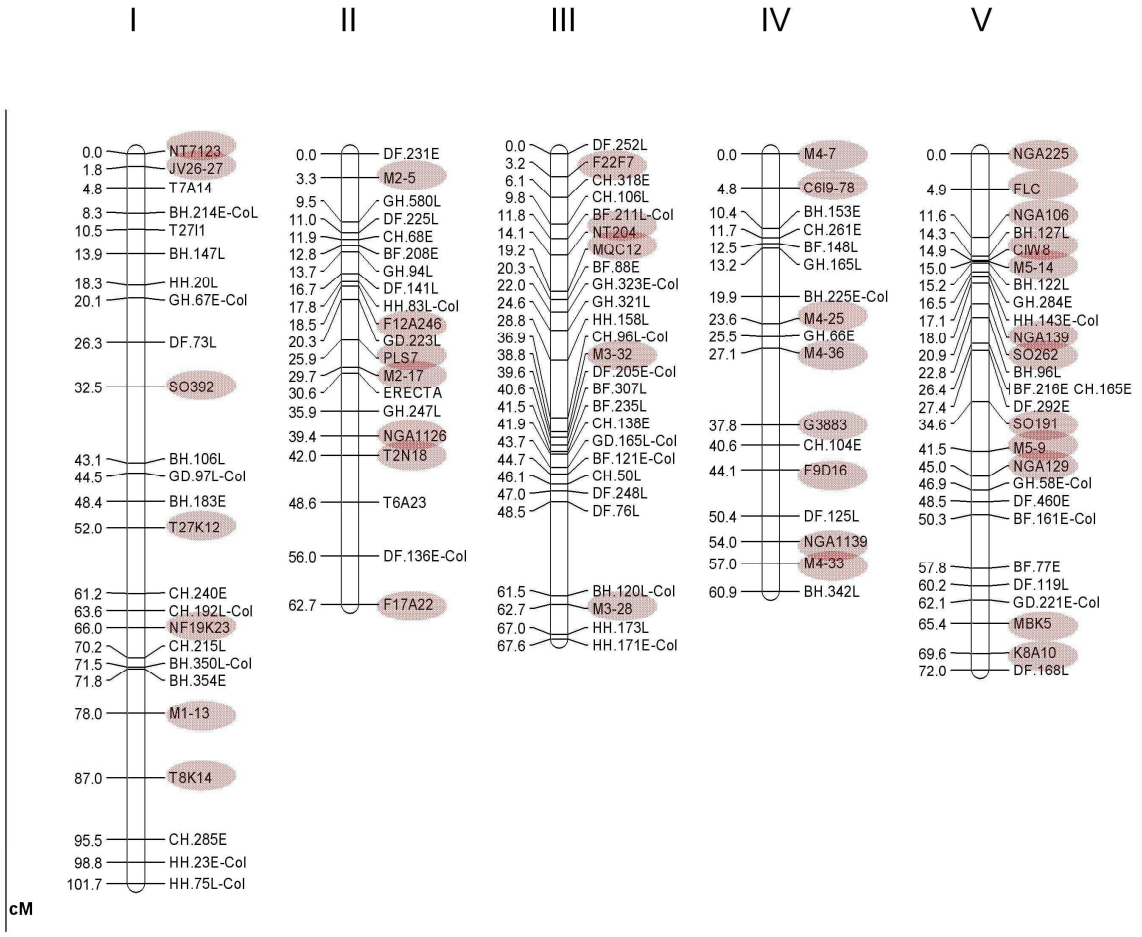


Figure III.1: Genetic map of the L*er*/Eri-1 RIL population genotyped with 115 markers. The map combined 76 AFLP markers used in the previous mapping (University of Wageningen, Wageningen, the Netherlands) and the 39 SSLP and microsatellites (displayed with rose ellipses). The figure shows the five chromosomes of *A. thaliana* and the genetic positions of the different markers in centi Morgan (cM).

Using the selected SSLP markers, we observed two RILs (RIL22 and RIL68) which scored different product sizes than those of the parent lines L*er* and Eri-1 (see Figure III.2). These unexpected results suggested strongly that these lines did not belong to the progeny of the cross between L*er* and Eri-1. Both lines were removed from the analysis when developing the genetic map.

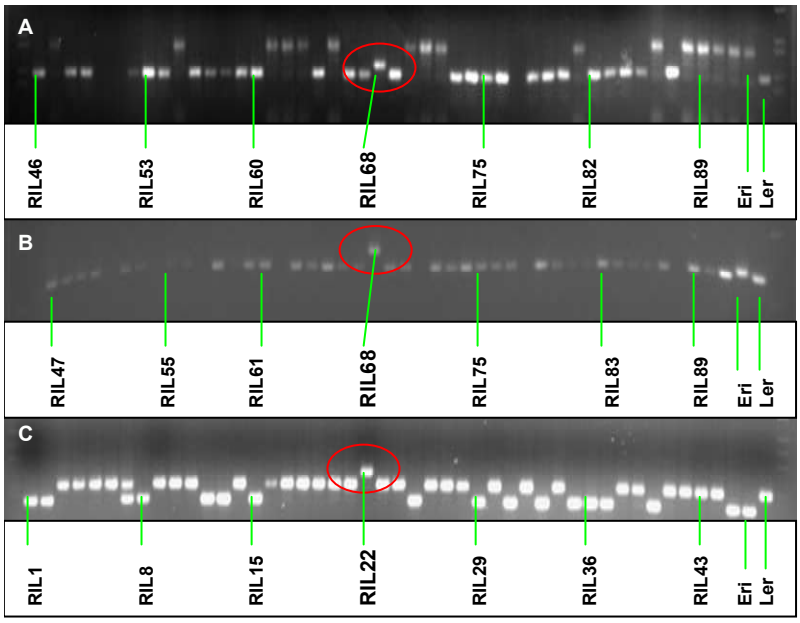


Figure III.2: **Images of fragment separation of PCR products from L*er*/Eri-1 RILs populations on a 3% agarose gel.**
A: RIL 46 to RIL 91 DNA plus parental line DNA of Eri-1 and L*er* were amplified with SSLP marker M3-32 in a PCR reaction and displayed on 3% agarose gel. The larger fragment in the gel shows the allelic value from Eri-1 whereas the smaller represents the allelic value from L*er*. RIL 68 shows a larger fragment size than both parental lines (red ellipse).
B: RIL 47 to RIL 91 DNA plus parental line DNA of Eri-1 and L*er* were amplified with SSLP marker M3-20 in a PCR reaction and displayed on 3% agarose gel. The bigger fragment in the gel shows the Eri-1 genetic background, the smaller fragment represents the L*er* background. RIL 68 shows again a different fragment size than both parental lines (red ellipse).
C: RIL 1 to RIL 45 DNA plus parental line DNA of Eri-1 and L*er* were amplified with SSLP marker M5-9 in a PCR reaction and displayed on 3% agarose gel. The bigger fragment in the gel shows the L*er* genetic background, the smaller fragment represents the Eri-1 background RIL 22 shows a different fragment size than both parental lines (red ellipse).

III.2.Phenotyping the RILs of the L*er*/Eri-1 mapping population

III.2.A. Genetic variation in flowering time and number of leaves

In both environments (CT and HL4) Eri-1 flowered 4-6 days later than L*er* (see Table III.1). The average time of flowering among the RILs occurred between the dates observed in L*er* and Eri-1, although transgression was present in both directions. Indeed, some RILs flowered before L*er* and some after Eri-1. Thus, variation of flowering time in this population was observed in both of the analysed environments. In the control environment, a 1.7-fold variation (from 21 to 35 days after sowing) and in the HL4 a 1.8-fold variation (from 31 to 55 days after sowing) was observed. Interestingly, the widest range of variation was found in high light with cold environment, where flowering time was delayed compared to the control environment. Heritabilities (broad sense; h^2) were calculated to determine the genetic contribution to the observed phenotypic variance. Overall the h^2 in both environments was high (0.77 and 0.76 for the CT and HL4 environments respectively). These high values suggest that the variations of flowering time observed in these environments are driven by a strong genetic determinism. Variation in the number of leaves (RLN, CLN and TLN) was also observed among the RILs in both tested environments. Consistently with FT, the number of leaves of the parental lines differed among the environments (with Control<HL4- Table III.1). In the control environment, the number of rosette leaves varied from 5 to 12 leaves, the cauline leaf number from 1-5 resulting in a variation from 6-16 of the total leaf number. Interestingly, only in a small subset of lines the number of leaves was increased in the HL4 environment. The heritability of RLN, CLN and TLN was even higher than that of FT (0.82, 0.79, and 0.84, respectively in the control environment and 0.81, 0.78, 0.84, in the HL4 environment). Here again, these high values of h^2 suggest variations of these traits are under a strong genetic determinism.

III.2.B. Genetic variation in pigmentation and photosynthetic related traits

Quantification of Chla, Chlb, Chla+Chlb and anthocyanin content was carried out for all RILs grown in CT and HL4 environments. In the CT environment no striking differences in chlorophylls and anthocyanin content between L*er* and Eri-1 were detected (see Table III.1). Only in the HL4 environment were anthocyanin contents different between L*er* and Eri-1, where Eri-1 showed double amount of anthocyanin compared to L*er*. Even in the absence of differences between the parental lines, variations in chlorophyll content (Chla, Chlb, Chla+Chlb in µg/ml) was observed among the RILs in both environments. The range of the Chla content varied from 4.86 to 7.62 µg/ml in the CT and from 1.20 µg/ml to 5.03 µg/ml in the HL4 environment. Similarly, the content of Chlb decreased in the HL4 compared to the CT environment (see Table III.1). In the CT environment, the anthocyanin content varied only from 0 to 2.92 µg/ml among the RILs (see Table III.1) whereas it varied greatly in the HL4 environment (from 1.24 to 128.31 µg/ml). In the latter environment, the average of anthocyanin content among the RILs occurred between the content of the parental accessions, and a wide transgression for values lower than L*er* was observed. Indeed, some RILs had a very low content of anthocyanins (even lower than the value for L*er*) and some others a very high content (almost similar to the value of Eri-1). This suggests that the alleles responsible for the high values of anthocyanin content observed in some RILs under the HL4 environment are originating from Eri-1 only.

It is worth noting that the chlorophyll content in HL4 decreased for both parental lines and the RILs, whereas the anthocyanin content increased at the same time. Heritabilities for Chla, Chlb, Chla+Chlb and anthocyanin content were 0.51, 0.55, 0.50, and 0.56 respectively, in the CT environment and 0.78, 0.70, 0.78 and 0.92 respectively, in the HL4 environment. In the HL4 environment the higher heritabilities observed for all pigmentation and photosynthetic related traits are probably due to the low ranges of variation observed for these traits in the CT environment. Nonetheless, the high estimated heritabilities indicated that variations of these traits are under strong genetic determinism. The amount of the total observed variation that could be explained by genetic variation indicated for all the quantified traits QTL are likely to be revealed.

Table III.1: Parental values and RILs, averages and ranges of FT etc. and heritability (h^2) in the control (CT) and in the high light and 4℃ (HL4) environments.

Trait	Environment	L*er* value	Eri-1 value	Average RILs	Range RILs	h^2
FT (DAS)	**CT**	24	30	26	21-35	0.77
	HL4	38	42	40	31-55	0.76
RLN	**CT**	5	11	8	5-12	0.82
	HL4	8	20	12	5-31	0.84
CLN	**CT**	2	3	2	1-5	0.66
	HL4	2	5	3	1-9	0.77
TLN	**CT**	7	14	10	6-16	0.81
	HL4	9	25	15	7-40	0.84
Chla (µg/ml)	**CT**	5.44	4.85	5.48	4.86-7.62	0.51
	HL4	2.42	2.19	2.77	1.20-5.03	0.78
Chlb (µg/ml)	**CT**	2.16	2.40	2.40	1.62-4.46	0.55
	HL4	0.52	0.48	0.57	0.24-1.00	0.70
Chla+Chlb (µg/ml)	**CT**	8.30	7.25	7.87	6.73-9.85	0.50
	HL4	2.93	2.67	3.34	1.45-5.94	0.78
Anthocyanin (µg/ml)	**CTl**	0.51	0.52	0.63	0-2.92	0.56
	HL4	56.13	117.15	33.23	1.24-128.31	0.92

III.2.C. Phenotypic correlations between flowering-, pigmentation- and photosynthetic related traits over the two studied environments

Pearson correlation coefficients (R^2) between the flowering and pigmentation related traits in the CT and HL4 environments are presented in Table III.2. All correlations between the traits were analysed within and between the CT and HL4 environment (see principle component analysis (PCA) plot in Figure III.3). Most variation was explained by PCA1 with 56.73%. PCA2 and PCA3 explained 11.92% and 8.73% of the variation, respectively. Clustering in the correlation structure in this mapping population reflected the fact that variation of clustered traits can be driven by similar genetic determinism.

In *A. thaliana*, a high correlation is usually observed between flowering time and total leaf number formed prior to flowering. In this new mapping population, strong positive and significant correlations (from 0.45 to 0.99 respectively) between FT and leaf number traits were observed in both tested environments. The highest positive correlation between all flowering related traits was detected in the HL4 environment, due to the presence of more extreme values in this condition. The high genetic correlation suggested that a common genetic determinism for FT and RLN, and FT and TLN could be expected. Between the environments the correlation of the RLN and TLN was found to be consistent. In the PCA plot the flowering related traits quantified in both environments clustered tightly together. This positive correlation again suggests a common genetic factor driving the variation of FT related traits, which is expected in both environments.

In the HL4 environment strong negative and significant correlations between the pigment- (anthocyanin content) and the photosynthetic related traits (chlorophyll content - Chla, Chlb, Chla+Chlb) were detected (from -0.64 to -0.76 respectively). In the PCA plot (Figure III.3) the anthocyanin content of the HL4 environment was isolated from the cluster of Chla, Chlb and the sum of both chlorophylls. In CT environment, no significant correlations between anthocyanin content and chlorophyll content were detected. In contrast with the FT related traits, the photosynthetic and pigmentation related traits did not cluster in both environments (see Figure III.3). However, the Chla content showed a weak positive correlation between the environments, whereas the Chlb content showed a negative correlation (see Table III.2). This denoted that the variations of photosynthetic and pigmentation related traits is dependent of the tested environments. Environmental interaction was also evident between FT related traits and anthocyanin content as well as

for FT related traits and chlorophyll content. Interestingly, significant correlations between FT related traits and the anthocyanin content was only found under HL4 condition (R^2: 0.54-0.66, see Table III.2). In both environments photosynthetic related traits were negative correlated and isolated from FT related traits in Figure III.3. The spatial distribution in the three first components of the PCA of the traits quantified in a mapping population reflected that the variation of these traits is driven either by similar or by diverse genetic determinism. For instance, collocated QTL are expected for FT related traits quantified in both environments, whereas specific QTL are expected for anthocyanin content in each environment.

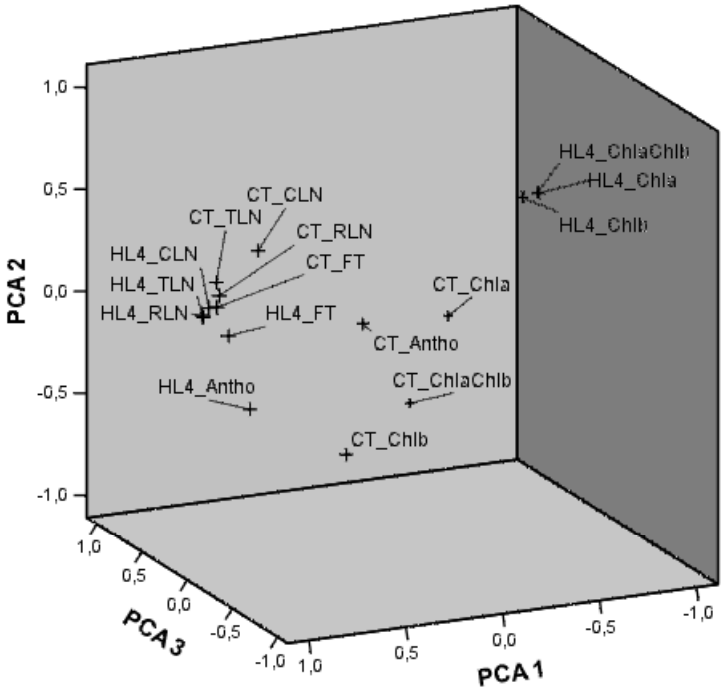


Figure III.3: **Principle component analysis plot of flowering, pigmentation and photosynthetic related traits in CT and HL4 environments.** Principle component analysis (PCA) plot of flowering time related traits (FT, RLN, CLN and TLN), pigmentation (Antho) and photosynthetic related traits (Chla, Chlb and Chla+Chlb) that were quantified in the control (CT) and high light/cold (HL4) environment in L*er* /Eri-1 RIL population.

Table III.2 <on the following page>: Pearson correlation between the traits in the analysed environments. The traits are plotted against each other and the correlation coefficient R^2 of the trait, as well as the 2-tailed significance is displayed.

Phenotypic correlations between quantified flowering and pigmentation related traits and CT and HL4 environments

		HL4-Chla	HL4-Chlb	HL4-Chla+Chlb	HL4-Antho	HL4-FT	HL4-RLN	HL4-CLN	HL4-TLN	CT-Chla	CT-Chlb	CT-Chla+Chlb	CT-Antho	CT-FT	CT-RLN	CT-CLN	CT-TLN
HL4-Chla	R^2	1	,932**	,998**	-,761**	-,627**	-,647**	-,630**	-,651**	,424**	-,381**	,165	-,064	-,585**	-,576**	-,330**	-,545**
HL4-Chlb	R^2	**	1	,952**	-,641**	-,579**	-,579**	-,554**	-,580**	,365**	-,270**	,181	-,046	-,539**	-,516**	-,282**	-,483**
HL4-Chla+Chlb	R^2	**	**	1	-,748**	-,625**	-,642**	-,624**	-,646**	,419**	-,366**	,170	-,062	-,583**	-,571**	-,325**	-,540**
HL4-Antho	R^2	**	**	**	1	,542**	,656**	,618**	,659**	-,313**	,365**	-,065	,049	,597**	,648**	,420**	,633**
HL4-FT	R^2	**	**	**	**	1	,849**	,826**	,858**	-,432**	,244**	-,265**	,144	,790**	,711**	,458**	,680**
HL4-RLN	R^2	**	**	**	**	**	1	,929**	,994**	-,414**	,218*	-,264**	,134	,864**	,834**	,609**	,832**
HL4-CLN	R^2	**	**	**	**	**	**	1	,955**	-,433**	,183	-,306**	,124	,828**	,819**	,655**	,835**
HL4-TLN	R^2	**	**	**	**	**	**	**	1	-,429**	,219*	-,278**	,127	,870**	,843**	,628**	,845**
CT-Chla	R^2	**	**	**	**	**	**	**	**	1	-,325**	,774**	,051	-,428**	-,461**	-,318**	-,436**
CT-Chlb	R^2	**	**	**	**	**	*		*	**	1	,348**	-,043	,198*	,168	,029	,137
CT-Chla+Chlb	R^2					**	**	**	**	**	**	1	,022	-,291**	-,344**	-,296**	-,340**
CT-Antho	R^2												1	,086	-,005	-,060	,016
CT-FT	R^2	**	**	**	**	**	**	**	**	**	*	**		1	,816**	,619**	,823**
CT-RLN	R^2	**	**	**	**	**	**	**	**	**		**		**	1	,672**	,975**
CT-CLN	R^2	**	**	**	**	**	**	**	**	**		**		**	**	1	,760**
CT-TLN	R^2	**	**	**	**	**	**	**	**	**		**		**	**	**	1

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

III.3. QTL mapping

Due to the correlation of traits between and within environments, common genetic determinism and consequently, common QTL are expected for (i) FT related traits in both environments, (ii) chlorophyll content in the HL4 environment (iii) FT related traits and chlorophyll content in the HL4 environment with opposite effects (iv) FT related traits in both environments with anthocyanin content in HL4 condition and (v) chlorophyll content with anthocyanin content in the HL4 environment with opposite effects. QTL were detected for all analysed traits for the L*er*/Eri-1 RIL population (except for anthocyanin content in the CT environment). The results of QTL mapping are summarized in Table III.3 and Figure III.4. Due to the G x E interaction for some traits, there were QTL detected specifically in an environment.

III.3.A. Detection of QTL involved in the variation of flowering, pigmentation and photosynthetic related traits in the CT and HL4 environments

Common QTL were mapped under both studied environments at the very top of chromosome V (from marker NGA225 to marker FLC; 0.0 to 4.9 cM respectively). At this locus, QTL for flowering time, RLN and TLN in CT and RLN and CLN in HL4 were detected with L*er* alleles leading to earlier flowering time. These QTL explained between 3.5% and 5.9% of the explained variance, whereas RLN in the HL4 environment explained up to 28.2% of the flowering variation. Another co-locating QTL detected in both environments was also mapped to chromosome V, around the marker CIW8 (14.9 cM). At this position QTL for all analysed traits (flowering, pigmentation and photosynthetic related traits) were mapped. In the case of flowering related traits and anthocyanin content the L*er* allele induced earlier flowering and decreased the anthocyanin content, whereas Chla content showed increased values. It is worth noting that anthocyanin content QTL at this position was the major QTL for that trait and was specific for the HL4 environment. It explained 32% of anthocyanin content variation. In contrast, QTL for flowering related traits in both environments were mapped at this locus. However, the percentage of explained variance for these QTL was highly dependent of the tested environments. The QTL detected in the CT environment only explained from 12.6% to 25.1%, whereas the

QTL for RLN in the HL4 explained up to 54% of the variation for flowering time. Interestingly, right above the QTL detected for Chla content in the CT condition, which mapped to the QTL close to marker CIW8 for all analysed traits, another major QTL for Chla content was mapped. In this case the L*er* allele decreased Chla content and the QTL explained up to 51.5% of the observed variation of Chla content in the CT condition. A third co-locating QTL for all flowering time related traits detected in both environments was mapped to chromosome V at the marker M5-9 (41.5 cM): At this locus, QTL for FT, RLN and TLN in the CT condition, as well as RLN in the HL4 environment were detected. These QTL for flowering related traits explained only 3.7% to 4.7% of the variation observed in the CT environment and up to 16% in the HL4 environment. Similarly to the other QTL presented above, the L*er* alleles led to an earlier flowering time phenotype.

QTL specific for an environment were also detected: Two QTL for all flowering and photosynthetic related traits were mapped only in the HL4 environment. On the top of chromosome III (from marker DF.252 to marker F22F7; 0 to 3.2 cM) the L*er* alleles increased values for all flowering time related traits and decreased chlorophyll contents. The explained variances for flowering time related traits ranged from 5.1% to 19%, variances for chlorophyll content from 8.7% to 8.8%. The second co-locating QTL for flowering related traits and chlorophyll content specific to the HL4 environment mapped to chromosome V in the chromosomal region flanked by the markers SO262 and BH.96L (20.9 to 22.8 cM). The explained variances for the flowering related traits ranged from 6.1% to 52.7%, where RLN again gave the highest percentage of the explained variance in the HL4. In all QTL of the cluster flowering was induced by the L*er* alleles and at the same time chlorophyll content levels were increased. In the CT environment two clusters of minor QTL, each co-locating for the rosette and total leaf numbers were detected. One of them mapped to chromosome II (in the vicinity of the marker F17A22; 62.7 cM), which explains only 2.9% to 3.2 % of variance of those traits and where the L*er* alleles decreased the number of leaves. The second minor QTL was mapped to chromosome V (marker DF.119L; 60.2 cM) and explained 3.8% to 3.95% of variance and where the L*er* alleles increased the number of leaves. Interestingly in both cases no QTL for flowering time (in DAS) was mapped to these positions. Some minor isolated QTL was detected depending on the trait and the environment. It is worth noting that the second QTL detected for anthocyanin content is on chromosome I close to the marker NF19K23 (66.0 cM). Similar to the other one for anthocyanin accumulation detected on chromosome V, the L*er* allele

decreased the anthocyanin content value (see Table III.3). Surprisingly, there was no co-locating QTL detected for chlorophyll content and anthocyanin content in HL4 environment, whereas the previous correlation test revealed a strong negative and significant correlation between both traits. On the other hand colocalization of flowering related QTL and chlorophyll content QTL were shown in the previously described region between markers SO262 and BH.96L (20.9 to 22.8 cM) specific for HL4 condition. The anthocyanin content QTL was detected just above this region around marker CIW8 (14.9 cM). Due to the high Pearson correlation one would expect that the QTL of these traits cluster together in the same region and do not separate in two regions. It is also worth noting that the explained variation for chlorophyll content QTL at marker SO262 and the variation for anthocyanin content QTL at marker CIW8 showed almost the same percentage.

In the progress of this work, I will focus on the three QTL where flowering related traits are colocalized with pigmentation and photosynthetic related traits. The clusters of QTL for different traits (see Figure III.4) refer to QTL3 (chromosome III; from DF.252 to FF2F7), QTL5-1 (chromosome V; CIW8) and QTL5-2 (chromosome V; from SO262 to BH.96L). These QTL were focused on because they were only detected in the HL4 environment and thus show QTL x E interaction. The QTL5-1 was detected in both environments, but colocalized with the anthocyanin QTL5-1 specific for the HL4 environment.

Table III.3: Characteristics of the detected QTL explaining flowering time, flowering time related traits, photosynthesis related traits and anthocyanin content (Antho) in L*er*/Eri-1 RIL population in CT and HL4 environment

Trait	Environment	QTL at nearest marker	Map position[1]	LOD score	% of explained variance	Additive allele effect[2]
FT	HL4	F22F7	III: 3.2	7.18	14.1	2.30
	HL4	CH.318E	III: 6.1	9.99	19.0	3.10
	CT	NGA225	V: 0.0	3.23	4.5	-0.75
	CT	CIW8	V: 14.9	8.08	12.6	-1.31
	HL4	SO262	V: 20.9	18.13	41.1	-3.20
	HL4	BH.96L	V: 22.8	18.28	41.4	-3.20
	CT	M5-9	V: 41.5	3.33	4.7	-0.65
RLN	CT	F17A22	II: 62.7	3.40	3.2	-0.34
	HL4	F22F7	III: 3.2	3.90	5.9	1.50
	CT	**NGA225**	**V: 0.0**	**3.24**	**3.5**	**-0.42**
	HL4	**NGA225**	**V: 0.0**	**7.85**	**28.2**	**-2.64**
	HL4	FLC	V: 4.9	12.88	42.1	-3.20
	CT	**CIW8**	**V: 14.9**	**17.20**	**25.1**	**-1.20**
	HL4	**CIW8**	**V: 14.9**	**18.00**	**54.0**	**-3.60**
	HL4	SO262	V: 20.9	17.60	52.2	-3.60
	HL4	BH.96L	V: 22.8	17.70	52.7	-3.60
	HL4	SO191	V: 34.6	6.23	23.0	-2.40
	CT	**M5-9**	**V: 41.5**	**3.88**	**4.2**	**-0.49**
	HL4	**M5-9**	**V: 41.5**	**4.17**	**16.0**	**-1.90**
	CT	DF.119L	V: 60.2	3.80	3.8	0.45
CLN	CT	BH.106L	I: 43.1	3.54	7.6	-0.17
	HL4	F22F7	III: 3.2	3.30	6.0	0.41
	HL4	NGA225	V: 0.0	3.23	5.9	-0.45
	CT	CIW8	V: 14.9	5.57	14.0	-0.28
	HL4	SO262	V: 20.9	11.78	25.5	-0.90
	HL4	BH.96L	V: 22.8	11.16	24.6	-0.90
TLN	CT	F17A22	II: 62.7	2.84	2.9	-0.41
	HL4	DF.252L	III: 0.0	3.39	5.1	1.50
	HL4	F22F7	III: 3.2	4.11	6.2	1.63
	CT	NGA225	V: 0.0	3.86	4.6	-0.62
	HL4	SO262	V: 20.9	3.50	5.4	-2.68
	HL4	BH.96L	V: 22.8	4.10	6.1	-2.64
	CT	M5-9	V: 41.5	3.25	3.7	-0.58
	CT	DF.119L	V: 60.2	3.95	4.6	0.62
Chla	HL4	F22F7	III: 3.2	3.92	8.7	-0.16
	CT		V: 11.4		51.5	-2.00
	CT	CIW8	V: 14.9	3.06	13.7	0.20
	HL4	BF.216E	V: 26.4	3.26	7.2	0.19
Chlb	HL4	DF.252L	III: 0.0	3.60	8.8	-0.02
	HL4	F22F7	III: 3.2	3.36	8.3	-0.02
	CT	MQC12	III: 19.2	4.16	15.0	0.14
	HL4	BH.96L	V: 22.8	2.56	6.9	0.03
Antho	HL4	NF19K23	I: 66.0	2.18	5.9	-14.5
	HL4	CIW8	V: 14.9	9.61	32.0	-33.00

[1]) Chromosome nb is given followed by marker position in cM
[2]) Positive values indicate that L*er* alleles increased the trait value
Values in bold indicate QTL detected in both environments

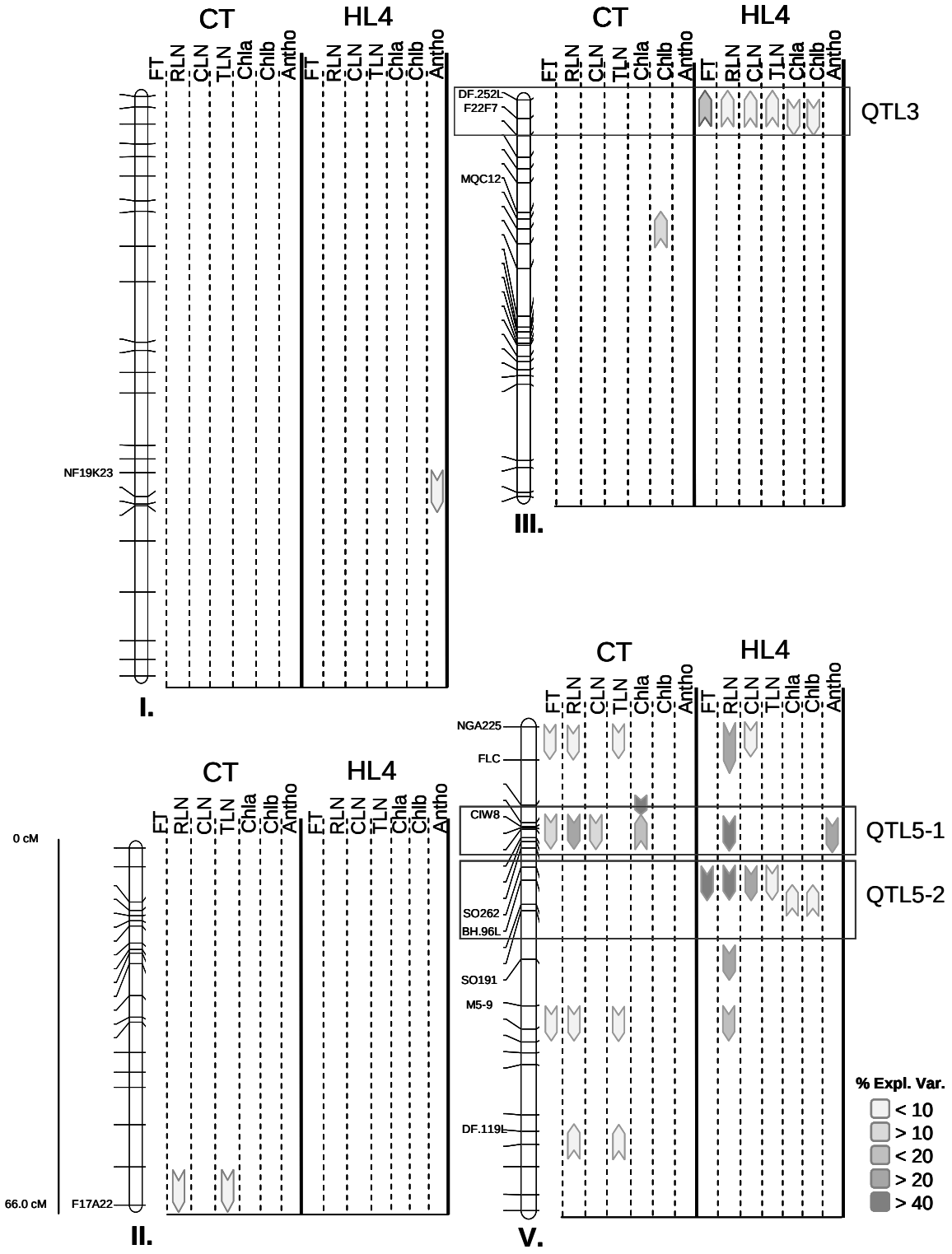

CT
HL4
FT
RLN
CLN
TLN
Chla
Chlb
Antho
NF19K23
I.
DF.252L
F22F7
MQC12
QTL3
III.
0 cM
66.0 cM
F17A22
II.
NGA225
FLC
CIW8
QTL5-1
QTL5-2
SO262
BH.96L
SO191
M5-9
DF.119L
V.
% Expl. Var.
< 10
> 10
< 20
> 20
> 40

Figure III.4 <on the previous page>: L*er*/Eri-1 linkage map showing the locations of QTL for the analysed traits in control (CT) and high light and cold (HL4) environment. Chromosomes I, II, III and V of *A. thaliana* are represented as bars with their number indicated on the bottom of each bar. Marker positions are indicated with lines. The arrows indicate the regions where significant QTL have been detected (LOD>2.1; see materials and methods) with corresponding marker on the left. The direction of the arrows indicates the allelic effect: upward, L*er* increases and Eri-1 decreases, downward, Eri-1 increases and L*er* decreases. The different shading of the arrows refers to the percentage of variance explained by each QTL (see legend). QTL3, QTL5-1 and QTL5-2 cluster are high lighted with grey boxes.

III.3.B. 2-way QTL interactions mapping

For most of the studied traits, the sum of R^2 of the main effect QTL could not explain all the variation which was observed in the RIL population, although the heritability of most quantified traits was high. This likely means that part of the genetic determinism has not been revealed with the main effect QTL detected. In order to encompass this problem, a further analysis with genome wide pair wise marker interaction was performed. The interaction affecting the variation of the studied traits was performed for both HL4 and CT environments separately (Figure III.5).

In both environments and for all analysed traits, significant interactions on chromosome V were detected. The strongest interaction, which explains 29% of the variation of FT in the CT condition was detected between the marker NGA225 (0.0 cM) and marker CIW8 (14.9 cM). This interaction was also significant in the HL4 condition, but rather weak explaining only 7.7% of the observed variance. At both positions main effect QTL were previously detected for FT, RLN and TLN. Additional interaction on chromosome V for both environments was detected between marker CIW8 (14.9 cM) and marker M5-9 (41.5 cM) in CT condition with an explained variance of 23% and in HL4 condition with an explained variance of 20%. A third strong interaction (27%) in the CT condition was detected between marker CIW8 (14.9 cM) and marker SO191 (34.6 cM). This interaction was not detected in the HL4 environment. Surprisingly, at the SO191 region no main effect QTL was detected in CT environment but only in HL4. Interactions between chromosomes were also detected between chromosome III and chromosome V for the traits FT, RLN and TLN. For the leaf numbers it was found in both environments between the marker NT204 (chromosome III- 14.1 cM) and CIW8 (chromosome V- 14.9 cM). The interacting loci mapped at the QTL5-1 locus and a bit below the QTL3 locus. Also on chromosome III, a weak interaction for FT in the HL4 condition was mapped. The previous mentioned region

around the marker NT204 (chromosome III- 14.1 cM) interacted with the QTL3 region at marker F22F7 (chromosome III- 3.2 cM). Additional interactions between chromosomes were also detected between QTL5-1 (CIW8) and chromosome I, II and IV. On chromosome I at marker NF19K23 (66.0 cM) an interaction explained 7.2 % for the variation in leaf numbers in CT condition. This interaction could not be shown in HL4 environment. The interaction between chromosome II (marker F12A246; 18.5 cM) was detected in both environments and explained 11% and 19% of the variance in CT and HL4 condition, respectively. Close to marker position F9D16 (44.1 cM) on chromosome IV again an interaction in both conditions was detected. It contributed 16% and 8.4% to the variation in RLN in CT and HL4 condition, respectively. A summary of all detected interactions for FT in both environments is displayed in Figure III.5.

For anthocyanin content only one major QTL and one minor QTL in the HL4 condition were previously mapped. Epistat revealed interactions which explained additional phenotypic variation and which could not be explained only by the previously main effect QTL detected. Two of the interactions on chromosome V and the interaction between chromosome III and the QTL5-1 region detected for FT were also identified for anthocyanin content. The interactions between the top chromosome V (NGA225) and M5-9 and the QTL5-1 region were detected in the HL4 environment and explained 14% and 8.5% of the phenotypic variance for anthocyanin content. The interaction on chromosome III and QTL5-1 for anthocyanin content was mapped below (MQC12; 19.2 cM; 6.1% explained variance) the interaction detected for FT (NT204; 14.1 cM). An additional interaction on chromosome V was detected in the HL4 environment for anthocyanin content between the QTL5-2 (SO262; 20.9 cM) and the bottom of Chr5 (K8A10; 69.6 cM). This interaction explained 18% of the variation in anthocyanin content in the RIL population in this environment. The region on chromosome III which already interacted with QTL5-1 showed a second interaction with chromosome II at the marker position M2-17 (29.7 cM) and explained the variation in anthocyanin content by 13%. An additional interaction between the QTL5-1 region (CIW8; 14.9 cM) and the region around marker T2N18 (42.0 cM) on chromosome II was detected in HL4 and explained 7.8% of the anthocyanin phenotype in that environment. The last minor interaction explaining only 6.6 % of anthocyanin content variation was found between CIW8 and F9D16 (44.1 cM) on chromosome IV. All these interactions for anthocyanin content detected in the HL4 are displayed in Figure III.5.

For FT in the CT condition a heritability of 77 % was detected. This percentage was not reached by the main QTL for this trait in this condition (21.8% in total). By including the 2-way interactions in the QTL model, the amount of the variation explained by the QTL was similar to the heritability. For anthocyanin content in the HL4 environment the heritability reached 92 %. Only 32 % and 5.9 % were explained by the main effect QTL detected in this condition. In total, 54% of the phenotypic variation was explained by the 2-way interactions for this trait. Thus, when the minimum effect and interacting QTL were considered jointly, the percentage of explained variance was similar to the estimated heritability for this trait. Remarkably, the heritability for RLN in the HL4 environment equalled to 84 %, but the percentage of explained phenotypic variance by the QTL was already much higher.

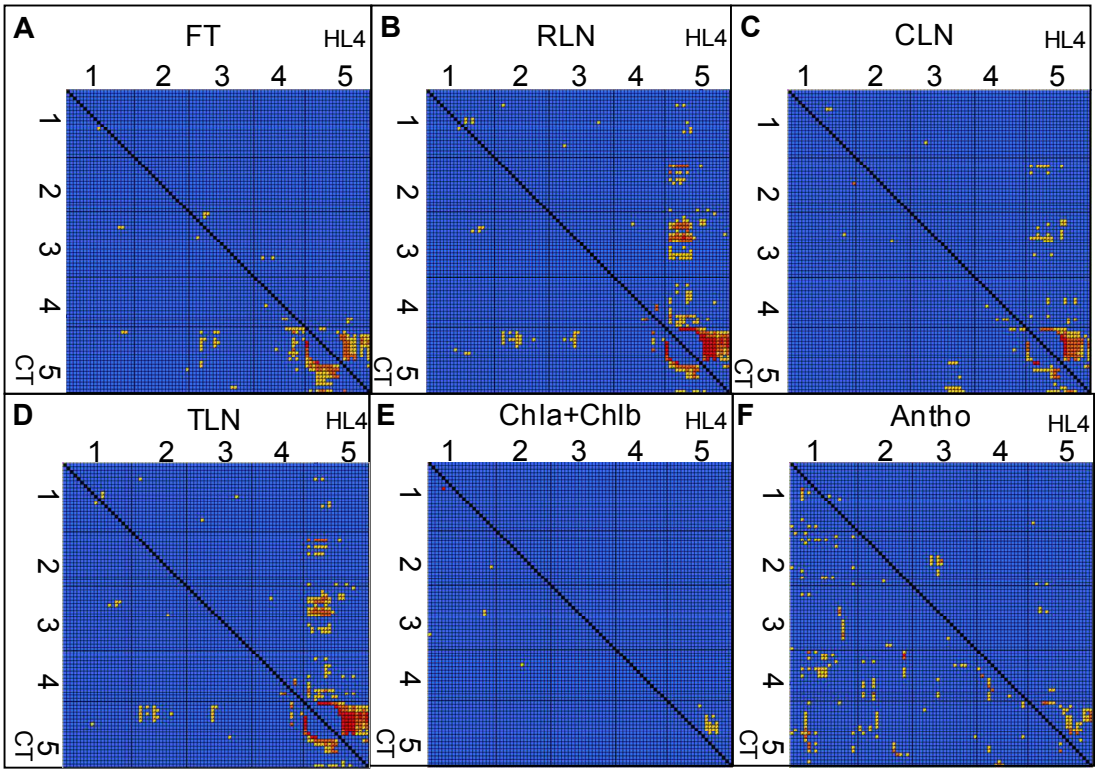


Figure III.5: Heat map of the log likelihood ratio (LLR) from pairwise epistatic interactions detected by Epistat for all flowering related traits and chlorophyll and anthocyanin content detected under CT and HL4 conditions. (A- FT for flowering time, B- RLN for rosette leaf number, C- CLN for cauline leaf number, D- TLN for total leaf number), photosynthetic related traits (E- chlorophyll a + chlorophyll b) and pigmentation-related traits (F-Antho for anthocyanin content) detected under control (CT) and high light/low temperature (HL4) conditions. In each panel (A-F) the significance of the occurrence of an interaction between each couple of markers in the five chromosomes is plotted. Presence of epistatic 2-way interactions with an LLR >6 and >12 are indicated in orange and red respectively. The values below the diagonal represent the interaction detected in the CT condition and the ones above the diagonal stand for the interactions in the HL4 condition.

III.3.C. QTL x environment interactions

Phenotypic variation over contrasted environments can also be under strong genetic determinism and this is commonly referred as GxE for genetic by environment interaction. Deciphering this genetic determinism can lead to better understanding of adaptation to specific environments. Thus, we tested every detected QTL within the three revealed QTL cluster for its QTL x E interaction by an analysis of variance between the tested environments (data not shown). No significant interaction between the QTL3 and the environment was revealed for the flowering related traits. One could expect a QTL x E interaction as most of the QTL were only mapped in HL4 condition. In contrast, the QTL3 detected for Chla and Chlb content in the HL4 environment showed an interaction to the environment. In both QTL5-1 and QTL5-2 on chromosome V nearly all mapped QTL showed a significant interaction with the environment. The only exception in QTL5-1 and QTL5-2 on chromosome V was the QTL for the Chla content for which no interaction with the environment was revealed.

III.4. Validation and characterization of the flowering QTL3

III.4.A. Selection of NILs in L*er* or Eri-1 genetic backgrounds to validate the presence and effect of flowering QTL3

The QTL mapping resulted in the detection of collocating QTL for flowering related traits located on chromosome III named QTL3 (see result above III.3.A.). A way to confirm the presence and the effect of this QTL would be to observe significant phenotypic differences between the parental accessions and near isogenic lines (NILs). The NILs carry the genetic background from the one parental accession with a small introgression of alleles from the other parental accession in the region of the QTL3. In addition to validate the effect of a QTL, near isogenic lines provide a suitable genetic material to fine map a validated QTL. Ideally, narrowing down the size of the introgressed region in the NIL by successive backcross with the corresponding accessions would lead to the identification of the gene(s) underlying the effect of the QTL. Such NILs were selected from successive backcrosses between selected RILs and either L*er* or Eri-1.

We selected appropriated lines from the remaining 110 L*er*/Eri-1 RIL population which after backcrossing either to the parental lines L*er* or Eri-1 or to another RIL would lead to the selection of lines with introgressions at only one, two or all three detected QTL. Backcrossing (BC1F1) the RIL 31 with the parental line L*er* resulted in a line with L*er* background and heterozygous introgressions on chromosome III from 0 to 5.6 Mbp (including the QTL3) (see Figure III.6). More heterozygous introgressions not expected to be involved in the variation of the studied traits were on chromosome IV and chromosome V. In the progeny of this line, a clean NIL referred to L*er* NIL QTL3 could be selected.

The RIL 38 was backcrossed to Eri-1 in order to produce NILs with Eri-1 genetic background. The BC1F1 line carried Eri-1 alleles over the full length of chromosome I, except on the top of the chromosome and chromosome II and smaller or bigger heterozygous introgressions on the chromosome III, chromosome IV and chromosome V – from the progeny of this line an almost clean line could be selected which is named Eri NIL QTL3. The heterozygous introgression at the top of chromosome III also from 0 to 5.6 Mbp included the QTL3 (see Figure III.6). The BC1F2 populations (192 plants) from each cross were grown in CT condition to select suitable near isogenic lines for the validation of QTL3. The progeny was genotyped with the SSLP markers which were previously used to build the genetic map of L*er*/Eri-1 and additional markers found to be polymorphic between L*er* and Eri-1 (see marker Table A.1 in appendix). Lines were selected according to the following criteria: (i) for marker elsewhere than in the QTL3 region, the allelic value should be similar than the one of the corresponding genetic background; (ii) for marker located within the QTL3 region, the allelic value should be from the other parent than the one of the genetic background (heterozygous or homozygous stage). None of the BC1F2 plants genotyped carried only an introgression in the QTL3 locus. In an attempt to select cleaner NILs the progenies (BC1F3; 25 plants) from each BC1F2 selected line were grown in the CT condition. All plants were not only genotyped but also phenotyped (flowering time and numbers of leaves) in order to test the co-segregation between the variation of these phenotypes and the allelic values of segregating markers. Due to the different allelic value of the lines and remaining heterozygous parts in the BC1F3 a phenotypic variation of the analysed traits was expected and observed. In the BC1F3 of the cross between RIL 31 and L*er*, a clean L*er* NIL QTL3 (L*er* NIL 97-3) could be selected (see Figure III.6). L*er* NIL 97-3 carried an Eri-1 introgression on the top of chromosome III at marker F14P3

(0.42Mbp) and a second Eri-1 introgression from marker F7O18 (1.23Mbp) until marker position NT204 (5.6Mbp – marker Table A.1 in appendix).

To narrow down the QTL3 region, the progeny of an Eri NIL QTL3 line with smaller L*er* introgression at chromosome III (from 0 to around 1.6 Mbp – Eri NIL 67-4) was selected and the BC1F4 (15 plants) of the Eri NIL 67-4 were additionally grown and phenotyped in the HL4 condition. All the lines were also genotyped. Based on the combination of genotyping and phenotyping data the clean Eri NIL 67-4-2 which only carried one introgression of the L*er* alleles on the very top of chromosome III and in a small region on chromosome IV was selected. Additionally Eri NIL 67-4-10 was chosen which was isogenic to NIL 67-4-2, except for heterozygous alleles at the top of chromosome III. The selection and genotype of both NILs are shown in Figure III.6.

The clean Eri NIL QTL3 (Eri NIL 67-4-2) carried a L*er* introgression at the top of chromosome III from marker T4P13-0.014Mbp (first marker on top chromosome III) to FTN3-175-1.75Mbp (see marker table A.1 in appendix) including the QTL3. The offspring of the heterozygous sister line Eri NIL 67-4-10 was used for later fine mapping experiments. Both Eri NILs carried a L*er* introgression on chromosome IV marker position M4-25 (6.97Mbp) where no QTL for flowering time were detected.

Both NILs, L*er* NIL and Eri NIL carried bigger and smaller introgressions including QTL3, respectively. The progeny of both clean NILs was grown in both environments under the same conditions which were used for the QTL mapping experiment in order to validate the effect of the QTL3.

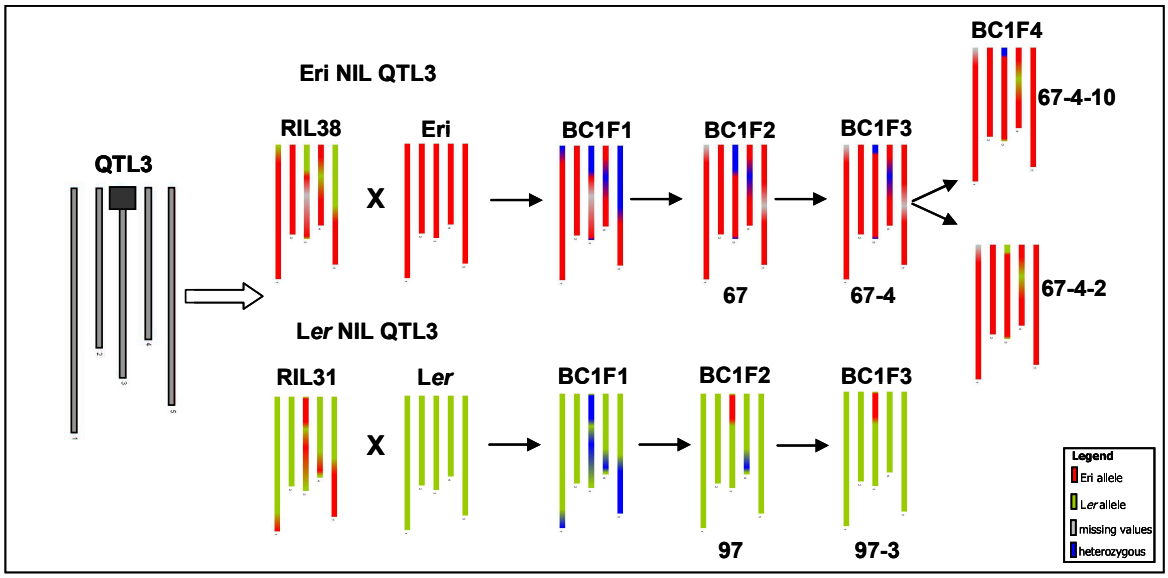


Figure III.6: Selection strategy and genotype of Eri NIL QTL3 and L*er* NIL QTL3. The five bars indicate the five chromosomes of *A. thaliana*, the dark grey box highlights the position of the QTL3, and the colours represent the allele (red-Eri-1, green –L*er*, blue heterozygous, grey-missing values). For developing the Eri NIL QTL3 (L*er* introgression at QTL3 in Eri-1 background) the RIL38 was backcrossed to the parental line Eri-1. In selection steps an almost clean Eri NIL QTL3 (NIL67-4-2) was selected. For developing the L*er* NIL QTL3 (Eri-1 introgression at QTL3 in L*er* background) the RIL31 was backcrossed to the parental line L*er*. In selection steps a clean L*er* NIL QTL3 (NIL97-3) was selected. The names of the NIL lines are printed below the respective genotypes.

III.4.B. Validation of flowering QTL3

The performed QTL detection and 2-way marker interaction study showed that the allelic values at the marker F22F7 (1.6 Mbp) on chromosome III could significantly explain parts of the phenotypic variation for the traits FT, RLN and TLN of the L*er*/Eri-1 RIL population grown in HL4 environment. At this position the L*er* alleles delayed the FT in HL4 for an average of about 5 days and increased the RLN and TLN on average for about 3 leaves compared to the Eri-1 alleles. In the CT environment the allelic value at marker F22F7 could not explain variation in leaf numbers and resulted in not significant changes. However also in this condition the L*er* alleles delayed the FT on average around 1.5 days and increased the leaf numbers around 0.5 leaves. The tendency that the L*er* alleles at QTL3 delayed FT was also present when phenotyping the NILs. The L*er* NIL97-3 (L*er* NIL QTL3) which carried the bigger Eri-1 introgression on top chromosome III showed an earlier flowering (in RLN) compared to L*er* (see Figure III.7). The leaf numbers in both

environments were significantly decreased, whereas the differences in the leaf numbers between the L*er* NIL97-3 and the L*er* parent were exacerbated in the HL4 environment (see next paragraph). The RLN and TLN were on average 1.5 leaves decreased in the L*er* NIL97-3 compared to L*er*. In the HL4 the RLN and TLN were on average 2 to 2.5 leaves decreased in this NIL compared to L*er*. The QTL mapping only highlighted significant effects of QTL3 in the HL4 environment, but interestingly the QTLxE analysis did not show any interaction between QTL3 and the tested environments, which indicated that QTL3 is probably independent from the environmental condition and would be detectable in both conditions if we would have more power for QTL detection. When validating QTL3 by phenotyping selected NILs, we could univocally confirm the effect of QTL3 in both environments. However, in the CT environment the effect of the QTL3 has a lower impact on FT related traits. While analysing heterogenous RILs the low effect of QTL3 in CT condition was perhaps masked by other flowering time QTL. Different allelic value combinations in the RILs could lead to a higher impact of other QTL related to flowering.
Additional to the L*er* NIL97-3 the Eri NIL67-4-2 (Eri NIL QTL3) which carried a smaller introgression of the L*er* allele on the very top chromosome III in an Eri-1 background also validated the late flowering caused by the L*er* allele at QTL3. In both environments the Eri NIL67-4-2 showed a much later flowering phenotype compared to Eri-1. In the CT environment the RLN and TLN of this NIL were increased on average around 6-7 leaves compared to Eri-1. Again, in the HL4 environment the effect was amplified and the RLN and TLN were increased on average around 12-15 leaves compared to Eri-1. According to a Student-Newman-Keuls test the leaf numbers of L*er*, Eri-1 and the selected NILs could be significantly divided into four different groups, with a likelihood of 95% (Figure III.7). The phenotypic values of Eri NIL67-4-2 were very interesting due to the fact that the allele of the earlier flowering parent L*er* even delays the flowering of the later flowering parent Eri-1. This later flowering can be only explained by a combination of the L*er* allele at QTL3 and another region or other regions in the Eri-1 background. This combination could already be seen in the RIL population. An interaction between the marker NT204 (chromosome III 5.6 Mbp), located a bit below the QTL3 region and CIW8 (chromosome V 7.5 Mbp) segregating marker of the QTL5-1 was previously detected (see above, paragraph III.3.B.). Indeed, almost all later flowering RILs carried the L*er* alleles at the QTL3 region and Eri-1 alleles at the QTL5-1 region. The interaction between the QTL3 and chromosome V will be described and discussed in a following chapter.

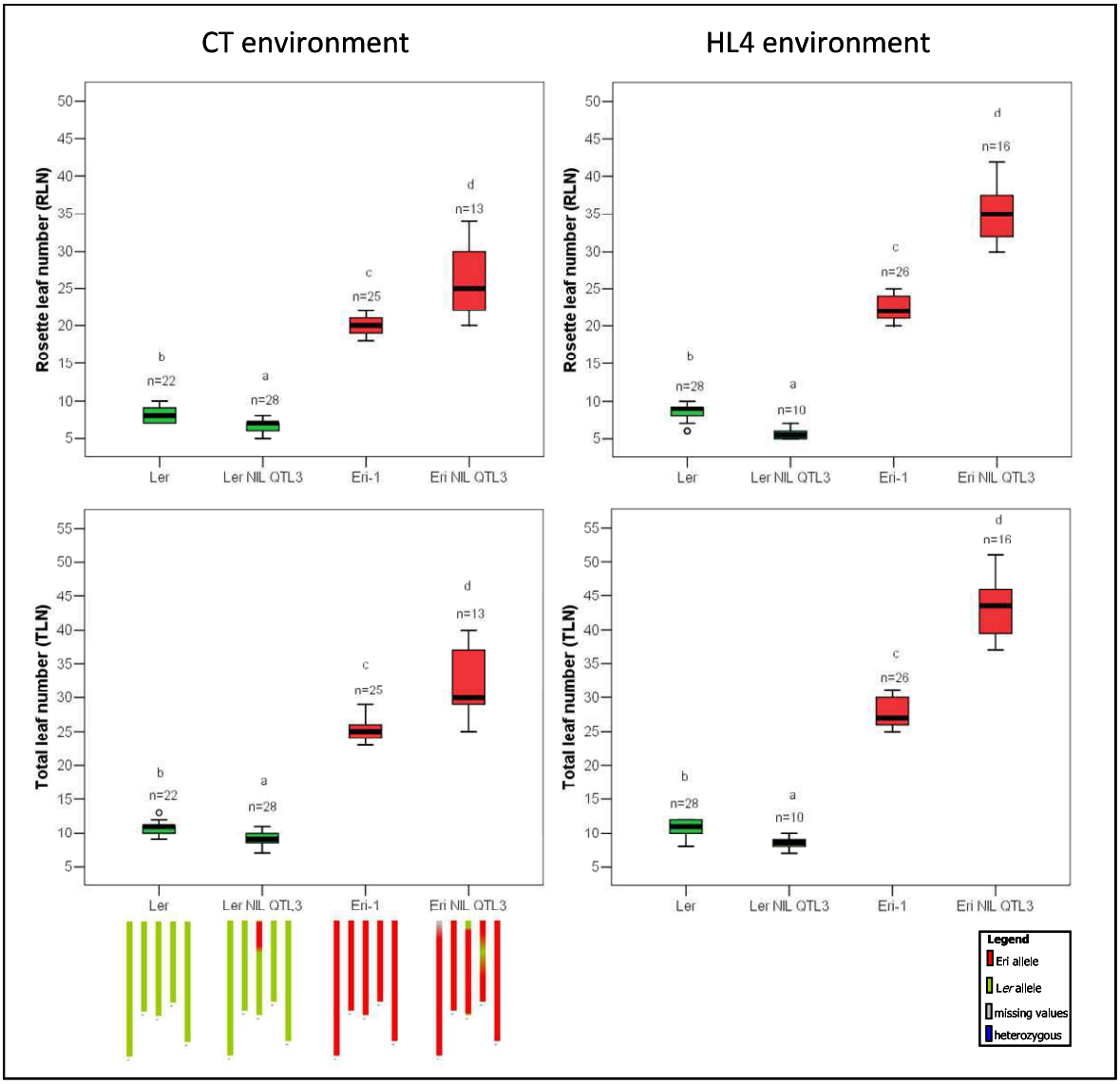


Figure III.7: Rosette leaf numbers and total leaf numbers of L*er*, Eri-1, and the selected NILs L*er* NIL QTL3 (L*er* NIL 97-3) and Eri NIL QTL3 (Eri NIL 67-4-2) in CT and HL4 environment. L*er*, Eri-1 and the two NILs, which were derived from backcrosses between the parental lines and RIL31 and RIL38, were grown in CT condition (left panel) and in HL4 environment (right panel). The boxplots represent the median surrounded by the lower and upper quartile and the lowest and highest trait value. The data was derived from an average of two independent experiments and the n indicates the number of plants that were used. The letters indicate the four different groups (a,b,c,d) identified by the Student-Newman-Keuls test (95% confidence interval). The genotypes of the lines are indicated below with L*er* alleles in green, Eri-1 alleles in red and missing values in grey.

III.4.C. Characterization of the validated flowering QTL3

During the validation of the QTL3 for flowering time, the phenotyping of the NILs showed the independency of QTL3 to the environment the plants were growing in. In both environments we saw a decreasing or increasing of the leaf numbers dependent on the L*er* or Eri NIL QTL3, compared to the parental lines L*er* and Eri-1. Interestingly HL4 condition intensified the effect of QTL3. The L*er* NIL QTL3 (L*er* NIL97-3) showed even less leaf numbers than L*er* grown in the CT condition, whereas the Eri NIL QTL3 (Eri NIL67-4-2) showed a much higher leaf number grown in the HL4 condition. The mapping experiment and the validation was performed in long day (LD) light conditions (16 hours light and 8 hours dark) to prolong the time span in which the plant is exposed to the high light. However, both NILs and the parental lines were additionally grown in short day (SD) light conditions (8 hours light and 16 hours dark) in moderate temperature to see if the prolongation time span of vegetative growth and the therefore deferred time of flowering, will influence the effect of the QTL3. Compared to L*er* the L*er* NIL97-3 showed still an earlier flowering time phenotype, whereas the Eri NIL67-4-2 showed a delayed flowering phenotype to Eri-1 (see Figure III.8; Schmalenbach personal communication). Thus the growth in different day lengths does not affect QTL3.

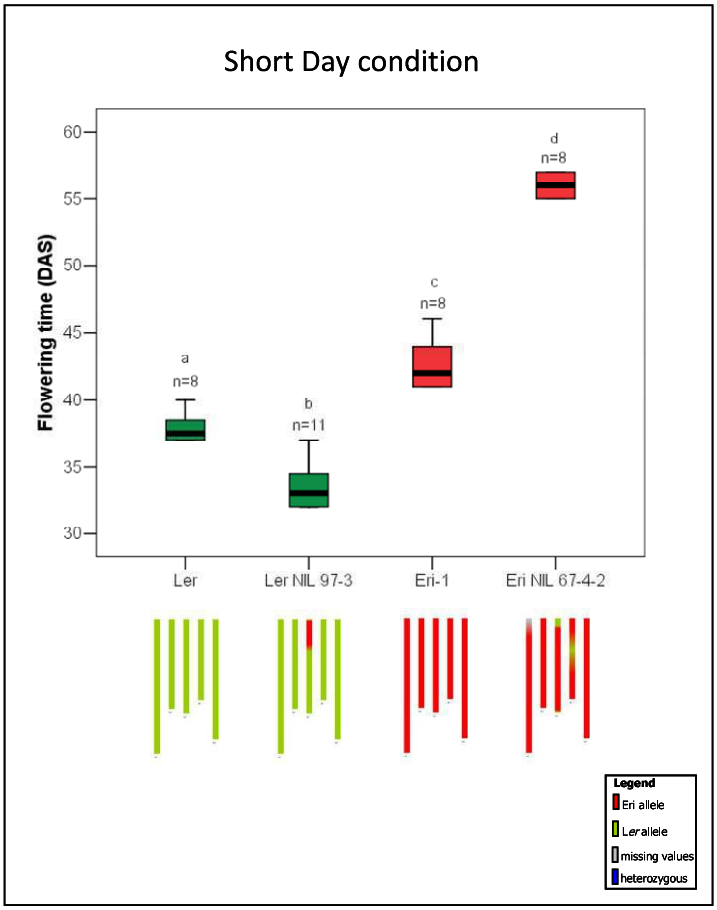


Figure III.8: Flowering time (DAS) of L*er*, Eri-1, and the selected NILs L*er* NIL QTL3 (L*er* NIL97-3) and Eri NIL QTL3 (Eri NIL67-4-2) in short day light condition. L*er*, Eri-1 and the two NILs, which were derived from backcrosses between the parental lines and RIL31 and RIL38, were grown in SD light (8 hours light and 16 hours dark) condition. The boxplots represent the median surrounded by the lower and upper quartile and the lowest and highest trait value. The n indicates the number of plants that were used in the experiment. The letters indicate the four different groups (a,b,c,d) identified by the Student-Newman-Keuls test (95% confidence interval). The genotypes of the lines are indicated below with L*er* alleles in green, Eri-1 alleles in red and missing values in grey.

III.4.D. Fine mapping of the flowering time QTL3

Fine mapping of the QTL3 region was performed to pinpoint the gene(s) which is (are) underlying the effect of this QTL. Validation of QTL3 with the Eri NIL67-4-2 could already show that the QTL3 is located at the very top of chromosome III between 0.014 Mbp (and maybe even above) to 1.75 Mbp. The region contains 708 genes according to the predicted genes in the reference sequence of Col-0. For the fine mapping experiment the progeny of the Eri NIL67-4-10 (isogenic sister line of 67-4-2 with heterozygous introgression on top chromosome III; see Figure III.6) and the parental lines were grown in CT condition to select lines carrying recombinant events within the QTL3 region. The recombinant lines were selected in the fifth generation of the first backcross of the RIL38 and Eri-1 (BC1F5) NIL67-4-10. While the plants were growing, flowering time and leaf numbers were scored for all plants. It is worth mentioning that a significant effect of the experiment (location in the greenhouse, time of the year, growth condition) was a

contributing factor of flowering time variation in the experiments. During the fine mapping experiment the RLN of the Eri NIL67-4-10-26 (L*er* introgression from T4P13 to FTN3-175; see marker Table A.1 in appendix) in an Eri-1 background (similar genotype to NIL67-4-2 in validation) was compared with the RLN of Eri-1. The leaf numbers of Eri-1 were lower than in previously carried out experiments (RLN average 14 leaves) and especially compared to the validation experiment (RLN average 20 leaves). Nevertheless, the tendency of flowering difference between Eri-1 and the Eri NIL QTL3 was similar and comparable to the validation experiment.

All the recombinant lines carrying the Eri-1 allele (homozygous or heterozygous stage) on the top of chromosome III (from marker T4P13 – 0.014 Mbp to F14P3 - 0.42 Mbp, respectively) produced similar numbers of rosette leaves as Eri-1 (line R6 to R10, see Figure III.9). Due to the fact, that the lines with heterozygous alleles showed similar RLN as the lines with homozygous Eri-1 alleles at top chromosome III, it can be concluded that the Eri-1 allele is dominant in promoting flowering (early) and decrease RLN. Recombinant lines with L*er* alleles or heterozygous alleles at the top of chromosome III (lines R1 to R5) presented two different phenotypes. R2 carried fixed L*er* alleles from marker T4P13 to F14P3 and showed a similar increased RLN than the control Eri NIL67-4-10-26, suggesting that QTL3 can be fine mapped between 0.014 Mbp and 0.42 Mbp, which further supports the aforementioned conclusion. However, the lines R1, R3, R4 and R5, which carried recombinant events in this region, showed an intermediate RLN phenotype between Eri-1 and the Eri NIL67-4-10-26. Because of this intermediate phenotype one could raise the hypothesis of the presence of two additive QTL affecting rosette leaf number. Only the combination of L*er* alleles at both QTL3 regions was responsible for the very late flowering phenotype of QTL3. These results are based on single recombinant plants. To conclude the hypothesis additional replicates of the recombinants have to be grown and analysed. To find the location of both QTL we compared the pheno-and genotypes of lines R1 to R5. Line R2 which showed a similar phenotype as the Eri NIL QTL3 had the same genotype as line R3, but line R3 showed an intermediate phenotype between Eri-1 and the NIL. This would suggest that one of the genes responsible for the effect of QTL3 would be located in the region flanked by markers F14P3 (0.42 Mbp) and F7O18 (1.23 Mbp). In the following text, the region between 0.014 Mbp to 0.42 Mbp will be referred as QTL3-1 and the region from 0.42 Mbp to 1.23 Mbp as QTL3-2 (see Figure III.9). Line R3 showed an intermediate phenotype to

Eri-1 and R2, whereas it carried the same allelic values as R2. At QTL3-1 R2 and R3 carried L*er* alleles and thus increased the leaf numbers compared to Eri-1. In the QTL3-2 region both lines carried recombinant events between the two flanking markers. Due to the lack of markers in between F14P3 and F7O18, the positions of the recombinant events in both lines were not ascertained. However, from the data available it appears that in R2 the recombinant event is more closely located to marker F7O18. That would mean that R2 carried L*er* alleles at the QTL3-2 compared to R3 and therefore delaying flowering time. The recombinant lines R4 and R5 with an intermediated phenotype showed similar data as R3 – fixed L*er* alleles at the QTL3-1 and heterozygous or Eri-1 alleles (in case of R5) at the QTL3-2 (see Figure III.9). Line R1 showed also an intermediate phenotype attributed to L*er* alleles at QTL3-2 and heterozygous alleles at the QTL3-1. This R1 line indicates that QTL3-1 is located between 0.014 Mbp and 0.20 Mbp. To summarize the first fine mapping results of QTL3: QTL3 can be split in two QTL (QTL3-1/QTL3-2) with additive effects to each other in which the L*er* allele is recessive and responsible for increased RLN. QTL3-1 is located between the markers T4P13 (0.014 Mbp) and the F4P13 (0.20 Mbp); QTL3-2 is located between the markers F14P3 (0.42 Mbp) and F7O18 (1.23 Mbp). The introgression that caused the increased RLN phenotype could be reduced to a region between 0.014 Mbp to 1.23 Mbp.

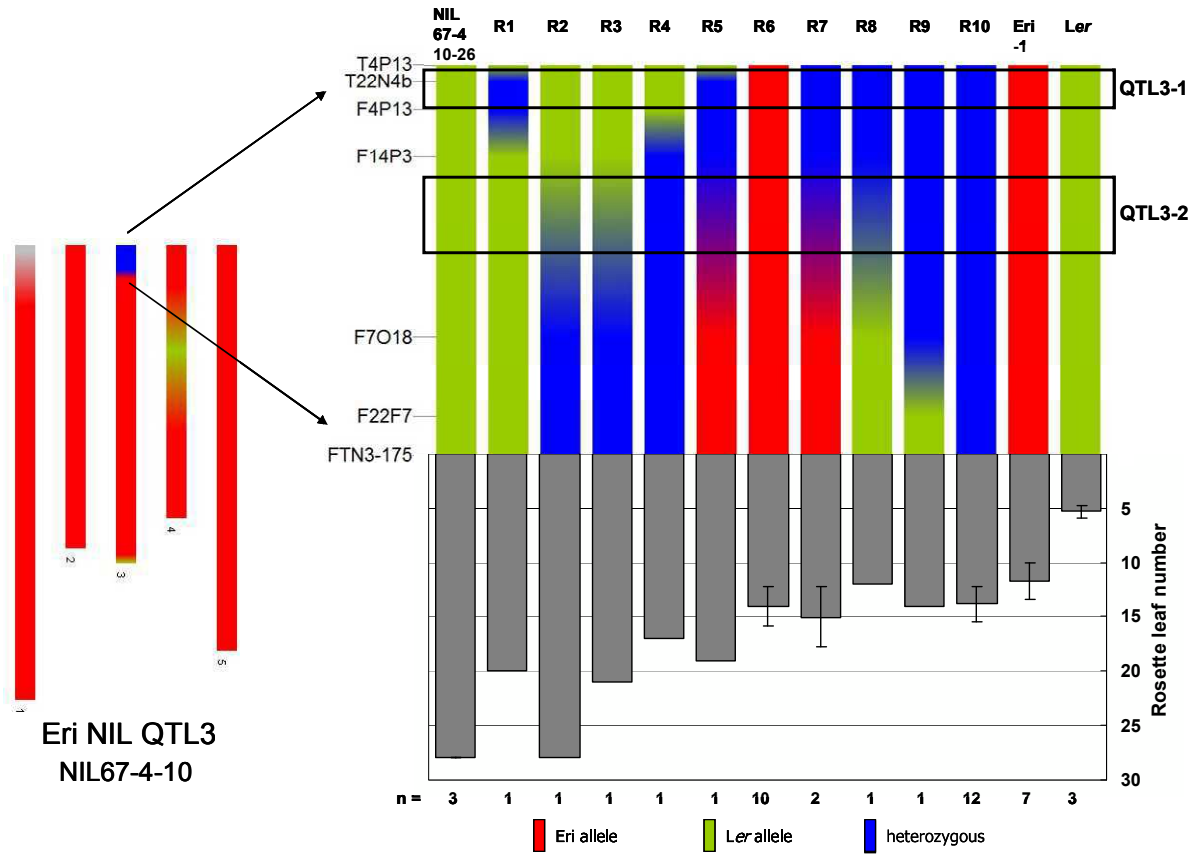


Figure III.9: Fine mapping of QTL3 with progeny (BC1F5) of Eri-NIL QTL3 (Eri NIL 67-4-10)
The rosette leaf numbers of the parents L*er* and Eri-1, Eri NIL67-4-10-26 and the recombinant lines are presented in the bottom panel on the right. The lines differ in their genotype on chromosome III which is schematically presented on the top right panel: green represents L*er*-allele, red Eri-1 and blue heterozygous alleles. The numbers (n) of genotypes and phenotypes analysed are presented below the phenotypic data. The physical distance between the markers T4P13 and FTN3-175 amounts 1.75 Mbp.

III.4.E. Reducing the size of the QTL with additional markers

As the first fine mapping analysis was mainly based on the values of few recombinant plants, the BC1F6 of the cross between RIL38 and Eri-1 – (progeny of Eri NIL 67-4-10-36 – recombinant line R7) were additionally grown in CT environment and analysed. In the first fine mapping experiment we excluded the region between the markers F7O18 (1.23Mbp) and FTN3-175 (1.75Mbp) as being responsible for flowering variation in L*er*/Eri-1 population. The Eri NIL 67-4-10-36 carried fixed Eri-1 alleles from F7O18 to FTN3-175 and a heterozygous introgression at the top of chromosome III until marker F7O18 (R7; see Figure III.9). A co-segregation analysis between rosette leaf numbers and allelic values at different markers indicated that the QTL is located between the markers

T4P13 and F7O18. As in the fine mapping experiment three phenotypic groups were identified. The first group contained early flowering plants with rosette leaves from 8 to 19 leaves (red parable in Figure III.10). The second group contained very late flowering plants with rosette leaf number up to 43 (green parable) and an intermediate phenotype from 19 to 25 leaves (black parable). Interestingly, in this experiment flowering variation was enhanced (with RLN reaching 43) compared to previous CT experiments.

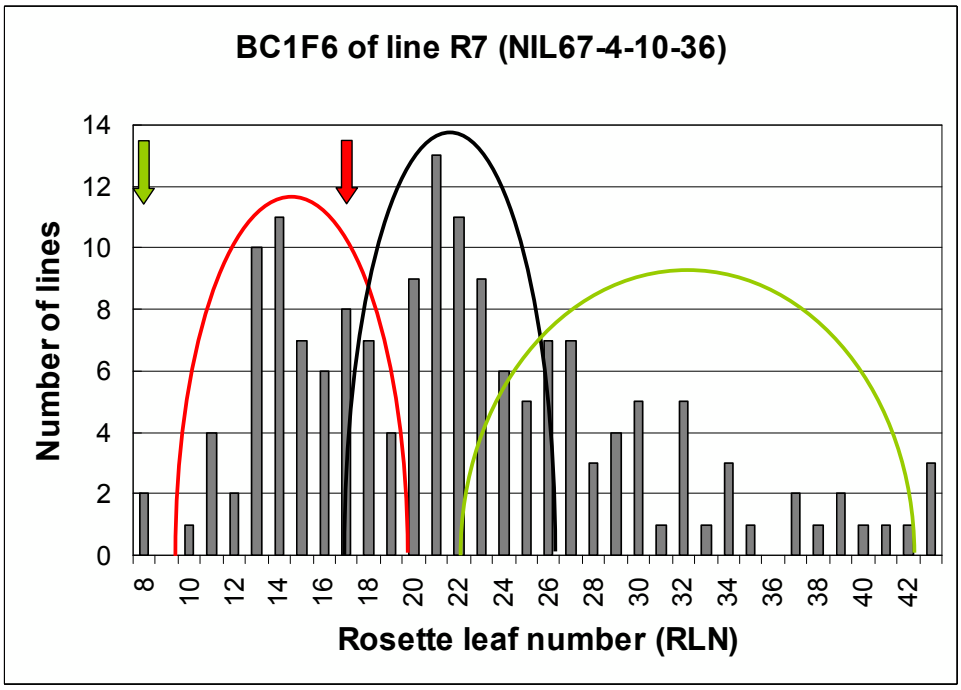


Figure III.10: Frequency distribution for RLN in progeny of the line R7 (NIL67-4-10-36) grown in CT condition. On the x-axes the rosette leaf numbers (RLN) are displayed. On the y-axes the number of lines are displayed. The parabolics indicate three segregating groups in the RLN leaf numbers. In red the group with less leaf numbers (early flowering plants), in green the group with a lot of rosette leaves (very late flowering plants) and in black the intermediate group. In total 170 plants of the BC1F6 were grown. The arrows indicate the RLN of the parental lines (green L*er* and red Eri-1).

To verify the hypothesis that two additive QTL were responsible for the flowering variation at QTL3, the segregating BC1F6 was genotyped with the already used markers and newly developed markers in that region. The associated gene of QTL3-2 was expected between the markers F4P13 (0.42 Mbp) and F7O18 (1.23 Mbp). In that region new markers based on the rare L*er* SNP strategy (explained in material and methods) were developed. SNPs specific for L*er* compared to five other accessions, in that region were used to design CAPS markers. As the Eri-1 genome was not sequenced at the time of the

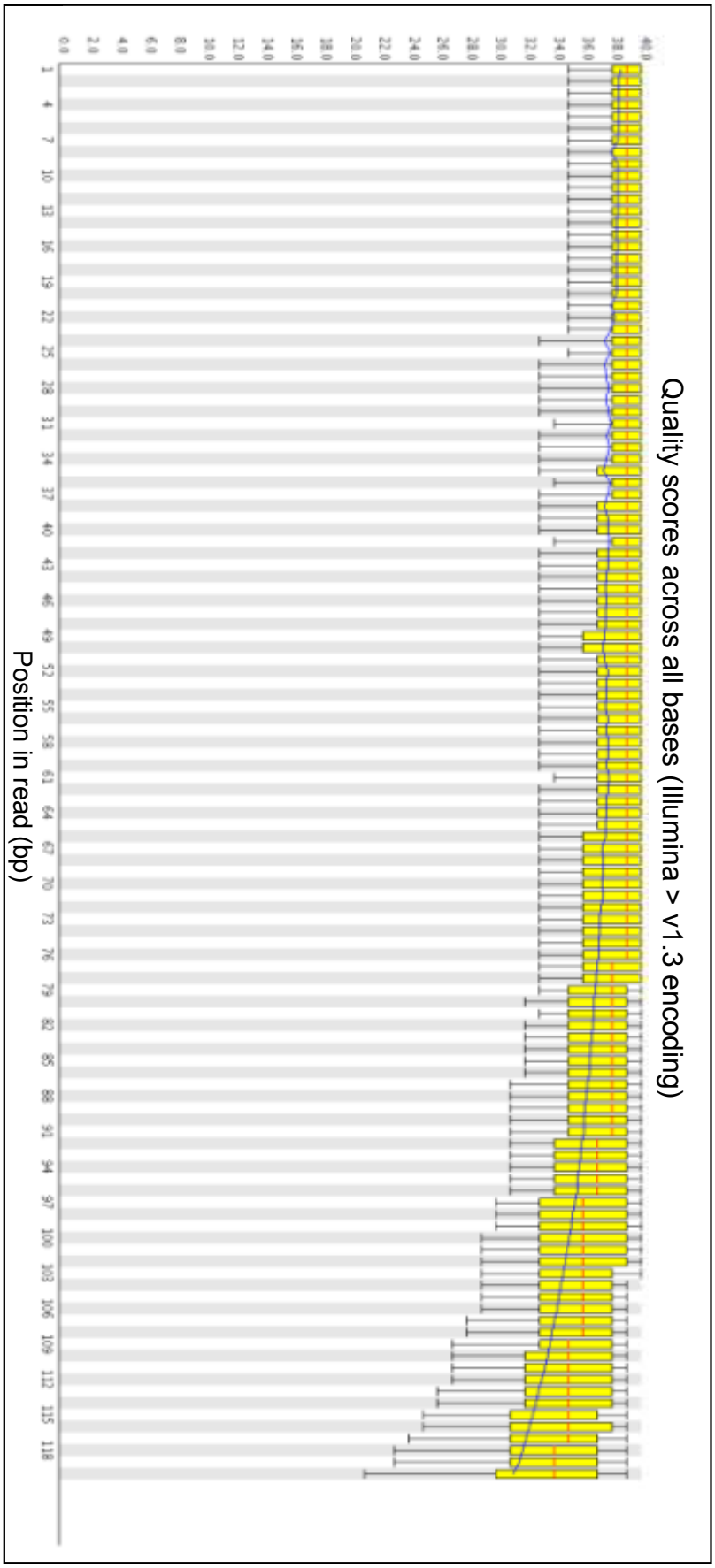
presented study, we hoped that the SNPs used in CAPS markers would be polymorphisms between L*er* and Eri-1. However, of 18 developed CAPS markers only 5 CAPS markers (see in marker Table A.1 in appendix) carried a different SNP in Eri-1 compared to L*er* in that region. With the 5 polymorphic CAPS markers the recombinant region between 0.42Mbp and 1.23Mbp, of QTL3-2 could be narrowed down. The CAPS markers amplified regions located on the edges of the recombinant region and limited the size of QTL3-2 region from 0.45 Mbp to 0.92 Mbp. To identify the SNP which is responsible for the flowering time variation in the L*er*/Eri-1 population and to develop new polymorphic markers, the accession Eri-1 was sequenced with an Illumina platform (see next paragraph and materials and methods).

III.4.F. Sequencing of Eri-1 accession

The sequencing of Eri-1 accession was performed with a Genome Analyser IIx from Illumina in the Max Planck Genome Centre Cologne. The Illumnia sequencing resulted in 32.987.304 single reads (additional information can be found in materials and methods section). After quality control and several optimization steps, the sequence reads could be mapped to the reference genome Col-0 (TAIR9, www.arabidopsis.org) and visualized with the IGV browser (http://www.broadinstitute.org/software/igv). By visualization, structural variation in the genome of Eri-1 compared to Col-0 could be highlighted. The coverage of the mapped sequence reads to the reference varied from ca. 0 - 40 with an average depth of 18. Deletions of DNA regions in Eri-1 in comparison to Col-0 could be identified from the absence of sequence reads of Eri-1 in that particular region. In addition to the Col-0 sequence other sequenced accessions can be compared to Eri-1 in the IGV browser. With this method structural variation between Eri-1, Col-0 and L*er* could be depicted (see paragraph III.4.G). The sequence reads were also analysed to call SNPs between L*er* and Eri-1 which could be used for marker development and further fine mapping of QTL3. The SNP calling pipeline is described in material and method section. The quality control of the sequencing was performed with the software FASTQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/). An overview of the range of quality values across all nucleotide bases at each position in the FastQ file is displayed in Figure III.11. The mean quality scores of the sequences ranged very high from 32 to 38,

whereas the score of 38 was stable until a read length of 80 bp, then the quality decreased to 32 with a read length of 118bp. Overall the quality of the Illumina reads of Eri-1 accession was very good. A quality score of 20 referred to as low quality sequencing data and all the reads below a quality score of 20 would be usually discarded from the analysis. Thus, in this sequencing experiment no read data were excluded from the analysis.

Figure III.11 <on the following page>: Per base sequence quality of sequencing Eri-1 illumina reads. The x-axis represents the read position; the y-axis represents the quality scores. The central red line is the median value and the yellow box represents the inter quartile range (25-75%). The upper and lower whiskers represent the 10% and 90% points. The blue line represents the mean quality.


Quality scores across all bases (Illumina > v1.3 encoding)
Position in read (bp)

After quality control the sequence reads were mapped to the reference genome Col-0 (TAIR9). Additional filtering checked the mapping quality and removed duplicates. In the end 397.836 SNPs could be called between Eri-1 and Col-0. The Figure III.12 displays sequence reads of Eri-1 in the region 485.494 bp to 489.804 bp on the top of chromosome III visualized by the IGV browser. This region is also included in the QTL3 region. Displayed by the gray arrow bars (single sequence reads) a sequence depth of around 20-40 could be obtained in that region. SNPs differences between Eri-1 and Col-0 which are detected in all reads in that region are indicated by coloured lines (red = T, green = A, blue = C and orange = G). No sequence gaps of Eri-1 sequences were presented in that region, therefore no deletion of genomic parts of Eri-1 was detected in that region.

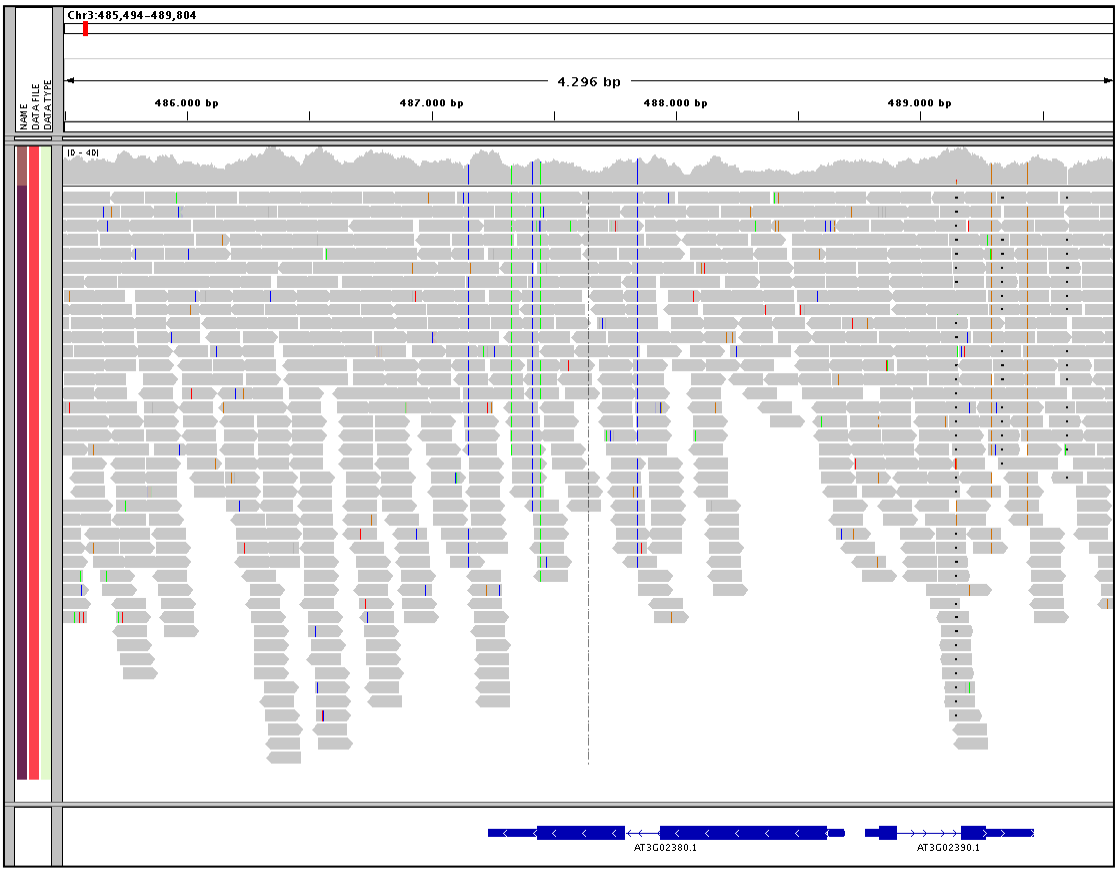


Figure III.12: IGV browser image of Illumina sequence reads of Eri-1 aligned to the region 485.494 bp – 489.804 bp of chromosome III of the reference genome Col-0 based on TAIR9 information. Reads are displayed in grey bars. The gene model AT3G02380 and AT3G02390 are located in the region, displayed below the reads in blue. Different SNPs compared between Eri-1 and Col-0 are indicated with coloured lines in the grey bars (red = T, green = A, blue = C and orange = G).

Figure III.13 displays a zoom in of the Figure III.12. The region 487.403 bp to 487.469 bp on chromosome III is presented. With zooming in the region, SNPs in the Eri-1 sequence can be compared in detail with the sequence of Col-0 and respective gene models.

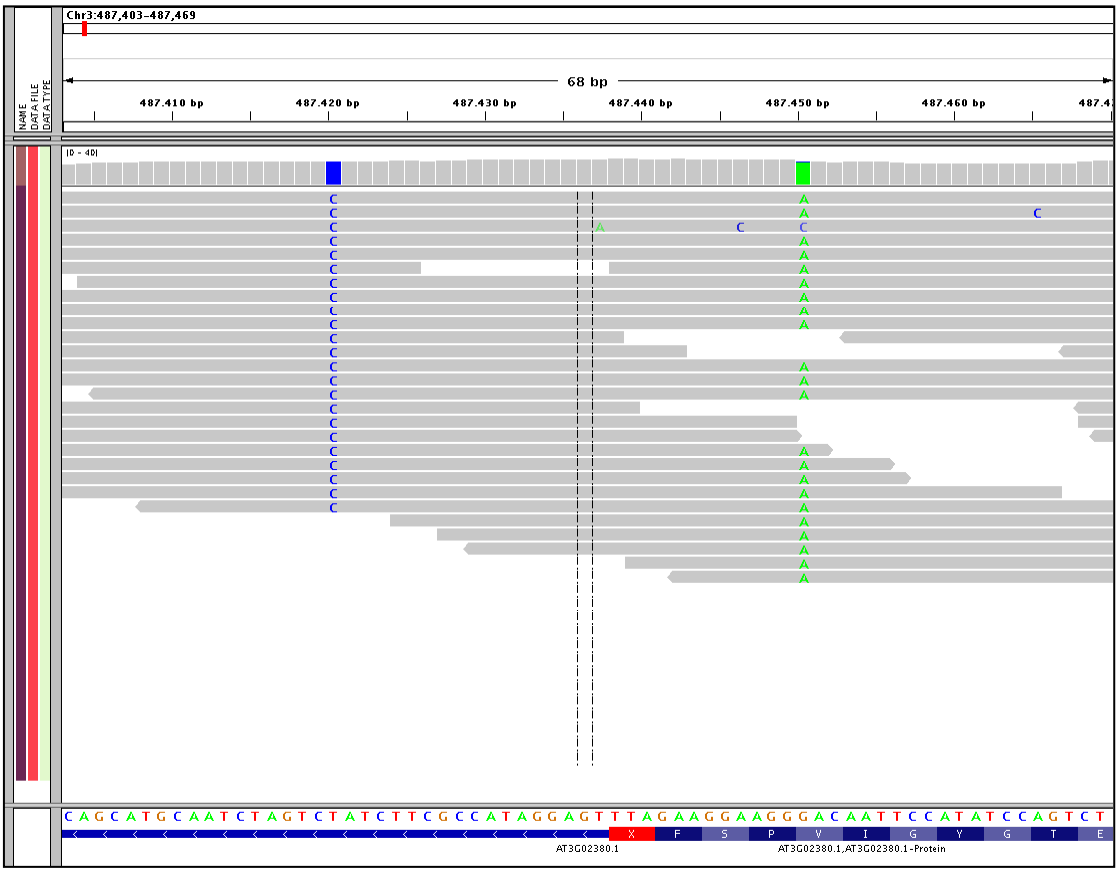


Figure III.13: Zoom in of Figure III.12. IGV browser image of Illumina sequence reads of Eri-1 aligned to the region 487.403 bp – 487.469 bp of chromosome III of the reference genome Col-0 based on TAIR9 information. The nucleotide sequence of the gene model of AT3G02380 in Col-0 is displayed at the bottom of the screenshot. In case of nucleotide changes in the Eri-1 sequence compared to Col-0 reference, the different nucleotide is displayed (here a C in Eri-1 reads and a T in the Col-0 reference, as well as an A in Eri-1 and a G in Col-0.

To produce polymorphic markers between L*er* and Eri-1 we identified SNP between L*er* and Eri-1. To compare all three accessions (Col-0, Eri-1 and L*er*) the public available SNP data of L*er* (www.1001genomes.com) were added to the SNP call analysis. The SNPs which were similar between L*er* and Eri-1, but different between Col-0 (163.042 SNPs) were removed from the analysis because they were not useful for later marker design. 184.924 different SNPs were detected between L*er* and Eri-1, whereas L*er* carried the same SNPs than Col-0. On the other hand 212.912 different SNPs were detected between L*er* and Eri-1, whereas Eri-1 carried the same SNP than Col-0. The high SNP number plus

indels which could be detected between the three accessions highlighted differences which can be found in natural occurring accessions. The high number of SNPs between L*er* and Eri-1 is promising for designing marker for fine mapping of detected QTL. Additionally, direct comparison of the structure of genes in L*er* and Eri-1 can be easily performed with the IGV browser in order to find polymorphisms in candidate genes

III.4.G. Candidate genes for flowering time – genes in the region of QTL3-1 and QTL3-2

After the fine mapping results the region of QTL3 was narrowed down to 900Kb. This region contained 373 genes, but only 4 of them were obvious flowering time genes. The genes are *COL2* (Ledger et al. 2001*)*, *MBD9* (Peng et al. 2006*)*, *VGT1* (Aluri and Büttner 2007*)* and *AGL4* (Pelaz et al. 2000*)*. *COL2* is homologous to the flowering time gene *CO* encoding zinc finger proteins (Putterill et al. 1995). *CO* is a main player of the photoperiod pathway and involved in the regulation of flowering under long days. The *CO* mutant phenotype showed a delayed flowering in long days (Putterill et al. 1995). *COL2* is located at 0.42 Mbp on chromosome III and included in the QTL3-2 region. The sequence region of $COL2^{Ler}$ and $COL2^{Eri\text{-}1}$ was compared with $COL2^{Col\text{-}0}$ and between each other in the IGV browser. No nucleotide differences or indel between L*er* and Eri-1 alleles for *COL2* gene and putative promoter region were detected. This result suggested that *COL2* can be excluded for being the gene responsible for the effect of the QTL3-2.

Using the same procedure we checked for differences between L*er* and Eri-1 in the other three candidates. First a comparison only of the SNP data for *MBD9*, located at 0.17 Mbp in the region of QTL3-1 resulted in only one SNP difference between L*er* and Eri-1. The SNP was located in the non-coding region (first intron) of the gene. A second comparison of the *MBD9* sequence of L*er*, Eri-1 and Col-0 in the IGV browser revealed structural differences in the first exon of Eri-1. Some parts of the exon did not show any or only 1-2 sequence read alignments of Eri-1 sequences compared to L*er* or Col-0. This absence of reads suggested that parts of the first exon in Eri-1 is deleted and led to protein changes and function. MBD9 is part of the vernalization pathway and known to enhance the expression of the flowering repressor *FLC* (Peng et al. 2006).

Comparison of the *VGT1* sequence between the three accessions did not give any nucleotide difference between L*er* and Eri-1, but several SNPs between Col-0 and both accessions. The gene is located at 0.72 Mbp on chromosome III and therefore in the region of QTL3-2. However a deletion in the third intron and fourth exon was detected in both accessions L*er* and Eri-1, indicating that differences in flowering time between them were not caused by *VGT1*.

The last candidate in the region is *AGL4* located at 0.46 Mbp in the vicinity of QTL3-2. Although only three differences in the non-coding region of this gene between L*er* and Eri-1 were detected, it appeared to be a good candidate for QTL3-2, displaying large deletions in L*er* in the putative promoter region. This deletion could lead to changes in the expression of $AGL4^{Ler}$. To confirm the presence of the deletions in $MBD9^{Eri\text{-}1}$ exon (candidate gene of QTL3-1) and $AGL4^{Ler}$ promoter (candidate gene of QTL3-2), the regions including the deletions need to be checked by PCR, as the absence of sequence reads could also be due to low quality sequencing. Additionally the expression of these genes needs to be tested by qRT-PCR.

III.5. Validation and characterization of QTL5

III.5.A. Selection of NILs in L*er* or Eri-1 genetic backgrounds to validate the presence and effect of 2 linked QTL on chromosome V

The QTL mapping resulted in the detection of two closely linked QTL located on chromosome V named QTL5-1 and QTL5-2 (see result above III.3.A.). A way to confirm the presence and the effects of these linked QTL would be to observe significant phenotypic differences between the parental accessions and NILs carrying introgression of alleles from one parental accession at QTL5-1 and/or QTL5-2 in an otherwise genetic background from the other parental accession. Such NIL have been selected from successive backcrosses between selected RILs and either L*er* or Eri-1.

III.5.AA. Selection of NILs in Eri-1 background

The RIL 103 (Figure III.14 upper panel) carried a L*er* introgression at both QTL5 regions in an almost clean Eri-1 genetic background and thus was appropriate for the development of a so called "Eri NIL QTL5" after backcrossing this RIL with Eri-1. In addition to the QTL5-1 and QTL5-2 loci (from 5.6 Mbp to 15 Mbp on chromosome V), the BC1F1 (Figure III.14 upper panel) carried additional introgression of Eri-1 alleles (heterozygous state) elsewhere in the genome: two small on top chromosome III (0.085 Mbp and 5.6 Mbp) and bigger regions on the top (0.0 Mbp to 7.0 Mbp) and the bottom (16.4 Mbp to 17.6 Mbp) of chromosome IV. In the BC1F2 progeny, a clean NIL which carried only the small heterozygous regions at 0.085 Mbp and 5.6 Mbp on chromosome III and a L*er* homozygous introgression at chromosome V including both QTL5 regions in an otherwise Eri-1 allelic background was selected (Figure III.14 upper panel). This line was called Eri NIL QTL5 (Eri NIL53).

In order to analyse QTL5-1 and QTL5-2 separately and to narrow down the QTL regions, the Eri NIL53-6 (isogenic to Eri NIL53, but homozygous Eri-1 on region 0.085 Mbp and still heterozygous on region 5.6Mbp on chromosome III – Figure III.14 upper panel) was used for a second backcross with the parental line Eri-1. The resulting and selected BC2F1 (Eri NIL53-6xE 3L, Figure III.14 upper panel) carried exclusively heterozygous introgression at the region including QTL5-1 and QTL5-2 in a clean Eri-1 background. In

the BC2F2, recombinant events allowed us to select NILs with smaller heterozygous introgressions at only QTL5-1 or QTL5-2 (Figure III.14 lower panel), enabling us to validate the effects of each QTL separately. The Eri NIL QTL5-1 53-6xE 3L-8 (abbreviated Eri NIL53-8, heterozygous introgression from marker NGA106 (5.4 Mbp) to marker NGA139 (8.43 Mbp)) was selected to validate the effects of QTL5-1 and to narrow down the region for fine mapping. The Eri NIL QTL5-2 53-6xE 3L-60 (abbreviated Eri NIL53-60, heterozygous introgressions between NGA139 (8.43 Mbp) and SO191 (15.0 Mbp) carried heterozygous alleles only at the QTL5-2 region and was selected for validation and further fine mapping of this QTL. Self pollination of these two lines (BC2F3) allowed us to generate similar introgressions of L*er* alleles at homozygous state in Eri-1 genetic background (Eri NIL QTL5-1 53-6xE 3L 8-118 abbreviated Eri NIL8-118 for QTL5-1 and the Eri NIL QTL5-2 53-6xE 3L 60-22 abbreviated Eri NIL60-22 for QTL5-2; Figure III.14 lower panel).

While selecting these NILs in the BC2F2, another line carrying solely an introgression of Eri-1 alleles at both QTL5 was selected and named Eri NIL QTL5 53-6xE 3L-7 (abbreviated Eri NIL53-7, Figure III.14, upper panel). In addition, Eri NIL QTL5-1 53-6xE 3L 8-96-27 (abbreviated Eri NIL8-96-27) with a small L*er* introgression in the QTL5-1 region from marker ICE5 (7.71 Mbp) to marker CDPK9 (7.95 Mbp) was also selected (Figure III.14 lower panel).

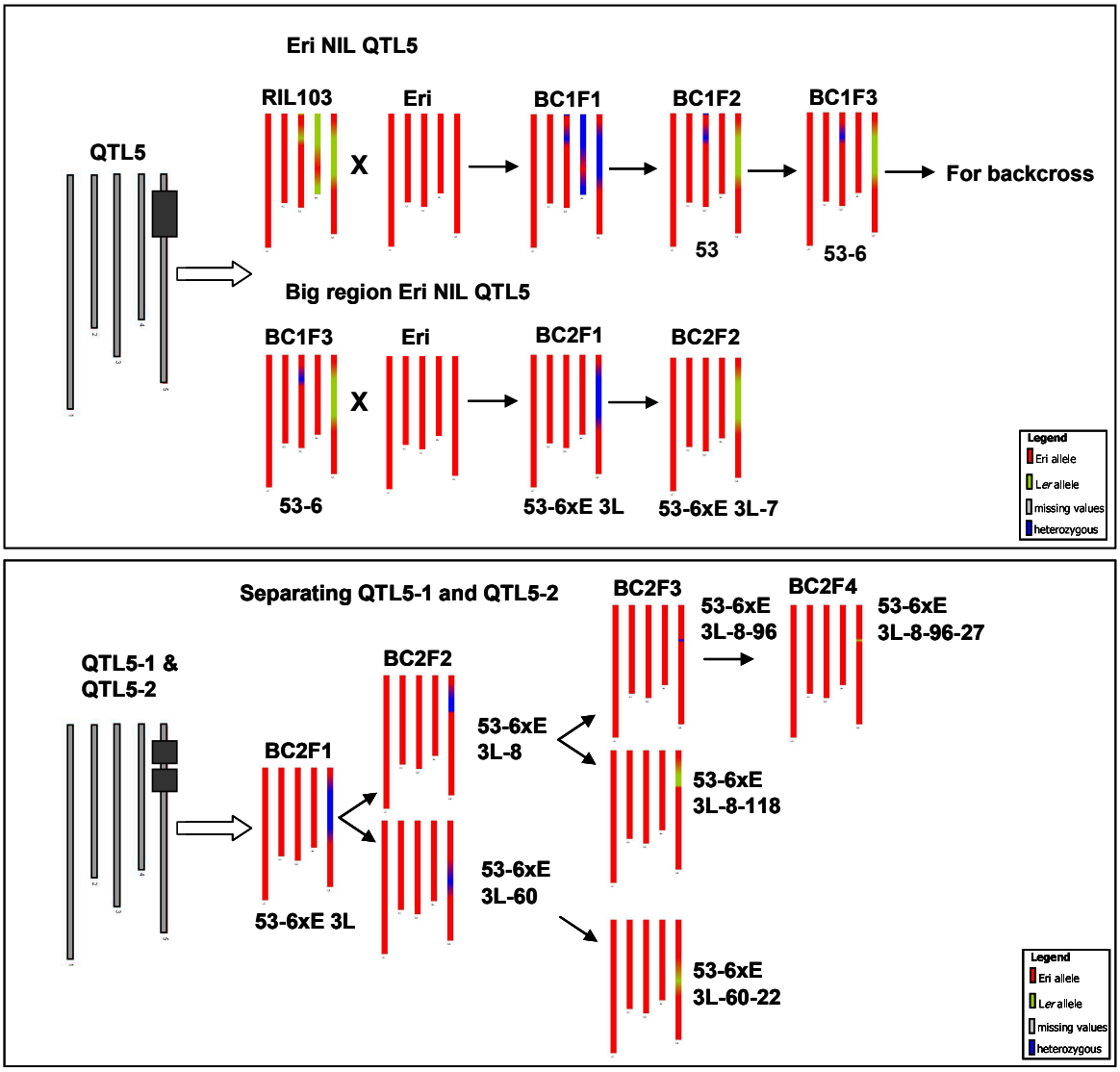


Figure III.14: Selection strategy and genotype of Eri-1 NIL QTL5, QTL5-1 and QTL5-2. The five bars indicate the five chromosomes of *A. thaliana* in both panels, in the upper panel the dark grey box highlights the position of the QTL5 and in the lower panel the grey boxes indicates the separation of QTL5-1 and QTL5-2. The colours represent the allele (red-Eri-1, green –L*er*, blue heterozygous, grey-missing values). For developing the Eri NIL QTL5 (L*er* introgression at QTL5 in Eri-1 background) the RIL103 was backcrossed to the parental line Eri-1. In selection steps a clean Eri NIL QTL5 (Eri NIL53-7) was developed. In separation steps clean NILs for both QTL5-1 and QTL5-2 were created (Eri NIL8-118 and Eri NIL60-22) and a NIL with a very small introgression of L*er* in Eri-1 background was developed (Eri NIL8-96-27). The names of the NILs are printed below or beside the respective genotypes.

III.5.AB. Selection of NILs in L*er* background

For developing a L*er* NIL QTL5 (a line with Eri-1 introgressions at the QTL5 in a L*er* allelic background) two RILs were crossed (RIL67 and RIL10) and an F2 plant 67x10-1 was then backcrossed to the parental line L*er* (see Figure III.15). The selected BC1F1 (L*er* NIL 67x10-1xL 10) resulted in a line with several heterozygous regions in a L*er* background. After two generations of self-pollination, four almost clean L*er* NILs QTL5 were selected (see Figure III.15). The L*er* NIL QTL5 41(3) carried an Eri-1 introgression on chromosome V from marker NGA106 (5.4Mbp) until marker MOK9 (13.67Mbp) (both markers fixed Eri-1) in an otherwise L*er* background with additional heterozygous regions on top chromosome I and top chromosome II. The L*er* NIL QTL5 37(1) carried the same Eri-1 introgression on chromosome V and an additional Eri-1 introgression on top chromosome I in a L*er* background. Both NILs carried introgression of Eri-1 alleles in a region including the QTL5-1 and the QTL5-2 (see Figure III.15).

In addition, two L*er* NILs QTL5 with smaller Eri-1 introgression were selected in the BC1F3. The L*er* NILs QTL5-1 84(1) and L*er* NIL QTL5-1 18(2) both carried the Eri-1 alleles between the markers NGA106 (5.4Mbp) and SO262 (9.8Mbp). In these two NILs, only the QTL5-1 carried the alleles from Eri-1 whereas the QTL5-2 was fixed with L*er* alleles. The L*er* NIL84(1) carried an additional Eri-1 introgression at top chromosome I in the L*er* background and a heterozygous region on chromosome III at marker NT204. The NIL18(2) was the cleanest L*er* NIL QTL5-1 with only Eri-1 introgression at QTL5-1 (see Figure III.15).

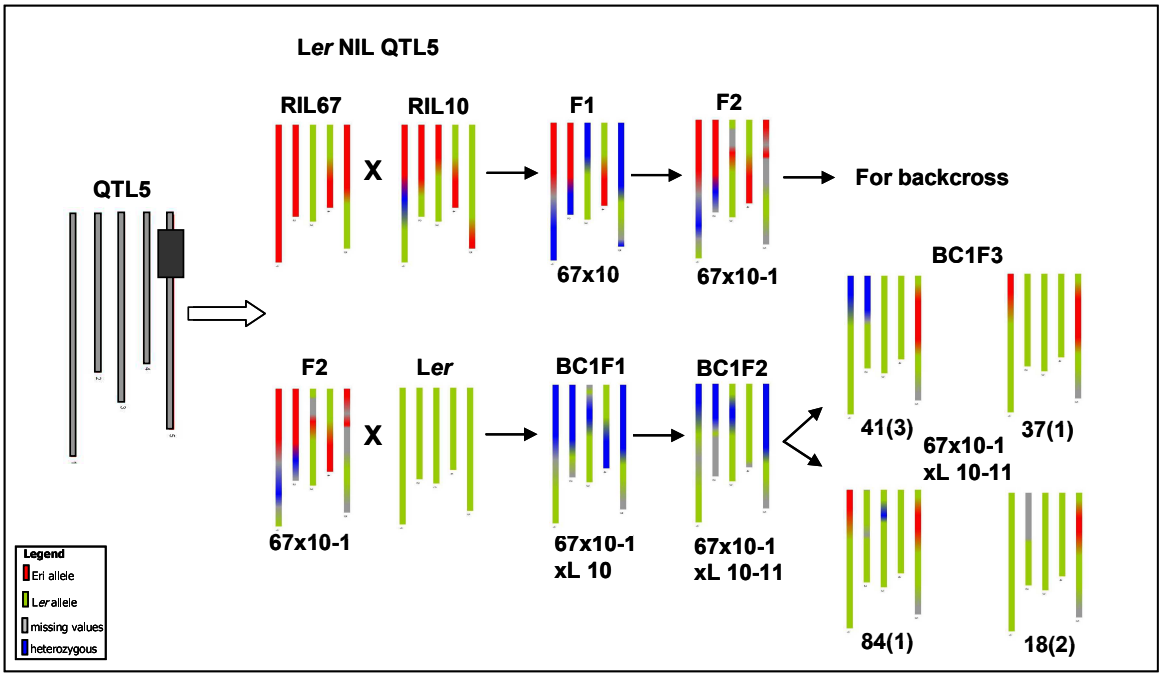


Figure III.15. Selection strategy and genotype of L*er* NIL QTL5 and QTL5-1. The five bars indicate the five chromosomes of *A. thaliana*, in the dark grey box highlights the position of the QTL5. The colours represent the allele (red-Eri-1, green –L*er*, blue heterozygous, grey-missing values). For developing the L*er* NIL QTL5 (Eri-1 introgression at QTL5 in L*er* background) the RIL67 and the RIL10 were crossed with each other and the second generation of the cross was backcrossed to the parent L*er*. In several selection steps two almost clean L*er* NIL QTL5 (NIL41(3), NIL37(1)) and two L*er* NIL QTL5-1 (NIL84(1) and NIL18(2)) were developed. The names of the NILs are printed below the respective genotypes.

III.6.A. Validation of the flowering and pigmentation effects of QTL5

QTL detection with the L*er*/Eri-1 RIL population revealed QTL for flowering related traits in both environments on chromosome V. Analysis of variance indicated that in CT condition the lines with Eri-1 alleles at QTL5 (big region including both linked QTL5-1 and QTL5-2) showed significant delay in FT on average about 3.5 days and increased the leaf numbers RLN and TLN on average for about 3 leaves compared to the lines with L*er* alleles at this locus. In the HL4 condition, the flowering time was delayed compared to the CT condition for all RILs and the flowering differences associated with the QTL5 were also increased. Compared to the L*er* alleles, the Eri-1 alleles at QTL5 delayed the flowering time on average 6 days and the RLN and TLN increased on average 7 and 9 leaves, respectively.

In order to validate the effect of closely linked QTL for flowering time related traits and anthocyanin content at QTL5, the progeny (8 plants) of Eri NIL53 was grown in the HL4 condition. These plants were grown in HL4 condition and their leaf numbers as well as their anthocyanin content were quantified. As expected (L*er* alleles present at QTL5), all these plants showed a decrease in RLN compared to Eri-1 (Figure III.16). In the genetic context of Eri-1 background the effect of L*er* alleles at this locus was even higher than expected from the QTL detection. L*er* alleles at the QTL5 region decreased the RLN on average around 13 leaves compared to Eri-1. Similarly to L*er* which has 8 RLN in this condition, the NIL53 flowered after producing on average 9.5 rosette leaves, whereas Eri-1 produced on average 23 rosette leaves before flowering (Figure III.16).

While the flowering time QTL5 was detected in both environments, the anthocyanin content QTL at the same position (nearest marker CIW8 7.5 Mbp) was only detected in the HL4 environment. RILs carrying the Eri-1 allele at this locus increased significantly their anthocyanin content (49 µg/ml on average) compared to those carrying the L*er* allele (17 µg/ml).

Anthocyanin content of Eri NIL53 grown in HL4 condition (here in 4 replicates) was also quantified to verify the effect of QTL5 (Figure III.16). Compared to Eri-1 which accumulated on average 120µg/ml anthocyanin, the Eri NIL53 with L*er* introgressions on QTL5 decreased this content to 60µg/ml on average. This observed effect in the NIL was again higher than the one expected from the QTL detection (17 and 49 µg/ml). The QTL detection as well as the 2-way marker interaction analysis revealed that more loci are

involved in the variation of anthocyanin content under HL4 condition than only QTL5. For all other loci, only Eri-1 alleles increased the value of this trait. The second anthocyanin QTL on chromosome I at marker position NF19K23 (22.9 Mbp) detected in the HL4 environment is an example for this. The presence of other loci increasing the anthocyanin content in the Eri NIL53 could explain the higher values in this line. It would explain the significantly higher anthocyanin accumulation of the Eri NIL53 with a clean Eri-1 background compared to the parental line L*er*.

In summary, the flowering time QTL5 (including QTL5-1 and QTL5-2) as well as the anthocyanin QTL5 could be validated by phenotyping the Eri NIL53 in the HL4 condition. As a next step to analyse the effect of QTL5 without additional loci affecting the anthocyanin variation in the Eri-1 background, L*er* NILs QTL5 with a clean L*er* background were selected (see Figure III.15) and analysed together with the selected smaller region NILs (Ler NILs QTL5-1 and Eri NILs QTL5-1 in another validation experiment (see paragraph III.6.C).

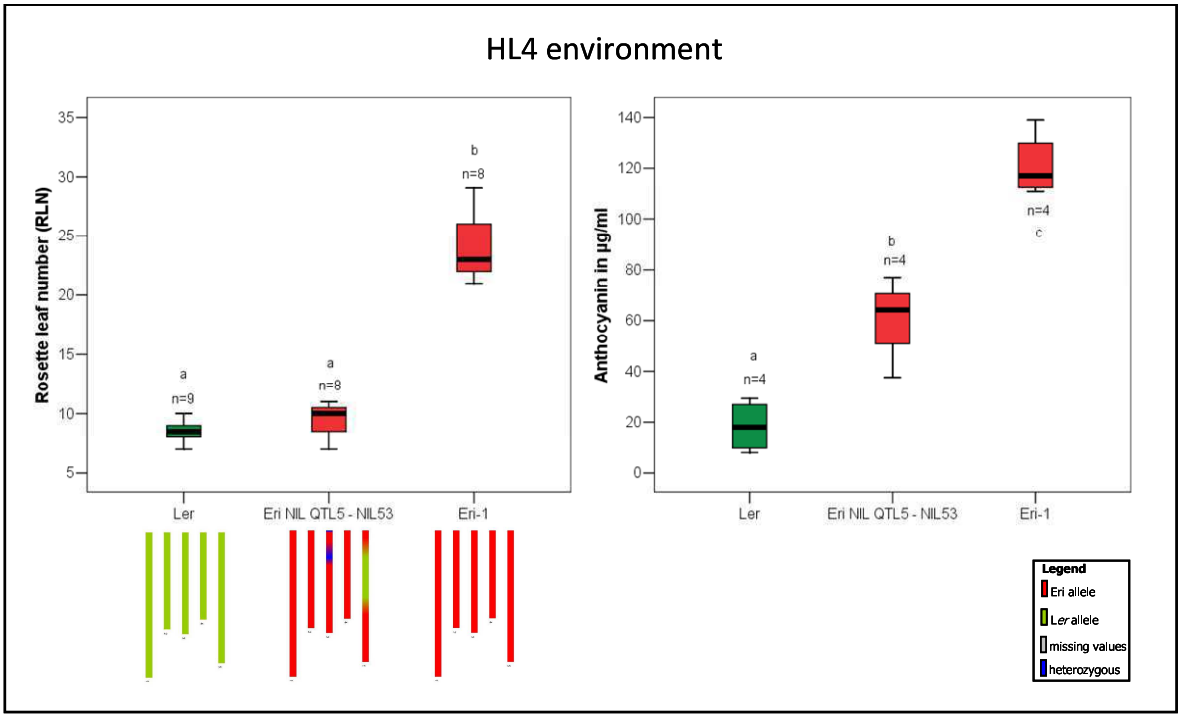


Figure III.16: Rosette leaf numbers and anthocyanin content of L*er*, Eri-1, and the selected Eri NIL QTL5 (NIL53) in HL4 environment. L*er*, Eri-1 and the NIL, which was derived from a backcross between the parental line Eri-1 and RIL103, were grown in high light/cold environment. The boxplots represent the median surrounded by the lower and upper quartile and the lowest and highest trait value. The n indicates the number of plants that were used for the test. The letters indicate the different groups (a and b) for rosette leaf numbers) and (a,b and c) for anthocyanin content, identified by the Student-Newman-Keuls test (95% confidence interval). The genotypes of the lines are indicated below with L*er* alleles in green, Eri-1 alleles in red and heterozygous alleles in blue.

III.6.B. Characterization of validated flowering QTL5

QTL5 for flowering related traits was detected in both environments tested. In the HL4 environment QTL5 has been validated using Eri NILs. To validate QTL5 also in the CT condition and to compare the effect of QTL5 in both environments, L*er* NILs QTL5 and Eri NILs QTL5 were analysed. The L*er* NILs QTL5, NIL41(3) and NIL37(1) were grown in both conditions. Eri NIL QTL5 NIL53-7 was also analysed together with both parents in both environments. The Eri NIL53-7 showed the same phenotypic responses for flowering related traits which were previously recovered in the QTL5 validation with Eri NIL53. Comparison of the L*er* NILs with L*er* showed significant differences in leaf numbers in the HL4 environment (Figure III.17, right panel). The Eri-1 alleles at the QTL5 (including both QTL5-1 and QTL5-2) in L*er* NIL41(3) and L*er* NIL37(1) showed significant

increases of around 5 RLN in the HL4 environment. Thus the flowering time QTL5 could be also validated in the L*er* genetic background in HL4 condition. In the CT environment the differences in leaf numbers between the L*er* NILs and the parental line L*er* were not significant (see Figure III.17 left panel). In accordance with results obtained in HL4 condition, the L*er* introgression in the Eri NIL53-7 was responsible for a decrease in leaf number (on average RLN=9.4 and TLN=12.7) compared to Eri-1 (RLN=19.6 and TLN=24.6) in the CT environment. Therefore the QTL5 for flowering time related traits in the CT environment could only be validated in the Eri NIL QTL5. In the HL4 environment for each line the leaf numbers were increased compared to CT environment. Eri-1 delayed flowering on average 2 leaves and L*er* on average 1 leaf. The leaf numbers of the L*er* NILs increased a bit more, between 2 and 3.8 leaves in the HL4 environment. Consequently these NILs grouped together in a different class than L*er* for RLN and TLN. All NILs used in this experiment carried introgressions which encompassed QTL5-1 and QTL5-2.

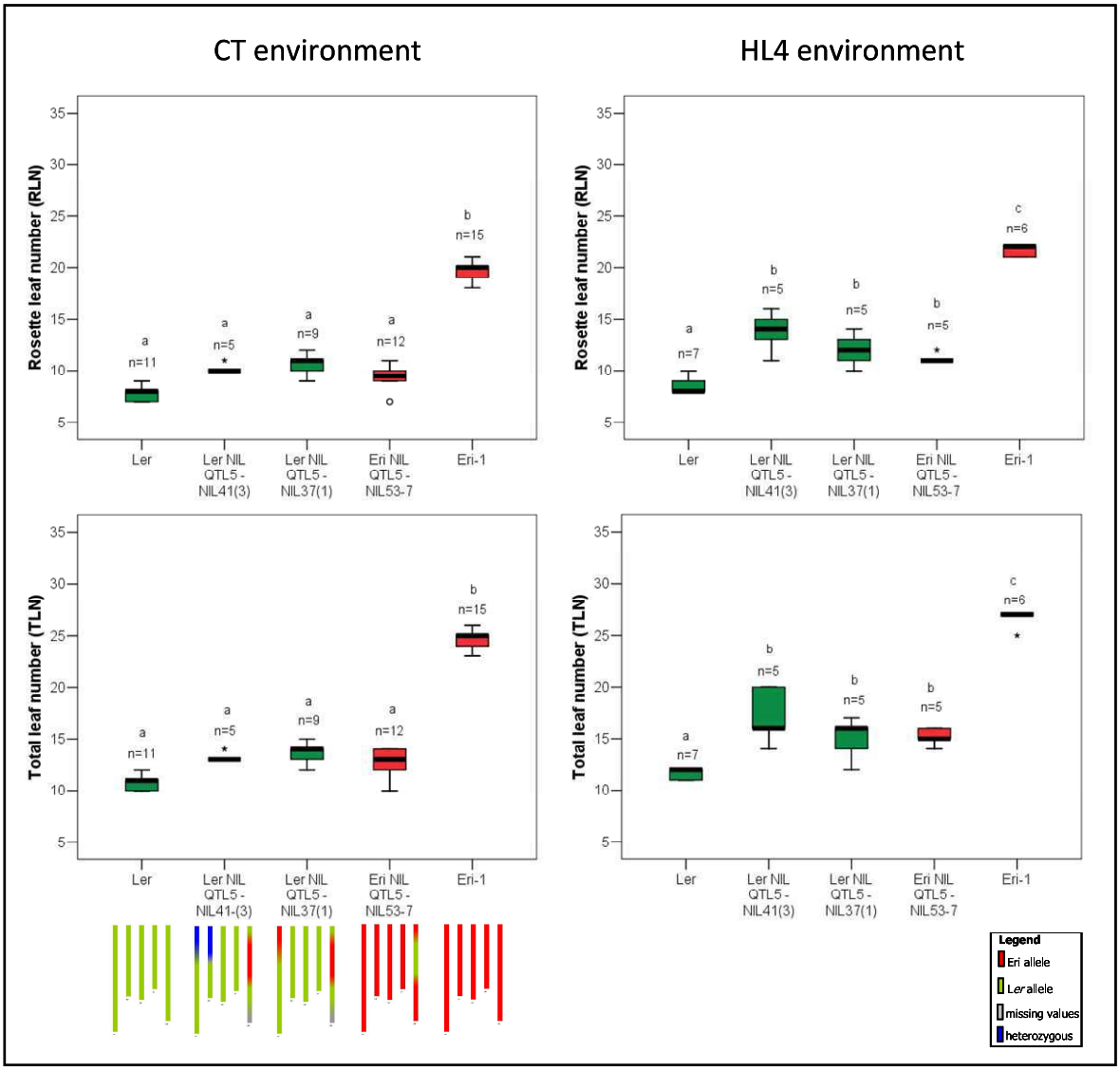


Figure III.17: Comparison of rosette leaf numbers and total leaf numbers of L*er*, Eri-1, and the selected NILs L*er* NILs QTL5 (NIL41(3) and NIL37(1)) and Eri NIL QTL5 (NIL53-7) in CT and HL4 environment. L*er*, Eri-1 and the three NILs, that were derived from a cross between RIL67 and RIL10, followed by a backcross with the parental lines L*er* and a backcross between RIL103 and the parental line Eri-1, were grown in CT condition (left panel) and in high light/cold environment (right panel). The boxplots represent the median surrounded by the lower and upper quartile and the lowest and highest trait value. The n indicates the number of plants that were used. The letters indicate the three different groups (a,b,c) identified by the Student-Newman-Keuls test (95% confidence interval). The genotypes of the lines are indicated below with L*er* alleles in green, Eri-1 alleles in red, heterozygous alleles in blue and missing values in grey.

III.6.C. Validation of separated flowering and anthocyanin QTL5-1 and flowering QTL5-2

An additional validation experiment in order to confirm QTL5-1 and QTL5-2 separately was performed (see Figure III.18). Selected L*er* NIL QTL5 (NIL37(1), L*er* NIL QTL5-1 (NIL18(2)), as well as Eri NIL QTL5 (NIL53-7), Eri NIL QTL5-1 (NIL8-118) and Eri NIL QTL5-2 (NIL60-22) were grown in 10-11 replicates in the HL4 environment and phenotyped for flowering related traits and anthocyanin content.

III.6.CA. Validation of flowering related trait QTL5-2 in HL4 condition

Analysing the RILs for flowering related traits resulted in a QTL referred as QTL5-2 (see III.3.A. above), mapped below QTL5-1. The associated marker to QTL5-2 was located at 20.9 cM (SO262 / 9.8 Mbp), closely linked to the marker CIW8 associated to the QTL5-1 (14.9 cM / 7.5 Mbp). The Eri-1 alleles at both QTL5-1 and QTL5-2 were responsible for a delayed flowering and increased leaf numbers in the RIL population. Because these markers were linked on the genetic map, we investigated if QTL5-2 was indeed present in addition to the QTL5-1 or a unique QTL. To elucidate the question, the L*er* NIL QTL5 - NIL37(1) (introgression at both QTL5-1 and QTL5-2) was compared to the L*er* NIL QTL5-1 - NIL18(2) (introgression only at QTL5-1) and to L*er*. In the Eri-1 genetic background the Eri NIL QTL5 – NIL53-7 (introgression at both QTL5-1 and QTL5-2) was compared to the Eri NIL QTL5-1 - NIL8-118 (introgression only at QTL5-1) and the selected Eri NIL QTL5-2 - NIL60-22 which carried the L*er* introgression only at QTL5-2 and to Eri-1 (Figure III.18). Even though the L*er* NIL QTL5-1 (NIL18(2)) showed on average a decreased rosette leaf number (less 2 leaves) compared to the L*er* NIL QTL5 (NIL37(1)), the difference was not significant and according to a Student-Newman-Keuls test (confidence interval 95%), both NILs were grouped together. Comparison of Eri NIL QTL5 (NIL53-7) with Eri NIL QTL5-1 (NIL8-118) resulted in no significant differences between the NIL QTL5 and the separated NIL QTL5-1; whereas comparison between NIL53-7 with Eri NIL QTL5-2 (NIL60-22) resulted in a significant difference between these two NILs. The Eri NIL QTL5-2 (NIL60-22) increased the RLN around 10-11 leaves compared to the Eri NIL QTL5 in the HL4 environment. Compared to the Eri-1 parent, Eri

NIL60-22 decreased the rosette leaf number about 1.7 leaves (Figure III.18). However the RLN differences between Eri NIL QTL5-1 and the parent Eri-1 were not significant and both lines fell into the same phenotypic group analysed by a Student-Newman-Keuls test. In summary, the L*er* alleles at the QTL5-2 did not significantly influence the flowering phenotype in both parental lines. This result did not validate the flowering time effects of QTL5-2 and strongly suggest the presence of a unique flowering time QTL on the top of chromosome V.

III.6.CB. Validation of flowering and anthocyanin QTL5-1 in HL4 condition

L*er* NIL QTL5 (NIL37(1)) and Eri NIL QTL5 (NIL53-7) carried introgressions at the full QTL5 (including both QTL5-1 and QTL5-2, see III.3.A.). L*er* NIL QTL5-1 (NIL18(2)) and Eri NIL QTL5-1 (NIL8-118) carried introgressions only at the QTL5-1 region. Comparison of the L*er* NILs with L*er* showed significant increases of 3 to 5 rosette leaves. The full L*er* QTL5 NIL37(1) increased the RLN about 5 leaves and the L*er* QTL5-1 L*er* NIL18(2) about 3 RLN (Figure III.18). Thus the flowering time QTL5-1 could be validated in the L*er* genetic background. All Eri NILs QTL5/QTL5-1 showed a decreased RLN compared to the parental line Eri-1 (Figure III.18). They flowered after developing only 9-10 rosette leaves and fell into the same group of the parental line L*er* in the Student-Newman-Keuls test (with a likelihood of 95%). This result confirmed the previous validation experiment of flowering time QTL5 and showed that only the narrowed QTL5-1 from 5.4 Mbp to 8.43 Mbp is responsible for the variation in flowering time of L*er*/Eri-1 population. Studied anthocyanin content test revealed that lines with increased leaf numbers also have increased anthocyanin content: Both L*er* NILs QTL5/QTL5-1 studied increased anthocyanin content on average to 24.5 µg/ml (NIL37(1)) and 16.6 µg/ml (NIL18(2)) compared to L*er* (3.05 µg/ml). These differences are not significant. According to the Students-Newman-Keuls test, both NILs group together with L*er* in class A. However the ´tendency` of higher anthocyanin accumulation in the L*er* NILs QTL5/QTL5-1 compared to L*er* parent is obvious, due to a second grouping in class B. Both Eri NILs QTL5/QTL5-1 decreased anthocyanin content on average between 24.5 µg/ml to 41.2 µg/ml compared to the Eri-1 mean value of 89 µg/ml (see Figure III.18). Overall, results from this experiment convincingly confirm the presence of the QTL5-1 in HL4

environment for flowering as well as for anthocyanin content in both genetic backgrounds tested.

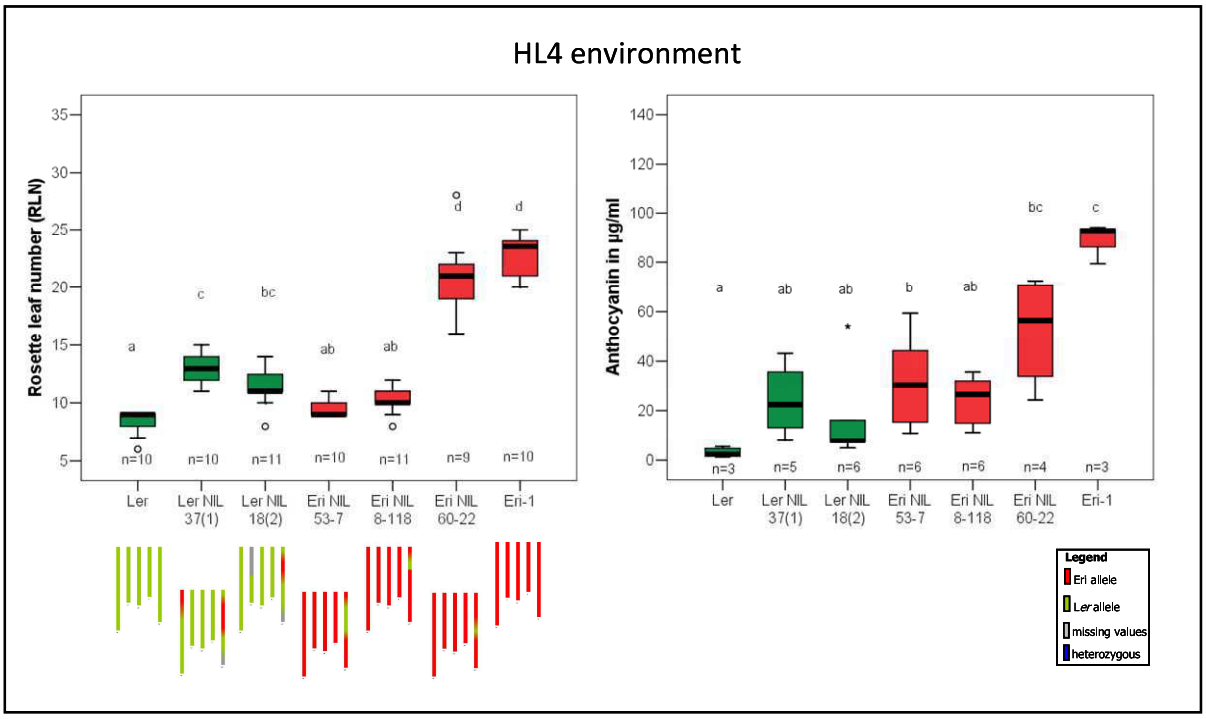


Figure III.18: Rosette leaf numbers and anthocyanin content of L*er*, Eri-1, and the selected L*er* NILs QTL5 NIL37(1); L*er* NILQTL5-1 NIL18(2) and Eri NIL QTL5 NIL53-7; Eri-NIL QTL5-1 NIL8-118 and Eri NIL QTL5-2 NIL60-22 in HL4 environment. L*er*, Eri-1, the L*er* NILs QTL5; QTL5-1 and Eri NILs QTL5, QTL5-1, QTL5-2, which either were derived from a cross between RIL67 and RIL10 and a following backcross with the parental line L*er* or from a backcross between the parental line Eri-1 and RIL103, were grown in high light/cold environment. The box plots represent the median surrounded by the lower and upper quartile and the lowest and highest trait value. The n indicates the number of plants that were used for the test. The letters indicate the different groups (a, b, c and d) for rosette leaf numbers) and (a,b and c) for anthocyanin content, identified by the Student-Newman-Keuls test (95% confidence interval). The genotypes of the lines are indicated below with L*er* alleles in green, Eri-1 alleles in red, heterozygous alleles in blue and missing values in grey.

III.6.D. Fine mapping of QTL5-1

Fine mapping of the QTL5-1 region was performed to narrow down the QTL region and to define the gene(s) which is (are) underlying the effect of the QTL. The validation of the QTL5-1 with smaller introgression NILs indicated already that QTL5-1 is located between marker NGA106 (5.4 Mbp) and marker NGA139 (8.43 Mbp).

The region contains over 1000 genes, according to the predicted genes in the reference sequence of Col-0. This region was deemed too large to name candidate genes already. Therefore it was required to scale down via fine mapping. To perform the fine mapping experiment, progeny (BC2F3) of the Eri NIL QTL5-1 NIL53-8 (see Figure III.14 in the previous paragraph) and Eri-1 were grown and lines carrying recombinant events within the QTL5-1 region were selected. While the plants where growing, flowering time and the leaf numbers were scored. In the QTL detection as well as in the validation experiment the presence of pleiotropic effects of QTL5-1 has been observed. QTL5-1 explained phenotypic variation for flowering time as well as anthocyanin content in the L*er*/Eri-1 population grown under HL4 conditions. Until now, we cannot conclude if these effects are due to a pleiotropic gene or linked genes in the region of QTL5-1. Because Anthocyanin quantification cannot be accurately estimated on a single plant, only flowering related traits were quantified in this fine mapping experiment (Figure III.19). After narrowing down the region, anthocyanin production would be quantified on smaller introgression lines to check if the effect on anthocyanin production is still present. The region of QTL5-1 was bounded by the flanking markers NGA106 (5.6 Mbp) and NGA139 (8.5 Mbp). To find recombinants in between these two markers, already existing markers with polymorphisms between L*er* and Eri-1, as well as newly developed markers (see marker list A.1 in appendix) were used to genotype the 185 segregating plants of the BC2F3 of Eri NIL QTL5-1 NIL53-8. During the fine mapping experiment the RLN was compared between recombinant Eri NILs QTL5-1 and NIL8-118 which carried a whole L*er* introgression between the two flanking markers and Eri-1. The NIL8-118 showed a significantly lower number of rosette leaves than the parental line (Figure III.19). Among the recombinant lines, only the R15, R22 and R23 produced a similar number of rosette leaves than Eri-1. These three lines harboured Eri-1 alleles at the homozygous state at the bottom of the region, from marker CIW8 (7.5 Mbp) to marker hua2-5n (7.85 Mbp). The recombinant lines which carried homozygous L*er* alleles in the 0.32 Mbp region (R1; R7-R11) produced a similar rosette leaf number as the Eri NILQTL5-1 NIL8-118. This result strongly suggested that the gene responsible for the effect of QTL5-1 is located within this region.

Lines with heterozygous alleles between CIW8 and hua2-5n (R2-R6; R14; R16-R21) presented an intermediate rosette leaf number between the RLN of NIL8-118 and that of Eri-1. This intermediate phenotype indicated that the alleles are semi-dominant. The lines

R12 and R13 carried a recombinant event in the above described region. Both lines presented low RLN of the Eri NIL8-118 and shared L*er* alleles between the marker ICE5 (7.71 Mbp) and hua2-5n (7.85 Mbp). Consequently the genes present in the region between CIW8 and ICE5 could be excluded as candidate genes. To summarize the fine mapping results of QTL5-1: The flowering QTL5-1 could be narrowed down to a region of 0.11 Mbp and the effect of Eri-1 allele is semi-dominant for the gene responsible for the QTL5-1 effect.

The recombinant line R16, which carried the Eri-1 alleles from the top of the introgression (marker NGA106) until marker CIW8 and heterozygous parts from CIW8 until the bottom of the introgression (CDPK9) was selected and used as a NIL with small introgressions (Eri NIL QTL5-1_{SM} NIL8-96). In the progeny of the NIL8-96, lines with L*er* alleles in the introgression have been selected. Eri NIL QTL5-1_{SM} (NIL8-96-27) carried a L*er* introgression from marker ICE5 (7.71 Mbp) until CDPK9 (7.95 Mbp).

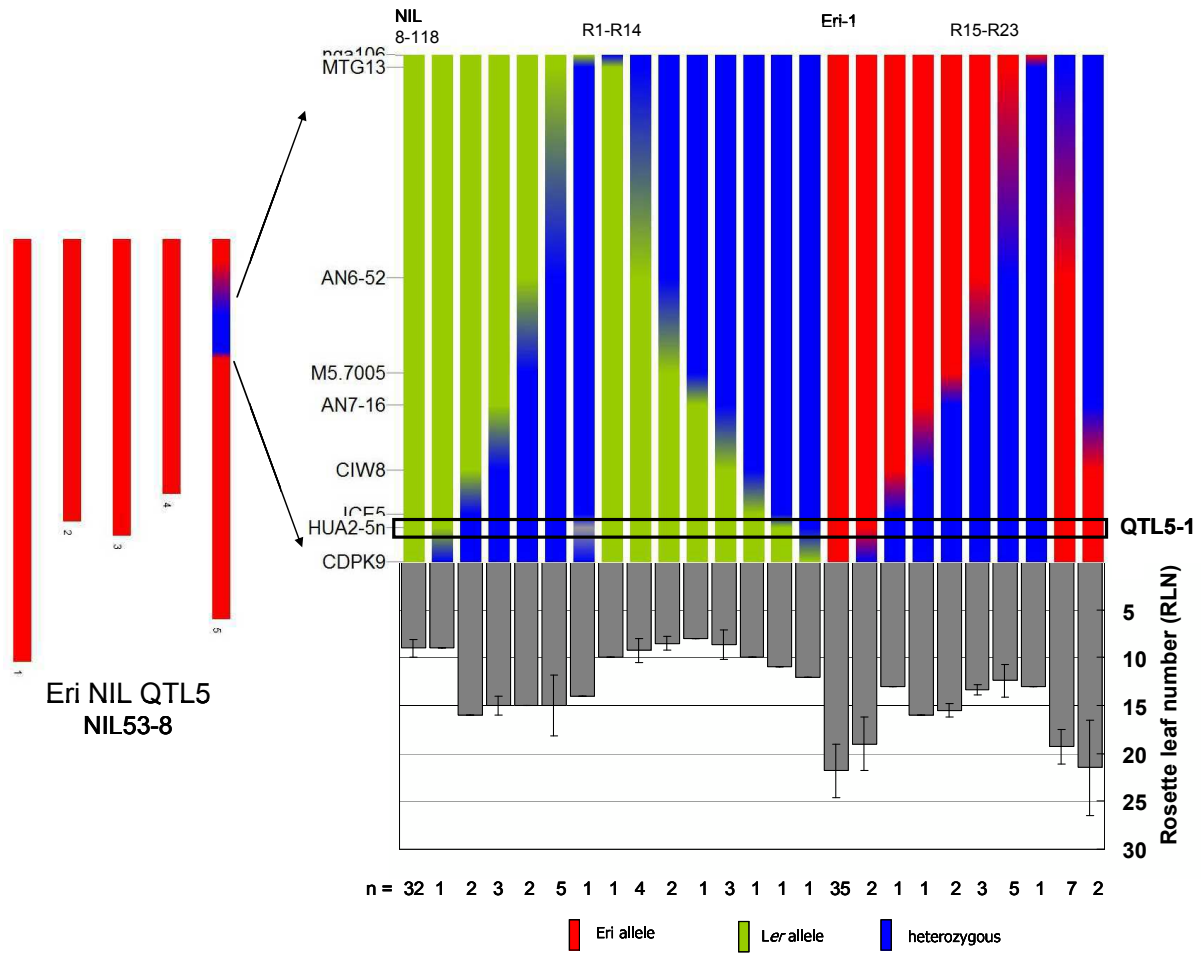


Figure III.19: Fine mapping of QTL5-1 with the progeny (BC2F3) of Eri NIL53-8
The rosette leaf numbers of the parental line Eri-1, NIL8-118 and the recombinant lines are presented in the bottom panel on the right. The lines differ in their genotype on chromosome V which is schematically presented on the top right panel: green represents L*er*-allele, red Eri-1 and blue heterozygous alleles. The numbers (n) of genotypes and phenotypes analysed are presented below the phenotypic data. The physical distance between the markers NGA106 and CDPK9 amounts 2.35 Mbp.

III.6.F. Candidate genes in the QTL5-1 region for flowering related traits and anthocyanin content

Fine mapping of QTL5-1 led to a narrowed region from 7.71 Mbp to around 7.82 Mbp on chromosome V. According to the reference sequence of Col-0 accession, this 110 kbp region contains 26 genes with 33 transcripts in total. Structural polymorphisms including gene copy number variation can be present in the sequence of L*er* and/or Eri-1 at this loci.

A candidate gene involved in the variation of anthocyanin content in the region of QTL5-1 is the gene *TTG1* (Walker et al. 1999) located at 8.37 Mbp on chromosome V. In the last selected Eri NIL QTL5-1_{SM} (NIL8-96-27, see Figure III.19) the gene is located in the region between marker CDPK9 (7.95 Mbp) and marker NGA139 (8.43 Mbp), whereas L*er* alleles are present at CDPK9 and Eri-1 alleles at NGA139. Therefore the gene could not be excluded during the fine mapping experiment. *TTG1* is known to be required for the accumulation of anthocyanin in leaves and stems and is also involved in trichome and root hair development (Bouyer et al. 2008). The gene produces a WD40 protein, which acts together with a MYB domain and bHLH (basic helix loop helix) domain in a complex to regulate anthocyanin biosynthesis (Gonzalez et al. 2008). Comparing the *TTG1* sequence between L*er* and Eri-1 via IGV browser, only two SNPs in the genomic region were detected. One SNP was highlighted in the first intron and the other one in the second intron. Polymorphisms of nucleotide sequence within a gene can lead to the presence of a QTL, either through changes altering the gene function in the coding region or due to changes in their expression level or timing in the promoter region. In the coding region of *TTG1* no nucleotide differences could be highlighted between L*er* and Eri-1, whereas in the putative promoter region several SNPs between L*er* and Eri-1 have been detected. In addition to these single nucleotide polymorphisms, Eri-1 carried a 151 bp (8.375.863 bp – 8.376.014 bp) deletion and L*er* a 37 bp (8.378.576 bp – 8.378.613 bp) in the putative promoter region of *TTG1* (according to the reference sequence of Col-0). These deletions could lead to changes in expression level or timing of *TTG1* between L*er* and Eri-1. Possible changes in the expression level of *TTG1* which could lead to the different anthocyanin accumulation between L*er* and Eri-1 were analysed and are described in paragraph III.6.H.

Another putative candidate gene for flowering related traits present in the sequence of Col-0 in this region is *HUA2* (Chen and Meyerowitz 1999). *HUA2* is a putative transcription factor and member of the floral homeotic *AGAMOUS* pathway. Single *hua* mutants are early flowering and have reduced levels of *FLC* mRNA (Doyle et al. 2005). *HUA2* activates *FLC* expression and enhances *AG* function (Chen and Meyerowitz 1999).
The Landsberg *erecta* parent used for the L*er*/Eri-1 RIL population emerged to carry the *hua2-5* point mutation in the gene *HUA2* (Doyle et al. 2005). By sequencing the *HUA2* gene in the L*er* strain which was used for developing the L*er*/Eri-1 population, this has

been confirmed. In contrast, the Eri-1 parent carried the wild-type *HUA2* allele. A CAPS marker designed for the polymorphic SNP was added during the fine mapping to take this into account. A perfect co-segregation of flowering time and allelic values at this marker was always noted during fine mapping experiment of QTL5-1 (see Figure III.19). Validation of *HUA2* as a candidate gene for QTL5-1 for rosette leaf number as well as for anthocyanin content was performed using a L*er* strain carrying the wild-type *HUA2* allele (L*er*-*HUA2*), the L*er* strain with the mutant *hua2-5* allele (parent of the RIL population; L*er*-*hua2-5*) and the Eri NILQTL5-1_{SM} NIL8-96-27. The Eri NIL QTL5-1_{SM} NIL8-96-27 carried a small L*er* introgression from marker ICE5 (7.71 Mbp) to marker CDPK9 (7.95 Mbp) including the *HUA2* gene in the region (7.82 Mbp) (Figure III.20). The three lines were analysed in the same HL4 environmental condition used for the QTL mapping experiment. L*er*-*HUA2* consistently had increased RLN and anthocyanin content values (on average 13 leaves and 18 µg/ml anthocyanin) compared to the parental mutant L*er*-*hua2-5* (on average 8 leaves and 3 µg/ml anthocyanin) (Figure III.20). L*er*-*hua2-5* and L*er*-*HUA2* are isogenic lines which only differ in the absence/presence of *HUA2*. With this comparison it could be confirmed that *HUA2* shows pleiotrophic effects of changing flowering related traits and anthocyanin accumulation.

Comparison of this new result with those of the second validation experiment of QTL5-1 (see paragraph III.6.C; Figure III.18), showed consistency in the leaf numbers and anthocyanin content of L*er* NIL QTL5 (NIL37(1);13 leaves; 24.5µg/ml anthocyanin) and L*er* NIL QTL5-1 (NIL18(2); 11 leaves; 17µg/ml anthocyanin) and the values of L*er*-*HUA2* (13 leaves; 18µg/ml anthocyanin). The L*er* NILs NIL37(1) and NIL18(2) carried both the Eri-1 alleles of *HUA2* (without the *hua2-5* mutation). This result strongly suggests that the polymorphism present at *HUA2* between Eri-1 and L*er* is responsible for the effect of the QTL5-1. The Eri NIL QTL5-1_{SM} NIL8-96-27 could support this assumption. The Eri NIL8-96-27 (9.5 leaves) which carried a small L*er* introgression from 7.71 Mbp to 7.95 Mbp which includes *hua2-5*^{Ler} produced statistically similar number of leaves as parental line L*er*-*hua2-5* (8 leaves) and less leaves than Eri-1 (22.8 leaves; Figure III.20). All the tested Eri NIL QTL5 in the previous validation experiment showed similar RLN numbers than Eri NIL8-96-27. The mutant *hua2-5*^{Ler} allele in the smaller L*er* introgression of Eri NIL8-96-27 decreased anthocyanin content compared to Eri-1. Eri NIL8-96-27 contained 41.1 µg/ml anthocyanins compared to Eri-1 with 88.8 µg/ml anthocyanins

(Figure III.20). The anthocyanin values of Eri NIL QTL5-1_{SM} NIL8-96-27 were consistent with the values generated in the previous validation experiment for Eri NIL QTL5 lines.

The *hua2-5*^{Ler} alleles decreased the leaf numbers on average about 13 leaves, indicating that *HUA2* is one of or the main gene responsible for the late flowering in Eri-1. The allelic effect of *HUA2* for anthocyanin content in the Eri-1 background was not as high as for leaf numbers. This indicates other genes are also involved in anthocyanin production to result in the high anthocyanin content of Eri-1. Taken together these results proposed *HUA2* as a good candidate for the observed QTL5-1 effect in the L*er*/Eri-1 RIL population. The *HUA2*$^{Eri-1}$ alleles had relatively weak allelic effects observed on RLN in the L*er* background of the L*er* NILs. This fits with the expectations from the report by Doyle et al. 2005.

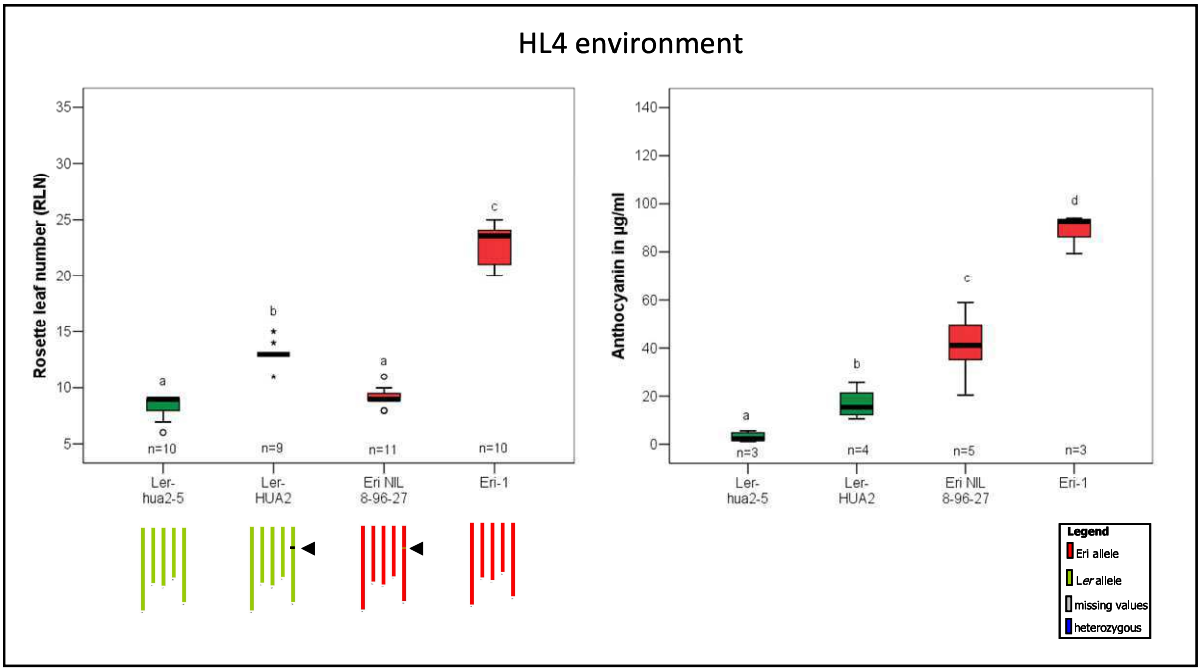


Figure III.20: Validation of *HUA2* as a candidate gene for QTL5-1 for flowering and anthocyanin content. The rosette leaf numbers and anthocyanin content of L*er-hua2*-5, L*er-HUA2*, Eri-1, and the selected Eri NIL QTL5-1_{SM} (NIL8-96-27) quantified in HL4 environment. Ler-hua2-5, Ler-HUA2, Eri-1 and Eri NIL QTL5-1_{SM}, which was derived from a cross between the parental line Eri-1 and RIL103, were grown in high light/cold environment. The boxplots represent the median surrounded by the lower and upper quartile and the lowest and highest trait value. The n indicates the number of plants that were used for the test. The letters indicate the different groups (a, b and c for rosette leaf numbers and a, b, c and d for anthocyanin content), identified by the Student-Newman-Keuls test (94% confidence interval). The genotypes of the lines are indicated below with L*er* alleles in green, Eri-1 alleles in red. The functional *Ler*^{HUA2} is indicated with a black point in the L*er-HUA2* strain.

The public available sequencing data of L*er* (1001genomes.org) does not contain the *hua2 -5*^{Ler} but the *HUA2*^{Ler}. Thus additional to the sequence comparison of *HUA2* in sequenced Eri-1 and L*er* (carrying *HUA2*^{Ler}) in the IGV browser, we sequenced *HUA2* gene in our used L*er-hua2-5*, L*er-HUA2* and Eri-1. Sequencing of the *HUA2* gene in L*er-hua2-5*, L*er-HUA2* and Eri-1 was performed in order to highlight all nucleotide polymorphisms in the coding region between L*er* and Eri-1. In addition to the premature stop codon which was generated through a SNP change in the eighth exon in the L*er-hua2 -5* parental line, two other non-synonymous SNP changes were detected between L*er* and Eri-1 in *HUA2*: In the third exon a SNP substitution from T to A resulted in an amino acid change from Serin in L*er* and also in Col-0 to Threonin in Eri-1 (see Figure III.21). The second SNP (T to C, in the fourth exon) encoded a Phenylalanin in Eri-1 compared to Serin in L*er* and Col-0. The changes in the protein sequence of *HUA2* could lead to changes in the functionality of the gene, for example enhancing or weakening of its function. In the ninth exon a synonymous SNP were detected between L*er* and Eri-1, whereas Eri-1 carried the same SNP as the reference Col-0. For all detected polymorphisms in the *HUA2* gene, L*er-HUA2* strain carried the same SNPs than the L*er-hua2-5* parental line, except for the mutant *hua2-5* polymorphic SNP which resulted in a premature stop codon. This could also be confirmed by sequence comparison with the IGV browser. The comparison also revealed a 70 bp (7.785.570 bp – 7.785.640 bp) deletion in the putative promoter region of *HUA2* in the Eri-1 accession.

```
           10        20        30        40        50        60        70
C:   MAPGRKRGASKAKAKGQLVLGDLVLAKVKGFPAWPAKISRPEDWDRAPDPKKYFVQFFGTEEIAFVAPPDIQAF
E:   MAPGRKRGASKAKAKGQLVLGDLVLAKVKGFPAWPAKISRPEDWDRAPDPKKYFVQFFGTEEIAFVAPPDIQAF
L:   MAPGRKRGASKAKAKGQLVLGDLVLAKVKGFPAWPAKISRPEDWDRAPDPKKYFVQFFGTEEIAFVAPPDIQAF
L-H: MAPGRKRGASKAKAKGQLVLGDLVLAKVKGFPAWPAKISRPEDWDRAPDPKKYFVQFFGTEEIAFVAPPDIQAF

        80        90       100       110       120       130       140
C:   TSEAKSKLLARCQGKTVKYFAQAVEQICTAFEGLQNHKSNALGDEDSLDATEPGLTKAEIVDGTDHIVIESERTD
E:   TSEAKSKLLARCQGKTVKYFAQAVEQICTAFEGLQNHKSNALGDEDSLDATEPGLTKAEIVDGTDHIVIESERTD
L:   TSEAKSKLLARCQGKTVKYFAQAVEQICTAFEGLQNHKSNALGDEDSLDATEPGLTKAEIVDGTDHIVIESERTD
L-H: TSEAKSKLLARCQGKTVKYFAQAVEQICTAFEGLQNHKSNALGDEDSLDATEPGLTKAEIVDGTDHIVIESERTD

       150       160       170       180       190       200       210       220
C:   NFNFRVDPCFPKLDENNGEERKAEIRKLDSSSFLESKVKTTSPVSESLEHSSFDPKIKKEDFDKGTDGSACNEHFG
E:   NFNFRVDPCFPKLDENNGEERKAEIRKLDSSSFLESKVKTTSPVSESLEHSSFDPKIKKEDFDKGTDGSACNEHFG
L:   NFNFRVDPCFPKLDENNGEERKAEIRKLDSSSFLESKVKTTSPVSESLEHSSFDPKIKKEDFDKGTDGSACNEHFG
L-H: NFNFRVDPCFPKLDENNGEERKAEIRKLDSSSFLESKVKTTSPVSESLEHSSFDPKIKKEDFDKGTDGSACNEHFG

       230       240       250       260       270       280       290
C:   NGQKKLANGKRIKKEAGGSDRKGEDTVHRDKSNNSHVPGGRTASGNSDSKKSKGLLTEKTSSKVSADKHENSPG
E:   NGQKKLANGKRIKKEAGGSDRKGEDTVHRDKSNNSHVPGGRTASGNSDSKKSKGLLTEKTSSKVSADKHENSPG
L:   NGQKKLANGKRIKKEAGGSDRKGEDTVHRDKSNNSHVPGGRTASGNSDSKKSKGLLTEKTSSKVSADKHENSPG
L-H: NGQKKLANGKRIKKEAGGSDRKGEDTVHRDKSNNSHVPGGRTASGNSDSKKSKGLLTEKTSSKVSADKHENSPG

       300       310       320       330       340       350       360       370
C:   IKVGVSGKKRRLESEQGKLAPRVDESSRAAKKPRCESADNKVKCEIDDGSDSTGTVSDIKREIVLGLGARGGNFQ
E:   IKVGVTGKKRRLESEQGKLAPRVDESSRAAKKPRCESADNKVKCEIDDGSDSTGTVSDIKREIVLGLGARGGNFQ
L:   IKVGVSGKKRRLESEQGKLAPRVDESSRAAKKPRCESADNKVKCEIDDGSDSTGTVSDIKREIVLGLGARGGNFQ
L-H: IKVGVSGKKRRLESEQGKLAPRVDESSRAAKKPRCESADNKVKCEIDDGSDSTGTVSDIKREIVLGLGARGGNFQ

        380       390       400       410       420       430       440
C:   YDKEAVAYTKRQRQTMEHATSPSFSGSRDKSGKGHLEQKDRSSPVRNVKAPAAQSLKKRRAVCIYDEDDDEDPK
E:   YDKEAVAYTKRQRQTMEHATSPSFSGSRDKSGKGHLEQKDRSSPVRNVKAPAAQSLKKRRAVCIYDEDDDEDPK
L:   YDKEAVAYTKRQRQTMEHATSPSFSGSRDKSGKGHLEQKDRSSPVRNVKAPAAQSLKKRRAVCIYDEDDDEDPK
L-H: YDKEAVAYTKRQRQTMEHATSPSFSGSRDKSGKGHLEQKDRSSPVRNVKAPAAQSLKKRRAVCIYDEDDDEDPK

       450       460       470       480       490       500       510       520
C:   TPLHGKPAIVPQAASVLTDGPKRANVCHSTSTKAKISAGSTESTEVRKFPLRKHCEDASRVLPSNAENSTNSLPVV
E:   TPLHGKPAIVPQAASVLTDGPKRANVCHSTSTKAKISAGSTESTEVRKFPLRKHCEDASRVLPSNAENSTNSLPVV
L:   TPLHGKPAIVPQAASVLTDGPKRANVCHSTSTKAKISAGSTESTEVRKFPLRKHCEDASRVLPSNAENSTNSLPVV
L-H: TPLHGKPAIVPQAASVLTDGPKRANVCHSTSTKAKISAGSTESTEVRKFPLRKHCEDASRVLPSNAENSTNSLPVV

        530       540       550       560       570       580       590
C:   KPINELPPKDVKQILQSPKMSPQLVLTNKHVAGQHKVVKSSVKVSGVVMAKKPQSDSCKEAVAGSDKISSSQSQP
E:   KPINELPPKDVKQILQSPKMSPQLVLTNKHVAGQHKVVKSSVKVSGVVMAKKPQSDSCKEAVAGSDKISSSQSQP
L:   KPINELPPKDVKQILQSPKMSPQLVLTNKHVAGQHKVVKSSVKVSGVVMAKKPQSDSCKEAVAGSDKISSSQSQP
L-H: KPINELPPKDVKQILQSPKMSPQLVLTNKHVAGQHKVVKSSVKVSGVVMAKKPQSDSCKEAVAGSDKISSSQSQP

       600       610       620       630       640       650       660       670
C:   ANQRHKSASVGERLTVVSKAASRLNDSGSRDMSEDLSAAMLDLNREKGSATFTSAKTPDSAASMKDLIAAAQA
E:   ANQRHKSASVGERLTVVSKAASRLNDSGSRDMSEDLSAAMLDLNREKGSATFTSAKTPDSAASMKDLIAAAQA
L:   ANQRHKSASVGERLTVVSKAASRLNDSGSRDMSEDLSAAMLDLNREKGSATFTSAKTPDSAASMKDLIAAAQA
```

Figure III.21: Comparison of the predicted protein sequences of *HUA2* from Col-0, Eri-1, L*er* –*hua2-5* and L*er-HUA2*. The predicted amino acid sequence is shown for Col-0 (C), Eri-1 (E), L*er hua2-5* (L) and L*er-HUA2* wild-type (L-H). The synonymous polymorphism between L*er*/L*er-HUA2* and Eri-1; Col-0 is highlighted by a green box, the non-synonymous polymorphisms are highlighted by red boxes. The non-synonymous polymorphism leading to the premature stop codon in L*er* and to the L*er hua2-5* mutant allele is indicated by a blue box.

```
L-H: ANQRHKSASVGERLTVVSKAASRLNDSGSRDMSEDLSAAMLDLNREKGSATFTSAKTPDSAASMKDLIAAAQA

                680       690       700       710       720       730       740
C:   KRKLAHTQNSIFGNLNPSFLSISDTQGRSHSPFMVQNASASAAISMPLVVQGHHQQGSSPSNHGHQSLSRNQIET
E:   KRKLAHTQNSIFGNLNPSFLSISDTQGRSHSPFMVQNASASAAISMPLVVQGHHQQGSSPSNHGHQSLSRNQIET
L:   KRKLAHTQNSIFGNLNPSFLSISDTQGRSHSPFMVQNASASAAISMPLVVQGHHQQGSSPSNHGHQSLSRNQIET
L-H: KRKLAHTQNSIFGNLNPSFLSISDTQGRSHSPFMVQNASASAAISMPLVVQGHHQQGSSPSNHGHQSLSRNQIET

      750       760       770       780       790       800       810
C:   DDNEERRLSSGHKSVGGSLSCSTEAAISRDAFEGMLETLSRTRESIGRATRLAIDCAKYGLASEVVELLIRKLESE
E:   DDNEERRLSSGHKSVGGSLSCSTEAAIFRDAFEGMLETLSRTRESIGRATRLAIDCAKYGLASEVVELLIRKLESE
L:   DDNEERRLSSGHKSVGGSLSCSTEAAISRDAFEGMLETLSRTRESIGRATRLAIDCAKYGLASEVVELLIRKLESE
L-H: DDNEERRLSSGHKSVGGSLSCSTEAAISRDAFEGMLETLSRTRESIGRATRLAIDCAKYGLASEVVELLIRKLESE

                830       840       850       860       870       880       890
C:   SHFHRKVDLFFLVDSITQHSHSQKGIAGASYVPTVQAALPRLLGAAAPPGTGASDNRRKCLKVLKLWLERKVFPE
E:   SHFHRKVDLFFLVDSITQHSHSQKGIAGASYVPTVQAALPRLLGAAAPPGTGASDNRRKCLKVLKLWLERKVFPE
L:   SHFHRKVDLFFLVDSITQHSHSQKGIAGASYVPTVQAALPRLLGAAAPPGTGASDNRRKCLKVLKLWLERKVFPE
L-H: SHFHRKVDLFFLVDSITQHSHSQKGIAGASYVPTVQAALPRLLGAAAPPGTGASDNRRKCLKVLKLWLERKVFPE

      900       910       920       930       940       950       960
C:   SLLRRYIDDIRASGDDATGGFSLRRPSRSERAVDDPIREMEGMLVDEYGSNATFQLPGFFSSHNFEDDEEDDDLP
E:   SLLRRYIDDIRASGDDATGGFSLRRPSRSERAVDDPIREMEGMLVDEYGSNATFQLPGFFSSHNFEDDEEDDDLP
L:   SLLRRYIDDIRASGDDATGGFSLRRPSRSERAVDDPIREMEGMLVDEYGSNATFQLPGFFSSHNFEDDEEDDDLP
L-H: SLLRRYIDDIRASGDDATGGFSLRRPSRSERAVDDPIREMEGMLVDEYGSNATFQLPGFFSSHNFEDDEEDDDLP

                980       990      1000      1010      1020      1030      1040
C:   TSQKEKSTSAGERVSALDDLEIHDTSSDKCHRVLEDVDHELEMEDVSGQRKDVAPSSFCENKTKEQSLDVMEPV
E:   TSQKEKSTSAGERVSALDDLEIHDTSSDKCHRVLEDVDHELEMEDVSGQRKDVAPSSFCENKTKEQSLDVMEPV
L:   TSQKEKSTSAGERVSALDDLEIHDTSSDKCHRVLEDVDHELEMEDVSGQRKDVAPSSFCENKTKEQSLDVMEPV
L-H: TSQKEKSTSAGERVSALDDLEIHDTSSDKCHRVLEDVDHELEMEDVSGQRKDVAPSSFCENKTKEQSLDVMEPV

      1050      1060      1070      1080      1090      1100      1110      1120
C:   AEKSTEFNPLPEDSPPLPQESPPPLPPLPPSPPPPSPPLPPSSLPPPPPAALFPPLPPPPSQPPPPPLSPPPSPPPPPP
E:   AEKSTEFNPLPEDSPPLPQESPPPLPPLPPSPPPPSPPLPPSSLPPPPPAALFPPLPPPPSQPPPPPLSPPPSPPPPPP
L:   AEKSTEFNPLPEDSPPLPQESPPPLPPLPPSPPPPSPPLPPSSLPPPPPAALFPPLPPPPSQPPPPPLSPPPSPPPPPP
L-H: AEKSTEFNPLPEDSPPLPQESPPPLPPLPPSPPPPSPPLPPSSLPPPPPAALFPPLPPPPSQPPPPPLSPPPSPPPPPP

      1130      1140      1150      1160      1170      1180      1190
C:   PPSQSLTTQLSIASHHQIPFQPGFPPPTYPLSHQTYPGSMQQDRSSIFTGDQIVQGPGNSSRGGLVEGAGKPEYFV
E:   PPSQSLTTQHSIASHHQIPIQPGFPPPTYPLSHQTYPGSMQQDRSSIFTGDQIVQGPGNSSRGGLVEGAGKPEYFV
L:   PPSQSLTTQHSIASHHQIPI_PGFPPPTYPLSHQTYPGSMQQDRSSIFTGDQIVQGPGNSSRGGLVEGAGKPEYFV
L-H: PPSQSLTTQHSIASHHQIPFQPGFPPPTYPLSHQTYPGSMQQDRSSIFTGDQIVQGPGNSSRGGLVEGAGKPEYFV

      1210      1220      1230      1240      1250      1260      1270
C:   QQSSSFSPAGVCSSREPSSFTSSRQLEFGNSDVLFNPEASSQNHRFQPSTPLSQRPMVRLPSAPSSHFSYPSHIQS
E:   QQSSSFSPAGVCSSREPSSFTSSRQLEFGNSDVLFNPEASSQNHRFQPSTPLSQRPMVRLPSAPSSHFSYPSHIQS
L:   QQSSSFSPAGVCSSREPSSFTSSRQLEFGNSDVLFNPEASSQNHRFQPSTPLSQRPMVRLPSAPSSHFSYPSHIQS
L-H: QQSSSFSPAGVCSSREPSSFTSSRQLEFGNSDVLFNPEASSQNHRFQPSTPLSQRPMVRLPSAPSSHFSYPSHIQS
```

Figure III.21 (continued): Comparison of the predicted protein sequences of *HUA2* from Col-0, Eri-1, L*er-hua2-5* and L*er-HUA2*.

```
     1280      1290      1300      1310      1320      1330      1340      1350
C:   QSQHSYTHPYPFPPQRDDARRYRNEEPWRIPSSGHSAENQNGAWIHGRNSHPGLPRVTDSFFRPPPERPPSGTM
E:   QSQHSYTHPYPFPPQRDDARRYRNEEPWRIPSSGHSAENQNGAWIHGRNSHPGLPRVTDSFFRPPPERPPSGTM
L:   QSQHSYTHPYPFPPQRDDARRYRNEEPWRIPSSGHSAENQNGAWIHGRNSHPGLPRVTDSFFRPPPERPPSGTM
L-H: QSQHSYTHPYPFPPQRDDARRYRNEEPWRIPSSGHSAENQNGAWIHGRNSHPGLPRVTDSFFRPPPERPPSGTM

          1360      1370      1380      1390
C:   NYQPSAASNLQAVPAIPGHTAPQMLPSRPDIPTVNCWRPA_
E:   NYQPSAASNLQAVPAIPGHTAPQMLPSRPDIPTVNCWRPA_
L:   NYQPSAASNLQAVPAIPGHTAPQMLPSRPDIPTVNCWRPA_
L-H: NYQPSAASNLQAVPAIPGHTAPQMLPSRPDIPTVNCWRPA_
```

Figure III.21 (continued): Comparison of the predicted protein sequences of *HUA2* from Col-0, Eri-1, L*er-hua2-5* and L*er-HUA2*.

III.6.G. Allelism test with L*er-hua2-5*, L*er-HUA2* and L*er* NILs QTL5-1 to confirm *HUA2* as a candidate for QTL5-1

The L*er* NIL QTL5-1 produced a similar number of rosette leaves as the L*er-HUA2* strain. An allelism test was performed to address the question whether the $HUA2^{Eri-1}$ wild-type alleles were responsible for the later flowering phenotype in L*er* NIL QTL5-1 or if a polymorphism at other loci in the region are responsible for the effect of QTL5-1. Crosses between L*er-hua2-5* and L*er-HUA2*, L*er-hua2-5* and L*er* NILQTL5-1 NIL84(1), L*er-HUA2* and L*er* NILQTL5-1 NIL18(2), L*er-HUA2* and L*er* NILQTL5 NIL41(3), as well as crosses between both L*er* NIL QTL5-1 NIL18(2) and NIL84(1) were then performed. As mentioned before, the L*er-hua2-5* parent of the L*er*/Eri-1 population carried mutant alleles of the gene *HUA2* ($hua2\text{-}5^{Ler}$; Doyle et al. 2005), the L*er-HUA2* strain carried $HUA2^{Ler}$ wild type alleles and all L*er* NILs carried the $HUA2^{Eri-1}$ wild-type alleles

Crossing of L*er-hua2-5* with the L*er-HUA2* strain resulted in a segregation for leaf numbers in the F2 generation, as expected from the involvement of both alleles of *HUA2* ($hua2\text{-}5^{Ler}$ and $HUA2^{Ler}$, see Figure III.22). The plants were genotyped with the CAPS marker hua2-5n designed specifically for the $hua2\text{-}5^{Ler}$ allele (see Materials and Methods). All plants which carried the $hua2\text{-}5^{Ler}$ alleles at the marker position showed an early flowering phenotype with on average 9.6 RLN and 2.8 CLN. This was on average higher than the control parental line L*er-hua2-5*, which produced around 8.5 RLN and 2.8 CLN, but this difference is not significant (Student-Newman-Keuls test, Figure III.22). Plants with both *HUA2* alleles ($hua2\text{-}5^{Ler}$ and $HUA2^{Ler}$) showed an intermediate flowering time

phenotype due to semi-dominant alleles of *HUA2*. This was in accordance with the results of the phenotyping results obtained during fine mapping experiments (paragraph III.6.D.). Results from this experiment convincingly showed that heterozygous plants at the QTL5-1 had an intermediate phenotype. Homozygous $HUA2^{Ler}$ alleles resulted in plants with similar leaf numbers than the L*er-HUA2* strain. All together, this result strongly suggests that the candidate gene *HUA2* is able to mimic the effects of the QTL5-1 and is underlying these effects. However the cross used does not involve the *HUA2* alleles of Eri-1. We then used the same approach by crossing L*er* NILs QTL5-1 with either L*er-HUA2* or L*er-hua2 -5* in order to convincingly prove that polymorphisms between L*er-hua2-5* and Eri-1 at *HUA2* locus are responsible for the effect of the QTL.

Crosses between L*er-hua2-5* with $hua2\text{-}5^{Ler}$ alleles and L*er* NILQTL5-1 NIL84(1) with $HUA2^{Eri\text{-}1}$ alleles were made. The variation of leaf numbers observed in the F2 plants derived from this cross were compared with the one of the F2 from the cross described previously (L*er-hua2-5* x L*er-HUA2*; see Figure III.22). Plants which carried homozygous $hua2\text{-}5^{Ler}$ alleles produced on average 9.6 RLN and 3.3 CLN similar to the homozygous $hua2\text{-}5^{Ler}$ plants of the control cross of L*er-hua2-5* x L*er-HUA2*. Plants with homozygous $HUA2^{Eri\text{-}1}$ alleles shared the same phenotypic group as the plants with homozygous $HUA2^{Ler}$ alleles. The heterozygous plants of the L*er-hua2-5* x L*er* NIL84(1) produced on average one leaf more than the heterozygous plants of the control cross, but still less than the plants with homozygous $HUA2^{Eri\text{-}1}$ alleles, which confirmed the semi-dominant effects of wild-type *HUA2* alleles. The L*er* NILs QTL5-1 NIL84(1) and NIL18(2), and L*er* NIL QTL5 NIL41(3)) produced on average one rosette leaf more than the L*er-HUA2* plants, but nevertheless according to a Stundent-Newman-Keuls test, they shared the same group C. The L*er* NILQTL5-1 NIL84(1) was an exception and fell into a different group D, indicating that in the L*er* NIL84(1) an additional factor was responsible for increasing the leaf numbers with an marginal effect. In addition to the Eri-1 alleles at the QTL5-1 (from marker NGA106 - 5.6Mbp - to marker NGA139 - 8.5Mbp- both markers inclusive), L*er* NILQTL5-1 NIL84(1) carried another Eri-1 introgression at chromosome I. The introgression was located on the top chromosome I from marker NT7123 (0.4Mbp) to marker JV26/27 (4.0Mbp) (both markers inclusive). It also carried a third introgression (heterozygous) which was located on chromosome III at marker position NT204 (5.6Mbp). Both additional introgressions in the L*er* NIL84(1) could be factors influencing the flowering behaviour of the NIL. Adding the allelic effects of a second factor increasing the

leaf numbers could also explain the minor changes of leaf numbers in the heterozygous plants of the F2 of L*er-hua2-5* x L*er* NIL84(1) cross, as well as the increased leaf numbers in the F2 of L*er* NIL18(2) x L*er* NIL84(1) cross (see Figure III.22).

We also crossed L*er* NILs QTL5 with the L*er-HUA2* strain. Crosses with $HUA2^{Eri-1}$ and $HUA2^{Ler}$ alleles did not show any segregation for the RLN phenotype in the F2 progenies of these crosses, whereas we noted segregation when crossing the L*er* NILs with L*er-hua2-5*. Taken together $HUA2^{Eri-1}$ and $HUA2^{Ler}$ alleles have the same effect on flowering related traits and polymorphisms between $HUA2^{Eri-1}$ and $hua2\text{-}5^{Ler}$ are the ones responsible for the effect of flowering QTL5-1 in the L*er*/Eri-1 population. We could confirm that *HUA2* is the gene influencing the main flowering variation between L*er* and Eri-1.

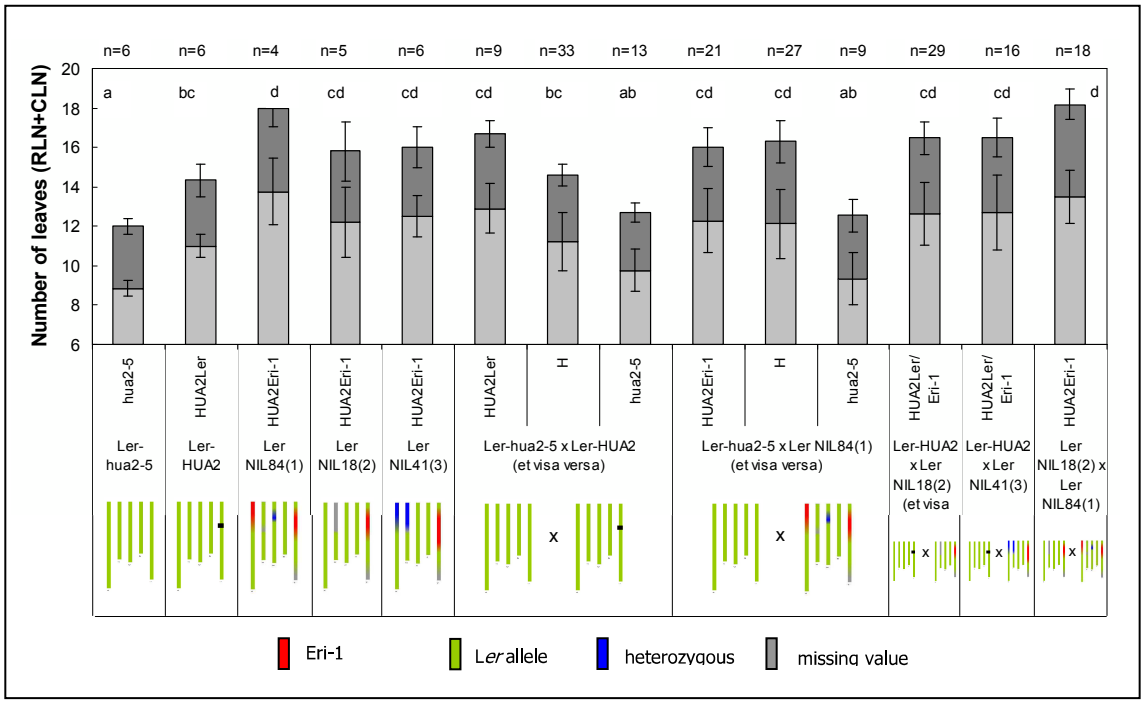


Figure III.22: Allelism test of F2 crosses containing *hua2-5*^{Ler} alleles and *HUA2*^{Ler} and *HUA2*$^{Eri-1}$ alleles for flowering time. The rosette leaf numbers (light grey) as well as the cauline leaf numbers (dark grey) of the 2nd generation of the crosses L*er-hua2-5* x L*er-HUA2*, L*er-hua2-5* x L*er* NIL84(1), L*er-HUA2* x L*er* NIL18(2), L*er-HUA2* x L*er* NIL41(3) and L*er* NIL18(2) x L*er* NIL84(1) are displayed in vertical bars (in mentioned order). Additional the RLN and CLN of the corresponding parental lines are displayed before on the left side of the figure. Below the phenotypic value, the alleles of *HUA2* occurring in the F2 of the crosses (*hua2-5, HUA2*^{Ler}, *HUA2*$^{Eri-1}$ or heterozygous H) are mentioned. The genotypic presentation of the lines is displayed at the bottom of the panel – the five chromosomes of *A. thaliana* are shown as coloured bars – the colours represent the alleles - L*er* alleles in green, Eri-1 in red, heterozygous alleles in blue and missing parts in grey. Above the phenotypic values the number (n) of plants used for the analysis is displayed. The letters below that indicate the different groups (a, b, c and d) for rosette leaf numbers, identified by the Student-Newman-Keuls test (90% confidence interval).

To confirm that *HUA2* is also the gene influencing anthocyanin content in the L*er*/Eri-1 population, the allelism test was also performed on the trait anthocyanin accumulation. The same plants which were phenotyped for leaf numbers were also quantified for anthocyanin content, but only in 2-7 replicates. When crossing L*er-hua2-5* and L*er-HUA2* (and visa versa), a clear segregation (due to the presence of both alleles $hua2\text{-}5^{Ler}$ and $HUA2^{Ler}$) of leaf numbers could be observed in the F2. This clear segregation could not be observed for anthocyanin accumulation. However, similar to previous anthocyanin quantifications, L*er-HUA2* (on average 22 µg/ml) accumulated significantly more anthocyanins than L*er-hua2-5* (on average 8 µg/ml). The F2 plants of the cross L*er-hua2-5* and L*er-HUA2* with homozygous $hua2\text{-}5^{Ler}$ alleles (in Figure III.23 indicated as an A) showed almost similar anthocyanin content (10 µg/ml) than the parental line L*er-hua2-5* (8 µg/ml). Plants carrying the homozygous $HUA2^{Ler}$ (indicated as B) increased the content to 13µg/ml anthocyanins. These differences were not significant due to large variation in anthocyanin accumulation in the F2 plants, but the `tendency` of higher anthocyanin levels in $HUA2^{Ler}$ plants could be observed. Significance was tested with a Student Newman-Keuls test (95% confidence interval). All lines fell into the same group and thus did not show significant differences.

This `tendency` occurred only when L*er-hua2-5* was used as a mother in the cross (see Figure III.23). The F2 progeny of L*er-HUA2* x L*er-hua2-5* showed for all three allelic variations of *HUA2* alleles ($hua2\text{-}5^{Ler}$, heterozygous and $HUA2^{Ler}$) a higher accumulation of anthocyanins than L*er-hua2-5*, on average 20 µg/ml similar to the L*er-HUA2* parent. Interestingly, the maternal effect of L*er-hua2-5* line could also be seen in the cross L*er-hua2-5* with Ler NIL 84(1). Only when L*er* was used as the mother in the cross did the F2 show a ´tendency` of anthocyanin variation dependant on the presence of *HUA2*. Plants with homozygous $HUA2^{Eri\text{-}1}$ alleles increased anthocyanin content from 34.5µg/ml to approximately 40 µg/ml compared to homozygous $hua2\text{-}5^{Ler}$ plants in this cross. Again these differences were not significant and represented only a trend of anthocyanin increase aroused by $HUA2^{Eri\text{-}1}$ alleles. The anthocyanin levels in the L*er-hua2-5* x L*er* NIL 84(1) cross were in total higher than the ones in the L*er-hua2-5* x L*er-HUA2* cross. Also the L*er* NIL84(1) (carrying $HUA2^{Eri\text{-}1}$) showed a higher anthocyanin content compared to L*er-HUA2* (carrying $HUA2^{Ler}$). To answer the question if this increase can be explained by a different factor in the L*er* NIL 84(1) background or with the polymorphisms between $HUA2^{Eri\text{-}1}$ and $HUA2^{Ler}$, the crosses between L*er* NIL QTL5 and L*er-HUA2* were analysed

for anthocyanin accumulation (Figure III.24). The cross with $HUA2^{Ler}$ and $HUA2^{Eri-1}$ (L*er*-*HUA2* x L*er* NIL 18(2) showed the same mean value and also variation in anthocyanin content as the cross with two $HUA2^{Eri-1}$ alleles (L*er* NIL 18(2) x L*er* NIL 84(1). This suggests that the higher value in anthocyanin content is probably given by a second factor in the L*er* NIL 84(1) background. This is also indicated by the lower anthocyanin value of L*er* NIL 18(2) compared to L*er* NIL 84(1).

Because all anthocyanin differences could not be significantly confirmed and only a tendency of accumulation connected with *HUA2* was observed, we can only suggest, not conclude, that *HUA2* is the gene underlying the anthocyanin content QTL5-1. In order to confirm this, more replicates need to be analysed.

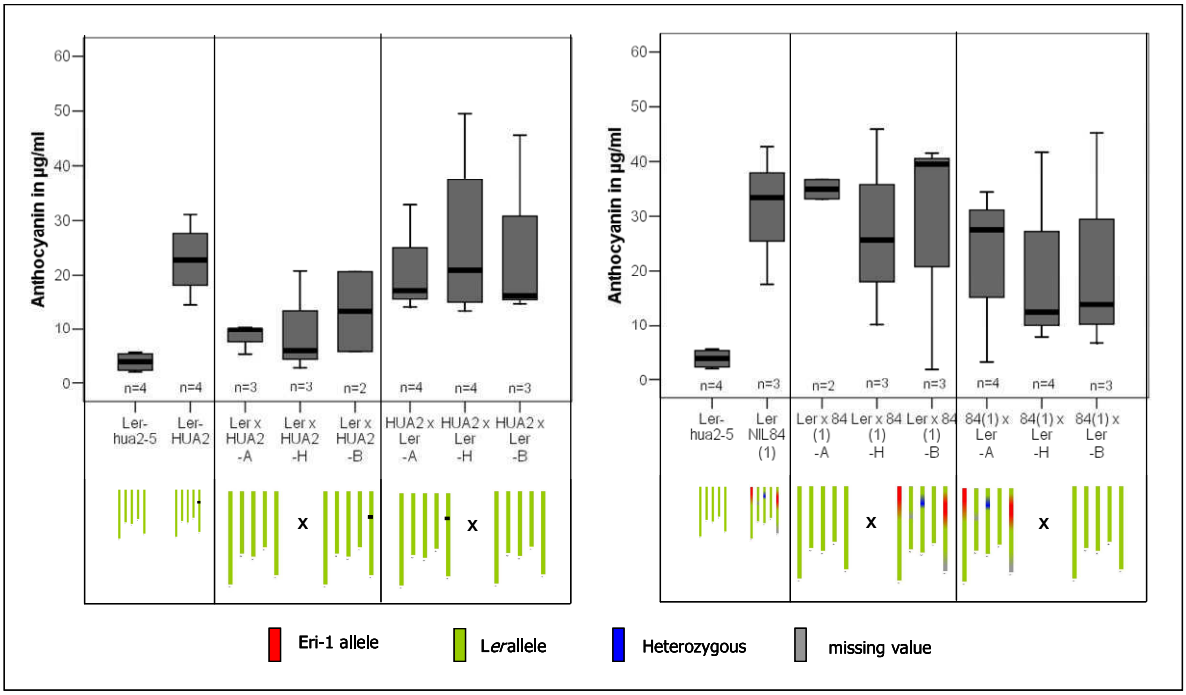


Figure III.23: Allelism test of F2 crosses containing *hua2-5*^{Ler} alleles and *HUA2*^{Ler} and *HUA2*$^{Eri-1}$ alleles for anthocyanin content. The anthocyanin content (grey bars) of the 2nd generation of the crosses L*er-hua2-5* x L*er-HUA2* (left panel) and L*er-hua2-5* x L*er* NIL84(1) (right panel), are displayed in box plots. The box plots represent the median surrounded by the lower and upper quartile and the lowest and highest trait value. The letters below the cross name indicate the allele of *HUA2* (A – *hua2-5*; B - *HUA2*^{Ler} or *HUA2*$^{Eri-1}$ and H for heterozygous. Additional the anthocyanin content value of the corresponding parental lines are displayed before on the left side of both panels. The genotypic presentation of the lines is displayed at the bottom of the panel – the five chromosomes of *A. thaliana* are shown as coloured bars – the colours represent the alleles - L*er* alleles in green, Eri-1 in red, heterozygous alleles in blue and missing parts in grey. Below the phenotypic values the number (n) of plants used for the analysis is displayed. Significance was tested with a Student Newman-Keuls test (95% confidence interval). All lines fell into the same group and thus did not show significant differences.

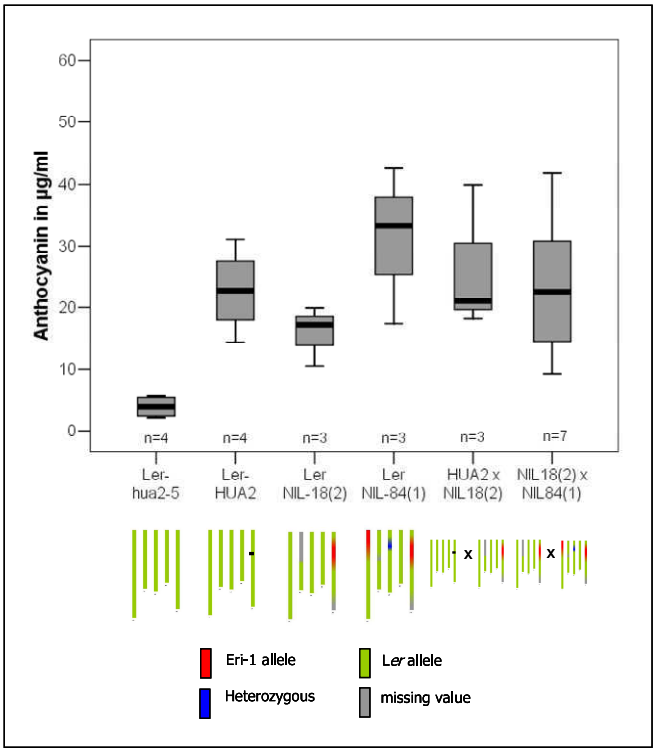


Figure III.24: Anthocyanin content of L*er-hua2-5*, L*er-HUA2*, L*er* NILs QTL5, F2 of the cross L*er-HUA2* x L*er* NIL 18(2) and F2 of L*er* NIL 18(2) x L*er* NIL 84(1). The anthocyanin content (grey boxes) are displayed as box plots. The box plots represent the median surrounded by the lower and upper quartile and the lowest and highest trait value. Additional the anthocyanin content value of the corresponding parental lines as well as L*er-hua2-5* are displayed before on the left side. The genotypic presentation of the lines is displayed at the bottom of the panel – the five chromosomes of *A. thaliana* are shown as coloured bars – the colours represent the alleles - L*er* alleles in green, Eri-1 in red, heterozygous alleles in blue and missing parts in grey.

III.6.H. Transcript level of genes involved in the regulation of anthocyanin production or biosynthesis of anthocyanins in the lines L*er-hua2-5*, L*er-HUA2* and Eri-1

The phenotypic comparison of L*er-HUA2* plants, L*er hua2-5* plants and L*er* NILs carrying $HUA2^{Eri-1}$ wild-type alleles suggested that *HUA2* explained the increased anthocyanin production in the L*er* NILs. The Figures III.18 and III.20 display the on average equal anthocyanin content of L*er-HUA2* (Figure III.20) and the L*er* NILs - L*er* NILQTL5 NIL37(1) and L*er* NILQTL5-1 NIL18(2) (Figure III.18). Additionally the increased anthocyanin content in L*er-HUA2* compared to L*er-hua2-5* showed that *HUA2* is involved in anthocyanin accumulation. Both lines are isogenic lines and only differ in the presence of *HUA2*. The process how *HUA2* contributes in the regulation of the anthocyanin production or in the biosynthesis of anthocyanins is not yet described. As an indication on which level it could act, the gene expression of 17 different genes (genes and used primer combinations are listed in marker table A.2 in appendix) involved in the regulation of anthocyanin accumulation or biosynthesis genes were tested via quantitative real time PCR (qRT-PCR) in Eri-1, L*er-hua2-5* and L*er-HUA2* plants. Plants of all three lines were

grown in HL4 condition to obtain the production of anthocyanins in the leaves. Plant material for RNA extraction was harvested at the same time as material for anthocyanin quantification was harvested (after 43-45 DAS). Of the 17 genes, 6 genes showed significant differences in transcript levels between Eri-1, L*er* and L*er-HUA2* (see Figure III.25). The other 11 genes showed minor or no differences in their expression level. The gene *TTG1* (*AT5G24520*), which was previously selected as a candidate gene in the QTL5-1 region for anthocyanin content in the L*er*/Eri-1 population, showed only minor differences in its transcript level between Eri-1, L*er* and the L*er-HUA2* strains (data not shown) and was thus excluded as a candidate for the QTL5-1. The expression of 6 differential expressed genes was the highest in the Eri-1 plants. The expression in L*er-hua2-5* was always low and the transcript level in L*er-HUA2* ranged in between the level of L*er-hua2-5* and Eri-1. An exception was the expression of *EGL3* (Ramsay et al. 2003). Eri-1 showed high expression of *EGL3*, whereas L*er* and also L*er-HUA2* had almost no expression of the gene. Because of the missing expression of *EGL3* in L*er-HUA2*, the gene was not expected to be influenced by *HUA2* but rather by another genetic factor common in both L*er* strains tested. *EGL3* encodes a bHLH transcription factor 1 (Zhang et al. 2003). The protein is functionally redundant with *GL3* and *TT8* and interacts with *TTG1* and the MYB proteins *GL1, PAP1* and *PAP2, CPC* and *TRY* (Schellmann et al. 2002, Zhu et al. 2009). The mutant of *EGL3* has reduced anthocyanin levels. Interestingly some interacting proteins with *EGL3* are *PAP1, PAP2* and *CPC* which all showed different transcript levels between Eri-1 (highest expression), L*er-hua2-5* (lowest expression) and L*er-HUA2* (expression level in between the two other lines). All three genes had a higher expression in the L*er-HUA2* line compared to the L*er-hua2-5* line, which indicates interplay between *HUA2* and the genes *PAP1, PAP2* and *CPC*. The molecular bases of influences or interactions which can act between *HUA2* and these genes have not yet been described.

The gene which encodes the Dihydroflavanol 4-reductase (*DFR*) was the only gene of the anthocyanin biosynthesis which showed expression differences in the Eri-1, L*er-hua2-5* and L*er-HUA2* lines during this screen. The other differentially expressed genes in that screen were all genes known to be involved in the regulation of anthocyanin production. The Dihydroflavanol 4-reductase catalyzes the conversion of dihydroquercetin to the colourless leucocyanidin and is the third step in the anthocyanin biosynthesis (see biosynthesis in introduction, summarized by Grotewold et al. 2006). The phenotype of described *dfr (tt3)* mutants showed an absent of anthocanins in leaves, stems and all other

tissues. Due to the reduced anthocyanin content phenotype of L*er-hua2-5* compared to Eri -1, transcript level differences in key genes of the anthocyanin biosynthesis like *DFR* were expected. The highest expression level differences between Eri-1, L*er-hua2-5* and L*er-HUA2* were found in the gene *PAP2 and TT8*. *PAP2* is a MYB domain containing transcription factor and homolog to *PAP1*. L*er-hua2-5* strain showed almost no expression of *PAP2* whereas L*er-HUA2* showed low level of expression and Eri-1 high expression levels. Almost similar levels between the three lines were found for the expression of *TT8*. L*er-hua2-5* strain showed almost no expression of *TT8* whereas L*er-HUA2* showed low level of expression and Eri-1 high expression levels.

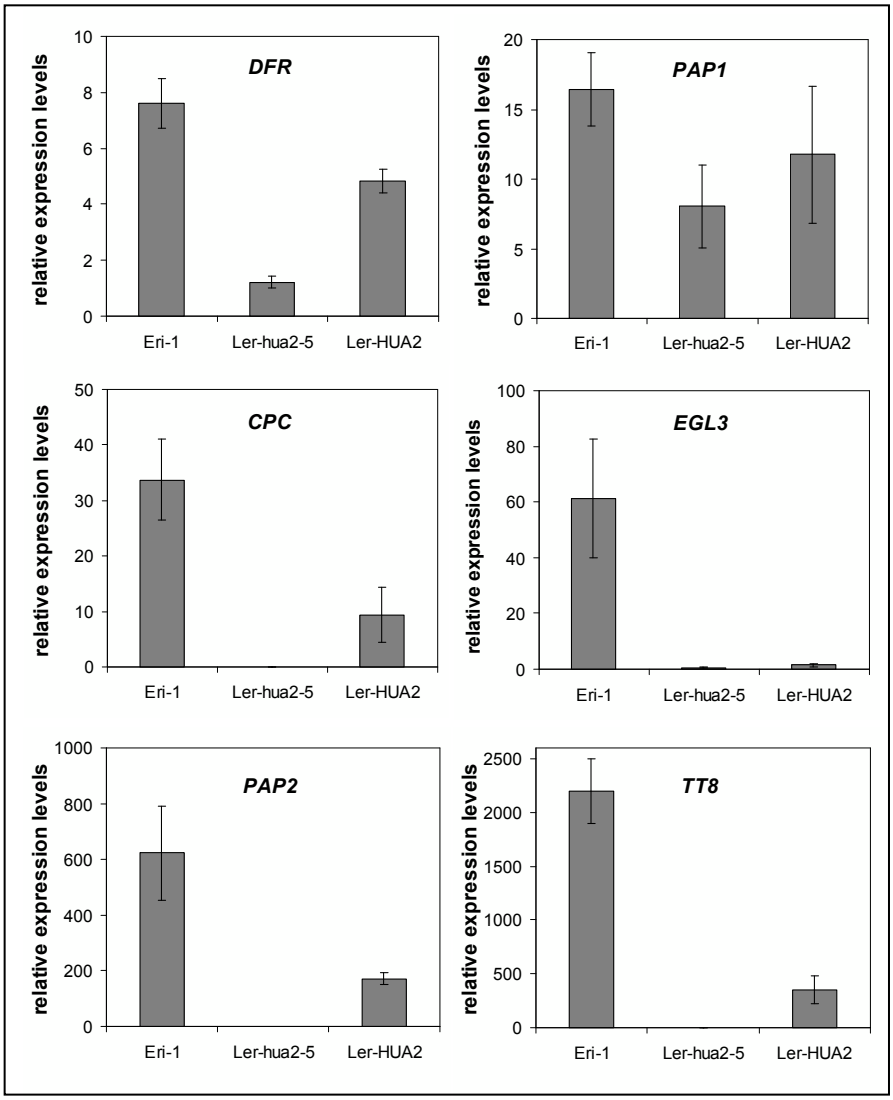


Figure III.25: Gene expression analysis of regulatory and biosynthesis anthocyanin genes. Transcript levels in Eri-1, L*er-hua2-5* and L*er-HUA2* grown in HL4 condition are displayed (y-axes) for the genes *DFR, PAP1, CPC, EGL3*, *PAP2* and *TT8*. Values are obtained by quantitative real time PCR and are the mean ± SD of 3 replicates expressed normalized to *Actin2*. Experiments were repeated twice with similar trends observed.

IV. Discussion

IV.1. Building a genetic map for the newly developed L*er*/Eri-1 population

Accuracy and saturation of a genetic map are crucial parameters for the precise localisation of a QTL position. It has been shown by Price (2006) that QTL mapping is quite accurate under these circumstances. The aim of developing a map is to fulfil these demands in order to be confident of the output of the following QTL detection.

Genotyping of the newly developed F_9 RIL population derived from the cross between L*er* and Eri-1 accessions was initiated at the Wageningen University by using AFLP markers. Unfortunately, using this genotyping approach, some of the regions were not covered by these markers. Additional markers were needed, *e.g.* microsatellites with known physical positions, in order to obtain a saturated genetic map with evenly distributed markers. The final genetic map was constructed by adding these markers. This approach is only feasible for species with a fully sequenced genome (*e.g. A. thaliana*, rice, maize, medicago, brachipodium, sorghum and more). Precision of the genetic map is very important, with levels of mistyping as low as possible and no missing data. Only univocal markers were selected to genotype the RILs, thus ensuring a very high level of precision to the assigned genetic position of each marker. This is important, since Zeng et al. (1999) showed that QTL detection may be affected by an uneven distribution of markers in the genome. Also, minor QTL can remain undetected because of the low density of markers. The genetic map obtained for this L*er*/Eri-1 RIL population has a length of 365 cM, which is in the same range as *A. thaliana* genetic maps obtained from different crosses (Alonso-Blanco et al. 1998; Loudet et al. 2002; El Lithy et al. 2006).

Moreover, by adding new markers all along the genome in all the lines from the population, we were able to detect the presence of two contaminated lines in this population. In fact, only alleles from L*er* or from Eri-1 were expected to be present at each position in each RIL. Presence of another allele (difference in band size on a 3% agarose gel in this study; see Figure III.2) was an indicator that the line bearing this unexpected allele had not been derived from the original cross. Indeed, during genotyping two lines were appointed as contaminants because their genome contained alleles other than the ones from L*er* or Eri-1. Such contaminants might show different phenotypes, change the variation in the trait response and thus influence QTL mapping. These lines can also

influence the genetic position of the markers in the map. For aforementioned reasons, these two lines were discarded from all the analyses performed in the present work.
Because the lines of the L*er*/Eri-1 population are over 99% homozygous, genotyping of the F_9-RIL is also valid for their progeny. Thus, the amount of seeds from each RIL is not limited and genotyping procedure not required anymore. This population is now publically available to be used by interested researchers. The genotyping information acquired during the present work has already been used for collaboration with Wageningen University (Ghandilyan et al. 2009).

IV.2. Effect of the environmental conditions on the correlations between the measured traits

The genetic and molecular basis of the developmental transition from vegetative to reproductive growth (flowering time related traits) has been intensively studied and deciphered in *A. thaliana* over the past decade. The analysis of artificially induced mutants led to identification of around 100 genes in pathways involved in flowering timing (Koornneef 1998, 2004, Mouradov 2002, Simpson and Dean 2002, Boss et al. 2004, Turck et al. 2008). However, there are still many open questions concerning the transition to flowering under certain environmental conditions.
We analysed flowering behaviour of one early (L*er*) and one late (Eri-1) flowering *A. thaliana* accession. Variation between segregating RILs derived from crosses between these two accessions showed substantial variation, as it was also observed in other crosses between L*er* x Cvi (Alonso-Blanco et al. 1998a), between Bay-0 x Sha (Loudet et al. 2002), L*er* x Col-0 (Jansen et al. 1995) and L*er* x Sha (El-Lithy et al. 2004), Col-5 x Kas-1 (Li et al. 2006) and L*er* x Fei-0 (Mendez-Vigo et al. 2010). Simon et al. (2008) even used the flowering time trait to validate the high power of QTL mapping in a set of fifteen integrated *A. thaliana* accessions. The timing of transition from vegetative to reproductive development is an adaptive trait, as it is essential for plants to complete flower development, pollination and seed production in favourable environmental conditions. Plants have developed mechanisms to perceive environmental cues as photoperiod, light intensity, temperature and other abiotic factors in order to alter their flowering time in response to these signals (Shindo et al. 2005, Werner et al. 2005 a,b,).

Arabidopsis is a facultative long day (LD) plant. Under LD conditions (16h light/8h dark), it flowers earlier than under short day (SD) conditions (Koornneef 1991). When plants are grown at 16 hours day length, the photosynthetic efficiency characterized in building up photosynthates, mainly transported as sucrose, is at high level. The sucrose is transported to developing organs such as roots, flowers, fruits or seeds. The outcome of higher carbohydrate production is also shown in increased growth rate of the plant. An increased growth rate results in development of chlorophylls, higher density of chlorophylls in the leaves and light harvesting proteins, all supporting photosynthetic activity (Labate 2004). The recurrent cycle ends if enough carbohydrates are produced to induce flowering. Ohto (2001) has shown similar effects of sugar on vegetative development and floral transition in *A. thaliana*. It is well known that sucrose promotes flowering in most species. Under LD conditions this event is reached even faster. Variations in the environment can decrease photosynthetic efficiency and result in sugar limited conditions in parts of the plants, decreasing biosynthetic activity and hence leaf expansion as well as flowering time. The effect of environmental variation was evident during our analysis in changing the light and temperature conditions, resulting in the subsequent delay in flowering time and increased number of leaves (see higher leaf number for the high light and cold environments, Figure III.7). Variation of the responses of every measured trait to high light and cold condition was observed in the studied population.

The described correlation between flowering time and leaf numbers found under LD control conditions during our analyses is in agreement with previous studies. The leaf number can be considered as a developmental measure for flowering time, whereas the number of days to flower is influenced by both the number of nodes and the developmental rate (Koornneef et al. 1991). The correlation still remains in the HL4 environment. In HL4 environment, the number of leaves is increased because of the delay in flowering time. Interestingly, in addition to the increased correlation in the HL4 environment, new correlations were found under this strong light and cold environment. In the high light and cold environment the delay of flowering time and increase of leaf numbers was correlated with content of anthocyanins solely. This strongly suggests that this condition affects the development of the plants. The combination of these abiotic factors (light intensity and temperature) and their impact on important steps during plant development has been previously studied and is described in the introduction. Carbohydrate content, especially Suc content, could be the connecting part of this correlation between flowering related

traits and pigmentation. The amount of sugars induced by high light intensity is a factor leading to higher anthocyanin accumulation. Sucrose induces PAP1, a transcription factor of the regulatory MYB-bHLH-WD40 transcription factor complex, binding at regulatory regions of the late anthocyanin biosynthesis genes (LBGs) and start the biosynthesis (Teng et al. 2005). It was already shown that sucrose and light co-ordinately regulate the anthocyanin accumulation in leaves (Teng et al. 2005, Jeong et al. 2010, reviewed by Das et al. 2011). The subsequent degradation of chlorophylls to limit the photosynthetic activity is also affected by sugar contents. Finally, flowering time is promoted by sucrose content. Nagira et al. (2006) have already observed that lower anthocyanin accumulation shortened flowering time.

The negative correlation between anthocyanin accumulation and consequential degradation of chlorophylls was also observed in our analysis under high light and cold condition. A negative correlation was observed for flowering time and photosynthetic related traits (chlorophylls) in this condition as well. This could be explained by high amounts of chlorophylls, the increasing of photosynthetic efficiency and the accumulation of carbohydrates which in turn promotes the transition to the reproductive phase as well as in case of too much energy produced leads to the degradation of chlorophylls. This supported the notion of Zhang (1997) that inhibition of light harvesting proteins can cause late flowering phenotypes. Labate (2004) showed that under high excitation, the expression of the light harvesting complex gene family is repressed and growth is reduced.

The correlations of traits were also translated at QTL level, because collocating QTL were mapped under HL4. To depict the genetic basis of anthocyanin accumulation in combination with delayed flowering time in the high light and cold scenario, selecting NILs would be of interest. At a population level, the quantification of biochemical traits is time consuming, while it is conceivable to measure them on a smaller number of selected lines. Results of this analysis can answer if the correlation is due to closely linked genes, meaning that the effects on anthocyanin accumulation might have had to overcome those caused by flowering time. An intriguing alternative would be that these are pleiotropic effects of the same gene. As the results from the phenotyping of the NILs can not answer this question, we performed the fine mapping (- cloning) experiments to elucidate this point. This is discussed in detail in chapter III.6.D and following pages.

IV.3. Major QTL mapped in L*er* /Eri-1 population

IV.3.A. Flowering time related QTL on chromosome V

It has been demonstrated that most of the late flowering accessions in *A. thaliana* carry active *FRI* and *FLC* loci, which act together genetically to delay flowering time (Koornneef et al. 2004, Shindo et al. 2006). Both genes are also involved in the response of vernalization treatment and therefore explain the winter-annual flowering strategy. The requirement for vernalization ensures the plant to overwinter as a rosette and to flower in early spring. In mild winters, a spring-annual flowering strategy without vernalization requirement would be an advantage. Several accessions carrying non functional alleles of *FRI* exhibit an early flowering phenotype (Johanson et al. 2000, Shindo et al. 2005). Early flowering strains from the laboratory (L*er*, Col-0) carry natural loss of function alleles of *FRI*. No QTL has been found in the L*er*/Eri-1 population collocating with the *FRI* gene, which is located on the top of chromosome IV. Since L*er* carries a loss of function allele of *FRI* caused by a 375 bp deletion and 31 bp insertion in the promoter and the first exon region (Johanson et al. 2000). This result indicates that Eri-1 also bears *FRI* defective alleles with similar effects to those of L*er*. In fact, visualizing the sequencing data of the Eri-1 *FRI* region with the IGV browser denotes no sequence reads in the 375 bp promoter and first exon region compared to Col-0. This suggests that Eri-1 carries the same deletion and therefore groups to the same haplotype group of *FRI* as L*er* (data not shown). The analysis of 192 *A. thaliana* accessions by Shindo et al. (2005) has shown that the L*er*-type deletion was the most common non-functional *FRI* allele. It has been found in four different *FRI* haplotypes in accessions from Western Europe, Russia, the United States and southern Sweden, where Eri-1 originates from. Natural allelic variation at *FRI* has been shown to explain up to 70% of the phenotypic variation for flowering time (Lempe et al. 2005, Shindo et al. 2005, and Werner et al. 2005b). However, while this gene is known to be a key factor of flowering time (Roux et al. 2006), it is not involved in the variation of flowering time observed in the L*er*/Eri-1 population.

In the L*er*/Eri-1 population large variation of flowering time was observed in both tested environments, with lines flowering earlier than L*er* and others flowering later than Eri-1. This transgression is also translated at the QTL level. Indeed, QTL for flowering time were detected for the two environments with L*er* alleles either decreasing or increasing the flowering time. Because a RIL is a “mixture” of genomic regions from both parental lines,

it can carry all the unfavourable alleles (reducing flowering time, coming either from L*er* or from Eri-1) and the resulting flowering time of such a RIL will be earlier than L*er*. Similarly, some RILs may contain alleles that increase flowering time and would be later flowering than Eri-1. We observed strong correlations for all flowering time related traits between the two environments. At QTL level, these correlations are strongly supported by three collocating QTL on chromosome V for all flowering related traits in both environments. Moreover, flowering time and related traits were also correlated within each environment, either only in CT or only in HL4. When detecting QTL for these traits, co-localised QTL for flowering related traits have been mapped on chromosome III under HL4 solely. The other two collocating QTL, including the QTL for all flowering related traits, were located on chromosome V, one under HL4, and another one under CT condition. In *A. thaliana* several quantitative analyses of flowering time have been performed so far and some QTL detected in this work collocated with QTL identified in these analyses (El-Lithy et al. 2006, Simon et al. 2008, O´Neill et al. 2008, Mendez-Vigo et al. 2010). Currently, there are nine genes that have been isolated underlying *A. thaliana* flowering time QTL (reviewed in Alonso-Blanco et al. 2009).

The loci *FRI* (Johanson et al. 2000) and *FLC* (Michaels and Amasino 1999) together account for almost three quarter of the flowering time variation in naturally occurring accessions. In the region of *FLC* at the top of chromosome V, collocating QTL for flowering time related traits in the L*er*/Eri-1 population were detected in both environments. Null alleles impairing gene function as found for *FRI* are rare for *FLC*. There are several polymorphisms and insertions between natural accessions of *A. thaliana* found in the first intron of *FLC* which are responsible for variation of flowering time. Intronic insertions at *FLC* typically lead to early flowering (Gazzani et al. 2003, Lempe et al. 2005, Shindo et al. 2005, 2006). Eleven *FLC* haplotypes have been found between accessions which can be differentiated in two major haplotype groups (Caicedo et al. 2004, Stinchcombe et al. 2004, Scarcelli et al. 2007). Two SNPs distinguish the two groups FLC^A and FLC^B. These two haplotypes are associated with flowering time in absence of a putatively functional *FRI* allele. Due to a 1233 bp insertion of a transposable element in the first intron, L*er* carries a loss of function allele of *FLC* (Gazzani et al. 2003, Michaels et al. 2003). On the other hand, Eri-1 belongs to the FLC^A haplotype group like Col-0, which carries a functional allele of *FLC*. Thus, *FLC* reveals to be a very suitable candidate gene

underlying the QTL on the upper region of chromosome V, being responsible for over 40% variation in flowering between L*er* and Eri-1 in both environmental conditions tested.

The major function of *FRI* is to up-regulate *FLC* and as *FRI* is not functional in the L*er*/Eri-1 population, this suggests the presence of other factor(s) enhancing *FLC* function. Shindo et al. (2005) have observed late flowering accessions with non-functional *FRI* haplotypes, showing high *FLC* levels. This data implicates variation in other genes than *FRI* causing enhanced *FLC* function. As previously mentioned, the strongest QTL (QTL5-1) explained 25% and 54% of flowering time variation in CT and HL4 condition, respectively. It was mapped on chromosome V and collocated with the *HUA2* locus. *HUA2* is a repressor of flowering and appears to function by enhancing the expression of several genes that delay flowering including *FLC, AG* and *MAF2-MAF5* (Chen and Meyerowitz 1999, Doyle et al. 2005). It was shown that *HUA2* contains a RPR domain, a motif found in proteins that function in RNA metabolism and pro-mRNA processing and splicing. Furthermore, it has been suggested that *HUA2* has transcription factor activity. Poduska et al. (2003) pointed out that *HUA2* (labeled *ART1* in their study, Wang et al. 2007 demonstrated later that *ART1* is *HUA2*) activates *FLC* independently of *FRI*, but with synergistic effects in case of an active *FRI* allele present. They showed that monogenic *FLC* lines (functional $FLC^{Sy\text{-}0}$ in a L*er* background) are early flowering and only by crossing with monogenic *HUA2* lines (functional $HUA2^{Sy\text{-}0}$ in a L*er* background) delayed flowering has been observed. The $HUA2^{Sy\text{-}0}$ allele appears to be a gain-of function allele exclusive for Sy-0 regarding its ability to activate *FLC*. One non-synonymous amino acid change in $HUA2^{Sy\text{-}0}$ is predominantly responsible for that gain of function allele. $HUA2^{Sy\text{-}0}$ is a rare allele and the SNP underlying this allele has not been found in other accessions yet.

The L*er* accession used for the development of the L*er*/Eri-1 population contains a premature stop codon in *HUA2* leading to a loss of function mutant *hua2-5* allele (Doyle et al. 2005). This lesion is located in the eights exon and results in removal of 280 amino acids from the C-terminus of the *HUA2* protein. In contrast, Eri-1 carries a functional allele of *HUA2*. The early flowering phenotype of the monogenic *FLC* lines (functional $FLC^{Sy\text{-}0}$ in a L*er* background) created by Poduska et al. (2003) is comparable with the early flowering phenotype of Eri-NIL8-96-27 (see Figure III.20). This NIL carries $hua2\text{-}5^{Ler}$ and functional $FLC^{Eri\text{-}1}$ alleles. The Eri-1 parent carries both functional alleles and is late flowering. Validation and fine mapping of QTL5-1 and an allelism test with the strains

L*er-hua2-5* and L*er-HUA2* confirmed that *HUA2* is the gene responsible for flowering time effects in L*er*/Eri-1 population. These results will be discussed in more detail in paragraph IV.4.B.

The third group of flowering time QTL in L*er*/Eri-1 collocating with already known genes involved in naturally occurring flowering variation was detected on the bottom of chromosome V. The *MAF2-MAF5* (for MADS Affecting Flowering 2 to 5) cluster has recently been implicated in flowering time variation and has been shown to be very polymorphic between accessions (Caicedo et al. 2009, Rosloski et al. 2010). This tandem of four MADS family genes are close homologues to *FLC and FLM (MAF1)*. Many of the QTL identified with different RILs in the past decades mapped to the same genomic region and overlapped with the location of *MAF2-MAF5* (Alonso-Blanco et al. 1998a, El-Lithy et al. 2004, 2006, Simon et al. 2008, Mendez-Vigo et al. 2010, Salomé et al. 2011). Doyle et al. (2005) mentioned that *HUA2* also interacts with the *MAF2-MAF5* and has shown that the expression of *MAF2* is also reduced by *hua2* mutations similar to *FLC*. In case L*er* was used as a parent of the RILs, like in the work presented here, L*er* alleles at *MAF2-MAF5* always delayed flowering time, indicating that L*er* carries functional *MAF*s and Eri-1 possibly loss of functional or less functional alleles of the *MAF*s. Sequence analysis of the Eri-1 accession (this work) revealed that Eri-1 carries deletions in the promoter, 5´UTR and 3´UTR of *MAF2* compared to L*er* and Col-0. These deletions and the presence of several SNPs in the first intron could be the reason for a loss of function (partial or total) allele in the Eri-1 accession. The QTL at the bottom of chromosome V explained only 4.6% of the flowering variation and was only detected in the CT environment. Interestingly, Salomé et al. (2011) identified that North Scandinavian accessions like Lov-5 show very strong *FLC* alleles with very late flowering phenotypes. These accessions are lacking the QTL near the *MAF* cluster because the late flowering *FLC* QTL appeared to mask *MAF2-MAF5* QTL effects. This could be the same case in the HL4 environment we tested. The effects in flowering variation of the *FLC* QTL and *HUA2* QTL in the HL4 are much stronger. They explain most of the variation in flowering compared to the CT environment and thus could also mask the effect of *MAF2-MAF5*. The output of multiple QTL mapping (MQM) during the QTL detection we performed is opposing this theory. In MQM the closest marker of a detected major QTL will be used as a cofactor during the detection. This enables to detect other QTL, even if they have a small effect. We did not

detect QTL collocating with *MAF2-MAF5* QTL in HL4 condition, although we selected *FLC* or *HUA2* as a cofactor.

To summarize the first paragraph: The overlap of L*er*/Eri-1 QTL on chromosome V with three already described flowering loci *FLC, HUA2* and *MAF2-MAF5* (Michaels and Amasino 1999, Doyle et al. 2005, Caicedo et al. 2009) suggests that these genes are candidate genes for the flowering variation seen in L*er*/Eri-1 population. The QTL located the upper part of chromosome V collocates with *FLC*, QTL5-1 with *HUA2* and the QTL on the lower part of chromosome V with *MAF2-MAF5*. However, the pleiotropic effect of QTL5-1 described in this study indicates that the already reported flowering time gene *HUA2* which is underlying the QTL5-1 also affects other mechanisms, thus altering the anthocyanin content in leaves under special environmental scenarios. This fact makes further analysis of already known flowering time genes in different conditions worthwhile in order to attribute other functions to these genes. It is also worth noting that the QTL5-1 on chromosome V detected in both environments is the major QTL for the flowering time related traits. Eri-1 alleles within the QTL at this position delayed flowering time.

The second QTL for flowering time on chromosome V was detected only in the high light and cold temperature environment, its positioning right below QTL5-1 so it was consequently labelled as QTL5-2. This QTL was not confirmed by using the selected NILs. The region between the two QTL5 is quite small and QTL5-2 explains almost the same variation of flowering in HL4 as QTL5-1 (52.2% and 54%, respectively). Furthermore, the additive effect of the Eri-1 allele appeared to be the same for both QTL. Taking everything together, we concluded that the QTL5-2 might be the same QTL as the QTL5-1. Additional evidence pointing to only one QTL was the separation of the anthocyanin content QTL (QTL5-1) from the chlorophyll content QTL (QTL5-2). This was unexpected due to previously performed correlation studies. Both traits showed a strong negative correlation, which is expected to lead to common QTL with opposite effects. Surprisingly, instead of one QTL cluster comprised of the anthocyanin and the chlorophyll content QTL, we detected two separate QTL in close vicinity to each other. A possible reason for detecting a false-positive QTL could be the low number of recombinant events between the two corresponding QTL. Only seven among the 110 RILs carried recombinant events between the two markers underlying these QTLs. All others either carried L*er* alleles or Eri-1 alleles at both marker positions. We assume that QTL5-1

combines QTL for flowering time related traits in both environments with pigmentation related traits and photosynthetic related in the HL4 environment.

IV.3.B. Flowering time related QTL on chromosome III

In the HL4 environment, another group of QTL involved in the variation of chlorophyll content and flowering related traits was mapped on the top part of chromosome III (labelled QTL3). In QTL mapping, QTL3 was only detected in the HL4 condition. In validation experiments of the QTL3 with clean NILs grown in both environments, it was shown that the effect of QTL3 in flowering variation was also detected in CT environment. The analysis of lines with L*er* introgression at QTL3 in clean Eri-1 backgrounds increased the effect of the flowering QTL3 in both environments tested. In QTL detection L*er* alleles at QTL3 increased RLN on average about 1.5 leaves whereas during the validation the Eri-NIL67-4-2 increased the RLN on average about 12.5 leaves. It seems that a combination of L*er* alleles at QTL3 and functional FLC^{Eri-1} and $HUA2^{Eri-1}$ alleles underlying the two major QTL for flowering traits on chromosome V, give rise for a very late flowering phenotype. This indicates cooperative action between QTL that may be in the same pathways to produce the same outcome. The late phenotype could then also be distinguished in the CT environment. In accordance with the phenomena observed in this study, Keurentjes et al. (2007) have verified that mapping populations derived from NILs have a higher statistical power to identify low effect QTL than RIL populations. The hypothesis of synergistic alleles of functional *HUA2*, functional *FLC* and the L*er* allele at QTL3 delaying the flowering time was verified with the early flowering phenotype of the L*er*-NILs QTL3 and the L*er*-*HUA2* strain. The L*er*-*HUA2* carrying the L*er* allele at QTL3 and a functional $HUA2^{Ler}$ is early flowering because it is lacking the full functional allele of *FLC*. The L*er*-NILs QTL3 carry weak alleles of FLC^{Ler} and non functional $hua2\text{-}5^{Ler}$ alleles and the Eri-1 allele at QTL3 and show even an earlier flowering phenotype than the L*er*-*hua2-5* strain. Almost similar synergistic effects between three genes have been demonstrated in the Sy-0 accession (Poduska et al. 2003, Wang et al. 2007). Alleles of FRI^{Sy-0}, FLC^{Sy-0} and $HUA2^{Sy-0}$ act together to delay flowering time. The combination of these three alleles of Sy-0 showed delayed flowering phenotype as well as a pleiotropic effect in shoot morphology by formation of aerial rosettes. As mentioned above, Wang et al. (2007) have indicated that a

rare allele of a gene in Sy-0, not found in other accessions, is responsible for the pleiotropic effect of aerial rosettes and very late flowering. Similarities are seen in the rare L*er* allele of QTL3. The QTL3 involved in flowering variation has already been detected in previous studies (El-Lithy et al. 2006, Tisné et al. 2010, Mendez-Vigo et al. 2010). It could be only detected if L*er* was a parent in the developed RIL populations, suggesting that L*er* carries a rare gain of function allele of the gene underlying QTL3. A very recent study of Salomé et al. in 2011 could not detect the QTL by analysing eighteen distinct accessions, because they did not include L*er* in their analysis. "Based on their presence in a single wild genotype and their relative large effects, some of the *A. thaliana* natural alleles are probably deleterious variants segregating in natural populations" (Alonso-Blanco et al. 2009). This could be also the case for the L*er* allele at QTL3. Candidate genes underlying QTL3 will be discussed in detail in the paragraph IV.4.A.

IV.3.C. Anthocyanin content QTL on chromosome I

Quantification of anthocyanin levels indicated a confined distribution in the CT environment with low levels of anthocyanins. Consequently, no QTL were detected under this condition for this trait. However, a vast distribution of anthocyanin levels was observed in the high light and cold environment and therefore QTL were detected. These QTL highlight the GxE interaction observed for this trait because they were detected only in the HL4 environment. Genes already known to be involved in the regulation of anthocyanin production are present close to the position of the QTL detected for anthocyanin content on chromosome I (closest marker NF19K23 - 22.9 Mbp). Between marker positions T27K12 (16.1 Mbp) and NF19K23 on chromosome I, *PAP1*, a positive regulator of anthocyanin pigment production (Borevitz et al. 2000) is present. *PAP1* encodes a R2R3 MYB protein. *PAP1* together with a bHLH protein (*GL3/EGL3/TT8*) and a WD40 protein (*TTG1*) acts in the MYB-bHLH-WD40 (MBW) complex that controls late genes of the anthocyanin biosynthesis pathway (Gonzalez et al. 2008, reviewed Hichri et al. 2011). The MBW complex binds the promoter of one of the late key genes (*DFR)* to start the biosynthesis (Zimmermann et al. 2004). The control of anthocyanin biosynthesis in *A. thaliana* is explained in detail in the Introduction. Up to date, only two studies analysing natural variation in anthocyanin accumulation with different *A. thaliana*

accessions have been published (Teng et al. 2005, Diaz et al. 2006). Both groups analysed the anthocyanin accumulation in *A. thaliana* induced by different stress factors. Teng et al. (2005) were growing plants with additional sucrose (Suc) in the medium. High energy levels lead to oxidative stress which is reduced by the production of anthocyanins. Another stress factor as demonstrated by Diaz et al. (2006) is the limitation of nitrogen in the soil. Anthocyanin production is generally induced by several abiotic and biotic factors (summarized in detail in the Introduction). In both cases, Diaz and colleagues detected QTL on chromosome I collocating with the one detected in this study for high light and cold temperature induced anthocyanin production. Underlying the QTL in their studies was the gene *PAP1*, thus being a likely candidate gene in our analysis as well. The L*er* allele at the QTL on chromosome I decreased the anthocyanin levels compared to the Eri-1 allele at this position. Sequence comparison between L*er* and Eri-1 for *PAP1* revealed that L*er* carries several deletions in the promoter, first intron, second exon and 3´UTR region. The first intron as well as the start of the second exon are deleted in L*er* leading to a weak or loss of function allele of *PAP1* in L*er* compared to Eri-1. The latter case is unlikely, because expression analysis of *PAP1* demonstrated that the gene is expressed in L*er* (see Figure III.25), but at much lower levels than $PAP1^{Eri-1}$. A loss of function $PAP1^{Ler}$ would be also in contrast with the results of Teng et al. (2005). In their study $PAP1^{Ler}$ was responsible for the Suc-induced anthocyanin production, whereas $PAP1^{Cvi}$ and $PAP1^{C24}$ resulted in loss of activity proteins due to mutations inside and downstream of the DNA binding domain of the *PAP1* protein. Combining these results and the results of Diaz et al. - using a Bay-0 x Sha population - give evidence that there is natural variation in *PAP1* responsible for part of variation in anthocyanin accumulation in nature.

IV.3.D. Anthocyanin content, chlorophyll content and flowering related QTL on chromosome V

The co-localisation of QTL for many different traits has two different conceivable reasons: 1) Pleiotropy: One gene underlies the QTL and is involved in different processes or 2) Linkage: there are two closely linked genes in the QTL region which separately act on the different traits. On chromosome V at marker position CIW8 the second and major QTL for anthocyanin content was detected in the high light and cold temperature environment. The

gene known to be involved in the production of anthocyanin located in the vicinity of the QTL5-1 is part of the TT (Transparent Testa) mutant phenotype, taking part in controlling the flavonoid biosynthesis pathway of *A. thaliana*. The *TTG1* locus encodes the WD40 protein (Walker et al. 1999) which is part of the above mentioned MBW complex and is required for the expression of late genes in the anthocyanin biosynthesis pathway (Zhang et al. 2003). It is also involved in trichome and root hair development (Bouyer et al. 2008). By sequence and expression analysis we excluded *TTG1* as a candidate gene for the QTL5-1 in accumulation of anthocyanin content. In addition, the sequence of $TTG1^{Ler}$ and $TTG1^{Eri\text{-}1}$ shows high similarities except for two SNPs in the first and second intron. Both accessions have sequence deletions nearby the promoter region which could lead to the low and almost equal expression of *TTG1* in L*er* and Eri-1. Keurentjes et al. (2007) and West et al. (2006) have stated that insertion or deletions (indels) close to promoter regions can affect gene expression. However, we assume *TTG1* to be functional in both accessions but carrying a weaker allele in both cases. *TTG1* is required as a stabilizing factor of the MBW complex (Baudry et al. 2004, Zhao et al. 2008). Until now, no other WD40 protein is known to be involved in the regulation of anthocyanin production in *A. thaliana*. Loss of function alleles of *TTG1* would probably result in almost no anthocyanin production, which is not the case in both accessions. Beside *TTG1* no other obvious candidate gene is present underlying the anthocyanin content QTL5-1. However, the possibility that linked genes (*HUA2* and another gene in the region of QTL5-1) are responsible for the variation seen in flowering time related traits and anthocyanin accumulation in L*er*/Eri-1 can not be excluded.

In this work we present that *HUA2* is the gene underlying QTL5-1 for the strongest flowering time effects in the L*er*/Eri-1 population. The pleiotrophy of *HUA2* involved in flowering time variation and anthocyanin content was tested by phenotyping anthocyanin accumulation among different L*er* strains carrying different alleles at *HUA2* ($hua2\text{-}5^{Ler}$ and $HUA2^{Ler}$). Phenotyping for anthocyanin accumulation showed increased levels in the strain carrying $HUA2^{Ler}$ allele compared to $hua2\text{-}5^{Ler}$, identifying *HUA2* to be involved in anthocyanin production of *A. thaliana*. Taken together, *HUA2* shows pleiotrophic function in delaying the flowering time and increasing anthocyanin content. Pleiotrophic effects of *HUA2* have previously been shown by Wang et al. (2007) where a special allele of *HUA2* ($HUA2^{Sy\text{-}0}$) affects multiple components of the plant life history strategy such as flowering time and plant morphology. This result and also a recent paper of Gou et al. (2011)

highlight the potential of multifunctional genes contributing to phenotypic novelty within species. In the publication they also presented the pleiotrophic effects of a transcription factor known to have an important role in flowering time and newly described in playing a role in negative regulation of anthocyanin accumulation. This is the first publication which illustrates a direct link between flowering and anthocyanin production and gives evidence that flowering time and anthocyanin production measured in our study may result from the activity of the same gene previously known to be involved in flowering time. The model explaining the involvement of *HUA2* in the regulation of anthocyanin production will be discussed in paragraph IV.6.

IV.4. Fine mapping of QTL3 and fine mapping and allelism test of QTL5-1

IV.4.A. Fine mapping of flowering QTL3 and sequence analysis of sequenced Eri-1 accession revealed candidate genes for upper QTL3-1 and lower QTL3-2

QTL regions usually contain several hundred genes. Therefore, after validation of QTL with NILs or HIFs, the region needs to be narrowed down. Fine mapping is a useful procedure to discover which gene(s) is/are underlying the effect of the QTL. QTL3 was narrowed down to a region of approximately 900 kb, from 0.014 Mbp to 0.92 Mbp. Fine mapping of the region with recombinant plants revealed the presence of two additive QTLs in the region of QTL3. QTL positions typically have large confidence intervals leading to one QTL being masked by another, when two closely linked loci are detected as a single QTL (Balasubramanian et al. 2009). This was the first reason why we did not detect both QTL separately during QTL detection. The second reason was the presence of only one anchored microsatellite marker (F22F7 1.6 Mbp) at the top of chromosome III. This marker was the last marker available at the top of chromosome III to genotype the RILs. Thus, the genotype above this marker was not known, but the small region was expected to be linked to the genotype at F22F7. The upper QTL3 (QTL3-1) is located between the markers T4P13 (0.014 Mbp) and F4P13 (0.20 Mbp), the lower QTL3 (QTL3-2) is located between the markers F14P3 (0.42 Mbp) and 3-926376 (0.92 Mbp). Interestingly, El-Lithy et al. (2006) detected a QTL (located on chromosome III, SNP105, 0.05 Mbp) for flowering time in the L*er* x An-1 RIL population at the same position we detected QTL3-1. In the L*er* x An-1 population the L*er* allele was responsible for the late flowering time

phenotype. In the same population Tisné et al. (2010) detected a QTL for rosette area at this position and an additional for leaf number at the marker NGA172 (0.79 Mbp). The second QTL collocated with QTL3-2 in the region of 0.42 Mbp to 0.92 Mbp. El-Lithy et al. also described QTL3 in 2006 (full region with both QTLs in our case) in the L*er* x Kas-2 and L*er* x Kond RIL populations. Alonso-Blanco detected a QTL for flowering time in the L*er* x Sha RIL population around the marker NGA172 (0.79 Mbp) (personal communication). The marker NGA172 was also used in the L*er* x Fei-0 population of Mendez-Vigo et al. (2010) and they mapped colocalised FT QTL, TLN QTL as well as rate of leaf production (RLF) QTL at this position. When we performed fine mapping of QTL3, the whole sequence data of Eri-1 was not yet available, resulting in difficulties to find polymorphisms for marker development and further fine mapping. Before the availability of ultra high resolution microarrays and new sequencing methods (Clark et al. 2007, Ossowski et al. 2008), marker analysis and polymorphism discovery were limited and time-consuming procedures. The rare L*er* polymorphism strategy we performed to find SNPs for CAPS marker development serves as an example. From eighteen markers, only five were useful because of the SNPs present between L*er* and Eri-1. The difficulties of developing and testing markers between L*er* and Eri-1 were rapidly overcome by sequencing the Eri-1 accession. Beside whole genome sequencing, a lot of studies have made use of high-density oligo-nucleotide arrays providing an alternative approach for polymorphism detection and the following genotyping with these SNPs (El-Lithy et al. 2006, Simon et al. 2008, O´Neill et al . 2008, Brachi et al. 2010). The advantage of whole genome sequencing is the possibility to locate indels and whole gene deletions in accessions in comparison to the Col-0 reference genome. By visualizing the Illumina sequence reads of Eri-1 aligned to the Col-0 reference in the IGV browser, indels and SNPs were detected in our analysis. The current sequencing of hundreds and more of genomes from *A. thaliana* accessions is under progress (Weigel and Mott 2000 – http://www.1001genomes.org) and represents a suitable dataset for genome wide annotation analysis of potential functional polymorphisms, apart from direct analysis of QTL candidates. We used the sequence data of L*er* (available online at http://www.1001genomes.org) and our sequenced Eri-1 data to compare both QTL3 regions in relation to Col-0. The whole QTL3 region contains 373 genes. Of these only four genes are known to be involved in flowering. One of these four collocated with QTL3-1, whereas the other three collocated with QTL3-2.

COL2 locates (0.42 Mbp) in the vicinity of QTL3-2. It is homologous to *CO* and encodes a zinc finger protein (Ledger et al. 1996, 2001). Analysing the sequence data of $COL2^{Eri\text{-}1}$ and $COL2^{Ler}$, no sequence polymorphisms or indels were detected between the two accessions. This lack of differences suggests that *COL2* is not the gene responsible for the flowering variation QTL3-2.

Another candidate is *MBD9* (0.17 Mbp; Peng et al. 2006), being the gene underlying QTL3-1. Comparison of the sequence data resulted in only one SNP difference between L*er* and Eri-1. The SNP is located in the first intron of the gene and probably does not change protein structure. However, the deletion in the first exon of $MBD9^{Eri\text{-}1}$ is a likely event to cause protein structure change. Regions in the first exon do not show any aligned sequence reads to the reference genome Col-0. The gaps in the sequences can either be explained by a deletion of this region in $MBD9^{Eri\text{-}1}$ or by inaccurate sequencing of this part. The second reason is unlikely, because Eri-1 sequencing results framing the gap at 173.396 bp – 173.779 bp region show accurate results with read depth of 25-28. The putative deletion in the first exon leads to a disruption of the gene and possibly to a loss of function allele of $MBD9^{Eri\text{-}1}$. Loss of function mutations are usually caused by insertions and deletions in the coding region of genes (Alonso-Blanco et al. 2005). It has been demonstrated that *FLC* expression is subject to epigenetic regulation through covalent modification of *FLC* chromatin (He and Amasino et al. 2005). *MBD9* has been shown to control and directly enhance the expression of the flowering repressor *FLC* via acetylation of histone H3 and H4 of *FLC* chromatin (Peng et al. 2006). Multiple chromatin associated domains are present in MBD9. These domains (two PHD motifs- Cys4-His-Cys3 zinc finger motifs, a BROMO domain and phenylalanine-tyrosine rich domains) are typically found in most chromatin associated proteins and corroborate the regulatory gene expression function of MBD9. *Mbd9* mutant plants flower early due to the reduction of *FLC* expression. Indirect control of gene expression through *MBD9* has been predicted by *FLC* DNA methylation (Yaish et al. 2009). They have shown that inactivation of *MBD9* causes DNA hypermethylation. Recurring to the three gene interaction hypothesis already mentioned in paragraph IV.III.C, a combination of $FLC^{Eri\text{-}1}$, $HUA2^{Eri\text{-}1}$ and L*er* alleles at QTL3 (Eri-NIL67-4-2) delayed flowering significantly, indicated by a TLN of 50-60. We propose *MBD9* as a proper candidate gene for QTL3-1. Both $HUA2^{Eri\text{-}1}$ and $MBD9^{Ler}$ promote the expression of $FLC^{Eri\text{-}1}$ leading to a much later flowering phenotype than seen in Eri-1. In Eri-1 only $HUA2^{Eri\text{-}1}$ is a functional promoter of *FLC* expression. The second

indication that *MBD9* is the gene underlying QTL3-1 is given by its pleiotroph function in shoot branching. In *mbd9* mutants axillary buds continuously grew out of rosette leaf axils. The opposite effect was observed in the EriNIL67-4-10-26, showing reduced shoot branching (see Figure IV.1). Peng et al. (2006) have suggested *MBD9* to be involved in a novel pathway of shoot branching, the branching phenotype not being related to auxin levels and MAX pathway (McSteen and Leyser 2005).

The third candidate flowering gene is *AGL4(SEP2)* (Pelaz et al. 2000). *AGL4* encodes a MADS box transcription factor involved in floral organ identity variation, causing variation of flower shape. The gene is located in the QTL3-2 region at position 0.46 Mbp. Sequence comparison for L*er* and Eri-1 revealed two big deletions in the putative promoter region of L*er*. Indels in and nearby promoter regions commonly result in change of functional alleles (Keurentjes et al. 2007, West et al. 2006). The L*er* allele at QTL3-2 delayed flowering. Deletions in the promoter region of $AGL4^{Ler}$ may result in a weak allele of *AGL4* with low expression being involved in a later flowering phenotype due to changes in flower morphogenesis. It is worth noting that beside the branching phenotype also the inflorescence phenotype of Eri-NIL67-4-10-26 differs in form (see Figure IV.1).

The last gene related to flowering in the QTL3 region is *VGT1* (Aluri and Büttner 2007). It encodes a vacuolar membrane-localised glucose transporter that can also transport fructose. Mutations in this gene have effects on seed germination and time to flowering. Sequence comparison between L*er* and Eri-1 revealed only identical SNPs between both accessions in relation to the reference genome Col-0. Both accessions also carried a deletion spanning parts of the third intron and fourth exon. The sequence similarities between $VGT1^{Ler}$ and $VGT1^{Eri\text{-}1}$ demonstrate that this gene is probably not underlying QTL3-2.

In summary, by sequencing the Eri-1 accession and comparing the visualized sequence reads between L*er* and Eri-1 we were able to highlight two candidate genes for QTL3-1 and QTL3-2 which gene function was possibly changed by deletions. To verify that gaps in the sequencing represent deletions and are not caused by low quality sequencing results, primer will be developed for testing amplification of the putative deletion region. If the deletion is verified either further fine mapping or direct genomic complementation of the candidates will be performed. Because only two candidates were identified in the region, the latter approach is to be preferred.

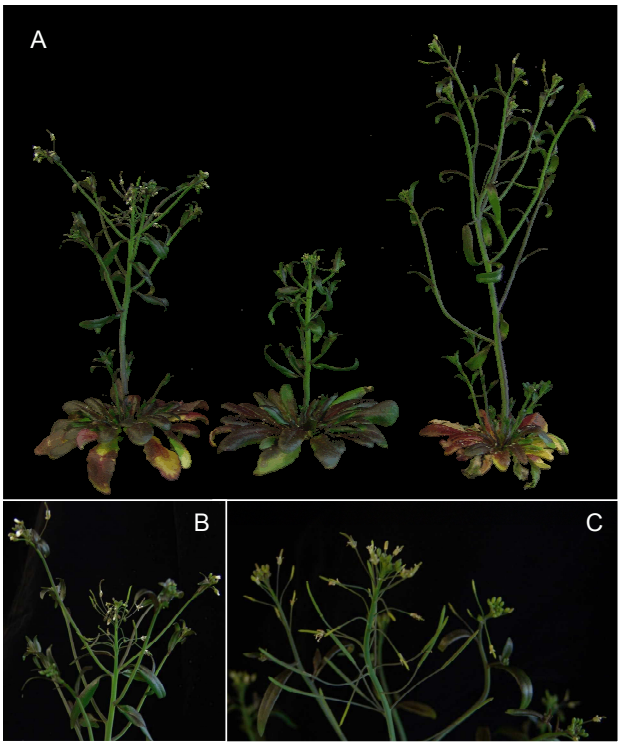


Figure IV.1: Phenotype of Eri NIL QTL3 – Eri NIL 67-4-10-26 grown in HL4 condition.

A) Eri NIL 67-4-10-26 plants at different time points of their development
B) inflorescence of Eri NIL 67-4-10-26

IV.4.B. Fine mapping of QTL5-1 and allelism test demonstrated that *HUA2* is the gene responsible for flowering time QTL5-1 and strongly supported that *HUA2* is also responsible for anthocyanin content variation in L*er*/Eri-1 population

Fine mapping of QTL5-1 resulted in a narrowed down region of 0.03 Mbp with 224 genes. *HUA2* was proposed as the best candidate gene in the region due to the presence of *hua2-5* mutant allele in L*er*. An allelism test using L*er*-*hua2-5* lines, L*er*-*HUA2* lines and L*er*NIL QTL5 lines was carried out in order to complement $hua2\text{-}5^{Ler}$ allele either with $HUA2^{Ler}$ (L*er*-*HUA2* strain) and/or with $HUA2^{Eri\text{-}1}$ (LerNILs QTL5). This experiment confirmed that the mutant phenotype was complemented with *HUA2* alleles and therefore, we concluded that *HUA2* is the gene explaining most of the flowering time variation at QTL5-1 in the L*er*/Eri-1 population. To date, no other accessions but L*er* have been found to contain the *hua2-5* lesion. However natural variation at different sites within the *HUA2* gene has been published (Wang et al. 2007). In this study, authors used the accession Sy-0 (from Isle of Skye in Great Britain), showing beside a very late flowering phenotype development of aerial rosettes in leaf axils. A non-synonymous SNP change in the third

exon of *HUA2* was shown to be responsible for this distinct phenotype and have not been detected in other accessions than Sy-0. $HUA2^{Sy\text{-}0}$ is a gain of function allele regarding its ability to activate *FLC* expression. Interestingly, $HUA2^{Col\text{-}0}$ and $HUA2^{Ler}$ conferred the same *FLC* expression as $HUA2^{Sy\text{-}0}$ suggesting that Sy-0 late flowering phenotype arises because of *FLC* up-regulation rather than from expansion of *FLC* expression.

In L*er*/Eri-1 population pleiotrophic effects of *HUA2* were also observed. We did show that the anthocyanin accumulation was significantly increased between L*er* carrying *hua2-5* and L*er* carrying *HUA2* under HL4 condition. The same increase was observed between L*er-hua2-5* plants and LerNILsQTL5 which carried L*er* alleles in the genomic background and $HUA2^{Eri\text{-}1}$. To verify that *HUA2* is involved in the anthocyanin accumulation observed in the L*er*/Eri population, we phenotyped the same plants used in the allelism test for anthocyanin content. Interestingly, $HUA2^{Ler}$ complemented $hua2\text{-}5^{Ler}$ and increased the anthocyanin content only when the L*er-hua2-5* mutant line was used as the mother plant for the crosses. The reciprocal crosses did not show higher anthocyanin levels after complementation, suggesting that a maternal effect influences anthocyanin accumulation. The maternal effect is not in contradiction with the QTL detection, because in this case L*er* (carrying *hua2-5* allele) was also used as the mother plant to build the RIL population between L*er* and Eri-1. However, the differences are not significant and only a "tendency" that *HUA2* is increasing the anthocyanin content in the allelism test is observed. Quantification of anthocyanin content is subject to variations at plant level but also at technical level. In order to have a precise estimation of this trait, a larger number of plants and replicates need to be measured. Beside the applied HL4 condition, several other abiotic and biotic factors can induce higher anthocyanin levels (Buer et al. 2010) in the climate chambers. Unequal fertilizing and watering, as well as herbivore attacks are imaginable as the reason for variability between plants of the same line. However, it is possible to circumvent this situation and acquire data showing significant differences in anthocyanin levels between the *hua2-5* and *HUA2* lines by screening large numbers of replicates. In the allelism test for anthocyanin content only 2-7 F2 plants per line were analysed. This number is too low to draw firm conclusion but nevertheless, a clear tendency that *HUA2* increased the anthocyanin levels between the lines was observed. There are more replicates available to verify or negate this tendency. Analysis of these replicates is already projected. In addition, phenotyping of L*er-hua2-5*, L*er-HUA2* and the LerNILsQTL5 for anthocyanin content, as well as the increased expression of regulatory anthocyanin genes in L*er-HUA2*

lines provided evidence that *HUA2* is also involved in anthocyanin regulation. The possible mechanism linking *HUA2* to the anthocyanin regulation pathway is proposed in paragraph IV.6. An up-regulation of genes involved in the regulation of anthocyanin production appears to be plausible. This up-regulation might be even increased in $HUA2^{Eri\ -1}$ leading to the comparably high accumulation in Eri-1 plants. $HUA2^{Eri-1}$ carries two non-synonymous SNP changes compared to L*er* and Col-0, leading to two amino acid changes in the protein. These protein changes might lead to a gain of function of *HUA2* in Eri-1 causing an increased accumulation of anthocyanins in HL4 conditions. A likewise situation has been observed in Sy-0 accession with one amino acid change leading to an up-regulation of *FLC* and the very late phenotype with aerial rosettes as explained above.

IV.5. Two-way QTL interactions mapping presents genetic networks in flowering time and anthocyanin content in L*er*/Eri-1 population

IV.5.A. Interactions involved in flowering time variation in L*er*/Eri-1 population

Variation in flowering time can not be fully explained by the major QTL on chromosome V, therefore genetic interaction analysis was performed. Two-way marker interactions suggested that the QTL5-1 region, for which *HUA2* was identified as the gene underlying the main flowering variation, is interacting with four other regions on chromosome V. The main interaction was found between *HUA2* and the QTL on top of chromosome V, where *FLC* is probably supporting the effect of the QTL. As previously mentioned, in case of a functional *FRI* allele, up-regulation of *FLC* expression is mainly supported by *FRI* in order to delay flowering (vernalization pathway). *FRI* is not functional in L*er* neither in Eri-1. In this situation *HUA2* is the main gene enhancing the expression of *FLC*. However, the overall effect of *HUA2* on flowering time could not be entirely explained through its interaction with *FLC*. *HUA2* must also interact with other components of flowering time regulation (Doyle et al. 2005). Another genetic interaction was detected between *HUA2* and the QTL on the bottom of chromosome V. The QTL at the bottom of chromosome V is collocating with the *MAF2-MAF5* cluster. Polymorphisms at the *HUA2* and *MAF2* genes between L*er* and Eri-1 can lead to an epistatic interaction. For the other two QTLs on chromosome V interacting with *HUA2* no obvious interacting partners can be proposed. *HUA2* region interacted genetically with regions from the other 4

chromosomes. An interaction appeared on chromosome IV at positions around 13.3 Mbp. This epistatic interaction between *HUA2* and this region could be explained with the presence of *AG* (10.4 Mbp) in that region. *HUA2* gene was first characterized in a screen for enhancers of *ag4*, a weak mutant allele of *AG* (Chen and Meyerowitz 1999). *HUA2* appears to play a role in processing *AG* pro-mRNA in the nucleus (Cheng et al. 2003). *AG* has several similarities with *FLC*. Both genes encode MADS-box transcription factors, (like *MAF2* and *FLM*) and contain a large intron. *HUA2* may also affect splicing of *FLC* and *MAF2*. Another candidate for the genetic interaction in that region could be explained by *HUA2* and *FCA*, which is a gene predicted to bind mRNA (MacKnight et al. 1997). Autonomous pathway genes *FCA* (chromosome IV 9.2 Mbp) and *FY* (chromosome V 4.3 Mbp) interact through *FCA*-WW (highly conserved tryptophans) and *FY*-PPLP (proline riche) domains. *HUA2* has a C-terminal end with 5 PPLP repeats, suggesting that it may interact with proteins containing WW domains like *FCA*. Since *HUA2* and *FCA* contain compatible interaction domains and both of them affect the same downstream gene *FLC*, we suggest that *HUA2* and *FCA* interact physically and are both involved in the mRNA processing of *FLC* or compete for the mRNA processing of *FLC*. Two other MADS Box transcriptions factors, *FLM* (*MAF1*, Werner et al. 2005) and *SVP* (Hartmann et al. 2000) are also likely to be interacting with *HUA2*. They are also MADS box genes and closely related to *FLC* both at the amino acid level and in exon/intron gene structure (Ratcliffe et al .2001, Scortecci et al. 2001). Both are also negative regulators in floral transition and *HUA2* affects the expression of both. Interestingly, *FLM* is located on chromosome I (28.9 Mbp). An interaction with *HUA2* was also mapped on chromosome I but above *FLM* at around 22.9 Mbp. Interaction between *HUA2* and *SVP* (chromosome II, 9.5 Mbp) can be hypothesized in the L*er*/Eri-1 population, because an epistatic interaction between chromosome II (7.3 Mbp) and *HUA2* QTL5-1 region was detected. The interaction found between *HUA2* and chromosome III could involve one of the candidate genes introduced in paragraph IV.4.A.

Overall, the 2-way marker interaction analysis revealed that the QTL5-1 for which *HUA2* has been identified, is interacting with a lot of chromosomal regions. In this interacting region, already described genes involved in flowering time are present. This analysis allowed us to propose a set of genes which are probably interacting with *HUA2* in order to produce the flowering time variation observed in the L*er*/Eri-1 population.

IV.5.B. Interactions involved in anthocyanin accumulation in L*er*/Eri-1 population

During QTL detection only one major QTL responsible for the variation of anthocyanin accumulation was detected under high light and cold temperature condition. This QTL did not explain all the phenotypic variation which was found in the population. An additional analysis of 2-way marker interaction revealed a couple of interactions across the whole *A. thaliana* genome but most of them mapped on chromosome V. The biosynthesis of anthocyanins is well studied and published (Stafford 1990, 1991, explained in introduction chapter). Anthocyanins are derived from a branch of the flavonoid pathway. Interestingly most of the structural genes of this pathway (see Figure I.3 in introduction) are located on chromosome V, where we found most of the epistatic interactions. Several transcription factors and complexes are regulatory parts of the anthocyanin biosynthesis and initiate the synthesis pathway. The main QTL5-1 for anthocyanin accumulation was detected on the chromosome V at marker CIW8 (7.5 Mbp). Phenotyping of L*er-hua2-5*, L*er-HUA2* and the LerNILsQTL5 for anthocyanin content, as well as the increased expression of regulatory anthocyanin genes in L*er-HUA2* lines suggested that *HUA2* is involved in anthocyanin regulation and is the gene underlying the QTL5-1 for anthocyanin content. For anthocyanin level, interactions with QTL5-1 and other regions on the chromosome V have been revealed. Additionally, epistatic interactions between QTL5-1 and two regions on chromosome II, one on chromosome III and one on chromosome IV have been detected. In each interacting region a gene involved in anthocyanin accumulation (either a structural gene of the anthocyanin biosynthesis or a gene encoding a transcription factor which has been shown to regulate the biosynthesis) is present.

One interaction was detected with a locus above the QTL5-1 at around 5.6 Mbp. At this locus the mutant *ANTHOCYANINLESS1* (*anl1*) has been described (Kubo et al. 2007). *anl1* encodes for a 3-0-glycosyltransferase which specially glucosylates the third position of the flavonoid C-Ring. It is the fifth step within the biosynthesis converting anthocyanidin to anthocyanin. The *TT19* locus is also located very close to this position (5.68 Mbp). *TT19* encodes the *Arabidopsis thaliana* glutathione transferase. Hsieh et al. 2007 have described green coloured *tt19* mutants. These mutants were able to accumulate flavonoids in the cytosol, but transport them into the vacuole was impaired, where the low pH affects flavonoid colouring. Both modificational steps, the glycosylation and the transport into the vacuole are necessary to produce coloured anthocyanins changing them

from colourless to violet. Both analysed accessions L*er* and Eri-1 accumulate anthocyanins in HL4 conditions, but in very different levels. This fact points out that the genes in the interacting regions, as well as the QTL candidates are most probably involved in the regulation of anthocyanin rather than in their biosynthesis. We already suggested *PAP1* and *HUA2* as the genes underlying the anthocyanin QTL1 and QTL5-1, respectively. *TTG1* is located in the region of QTL5-1. We already discussed the possibility of *TTG1* being a candidate gene for anthocyanin accumulation under QTL5-1 (see paragraph IV.3.D). *TTG1* encodes a WD40 protein, acting together with a MYB factor and bHLH (basic helix loop helix) protein in a complex to control the anthocyanin biosynthesis (Zhang et al. 2003, Baudry et al. 2004, Zimmermann et al. 2004, Gonzalez et al. 2008). The encoding genes whose products are necessary to build the regulatory complex of the MYB-bHLH-WD40 (MBW) are also candidates for the detected 2-way interactions for anthocyanin production: One main interaction was detected between QTL5-1 and the *TT2* locus on chromosome V, which is located at the marker position MOK9 (13.8 Mbp). *TT2* encodes a R2R3 MYB domain, which has been described to be part of the MBW transcription factor complex (Gonzalez et al. 2009). The complex can be formed by different redundant MYB and bHLH factors and the *TTG1* WD40 protein (reviewed by Hichri 2011). Other R2R3 MYB proteins are for example *PAP1 or PAP2 or GL1,* located on chromosome I and chromosome III. We already suggested *PAP1* as the gene underlying the QTL1 for variation in anthocyanin levels due to a putative weak $PAP1^{Ler}$ allele. No interaction was detected between QTL5-1 and chromosome I. The third part of the transcriptional complex is the bHLH domain. Another interaction between QTL5-1 region and marker region M5-9 (17.3 Mbp) was found. In this region the *GL3* locus is located. The *GL3* locus encodes for a bHLH domain protein. Other bHLH factors are *TT8*, located at chromosome IV 6.1Mbp and *EGL1/3*, located at chromosome I. QTL5-1 is interacting with the chromosome IV but below *TT8* at around 13 Mbp. The MBW complex binds to the promoter of late anthocyanin biosynthesis genes like *DFR* (dihydro-flavanol 4 reductase) via the MYB factor. The *TT3* locus encodes the *DFR* which catalyzes the conversion of dihydroquercitin to the colourless leucocyanidin. It is the third step in the anthocyanin biosynthesis (see Introduction). The phenotype of described *tt3* mutants showed an absent of anthocyanins in leaves, stems and all other tissues. The *TT3/DFR* locus is located near the *GL3* locus in the M5-9 region which is also interacting with QTL5-1. For the interaction on the very bottom of chromosome V at the marker position MBK5 (25.5 Mbp) no putative candidate gene

was found. The main interacting region of QTL5-1 showed interactional effects to the chromosome III (NT204-5.6 Mbp, MQC12-7.1 Mbp) and a 3-way interaction with the chromosome II (PLS7- 9.8 Mbp). Below the region on chromosome II, a negative regulator of anthocyanin regulation is located. The *CPC* gene (Zhu et al. 2009) belongs to the R3 MYB genes, which are lacking the R2 domain and thus are not able to bind to DNA. The *CPC* protein negatively controls anthocyanin production by binding the bHLH-WD complex and thus impedes binding to the promoter of *DFR*. In the vicinity of NT204 and MQC12 on chromosome III no obvious candidate was identified.

In summary, for development of the MBW complex necessary to start anthocyanin production, several genes encoding for proteins that act together are required. The position of these interacting genes overlaps in most cases with genetic interactions detected for anthocyanin accumulation in the L*er*/Eri-1 population under HL4 condition. Whether the genes involved in the formation of the MBW complex are the ones underlying the detected interaction is still elusive. In addition, detected interactions showed that QTL5-1 for which we propose *HUA2* is playing a central role in these interactions. A model showing how *HUA2* could act in the anthocyanin content variation in L*er*/Eri-1 will be proposed in the next paragraph.

IV.6. Proposed function of *HUA2* in the regulation of anthocyanin production via MBW complex

Sequence analysis of the L*er* and Eri-1 genome by IGV visualization revealed deletions in promoter regions and different SNPs in some genes encoding parts of the MBW transcription factor complex. The *TTG1* allele shows a deletion in the promoter region of both L*er* and Eri-1 when compared to Col-0. This leads to a lower amount of transcript in these accessions. However, the expression is still detectable in both L*er* and Eri-1, indicating that the protein is still produced with these weak alleles of *TTG1*. One of the MYB factors of the MBW complex responsible for binding the promoter of *DFR* is *PAP1*. The $PAP1^{Ler}$ seems to be either a loss of function allele or a very weak allele, due to the fact that most of the gene is deleted. There are two deletions in the promoter region as well as two deletions (in the first intron and parts of the second exon, and in the second intron). Low expression of $PAP1^{Ler}$ suggests the latter possibility that some parts of the gene are

still functional. *PAP1*$^{Eri-1}$ is fully functional and shows a high expression via qRT-PCR. We observed an interesting situation in the L*er* strain carrying a functional *HUA2* allele. *PAP1* expression is increased compared to L*er-hua2-5* and reaches intermediate *PAP1* transcript levels (between L*er-hua2-5* and Eri-1), indicating that *HUA2* enhances the expression of the weak *PAP1*^{Ler}. As previously explained, *HUA2* has been shown to be involved in the regulation of nuclear pre-mRNA. Firstly, because of the presence of a RPR domain characteristic for genes involved in splicing and secondly, due to the presence of nuclear localisation sequences targeting *HUA2* to the nucleus e.g. in order to process *PAP1* mRNA. *PAP2,* a second MYB factor which could take over the decreased function of *PAP1*^{Ler} for combining a functional MBW complex, showed almost no expression in *PAP2*^{Ler}. Sequence comparison revealed one SNP in the coding region of *PAP2*^{Ler} leading to an amino acid change from glycine to glutamate in L*er*. One non-synonymous SNP is not the reason in most cases for changes of expression. However, Loudet et al. 2007 have shown that one single amino acid substitution in an enzyme of the assimilatory sulphate reduction pathway was responsible for decreased activity. *PAP2*^{Ler} could also be decreased in activity. Similar to *PAP1*^{Ler} the expression of *PAP2*^{Ler} was enhanced in the presence of an active *HUA2*.

The bHLH domain of the MBW complex is responsible for the activation of expression of a target gene. *TT8* bHLH sequence carries deletions in the putative promoter region and in the big 5th intron of both L*er* and Eri-1. Additional Eri-1 carries a 2140bp deletion which removes most of the 7th last exon and the full 3´UTR apart from the C-terminus of *TT8*$^{Eri-1}$. Rather than to see a decrease of *TT8*$^{Eri-1}$ expression in Eri-1 parental line, transcript levels are still much higher than in L*er-hua2-5* parent. The interacting domains for MYB factors and WD40 protein are not affected by the deletions in Eri-1. However most of the C-terminal regions are deleted. This region is known to participate in homodimer or heterodimer formation (Zhang et al. 2003, Pattanaik et al. 2008). Some of the bHLH factors with homo-or heterodimer formation can bind the G-Box of the target gene on their own without building a dimer with a MYB factor. It is depending on the transcription factor itself and the target gene. In Arabidopsis the complex of *PAP1/PAP2/TT8* is necessary to bind the promoter of *DFR* (reviewed Hichri et al. 2011) therefore lacking the C-terminus of *TT8* is probably not much corruptive in terms of function of the *PAP1/PAP2/TT8* complex. *TT8* expression was also enhanced in L*er-HUA2* strain compared to L*er-hua2-5*. *EGL3* is another bHLH factor which could complete the MBW

complex. The expression of *EGL3*^{Ler} could not be enhanced with the presence of *HUA2*. Besides governing the expression of anthocyanin structural key genes, the members of the MBW complex also regulate their own expression in a complex circuit. *TT8* interacts with *TTG1* and *PAP1* to regulate its own transcription (Tohge et al. 2005, Baudry et al. 2006) meaning that, in the presence of a fully functional MBW complex like in Eri-1, the expression of all members will be increased automatically. Down-regulation of the complex could be the consequence of the presence of functional *CPC,* due to its binding to the bHLH factor. Since *CPC* is an R3 MYB protein and does not bind DNA, it avoids binding of the MBW complex to the *DFR* promoter (Zhu et al. 2009). In the Eri-1 accession the down regulation by *CPC* is blocked, because almost full *CPC* genomic region is deleted in Eri-1. The higher expression levels of *CPC* in Eri-1 are in contrast with this result. It could be that the gene itself will be expressed, but the R3 binding domain of the protein which connects to TT8 is changed due to deletions and therefore CPC is not able to bind to TT8 and not capable to down-regulate the anthocyanin biosynthesis neither.
Gathering all the above mentioned hypotheses, we provide a model in Figure IV.2 in order to explain the variation of anthocyanin variation in L*er* and Eri-1 accessions as well as in the selected NILs. In the absence of *HUA2* low expression of *PAP1*^{Ler} and *TT8*^{Ler} lead to weak or rather no combining of the members of the MBW transcription factor complex. Without a MBW complex the late anthocyanin biosynthesis genes are not transcribed and thus no anthocyanin will be accumulated in L*er* (Top A). In the presence of *HUA2* the expression of *PAP1*^{Ler} and *TT8*^{Ler} will be enhanced which leads to a binding of TT8 to TTG1 and PAP1. The combined complex binds to the promoter of *DFR* which initiates the second part of anthocyanin biosynthesis. The synthesis stops when CPC binds TT8 and separates the MBW complex from the promoter of *DFR* (this is the case of *Ler-HUA2*; Bottom A). Eri-1 shows very high accumulation of anthocyanins in rosette leaves, due to a full functional *PAP1*$^{Eri-1}$ allele and a strong *TT8*$^{Eri-1}$ allele. The expression of both genes is even enhanced by *HUA2*. In Eri-1, CPC cannot compete with R2R3 MYB PAP1$^{\text{Eri-1}}$ for the binding site of bHLH TT8$^{\text{Eri-1}}$ and thus the MBW complex remains bound to the promoter of *DFR* which promotes continuously biosynthesis of anthocyanins (Top B). The Eri-NIL8-96-27 carries full functional *PAP1*$^{Eri-1}$ and *TT8*^{Eri}, but lacks the enhancer of both genes – *HUA2*. The anthocyanin levels are still high, but do not reach the high levels of Eri-1. All four combinations are hypotheses and need to be analysed with further confirmations.

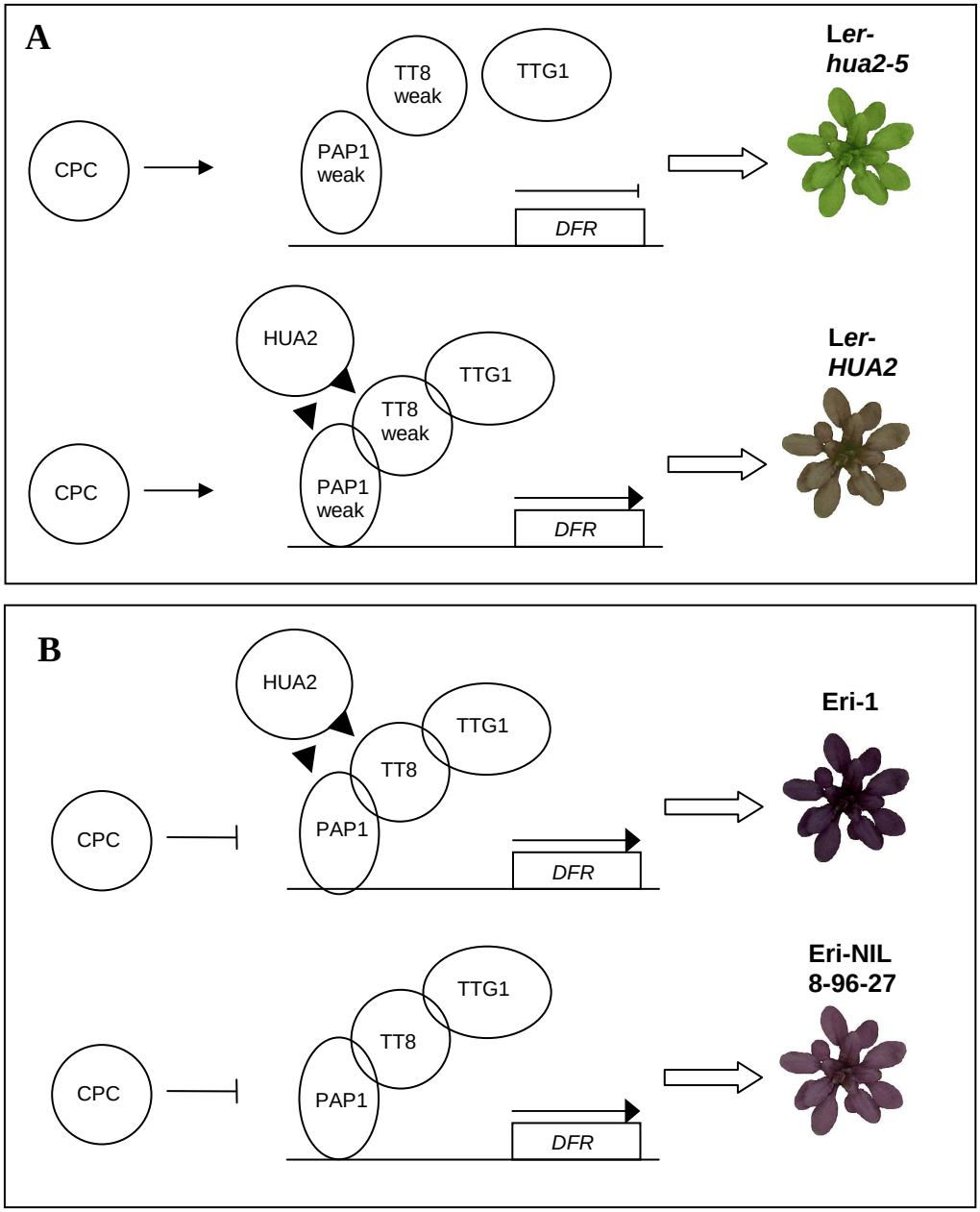


Figure: IV.2: Working model depicting the possible role of HUA2 together with PAP1/TT8/TTG1 complex in transcription activation of a structural anthocyanin biosynthesis gene (*DFR*). Activation would lead to high levels of anthocyanin accumulation in rosette leaves (colours are exaggerated for better distinguish) until down-regulation of CPC occurs. A) Comparison of the procedure between L*er* and L*er-HUA2.* B) Comparison of the procedure in Eri-1 and Eri-NIL8-96-27. Description can be followed in the running text.

References

Adir, N., Zer, H., Shochat, S., and Ohad, I. (2003). Photoinhibition - a historical perspective. *Photosyn. Res*, **76**, 343-370.

Albert, N. W., Lewis, D. H., Zhang, Huaibi, Irving, L. J., Jameson, P. E., and Davies, K. M. (2009). Light-induced vegetative anthocyanin pigmentation in Petunia. *J. Exp. Bot.*, **60**, 2191 -2202.

Alcázar, R., García, A. V., Parker, J. E., and Reymond, M. (2009). Incremental steps toward incompatibility revealed by Arabidopsis epistatic interactions modulating salicylic acid pathway activation. *PNAS.* **106**, 334 -339.

Al-Shebaz and Ókane, (2002). Taxonomy and phylogeny of Arabidopsis (Brassicaceae). *The Arabidopsis Book*.

Alsonso-Blanco C., Koornneef M., and van Ooijen JW. (2006) QTL analysis. *Meth. Mol. Biol.* **323**: 79-99.

Alonso, J. M. and Ecker, J. R. (2006). Moving forward in reverse: genetic technologies to enable genome-wide phenomic screens in Arabidopsis. *Nat. Rev. Genet.*, **7**, 524-536.

Alonso, J. M., Stepanova, A. N., Leisse, T. J., Kim, C. J., Chen, H., Shinn, P., Stevenson, D. K., Zimmerman, J., Barajas, P., Cheuk, R., u. a. (2003). Genome-Wide Insertional Mutagenesis of Arabidopsis thaliana. *Science*, **301**, 653 -657.

Alonso-Blanco, C., Aarts, M. G. M., Bentsink, L., Keurentjes, J. J. B., Reymond, M., Vreugdenhil, D., and Koornneef, M. (2009). What has natural variation taught us about plant development, physiology, and adaptation? *Plant Cell*, **21**, 1877 -1896.

Alonso-Blanco, C., El-Assal, S. E.-D., 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., Gomez-Mena, C., Llorente, F., Koornneef, M., Salinas, Julio, and Martínez-Zapater, J. M. (2005). Genetic and molecular analyses of natural variation indicate CBF2 as a candidate gene for underlying a freezing tolerance Quantitative Trait Locus in Arabidopsis. *Plant Physiol.*, **139**, 1304 -1312.

Alonso-Blanco, C., Koornneef, M., and Stam, P. (1998). The use of Recombinant Inbred Lines (RILs) for genetic mapping. In, Martinez-Zapater, José M. and Salinas, Julio (hrsg), *Arabidopsis Protocols*. Humana Press, Totowa, NJ, p. 137-146.

Alonso-Blanco, C., Mendez-Vigo, B., and Koornneef, M. (2005). From phenotypic to molecular polymorphisms involved in naturally occurring variation of plant development. *Int. J. Dev. Biol*, **49**, 717-732.

Aluri, S. and Büttner, M. (2007). Identification and functional expression of the Arabidopsis thaliana vacuolar glucose transporter 1 and its role in seed germination and flowering. *PNAS.* **104**, 2537 -2542.

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

Apel, K. and Hirt, H. (2004). Reactive oxygen species: metabolism, oxidative stress, and signal transduction. *Annu Rev Plant Biol*, **55**, 373-399.

Aron, P. M. and Kennedy, J. A. (2008). Flavan‐3‐ols: Nature, occurrence and biological activity. *Mol. Nutr. Food Res.*, **52**, 79-104.

Asada, K. (1999). The water-water cycle in chloroplasts: Scavenging of active oxygens and dissipation of excess photons. *Annu. Rev. Plant. Physiol. Plant. Mol. Biol.*, **50**, 601-639.

Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y., Meng, D., Platt, A., Tarone, A. M., Hu, T. T., u. a. (2010). Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. *Nature*, **465**, 627-631.

Balasubramanian, S., Schwartz, C., Singh, A., Warthmann, N., Kim, M. C., Maloof, J. N., Loudet, O., Trainer, G. T., Dabi, T., Borevitz, J. O., u. a. (2009). QTL mapping in new Arabidopsis thaliana advanced intercross-recombinant inbred lines. *PLoS ONE*, **4**, e4318.

Balasubramanian, S., Sureshkumar, S., Agrawal, M., Michael, T. P., Wessinger, C., Maloof, J. N., Clark, R., Warthmann, N., Chory, J., and Weigel, D. (2006). The PHYTOCHROME C photoreceptor gene mediates natural variation in flowering and growth responses of Arabidopsis thaliana. *Nat. Genet.*, **38**, 711-715.

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.

Beck, J. B., Schmuths, S, H., and Schaal, B. A. (2008). Native range genetic variation in Arabidopsis thaliana is strongly geographically structured and reflects Pleistocene glacial dynamics. *Mol. Ecol.*, **17**, 902-915.

Benfey, P. N. and Mitchell-Olds, T. (2008). From genotype to phenotype: Systems biology meets natural variation. *Science*, **320**, 495 -497.

Bennett, M. D., Leitch, I. J., Price, H. J., and Johnston, J. S. (2003). Comparisons with Caenorhabditis (100 Mb) and Drosophila (175 Mb) using flow cytometry show genome size in Arabidopsis to be 157 Mb and thus 25 % larger than the Arabidopsis genome initiative estimate of 125 Mb. *Annals of Botany*, **91**, 547 -557.

Bentsink, L., Hanson, J., Hanhart, C. J., Blankestijn-de Vries, H., Coltrane, C., Keizer, P., El-Lithy, M., Alonso-Blanco, C., de Andrés, M. T., Reymond, M., u. a. (2010). Natural variation for seed dormancy in Arabidopsis is regulated by additive genetic and molecular pathways. *PNAS.* **107**, 4264-4269.

Bentsink, L., Jowett, J., Hanhart, C. J. and Koornneef, M. (2006). Cloning of DOG1, a quantitative trait locus controlling seed dormancy in Arabidopsis. *PNAS*. **103**, 17042-17047.

Bergelson, J. and Roux, F. (2010). Towards identifying genes underlying ecologically relevant traits in Arabidopsis thaliana. *Nat. Rev. Genet*, **11**, 867-879.

Bernhardt, C., Lee, M. M., Gonzalez, A., Zhang, F., Lloyd, A., and Schiefelbein, J. (2003). The bHLH genes GLABRA3 (GL3) and ENHANCER OF GLABRA3 (EGL3) specify epidermal cell fate in the Arabidopsis root. *Development*, **130**, 6431-6439.

Bhattacharya, A., Sood, P., and Citovsky, V. (2010). The roles of plant phenolics in defence and communication during Agrobacterium and Rhizobium infection. *Mol. Plant P*, **11**, 705-719.

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.

Boss, P. K. (2004). Multiple pathways in the decision to flower: enabling, promoting, and resetting. *Plant Cell*, **16**, S18-S31.

Bouyer, D., Geier, F., Kragler, F., Schnittger, Arp, Pesch, M., Wester, K., Balkunde, R., Timmer, J., Fleck, C., and Hülskamp, Martin (2008). Two-dimensional patterning by a trapping/depletion mechanism: The role of TTG1 and GL3 in Arabidopsis trichome formation. *PLoS Biol*, **6**, e141.

Brachi, B., Faure, N., Horton, M., Flahauw, E., Vazquez, A., Nordborg, M., Bergelson, J., Cuguen, Joel, and Roux, F. (2010). Linkage and association mapping of Arabidopsis thaliana flowering time in nature. *PLoS Genet*, **6**, e1000940.

Brock, M. T., Stinchcombe, J. R., and Weinig, C. (2009). Indirect effects of FRIGIDA: floral trait (co)variances are altered by seasonally variable abiotic factors associated with flowering time. *J. Evolution Biol.*, **22**, 1826-1838.

Buer, C. S., Imin, N., and Djordjevic, M. A. (2010). Flavonoids: New roles for old molecules. *J Integr Plant Biol*, **52**, 98-111.

Burow, M., Halkier, B. A., and Kliebenstein, D. J. (2010). Regulatory networks of glucosinolates shape Arabidopsis thaliana fitness. *Curr. Opin. Plant Biol*, **13**, 348-353.

Caicedo, A. L., Richards, C., Ehrenreich, I. M., and Purugganan, M. D. (2009). Complex rearrangements lead to novel chimeric gene fusion polymorphisms at the Arabidopsis thaliana MAF2-5 flowering time gene cluster. *Mol. Biol. Evol.*, **26**, 699 - 711.

Caicedo, A. L., Stinchcombe, J. R., Olsen, K. M., Schmitt, J., and Purugganan, M. D. (2004). Epistatic interaction between Arabidopsis FRI and FLC flowering time genes generates a latitudinal cline in a life history trait. *PNAS.* **101**, 15670 -15675.

Casal, J. J., Fankhauser, C., Coupland, G., and Blázquez, M. A. (2004). Signalling for developmental plasticity. *Trends Plant Sci.*, **9**, 309-314.

Chalker-Scott L. (1999) Environmental significance of anthocyanins in plant stress response. *Photochem. Photobiol***. 70**, 1-9.

Charmet, G. (2000). Power and accuracy of QTL detection: simulation studies of one-QTL models. *Agronomie*, **20**, 15.

Chase, K., Adler, F. R., and Lark, K. G. (1997). Epistat : A computer program for identifying and testing interactions between pairs of quantitative trait loci. *Theor. Appl. Genet.*, **94**, 724-730.

Chen, W.J., Chang, S.H., Hudson, M.E., Kwan, W.K., Li, J., Estes, B., Knoll, D., Shi, L. and Zhu, T. (2005). Contribution of transcriptional regulation to natural variations in Arabidopsis. *Genome Biol.,* **6**, R32.

Chen, X and Meyerowitz, E. M. (1999). HUA1 and HUA2 are two members of the floral homeotic AGAMOUS pathway. *Mol. Cell*, **3**, 349-360.

Cheng, Y., Kato, N., Wang, W., Li, J., and Chen, X. (2003). Two RNA binding Proteins, HEN4 and HUA1, act in the processing of AGAMOUS pre-mRNA in Arabidopsis thaliana. *Dev. Cell*, **4**, 53-66.

Christie, P. J., Alfenito, M. R., and Walbot, V. (1994). Impact of low-temperature stress on general phenylpropanoid and anthocyanin pathways: Enhancement of transcript abundance and anthocyanin pigmentation in maize seedlings. *Planta*, **194**, 541-549.

Churchill, G. A. and Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics*, **138**, 963-971.

Clark, R. M., Schweikert, G., Toomajian, C., Ossowski, S., Zeller, G., Shinn, P., Warthmann, N., Hu, T. T., Fu, G., Hinds, D. A., u. a. (2007). Common sequence polymorphisms shaping genetic diversity in Arabidopsis thaliana. *Science*, **317**, 338-342.

Cominelli, E., Gusmaroli, G., Allegra, D., Galbiati, M., Wade, H. K., Jenkins, G. I., and Tonelli, C. (2008). Expression analysis of anthocyanin regulatory genes in response to different light qualities in Arabidopsis thaliana. *Plant Physiol.*, **165**, 886-894.

Cone, K. C., Cocciolone, S. M., Moehlenkamp, C. A., Weber, T., Drummond, B. J., Tagliani, L. A., Bowen, B. A., and Perrot, G. H. (1993). Role of the regulatory gene pl in the photocontrol of maize anthocyanin pigmentation. *Plant Cell,* **5**, 1807 - 1816.

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.

Darvasi, A. and Soller, M. (1994). Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. *Genetics*, **138**, 1365-1373.

Das, P. K., Geul, B., Choi, S.-B., Yoo, S.-D., and Park, Y.-I. (2011). Photosynthesis-dependent anthocyanin pigmentation in Arabidopsis. *Plant Signal Behav*, **6**. 23-25.

Dellaporta, S. L., Wood, J., and Hicks, J. B. (1983). A plant DNA minipreparation: Version II. *Plant Mol Biol Rep*, **1**, 19-21.

Delseny, M., Han, B., and Hsing, Y. I. (2010). High throughput DNA sequencing: The new sequencing revolution. *Plant Sci.*, **179**, 407-422.

Diaz, C., Saliba-Colombani, V., Loudet, O., Belluomo, P., Moreau, L., Daniel-Vedele, F., Morot-Gaudry, J.-F., and Masclaux-Daubresse, C. (2006). Leaf yellowing and anthocyanin accumulation are two genetically independent strategies in response to nitrogen limitation in Arabidopsis thaliana. *Plant Cell Physiol.*, **47**, 74 -83.

Dixon, R. A. (2005). Engineering of plant natural product pathways. *Curr. Opin. Plant Biol.*, **8**, 329-336.

Dixon, R. A., Xie, D., and Sharma, S. B. (2005). Proanthocyanidins – a final frontier in flavonoid research? *New Phytologist*, **165**, 9-28.

Doyle, M. R., Bizzell, C. M., Keller, M. R., Michaels, S. D., Song, J., Noh, Y.-S., and Amasino, R. M (2005). HUA2 is required for the expression of floral repressors in Arabidopsis thaliana. *Plant J*, **41**, 376-385.

Ehrenreich, I. M., Hanzawa, Y., Chou, L., Roe, J. L., Kover, P. X., and Purugganan, M. D. (2009). Candidate gene association mapping of Arabidopsis flowering time. *Genetics*, **183**, 325-335.

Ehrenreich, I. M. and Purugganan, M. D. (2006). The molecular genetic basis of plant adaptation. *Am. J. Bot.*, **93**, 953 -962.

El-Lithy, M. E., Clerkx, E. J. M., Ruys, G. J., Koornneef, M., and Vreugdenhil, D. (2004). Quantitative trait locus analysis of growth-related traits in a new Arabidopsis recombinant inbred population. *Plant Physiol*, **135**, 444-458.

El-Lithy, M. E., Bentsink, L., Hanhart, C. J., Ruys, G. J., Rovito, D., Broekhof, J. L. M., van der Poel, H. J. A., van Eijk, M. J. T., Vreugdenhil, D., and Koornneef, M. (2006). New Arabidopsis Recombinant Inbred Line populations genotyped using SNPWave and their use for mapping flowering-time quantitative trait loci. *Genetics*, **172**, 1867 -1876.

Fambrini, M., Mariotti, L., Parlanti, S., Picciarelli, P., Salvini, M., Ceccarelli, N., and Pugliesi, C. (2011). The extreme dwarf phenotype of the GA-sensitive mutant of sunflower, dwarf2, is generated by a deletion in the ent-kaurenoic acid oxidase1 (HaKAO1) gene sequence. *Plant Mol. Biol*, **75**, 431-450.

Foyer, C. H. and Noctor, G. (2011). Ascorbate and glutathione: The heart of the redox hub. *Plant Physiol*, **155**, 2-18.

Gazzani, S., Gendall, A. R., Lister, C., and Dean, C. (2003). Analysis of the molecular basis of flowering time variation in Arabidopsis accessions. *Plant Physiol*, **132**, 1107-1114.

Ghandilyan, A., Ilk, N., Hanhart, C., Mbengue, M., Barboza, L., Schat, H., Koornneef, M., El-Lithy, M., Vreugdenhil, D., Reymond, M., u. a. (2009). A strong effect of growth medium and organ type on the identification of QTLs for phytate and mineral concentrations in three Arabidopsis thaliana RIL populations. *J Exp Bot*., **60**, 1409 -1425.

Gonzalez, A., Zhao, M., Leavitt, J. M., and Lloyd, A. M. (2008). Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in Arabidopsis seedlings. *Plant J.*, **53**, 814-827.

Gou, J.-Y., Felippes, F. F., Liu, C.-J., Weigel, D., and Wang, J.-W. (2011). Negative regulation of anthocyanin biosynthesis in Arabidopsis by a miR156-targeted SPL transcription factor. *Plant Cell*. **23**, 1512-1522.

Greenberg, J. T. (1996). Programmed cell death: A way of life for plants. *PNAS.* **93**, 12094-12097.

Grotewold, E. (2006). The genetics and biochemistry of floral pigments. *Annu Rev Plant Biol*, **57**, 761-780.

Harborne, J. B. and Williams, C. A. (2000). Advances in flavonoid research since 1992. *Phytochem.*, **55**, 481-504.

Hartmann, U., Höhmann, S., Nettesheim, K., Wisman, E., Saedler, H., and Huijser, P. (2000). Molecular cloning of SVP: A negative regulator of the floral transition in Arabidopsis. *Plant J.*, **21**, 351-360.

Hatier, J.-H. B. and Gould, K. S. (2009). Anthocyanin Function in Vegetative Organs. In, *Anthocyanins*. Springer New York, New York, NY, 1-19.

He, F., Kang, D., Ren, Y., Qu, L.-J., Zhen, Y., and Gu, H. (2007). Genetic diversity of the natural populations of Arabidopsis thaliana in China. *Heredity*, **99**, 423-431.

He, R., Drury, G. E., Rotari, V. I., Gordon, A., Willer, M., Farzaneh, T., Woltering, E. J., and Gallois, P. (2008). Metacaspase-8 modulates programmed cell death induced by ultraviolet light and H2O2 in Arabidopsis. *J. Biol. Chem*, **283**, 774 -783.

He, Y. and Amasino, R. M. (2005). Role of chromatin modification in flowering-time control. *Trends Plant Sci*, **10**, 30-35.

Hichri, I., Barrieu, F., Bogs, J., Kappel, C., Delrot, S., and Lauvergeat, V. (2011). Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. *J Exp Bot*. **62**, 2465-2483.

Hoch, W. A., Singsaas, E. L., and McCown, B. H. (2003). Resorption protection. anthocyanins facilitate nutrient recovery in autumn by shielding leaves from potentially damaging light levels. *Plant Physiol*, **133**, 1296 -1305.

Hoffmann, M.H., (2002) Biogeographie of Arabidopsis thaliana (L.) Henynh. (Brassicaceae). *J.Biogeogr*. **29**, 125-134.

Holton, T. and Cornish, E. (1995). Genetics and biochemistry of anthocyanin biosynthesis. *Plant Cell*, **7**, 1071-1083.

Hsieh, K. and Huang, A. H. C. (2007). Tapetosomes in Brassica tapetum accumulate endoplasmic reticulum–derived flavonoids and alkanes for delivery to the pollen surface. *Plant Cell*, **19**, 582 -596.

Jansen, R. C., Ooijen, J. W., Stam, P., Lister, C., and Dean, C. (1995). Genotype-by-environment interaction in genetic mapping of multiple quantitative trait loci. *Theoret. Appl. Genetics*, **91**, 33-37.

Jeong, S.-W., Das, P. K., Jeoung, S. C., Song, J.-Y., Lee, H. K., Kim, Y.-K., Kim, W. J., Park, Y. I., Yoo, S.-D., Choi, S.-B., u. a. (2010). Ethylene suppression of sugar-induced anthocyanin pigmentation in Arabidopsis. *Plant Physiol.*, **154**, 1514-1531.

Johanson, U., West, J., Lister, Clare, Michaels, S., Amasino, R., and Dean, C. (2000). Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. *Science*, **290**, 344 -347.

Jørgenson, S. and Mauricio, R. (2004). Neutral genetic variation among wild North American populations of the weedy plant Arabidopsis thaliana is not geographically structured. *Mol. Ecol.*, **13**, 3403-3413.

Kaul S., Koo H.L., Jenkins J. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408 (6814)**: 796-815.

Kendrick R.E., Kronenberg G. H. M. (1994). Photomorphogenesis in plants. *Kluwer Academic Publishers*, Dordrecht, The Netherlands.

Keurentjes, J. J. B., Bentsink, L., Alonso-Blanco, C., Hanhart, C. J., Blankestijn-De Vries, H., Effgen, S., Vreugdenhil, D., and Koornneef, M. (2007). Development of a Near-Isogenic Line population of Arabidopsis thaliana and comparison of mapping power with a Recombinant Inbred Line population. *Genetics*, **175**, 891 -905.

Kim, S., Plagnol, V., Hu, T. T., Toomajian, C., Clark, R. M., Ossowski, S., Ecker, J. R., Weigel, D., and Nordborg, M. (2007). Recombination and linkage disequilibrium in Arabidopsis thaliana. *Nat. Genet*, **39**, 1151-1155.

Koes, R., Verweij, W., and Quattrocchio, F. (2005). Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. *Trends Plant Sci.*, **10**, 236-242.

Koornneef, M., Reymond, M., Alonso-Blanco, C. (2011). Natural variation in Arabidopsis thaliana. Genetics and genomics of the Brassicaceae. *Plant Genetics and Genomics. Crops and Models.* Volume **9**, 123-152.

Koornneef, M, Alonso-Blanco, C, Blankestijn-de Vries, H, Hanhart, C J, and Peeters, A. J. (1998). Genetic interactions among late-flowering mutants of Arabidopsis. *Genetics*, **148**, 885-892.

Koornneef, M., Hanhart, C.J., and Veen, J. H. (1991). A genetic and physiological analysis of late flowering mutants in Arabidopsis thaliana. *Mol. Gen. Genet.*, **229**. 57-66.

Koornneef, M., Alonso-Blanco, C., and Vreugdenhil, D. (2004). Naturally occurring genetic variation in Arabidopsis thaliana. *Annu Rev Plant Biol*, **55**, 141-172.

Kowalski, S. P., Lan, T. H., Feldmann, K. A., and Paterson, A. H. (1994). QTL mapping of naturally-occurring variation in flowering time of Arabidopsis thaliana. *Mol. Gen. Genet*, **245**, 548-555.

Kroymann, J. and Mitchell-Olds, T. (2005). Epistasis and balanced polymorphism influencing complex trait variation. *Nature*, **435**, 95-98.

Kubo, H., Nawa, N., and Lupsea, S. A. (2007). Anthocyaninless1 gene of Arabidopsis thaliana encodes a UDP-glucose:flavonoid-3-O-glucosyltransferase. *J Plant Res*, **120**, 445-449.

Kuittinen, H., Mattila, A., and Savolainen, O. (1997). Genetic variation at marker loci and in quantitative traits in natural populations of Arabidopsis thaliana. *Heredity*, **79**, 144-152.

Labate, M., Ko, K., Ko, Z., Pinto, L., Real, M., Romano, M., Barja, P., Granell, A., Friso, G., Wijk, K., u. a. (2004). Constitutive expression of pea Lhcb 1-2 in tobacco affects plant development, morphology and photosynthetic capacity. *Plant Mol. Biol.*, **55**, 701-714.

Laitinen, R. A. E., Schneeberger, K., Jelly, N. S., Ossowski, S., and Weigel, D. (2010). Identification of a spontaneous frame shift mutation in a nonreference Arabidopsis accession using whole genome sequencing. *Plant Physiol.*, **153**, 652 -654.

Lam, E. (2004). Controlled cell death, plant survival and development. *Nat. Rev. Mol. Cell. Biol.*, **5**, 305-315.

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

Ledger, S., Strayer, C., Ashton, F., Kay, S. A., and Putterill, J. (2001). Analysis of the function of two circadian ‐ regulated CONSTANS ‐ LIKE genes. *Plant J.*, **26**, 15-22.

Lempe, J., Balasubramanian, S., Sureshkumar, S., Singh, A., Schmid, M., and Weigel, D. (2005). Diversity of flowering responses in wild Arabidopsis thaliana strains. *PLoS Genet*, **1**, e6.

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*, **57**, 405-430.

Leyva, A., Jarillo, J. A., Salinas, J., and Martinez-Zapater, J. M. (1995). Low temperature induces the accumulation of phenylalanine ammonia-lyase and chalcone synthase mRNAs of Arabidopsis thaliana in a light-dependent manner. *Plant Physiol.*, **108**, 39-46.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.

Li, Yan, Huang, Y., Bergelson, J., Nordborg, M.,and Borevitz, J. O. (2010). Association mapping of local climate-sensitive quantitative trait loci in Arabidopsis thaliana. *PNAS.* **107**, 21199-21204.

Li, Yan, Roycewicz, P., Smith, E., and Borevitz, J. O. (2006). Genetics of local adaptation in the laboratory: Flowering time quantitative trait loci under geographic and seasonal conditions in Arabidopsis. *PLoS ONE*, **1**, e105.

Lillo, C., Lea, U. S., and Ruoff, P. (2008). Nutrient depletion as a key factor for manipulating gene expression and product formation in different branches of the flavonoid pathway. *Plant Cell Environ.*, **31**, 587-601.

Lokhande, S. D., Ogawa, K., Tanaka, A., and Hara, T. (2003). Effect of temperature on ascorbate peroxidase activity and flowering of Arabidopsis thaliana ecotypes under different light conditions. *Plant Physiol.*, **160**, 57-64.

Loreti, E., Povero, G., Novi, G., Solfanelli, C., Alpi, A., and Perata, P. (2008). Gibberellins, jasmonate and abscisic acid modulate the sucrose ‐ induced expression of anthocyanin biosynthetic genes in Arabidopsis. *New Phytologist*, **179**, 1004-1016.

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

Loudet, O., Saliba-Colombani, V., Camilleri, C., Calenge, F., Gaudon, V., Koprivova, A., North, K. A., Kopriva, S., and Daniel-Vedele, F. (2007). Natural variation for sulfate content in Arabidopsis thaliana is highly controlled by APR2. *Nat. Genet*, **39**, 896-900.

Macknight, R., Bancroft, I., Page, T., Lister, C., Schmidt, R., Love, K., Westphal, L., Murphy, G., Sherson, S., Cobbett, C., u. a. (1997). FCA, a gene controlling flowering time in Arabidopsis, encodes a protein containing RNA-binding domains. *Cell*, **89**, 737-745.

McElver, J., Tzafrir, I., Aux, G., Rogers, R., Ashby, C., Smith, K., Thomas, C., Schetter, A., Zhou, Q., Cushman, M. A., u. a. (2001). Insertional mutagenesis of genes required for seed development in Arabidopsis thaliana. *Genetics*, **159,** 1751–1763.

McSteen, P. and Leyser, O. (2005). Shoot Branching. *Annu. Rev. Plant Biol.*, **56**, 353-374.

Méndez-Vigo, B., de Andrés, M. T., Ramiro, M., Martínez-Zapater, J. M., and Alonso-Blanco, C. (2010). Temporal analysis of natural variation for the rate of leaf production and its relationship with flowering initiation in Arabidopsis thaliana. *J. Exp. Bot*, **61**, 1611-1623.

Meurer, J., Plucken, H., Kowallik, K. V., and Westhoff, P. (1998). A nuclear-encoded protein of prokaryotic origin is essential for the stability of photosystem II in Arabidopsis thaliana. *EMBO J*, **17**, 5286-5297.

Michaels, S D and Amasino, R M (1999). FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell*, **11**, 949-956.

Michaels, Scott D., He, Y., Scortecci, K. C., and Amasino, R. M. (2003). Attenuation of FLOWERING LOCUS C activity as a mechanism for the evolution of summer-annual flowering behavior in Arabidopsis. *PNAS*. **100**, 10102-10107.

Mitchell-Olds, T. and Schmitt, J. (2006). Genetic mechanisms and evolutionary significance of natural variation in Arabidopsis. *Nature*, **441**, 947-952.

Monforte, A. J. and Tanksley, S. D. (2000). Development of a set of near isogenic and backcross recombinant inbred lines containing most of the *Lycopersicon hirsutum* genome in a *L. esculentum* genetic background: A tool for gene mapping and gene discovery. *Genome*, **43**, 803-813.

Mouradov, A., Cremer, F., and Coupland, G. (2002). Control of flowering time: interacting pathways as a basis for diversity. *Plant Cell*, **14**, S111-130.

Murashige, T. and Skoog, F. (1962). A revised medium for rapid growth and bio assays with Tobacco tissue cultures. *Physiol. Plantarum*, **15**, 473-497.

Nagira, Y., Shimamura, K., Hirai, S., Shimanuki, M., Kodama, H., and Ozeki, Y. (2006). Identification and characterization of genes induced for anthocyanin synthesis and chlorophyll degradation in regenerated torenia shoots using suppression subtractive hybridization, cDNA microarrays, and RNAi techniques. *J. Plant Res*, **119**, 217-230.

Neill, S. O., Gould, K. S, Kilmartin, P. A., Mitchell, K. A., and Markham, K. R. (2002). Antioxidant activities of red versus green leaves in Elatostema rugosum. *Plant Cell Environ.*, **25**, 539-547.

Noctor, G. and Foyer, C. H. (1998). Ascorbate and Glutathion: Keeping active oxygen under control. *Annu. Rev. Plant. Physiol. Plant. Mol. Biol.*, **49**, 249-279.

Nordborg, M., Hu, T. T., Ishino, Y., Jhaveri, J., Toomajian, C., Zheng, H., Bakker, E., Calabrese, P., Gladstone, J., Goyal, R., u. a. (2005). The pattern of polymorphism in Arabidopsis thaliana. *PLoS Biol*, **3**, e196.

Nozzolillo C., Isabelle P., Andersen Ø.M., and Abou-Zaid M. (2002). Anthocyanins of jack pine (Pinus banksiana) seedlings. *Can. J. Botany.* **80**, 796-801.

O'Neill, C. M., Morgan, C., Kirby, J., Tschoep, H., Deng, P. X., Brennan, M., Rosas, U., Fraser, F., Hall, C., Gill, S., u. a. (2008). Six new recombinant inbred populations for the study of quantitative traits in Arabidopsis thaliana. *Theor Appl Genet*, **116**, 623-634.

Ohto, M.-aki, Onai, K., Furukawa, Y., Aoki, E., Araki, T., and Nakamura, K. (2001). Effects of sugar on vegetative development and floral transition in Arabidopsis. *Plant Physiol.*, **127**, 252-261.

Olsen, K. M., Slimestad, R., Lea, U. S., Brede, C., Løvdal, T., Ruoff, P., Verheul, M., and Lillo, C. (2009). Temperature and nitrogen effects on regulators and products of the flavonoid pathway: experimental and kinetic model studies. *Plant Cell Environ.*, **32**, 286-299.

Ooijen, J. (1992). Accuracy of mapping quantitative trait loci in autogamous species. *Theoret. Appl. Genetics*, **84,** 803-811.

Ossowski, S., Schneeberger, K., Clark, R. M., Lanz, C., Warthmann, N., and Weigel, D. (2008). Sequencing of natural strains of Arabidopsis thaliana with short reads. *Genome Res*, **18**, 2024-2033.

Page, D. R. and Grossniklaus, U. (2002). The art and design of genetic screens: Arabidopsis thaliana. *Nat. Rev. Genet.*, **3**, 124-136.

Paran, I. and Zamir, D. (2003). Quantitative traits in plants: beyond the QTL. *Trends Genet.*, **19**, 303-306.

Pattanaik, S., Xie, C. H., and Yuan, L. (2007). The interaction domains of the plant Myc-like bHLH transcription factors can regulate the transactivation strength. *Planta*, **227**, 707-715.

Pelaz, S., Ditta, G. S., Baumann, E., Wisman, E., and Yanofsky, M. F. (2000). B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature*, **405**, 200-203.

Peng, M., Cui, Y., Bi, Yong - Mei, and Rothstein, S. J. (2006). AtMBD9: a protein with a methyl - CpG - binding domain regulates flowering time and shoot branching in Arabidopsis. *Plant J.*, **46**, 282-296.

Pérez-Pérez, J. M., Esteve-Bruna, D., and Micol, J. L. (2010). QTL analysis of leaf architecture. *J. Plant Res*, **123**, 15-23.

Petroni, K., Cominelli, E, Consonni, G., Gusmaroli, G, Gavazzi, G., and Tonelli, C (2000). The developmental expression of the maize regulatory gene Hopi determines germination-dependent anthocyanin accumulation. *Genetics*, **155**, 323-336.

Piazza, P., Procissi, A., Jenkins, G. I., and Tonelli, Chiara (2002). Members of the c1/pl1 regulatory gene family mediate therResponse of maize aleurone and mesocotyl to different light qualities and cytokinins. *Plant Physiol.*, **128**, 1077 -1086.

Picó, F. X., Méndez-Vigo, B., Martínez-Zapater, J. M., and Alonso-Blanco, C. (2008). Natural genetic variation of Arabidopsis thaliana is geographically structured in the Iberian Peninsula. *Genetics*, **180**, 1009 -1021.

Pieper, B. (2009). The genetic and molecular basis of natural variation for plant growth and related traits in *Arabidopsis thaliana*. Wageningen University. Wageningen. p. 13-14.

Poduska, B., Humphrey, T., Redweik, A., and Grbic, V. (2003). The Synergistic Activation of FLOWERING LOCUS C by FRIGIDA and a New Flowering Gene AERIAL ROSETTE 1 Underlies a Novel Morphology in Arabidopsis. *Genetics*, **163**, 1457 -1465.

Porra, R. J. (2002). The chequered history of the development and use of simultaneous equations for the accurate determination of chlorophylls a and b. *Photosyn. Res*, **73**, 149-156.

Price, A. H. (2006). Believe it or not, QTLs are accurate! *Trends Plant Sci*, **11**, 213-216.

Prinzenberg, A. E., Barbier, H., Salt, D. E., Stich, B., and Reymond, M. (2010). Relationships between growth, growth response to nutrient supply, and ion content

using a Recombinant Inbred Line Population in Arabidopsis. *Plant Physiol.*, **154**, 1361 -1371.

Putterill, J., Robson, F., Lee, K., Simon, R., and Coupland, G. (1995). The CONSTANS gene of Arabidopsis promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. *Cell*, **80**, 847-857.

Qi, R. and John, P. C. L. (2007). Expression of genomic AtCYCD2;1 in Arabidopsis induces cell division at smaller cell sizes: implications for the control of plant growth. *Plant Physiol*, **144**, 1587-1597.

Ramsay, N. A., Walker, A. R., Mooney, M., and Gray, J. C. (2003). Two basic-helix-loop-helix genes (MYC-146 and GL3) from Arabidopsis can activate anthocyanin biosynthesis in a white-flowered Matthiola incana mutant. *Plant Mol. Biol*, **52**, 679-688.

Ratcliffe, O. J., Nadzan, G. C., Reuber, T. L., and Riechmann, J. L. (2001). Regulation of flowering in Arabidopsis by an FLC homologue. *Plant Physiol*, **126**, 122-132.

Reymond, M., Muller, B., Leonardi, A., Charcosset, A., and Tardieu, F. (2003). Combining quantitative trait Loci analysis and an ecophysiological model to analyze the genetic variability of the responses of maize leaf growth to temperature and water deficit. *Plant Physiol*, **131**, 664-675.

Richmond, T. and Somerville, S. (2000). Chasing the dream: Plant EST microarrays. *Curr. Opin. Plant Biol.*, **3**, 108-116.

Rolland, F., Baena-Gonzalez, E., and Sheen, J. (2006). Sugar sensing and signalling in plants: Conserved and novel mechanisms. *Annu. Rev. Plant Biol.*, **57**, 675-709.

Rook, F., Hadingham, S. A., Li, Yunhai, and Bevan, M. W. (2006). Sugar and ABA response pathways and the control of gene expression. *Plant Cell Environ*, **29**, 426-434.

Roosens, N. H. C. J., Willems, G., and Saumitou-Laprade, P. (2008). Using Arabidopsis to explore zinc tolerance and hyperaccumulation. *Trends Plant Sci*, **13**, 208-215.

Rosloski, S. M., Jali, S. S., Balasubramanian, S., Weigel, D., and Grbic, V. (2010). Natural diversity in flowering responses of Arabidopsis thaliana caused by variation in a tandem gene array. *Genetics*, **186**, 263 -276.

Rossel, J. B., Wilson, I. W., and Pogson, B. J. (2002). Global changes in gene expression in response to high light in Arabidopsis. *Plant Physiol.*, **130**, 1109 -1120.

Roux, F., Touzet, P., Cuguen, J., and Le Corre, V. (2006). How to be early flowering: An evolutionary perspective. *Trends Plant Sci*, **11**, 375-381.

Rozen, S. and Skaletsky, H. (1999). Primer3 on the www for general users and for biologist programmers. In, *Bioinformatics Methods and Protocols*. Humana Press, New Jersey, p. 365-386.

Salomé, P. A., Bomblies, K., Laitinen, R. A. E., Yant, L., Mott, R., and Weigel, D. (2011). Genetic architecture of flowering time variation in Arabidopsis thaliana. *Genetics*. **188**, 421-433.

Salvi, S. and Tuberosa, R. (2005). To clone or not to clone plant QTLs: present and future challenges. *Trends Plant Sci*, **10**, 297-304.

Sarma, A. D. and Sharma, R. (1999). Anthocyanin-DNA copigmentation complex: Mutual protection against oxidative damage. *Phytochemistry*, **52**, 1313-1318.

Scarcelli, N. and Kover, P. X. (2009). Standing genetic variation in FRIGIDA mediates experimental evolution of flowering time in Arabidopsis. *Mol. Ecol.*, **18**, 2039-2049.

Schellmann, S., Schnittger, A., Kirik, V., Wada, T., Okada, K., Beermann, A., Thumfahrt, J., Jürgens, G., and Hülskamp, M. (2002). TRIPTYCHON and CAPRICE mediate lateral inhibition during trichome and root hair patterning in Arabidopsis. *EMBO J*, **21**, 5036-5046.

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.

Schmid, K. J., Törjék, O., Meyer, R., Schmuths, H., Hoffmann, M. H., and Altmann, T. (2006). Evidence for a large-scale population structure of Arabidopsis thaliana from genome-wide single nucleotide polymorphism markers. *Theor Appl Genet*, **112**, 1104-1114.

Schneeberger, K., Ossowski, S., Lanz, C., Juul, T., Petersen, A. H., Nielsen, K. L., Jorgensen, J.-E., Weigel, D., and Andersen, S. U. (2009). SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nat. Meth.*, **6**, 550-551.

Scortecci, K. C., Michaels, S. D., and Amasino, R. M. (2001). Identification of a MADS‐box gene, FLOWERING LOCUS M, that represses flowering. *Plant J.*, **26**, 229-236.

Shindo, C., Aranzana, M.J., Lister, C., Baxter, C., Nicholls, C., Nordborg, M., and Dean, C. (2005). Role of FRIGIDA and FLOWERING LOCUS C in determining variation in flowering time of Arabidopsis. *Plant Physiol.*, **138**, 1163-1173.

Shindo, C., Lister, C., Crevillen, P., Nordborg, M., and Dean, C. (2006). Variation in the epigenetic silencing of FLC contributes to natural variation in Arabidopsis vernalization response. *Genes Dev*, **20**, 3079-3083.

Simon, M., Loudet, O., Durand, S., Bérard, A., Brunel, D., Sennesal, F.-X., Durand-Tardif, M., Pelletier, G., and Camilleri, C. (2008). Quantitative trait loci mapping in five new large recombinant inbred line populations of Arabidopsis thaliana genotyped with consensus single-nucleotide polymorphism markers. *Genetics*, **178**, 2253-2264.

Simpson, G. G. and Dean, C. (2002). Arabidopsis, the Rosetta stone of flowering time? *Science*, **296**, 285-289.

Solfanelli, C., Poggi, A., Loreti, E., Alpi, A., and Perata, P. (2006). Sucrose-specific induction of the anthocyanin biosynthetic pathway in Arabidopsis. *Plant Physiol.*, **140**, 637 -646.

Somerville, C. and Koornneef, M. (2002). A fortunate choice: the history of Arabidopsis as a model plant. *Nat. Rev. Genet.*, **3**, 883-889.

Stafford, H. A. (1991). Flavonoid Evolution: An Enzymic Approach. *Plant Physiol.*, **96**, 680 -685.

Stenøien, H. K., Fenster, C. B., Tonteri, A., and Savolainen, O. (2005). Genetic variability in natural populations of Arabidopsis thaliana in northern Europe. *Mol. Ecol.*, **14**, 137-148.

Stinchcombe, J. R., Weinig, C., Ungerer, M., Olsen, K. M., Mays, C., Halldorsdottir, S. S., Purugganan, M. D., and Schmitt, J. (2004). A latitudinal cline in flowering time in Arabidopsis thaliana modulated by the flowering time gene FRIGIDA. *PNAS.*, **101**, 4712-4717.

Stitt, M., Lunn, J., and Usadel, B. (2010). Arabidopsis and primary photosynthetic metabolism - more than the icing on the cake. *Plant J*, **61**, 1067-1091.

Sureshkumar, S., Todesco, M., Schneeberger, K., Harilal, R., Balasubramanian, S., and Weigel, D. (2009). A genetic defect caused by a triplet repeat expansion in Arabidopsis thaliana. *Science*, **323**, 1060 -1063.

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.

Tisné, S., Schmalenbach, I., Reymond, M., Dauzat, M., Pervent, M., Vile, D., and Granier, C. (2010). Keep on growing under drought: genetic and developmental bases of the response of rosette area using a recombinant inbred line population. *Plant Cell Environ.*, **33**, 1875-1887.

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

Tonsor, S. J., Alonso-Blanco, C., and Koornneef, M. (2005). Gene function beyond the single trait: natural variation, gene effects, and evolutionary ecology in Arabidopsis thaliana. *Plant Cell Environ.*, **28**, 2-20.

Torii, K. U., Mitsukawa, N., Oosumi, T., Matsuura, Y., Yokoyama, R., Whittier, R. F., and Komeda, Y. (1996). The Arabidopsis ERECTA gene encodes a putative receptor protein kinase with extracellular leucine-rich repeats. *Plant Cell.*, **8**, 735 - 746.

Törjék, O., Meyer, R. C., Zehnsdorf, M., Teltow, M., Strompen, G., Witucka-Wall, H., Blacha, A., and Altmann, T. (2008). Construction and analysis of 2 reciprocal Arabidopsis introgression line populations. *J. Hered*, **99**, 396-406.

Turck, F., Fornara, F., and Coupland, G. (2008). Regulation and identity of Florigen: FLOWERING LOCUS T moves center stage. *Annu. Rev. Plant Biol.*, **59**, 573-594.

Ulm, R., Baumann, A., Oravecz, A., Máté, Z., Ádám, É., Oakeley, E. J., Schäfer, E., and Nagy, F. (2004). Genome-wide analysis of gene expression reveals function of the bZIP transcription factor HY5 in the UV-B response of Arabidopsis. *PNAS.*, **101**, 1397 -1402.

Van Ooijen, J.W. (2004) MapQTL® 5.0, Software for the mapping of quantitative trait loci in experimental populations. Kyazma B.V., Wageningen, Netherlands

Van Ooijen, J.W. (2006) JoinMap® 4.0, Software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen, the Netherlands.

Verbruggen, N., Hermans, C., and Schat, H. (2009). Molecular mechanisms of metal hyperaccumulation in plants. *New Phytol*, **181**, 759-776.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Lee, T. van de, Hornes, M., Friters, A., Pot, J., Paleman, J., Kuiper, M., u. a. (1995). AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Res.*, **23**, 4407-4414.

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. *Plant Cell*, **11**, 1337 -1350.

Wang, Q., Sajja, U., Rosloski, S., Humphrey, T., Kim, M. C., Bomblies, K., Weigel, D., and Grbic, V. (2007). HUA2 caused natural variation in shoot morphology of A. thaliana. *Curr. Biol.*, **17**, 1513-1519.

Weigel, D. and Mott, R. (2009). The 1001 genomes project for Arabidopsis thaliana. *Genome Biol*, **10**, 107.

Weigel, D. and Nordborg, M. (2005). Natural variation in Arabidopsis. How do we find the causal genes? *Plant Physiol*, **138**, 567-568.

Weinig, C., Ungerer, M. C., Dorn, L. A., Kane, N. C., Toyonaga, Y., Halldorsdottir, S. S., Mackay, T. F. C., Purugganan, M. D., and Schmitt, J. (2002). Novel loci control variation in reproductive timing in Arabidopsis thaliana in natural environments. *Genetics*, **162**, 1875 -1884.

Werner, J. D., Borevitz, J. O., Uhlenhaut, N. H., Ecker, J. R., Chory, J., and Weigel, D. (2005). FRIGIDA-independent variation in flowering time of natural Arabidopsis thaliana accessions. *Genetics*, **170**, 1197-1207.

Werner, J. D., Borevitz, J. O., Warthmann, N., Trainer, G. T., Ecker, J. R., Chory, J., and Weigel, D. (2005). Quantitative trait locus mapping and DNA array hybridization identify an FLM deletion as a cause for natural flowering-time variation. *PNAS.*, **102**, 2460-2465.

West, M. A. L., van Leeuwen, H., Kozik, A., Kliebenstein, D. J., Doerge, R. W., St Clair, D. A., and Michelmore, R. W. (2006). High-density haplotyping with microarray-based expression and single feature polymorphism markers in Arabidopsis. *Genome Res*, **16**, 787-795.

Willmann, M. R., Mehalick, A. J., Packer, R. L., and Jenik, P. D. (2011). MicroRNAs regulate the timing of embryo maturation in Arabidopsis. *Plant Physiol*, **155**, 1871-1884.

Winkel-Shirley, B. (2001). Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol.*, **126**, 485-493.

Yaish, M. W. F., Peng, M., and Rothstein, S. J. (2009). AtMBD9 modulates Arabidopsis development through the dual epigenetic pathways of DNA methylation and histone acetylation. *Plant J.*, **59**, 123-135.

Yang, W., Ren, S., Zhang, X., Gao, M., Ye, S., Qi, Y., Zheng, Y., Wang, J., Zeng, L., Li, Q., u. a. (2011). BENT UPPERMOST INTERNODE1 encodes the class II formin FH5 crucial for actin organization and rice development. *Plant Cell*, **23**, 661-680.

Yano, M (2001). Genetic and molecular dissection of naturally occurring variation. *Curr. Opin. Plant Biol*, **4**, 130-135.

Yano, M, Katayose, Y., Ashikari, M., Yamanouchi, U., Monna, L., Fuse, T., Baba, T., Yamamoto, K., Umehara, Y., Nagamura, Y., u. a. (2000). Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene CONSTANS. *Plant Cell*, **12**, 2473-2484.

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

van Zanten, M., Snoek, L. B., Proveniers, M. C. G., and Peeters, A. J. M. (2009). The many functions of ERECTA. *Trends Plant Sci*, **14**, 214-218.

Zeng, Z.-B., Kao, C.-H., and Basten, C. J. (1999). Estimating the genetic architecture of quantitative Traits. *Genet. Res.*, **74**, 279-289.

Zhang, F., Gonzalez, A., Zhao, M., Payne, C. T., and Lloyd, A. (2003). A network of redundant bHLH proteins functions in all TTG1-dependent pathways of Arabidopsis. *Development*, **130**, 4859 -4869.

Zhang, H., Goodman, H. M., and Jansson, S. (1997). Antisense inhibition of the photosystem I antenna protein Lhca4 in Arabidopsis thaliana. *Plant Physiol.*, **115**, 1525 -1531.

Zhang, Y., Zheng, S., Liu, Z., Wang, L., and Bi, Y. (2011). Both HY5 and HYH are necessary regulators for low temperature-induced anthocyanin accumulation in Arabidopsis seedlings. *Plant Physiol*, **168**, 367-374.

Zhao, K., Aranzana, M. J., Kim, S., Lister, C., Shindo, C., Tang, C., Toomajian, C., Zheng, H., Dean, C., Marjoram, P., u. a. (2007). An Arabidopsis example of association mapping in structured Samples. *PLoS Genet*, **3**, e4.

Zhao, M., Morohashi, K., Hatlestad, G., Grotewold, E., and Lloyd, A. (2008). The TTG1-bHLH-MYB complex controls trichome cell fate and patterning through direct targeting of regulatory loci. *Development*, **135**, 1991-1999.

Zhu, H.-F., Fitzsimmons, K., Khandelwal, A., and Kranz, R. G. (2009). CPC, a single-repeat R3 MYB, is a negative regulator of anthocyanin biosynthesis in Arabidopsis. *Mol Plant*, **2**, 790-802.

Zhu, J., Dong, C.-H., and Zhu, J.-K. (2007). Interplay between cold-responsive gene regulation, metabolism and RNA processing during plant cold acclimation. *Curr. Opin. Plant Biol.*, **10**, 290-295.

Zimmermann, I. M., Heim, M. A., Weisshaar, B., and Uhrig, J. F. (2004). Comprehensive identification of Arabidopsis thaliana MYB transcription factors interacting with R/B-like BHLH proteins. *Plant J*, **40**, 22-34.

Table A.1: Marker used for building genetic map of L*er*/Eri-1 and validation and fine mapping of the detected QTL

Marker Type	Marker Name	Chr	Physical (Mb)	Genetic (cM)	Annealing T (℃)	Forward Primer Sequence	Reverse Primer Sequence	Enzyme	Experiment used
SSLP	NT7123	I	0.4	0	60	GTG TCC TTT TTT CTC AAC GAT G	CAT GCA CGT ACG ATT TGT TTA AC		genetic map
SSLP	JV26/27	I	4	1.8	55	CAA GAG ATT GCA ACA TCC ACA	AAG CTC CTT GGA TCC GAT TT		genetic map
SSLP	SO392	I	10.9	32.5	55	GTT GAT CGC AGC TTG ATA AGC	TTT GGA GTT AGA CAC GGA TCT G		genetic map
SSLP	T27K12	I	16.1	52	55	GGA GGC TAT ACG AAT CTT GAC A	GGA CAA CGT CTC AAA CGG TT		genetic map
SSLP	CIW1	I	18.3		50	ACA TTT TCT CAA TCC TTA CTC	GAG AGC TTC TTT ATT TGT GAT		validation and fine mapping
SSLP	NF19K23	I	22.9	66	55	GAA TTC TGT AAC ATC CCA TTT CC	GGT CTA ATT GCC GTT GTT GC		genetic map
CAPS	PAP2-CP-F + PAP2-BP-R	I	24.76		55	GGA ACA CCC ATC TGA GTA AA	TAG TGT TTC TCA CCG TTT GTT	BpmI	fine mapping
SSLP	M1-13	I	25.5	78	56.5	CAA CCA CCA GGC TC	GTC AAA CCA GTT CAA TCA		genetic map
SSLP	T8K14	I	29.9	87	52	GGA CAA GAA CCT CAT ACC	AGA GAT AAG GAC GTG GTA G		genetic map
SSLP	M2-5	II	0.2	3.3	50	TGA GAG GGA CAG ATA GGA A	ATC AAA AGG GAT ACT GAC AA		genetic map
SSLP	F12A24b	II	7.3	18.5	55	GGT GTG ATG TCG ACC GGT AAA G	TGC ACA ACG TGC TCT CCA TG		genetic map
SSLP	PLS7	II	9.8	25.9	57	GAT GAA TCT TCT CGT CCA AAA T	GAC AAA CTA AAC AAC ATC CTT CTT		genetic map

SSLP	M2-17	II	10.7	29.7	50	GAT TCC ACC ATA TGT GGA T	CTT CGT CAC TGC CAA TAC		genetic map
SSLP	NGA1126	II	11.7	39.4	58	GCA CAG TCC AAG TCA CAA CC	CGC TAC GCT TTT CGG TAA AG		genetic map
SSLP	T2N18	II	15.6	42	55	TTT ACG AAT AGG ATT GGG TTT CAT C	ATG ATT CTC TTT ATG GCT CCT CAG C		genetic map
SSLP	F17A22	II	19.6	62.7	55	ACA CAC GAA TAT TGA TTG TCT AAG G	TCA CTT GTC GGT TTG TGT GG		genetic map
CAPS	T4P13	III	0.014		60	GTG GGT GGC AAA TTG GTG AAA GT	AGA ATG GGC CGA AAA CAG TGA AA	RsaI	validation and fine mapping
SSLP	T22N4b	III	0.085		48	TTT CCC TTG ATG ATT AGT CG	TAC TAT ATC AGC CGC TAA CTA G		validation and fine mapping
SSLP	F4P13	III	0.2		48	TCG CAA CCA ACC TAC AAG	CAT GAG ACA ATC TAA CAA GGC		validation and fine mapping
SSLP	F28J7	III	0.33		55	AAA TAC CAC ATC TTG CTT GAG	AGA TTC TTA GAG ATG ACG TGG		validation and fine mapping
SSLP	F14P3	III	0.42		48	ACT CCA ACA ACC AGC CAG	AAA TGT TGC CAT AAC TCG G		validation and fine mapping
CAPS	3-452814	III	0.45		55	AAC AAA ATG AAT ACA TGC GTA A	ATC ATC TTC TCA AAA TCA AAG G	BsmFI	fine mapping
CAPS	3-926376	III	0.92		55	GCA AAG AAC TGG AAG GTA TG	TTG CTG TGT TGA AAG ATT TG	ApoI	fine mapping
CAPS	3-951148	III	0.95		55	GCC TAG AGG AGT GTG TGT GT	ACA AGA GCA AAA TGG AAA AA	FokI	fine mapping
CAPS	3-1066606	III	1.06		55	GAA AAC GCC AAA ATC TAA AG	TAG TAC CCA TCA TTC GGT TC	MboII	fine mapping

CAPS	3-1112663	III	1.11		55	GGG AGT TGG ATA CAG ATC AA	GGT TTC TCT TTT GCG TTT TA	AflII	fine mapping
SSLP	F7O18	III	1.23		50	GTT CTC AAG AGT GCC CTA AAG	CCG GAG TAG ATT CTG ATG G		validation and fine mapping
CAPS	FLK7	III	1.25		55	CTC TTC AAG ACG ATA GGG TTG T	CAG CAT AAG ACA GTG GAA TCT G	AclI	fine mapping
SSLP	T12H1	III	1.54		48	ACA CAC AGT GAA GAA ACA CG	TGG AAG CTC ACT CGA TTG		validation and fine mapping
SSLP	F22F7	III	1.6	3.2	55	AAA GCT TCA TAT TGT CAC CAC	AGA CTT TAA TGC TAC TCC AAG G		genetic map
CAPS	FTN3-175	III	1.75		55	TGA ATG TGT CGG TCT ACA AG	ACC GAT GAG TAC TGA TGA CC	BsmI	fine mapping
SSLP	F24P17	III	1.95		48	TCT TCA GCT CGC CTT AAG	AAG ACC TTC AAG TAC ATG CC		validation and fine mapping
SSLP	F17A9	III	2.22		50	TTC CTC CAC CAT ACC TGT C	TGC TCC ACT TTA TTT GTG TTA G		validation and fine mapping
CAPS	FTN3-245	III	2.45		55	GAA CTG ATG AAG AAG AAA GGG	CGG AAT CTC CTC GTT GAC	Tsp509I	fine mapping
SSLP	CHIB	III	3.9		55	ATG AGA AGC TAT AAT TTT TTC AAT A	CTC ATA TAT ACA AAG AAC TAC TAT AC		validation and fine mapping
SSLP	NT204	III	5.6	14.1	55	TGG AAG CTC TAG AAA CGA TCG	ACC ACC TAA ACC GAG AAT TGG		genetic map
SSLP	MQC12	III	7.1	19.2	55	AGC TCC TCC TCC TAA ACC G	CAT TTG CCA GTG TCG CTA		genetic map
SSLP	M3-32	III	11.2	38.8	55	GCA CTT GCA GCT TAA CTT	CGT GAC TGT CAA ACC G		genetic map

SSLP	M3-28	III	20.5	62.7	45	TAC AAG TCA TAA TAG AGG C	GGG TTT AGC ATT TAG C		genetic map
SSLP	M4-7	IV	1	0	55	GGC AGA GTT TAG GGG AAG C	CTT TCA AAT TCA ATG GGA CG		genetic map
SSLP	24C21	IV	1.4		50	TGA TCT GTT CTG TTG AGT TC	CAA GGA TAC GGT GTT AGA G		validation and fine mapping
SSLP	C6L9-78	IV	2.7	4.8	60	TGC TTT GTG AAA GTC TCT CAT GCC	CCC TTT GAT TGC TCA GTG ATA TCG		genetic map
SSLP	M4-25	IV	7	23.6	50	GAA TGG TTG TTG ATA GTT GA	AAA TTT CAG GAG GTG ATA GA		genetic map
SSLP	M4-36	IV	7.6	27.1	50	TAT TTC GGT GAA GAC AGA A	GCC TTG CAT AAA CCA A		genetic map
SSLP	G3883	IV	10.6	37.8	50	TGT TTC AGA GTA GCC AAT TC	CAT CCA TCA AAC AAA CTC C		genetic map
SSLP	F9D16	IV	13.3	44.1	50	CTG CTA TGT TGT GAC TTG TG	AAG TCG GTA ACG TAG TTC C		genetic map
SSLP	NGA1139	IV	16.4	54	55	TTT TTC CTT GTG TTG CAT TCC	TAG CCG GAT GAG TTG GTA CC		genetic map
SSLP	M4-33	IV	17.6	57	55	TTC TTT GAC ACG CAA ACA	TGG TGAC ATA GAC CCA ATG		genetic map
SSLP	NGA225	V	1.5	0	55	GAA ATC CAA ATC CCA GAG AGG	TCT CCC CAC TAG TTT TGT GTC C		genetic map
SSLP	NGA158	V	1.69		55	ACC TGA ACC ATC CTC CGT C	TCA TTT TGG CCG ACT TAG C		validation and fine mapping
SSLP	MHFD	V	1.9		50	AAA AAC CCA AAC TTT CTA TTT ATA C	ACT TCG CTT CAA GTA AAG AGG		validation and fine mapping

SSLP	FLC	V	3.1	4.9	58	CAT TGG ATA ACT AAT CTT TGA GC	CAG GCT GGA GAG ATG ACA AAA		genetic map
SSLP	AN5-360	V	3.6		55	GTT TTG ATC AGA CGG TAT GC	AAG CTC GTA TAG CTT GTT GC		fine mapping
SSLP	MTG13	V	5.46		54	AGG GAG AAA GAG AGA CAG AG	TTT CCA TTT AGT TCC TTC AG		validation and fine mapping
SSLP	NGA106	V	5.6	11.6	58	TGC CCC ATT TTG TTC TTC TC	GTT ATG GAG TTT CTA GGG CAC G		genetic map
SSLP	M5.07005	V	7		55	AGT GTG AAA CAA GGA CGA	TTT GAC TGA GAA AAC GAG AAG		validation and fine mapping
CAPS	AN7-16	V	7.16		55	GAA AAG TCG GTG TCG TTA AG	GGT TGT TTG ATT TGG TCA AC	RsaI	fine mapping
SSLP	CIW8	V	7.5	14.9	48	TAG TGA AAC CTT TCT CAG AT	TTA TGT TTT CTT CAA TCA GTT		genetic map
SSLP	M5-14	V	7.5	15	50	AAC AAC CCT ATC TTC TTC TG	TGT GAC CCC TTA CTC AAT A		genetic map
SSLP	AN7-5	V	7.5		55	CAT CTG TGT TGA CGC TTT AC	AGT CCG TAC ATT TCA TCA GG		fine mapping
SSLP	ICE5	V	7.71		55	CTT GCA ACC GCC AAC TCA ATC G	CCT GTC TCG CTC CCG CAC G		validation and fine mapping
CAPS	hua2.5n*	V	7.85		58	GTA GCA CCT TCC TCT TTC TGT	TGG ACA ATT TGA TCA CCC TG	MseI	validation and fine mapping
SSLP	CDPK9	V	7.95		48	TCA ATC ATT GTC CAA AAC TTG G	GAA ACT GAC TTG GAG AAG GCA		validation and fine mapping
SSLP	ICE2	V	8.43		50	CTC GGG TCA AAA TTA GGG TTT CG	GCT ACC AGA TCC GAT GGT AAG ATG		validation and fine mapping

SSLP	NGA139	V	8.5	18	55	GGT TTC GTT TCA CTA TCC AGG	AGA GCT ACC AGA TCC GAT GG		genetic map
SSLP	T1N24	V	9.01		49	ATC ATC AAT TAG TAG AGG TA	TAC TAC TCG AAC TCA GAT		validation and fine mapping
SSLP	ICE15	V	9.08		55	CCT GCT CCT CTT TCC TTC TGC TC	CAC ACT AAA TTA TCT ATC AAC T		validation and fine mapping
SSLP	F21E10	V	9.28		50	AAC TGT CTC TTC TCC ATA	CTT AAA CAT AAA GCA AAC ATC		validation and fine mapping
SSLP	F2P16	V	9.48		51	TTT CAA TAA TGG ATC TCG G	AGC CAA TAG TCT TCT TCC C		validation and fine mapping
SSLP	SO262	V	9.8	20.9	50	ATC ATC TGCC CAT GGT TTT T	TTG CTT TTT GGT TAT ATT CGG A		genetic map
CAPS	AN5-108	V	10.8		55	CTC TGT GTG TTG AGG GAG AT	AAT TTA GCA GTC GGA ATT TG	HaeIII	fine mapping
CAPS	AN5-1155	V	11.55		55	CGC TAC TTG GCT TGT TAT TT	ACA ATC CGA AGT AGC AAC AG	Tsp509I	fine mapping
SSLP	AN5-120	V	12		55	ATC AGA ACT CCG CAA TTA AG	GAG CTT TGA TAC ATG CAA		fine mapping
SSLP	M5-12620	V	12.62		55	TAA AAT ACA ATT TGA AAA CGT A	AAC AAT ATT CTA AAA CAA GGT AA		validation and fine mapping
CAPS	AN5-133	V	13.3		55	CCA AAC TGG ATT AGA TGG TG	AGT AAT GTC TAA CGG CCA AG	BfaI	fine mapping
SSLP	MOK9	V	13.67		59	CGA CGT TTT GAT TCA GCT TG	CTT CGA AAC TAC TTG AGC AGA AAC		validation and fine mapping
SSLP	MOK10	V	13.7		59	GTT GGG CCT TTT GTT TTG AAA G	GAT TCA TGG AAT CGA ACA CTG C		validation and fine mapping

SSLP	M5-22	V	13.9		50	AGA ACA AGT TAG GTG GCT	GGG ACA AGA ATG GAG T		validation and fine mapping
SSLP	PHYC-Pro*	V	14		57	TTG GTG TTT CGG TCT TTT CC	TGG AAC GTT CCT CCT TAG TGG		validation and fine mapping
SSLP	T30G6	V	14.2		55	ATG GTA AAC AAT TAT GTC ACG TAT G	CAT TGG TTC TCC CAT CGT CTC TCT AG		validation and fine mapping
SSLP	MNJ8	V	14.74		55	CCA TGT CTT GCA CAA CTC	TCT GCA ATA AAA ATT GAT GG		validation and fine mapping
SSLP	K12B20	V	14.95		50	TCG AAA AAC TGT CAG AAG	GCC TGC ATG ACA TGA ATC		validation and fine mapping
SSLP	T31G3	V	14.99		55	CTC CAC CAA TCA TGC AAA TG	TGA TGT TGA TGG AGA TGG TCA		validation and fine mapping
CAPS	5-14994339	V	14.99		55	TGA GTT TGA CTA TTT CGG AGA	ATA GAC ACA TGG CAA ATG CT	MspI	fine mapping
SSLP	SO191	V	15	34.6	50	CTC CAC CAA TCA TGC AAA TG	TGA TGT TGA TGG AGA TGG TCA		genetic map
SSLP	M5-9	V	17.3	41.5	60	CGT CAT TTT TCG CCG CTC T	CAT GGT GGC GCG TAG CTT A		genetic map
SSLP	NGA129	V	20.1	45	60	CAC ACT GAA GAT GGT CTT GAG G	TCA GGA GGA ACT AAA GTG AGG G		genetic map
SSLP	MBK5	V	25.5	65.4	50	GAG CAT TTC ACA GAG ACG	TCA TAC CTA CAT ATT GCC CT		genetic map
SSLP	M5-19	V	25.9		55	AAC GCA TTT GCT GTT TCC CA	ATG GTT ATC TCA TCT GGT CT		validation and fine mapping
SSLP	K8A10	V	26.8	69.6	58	AAT GCC AAG GAT CAA AAG TGT T	GAT GAT CGG AGG AAA ATG AAA A		genetic map

Table A.2: Marker used for sequencing candidate regions and qRT-PCR primer for expression analyses

Gene model	Marker Name	Chr	Physical (Mb)	Annealing T (℃)	Forward Primer Sequence	Reverse Primer Sequence	Experiment used
ETC1 *AT1G01380*	ETC1	I	0.14	60	AAG CTT GTC GGT GAA AGG TG	CCT CTC AAT CTC TTC TGC TGT TC	qRT-PCR
PAP1 *AT1G56650*	PAP1	I	21.23	60	AAA TGG CAC CAA GTT CCT GT	TCA GAG CTA AGT TTT CCT CTC TTG AT	qRT-PCR
EGL3 *AT1G63650*	EGL3	I	23.59	60	TTG GCA CGA CCG AAC ATA	TTG ATA GTC TGA TCT TGT CGA TAT TGT	qRT-PCR
MYB113 *AT1G66370*	Myb113-AP	I	24.75	55	AAG AGT CTG GCT GCT AAG TG	AGG ACC ACC TAA ACA TTG TG	Sequencing
	Myb113-BP	I	24.75	55	GAT CGC GGT TCA ATA CTA GA	ATT GAA CTT TCT GAC GTT GG	Sequencing
MYB114 *AT1G66380*	Myb114-AP	I	24.75	55	TTA GCA AAT AAA GCA CGA CA	GAT TTA GCC CTA CAA AAC CA	Sequencing
	Myb114-BP	I	24.75	55	ACA CGT GCG TAT GTG TGT AT	ACT CAC ACA GAA ACG GAA AC	Sequencing
	PAP2-C2P	I	24.75	55	GTG GTT GGA GAA TTT ACT GG	CCA AAA GCC AAA ATT TAG AA	Sequencing
MYB113 *AT1G66370*	Myb113	I	24.75	60	GGC AAA TGG CAT CGA GTT	CAA CCA TCT AAG TCT ACA ACT CTT TCG	qRT-PCR
MYB114 *AT1G66380*	Myb114	I	24.75	60	GTC TCT TGA GGC AGT GTA TTG GT	TTT TCC TGC ACC GAT TTA GC	qRT-PCR
PAP2 *AT1G66390*	1-24760748	I	24.76	55	ACA CCT TAC CGC TTT TAT GA	AAC CCT AAC AAC AAT GGC TA	Sequencing
	1-24761596	I	24.76	55	TTT TTG CGC TAT AAC TTT CTT	TAC TCA TGT GCC GTG ATA GA	Sequencing

PAP2 AT1G66390	1-24762547	I	24.76	55	TAT TGT GGG AAT TGA AGT CG	GAA TTG AGT TCG AGG CTT TA	Sequencing
	PAP2-AP + PAP2-A3P	I	24.76	55	CGA TCC AAA GAC TCA ATG TT	ACA TAC CAG CTC TCA AAG GA	Sequencing
	PAP2-A3P + PAP2-B2PF	I	24.76	55	ATA CCT ATT CGC TCA CAT GC	TCT TCG TTT AAA TAA ACT GCA	Sequencing
	PAP2-BP + PAP2-B3P	I	24.76	55	TCT TTT TGT TTT AAG GAG CAA	TGG CTC TTG AAA TAA CCA TT	Sequencing
	PAP2-C3P	I	24.76	55	AAT GGT TAT TTC AAG AGC CA	CCA GTA AAT TCT CCA ACC AC	Sequencing
PAP2 AT1G66390	PAP2*	I	24.76	60	TGG GCT AAA TCG ATG CAG AAA GA	CCA GCA ATC AAG GAC CAC CTA T	qRT-PCR
TCL1 AT2G30432	TCL1	II	12.96	60	GAA ATG GGA GTT TAT CAA TAT GAC C	ACG TCC CAC CAC TCT TCT TG	qRT-PCR
ETC2 AT2G30420	ETC2	II	12.96	60	GAC CGA ACA AGA AGA AGA TCT CA	TCA TTT GCC TTT CTT CCT ACG	qRT-PCR
CPC AT2G46410	CPC	II	19.04	60	AAG GCT TCT TGT TCC GAA GAG	CCT GTC GCC ACC GAG TTT AT	qRT-PCR
MYB12 AT2G47460	Myb12	II	19.47	60	CCA CTT TGG GAA ACA GGT GGT CAC T	GTT GTG GAG TTT ACG GCT GA	qRT-PCR
COL2 AT3G02380	COL2a	III	0.42	59	TTT TGC CAA CAA TCA TTG C	CGA ATT GGT TGT CCA TGC	Sequencing
	COL2b	III	0.42	59	CCG ATT CTT CCA TTA TCT GC	CTC TGT ATC TCA GGA CCC TAG C	Sequencing
	COL2c	III	0.42	59	CCT TAA AGG ACC TAA ACC ATA GC	TAG GAA CAT ATC TGA TTA AAC TCG TAC T	Sequencing

COL2 AT3G02380	COL2d	III	0.42	59	CGG TAG CAA CTT AAC AAT GAG A	GAT AAA GAA GAT GGA GAA ACA GAG A	Sequencing
	COL2e	III	0.42	59	TTC TCT TTC TCT TTC AAT ATC TTG TG	ATT ACG GAG GAA GAA GAA GTT GT	Sequencing
	COL2f	III	0.42	59	GAG CCC TAG AGT TGC CTT T	CGA ATC GTC CGT GGT AA	Sequencing
	COL2g	III	0.42	59	AAG AAC CAA GAT CGT GGA TT	TTA CAA CAA CAA CAA CAA CAA CA	Sequencing
	COL2h	III	0.42	59	CAG TCC AGT CCA TAA GTT ACC A	GGA TAA GAT TCT CAG AGA TTG CTA	Sequencing
COL2 AT3G02380	COL2i	III	0.42	59	CCA AAA CAG TAA AAG TAA ATA GCA A	AGA TAA TGG AAG AAT CGG AAC T	Sequencing
FLK AT3G04610	FLK1	III	1.25	59	TCT TTC TCC GCC GAA CT	TGG TGA ACA AGG GTC TCA T	Sequencing
	FLK2	III	1.25	59	CGG GAC TAA TGT GTG TTA TCT G	ATC ACC TTT GCG ACC AA	Sequencing
	FLK3	III	1.25	59	TAT TCC GTA TGC TGG TTC CT	GCC GAT TAA CTT CCT CCT TT	Sequencing
	FLK4	III	1.25	59	GCA TCA AGC GAA ACC AT	TAG CAG CCA TTA CCA ACC T	Sequencing
	FLK5	III	1.25	59	CCT GGA TGA TTG ATG TCT TG	GAG ATG ATT CGG GCT CTT C	Sequencing
	FLK6	III	1.25	59	TCG CAG GTT ATG GTT TCT G	TTC TGA GAT GTG ACG CAA TC	Sequencing
	FLK7	III	1.25	59	CTC TTC AAG ACG ATA GGG TTG T	CAG CAT AAG ACA GTG GAA TCT G	Sequencing

FLK AT3G04610	FLK8	III	1.25	59	GGT GTG GGG ATT TGT TTG	CGA TTA GGC ATC TAG CGT TT	Sequencing
	FLK9	III	1.25	59	GGC TTG TGA ACG CAT CTA C	CAT TGT GTG TAG ACT GTG TGG TA	Sequencing
	FLK10	III	1.25	59	CTA CCA CCT TCA ACG AGA TTG	CTT CCA TCG CCA TCT TTC	Sequencing
	FLK11	III	1.25	59	CGC AGA GAG ACT GAA AGA AGA	CCT CGT TGT GTT GCT GAA	Sequencing
	FLK12	III	1.25	59	GAG GTA CAT TCA GCA ACA CAA C	AGC GGC AAC TTC TTC TTG	Sequencing
	FLK13	III	1.25	59	TGT TGG GCA AGT CAG ACA	TGA TTT GGA AGT GGG ATA GG	Sequencing
	FLK14	III	1.25	59	GTT CAA TGA CCT ATC CCA CTT C	AAT TCA GAG TCC AGA TAA CAC ACA	Sequencing
Actin2 AT3G18780	ACT2*	III	6.47	60	GAT TCA GAT GCC CAG AAG TCT TGT	TGG ATT CCA GCA GCT TCC AT	qRT-PCR
ETC3 AT4G01060	ETC3	IV	0.46	60	TGC ATA AGC TTG TCG GTG AC	CTC AAT TTC TCC AGC GGT TC	qRT-PCR
TT8 AT4G09820	TT8	IV	6.18	60	TGA ATC AAC CCA TAC GTT AGA CA	GGG GTG TGA CAT GAG AAG TGT	qRT-PCR
ANL1 AT5G17050	NL1	V	5.6	55	CGG GTT TTT ATT ATT GTT GG	GAG ATA GCA AAT TTG GGT TG	Sequencing
	NL2	V	5.6	55	TTC CTC AAG ATT TTC TGA GC	GCT TAC GAA GCT GTC TCT TC	Sequencing
	NL3	V	5.6	55	TTT GTC CAA AAT CAT TTC AC	TTT GGA TAG GAC AAG AGA GC	Sequencing

ANL1 AT5G17050	NL4	V	5.6	55	GTA CAC CAC CCG ATA CAC TC	AAA GAT GCT TCA TCA AAT GG	Sequencing
	NL5	V	5.6	55	CGT CAA TGT AGG ATC CAA AT	AGC AAA CTC ACT CTC TGC TC	Sequencing
	NL6	V	5.6	55	AAA AGA CAA AAT CAC AAG TCG	CTT CTC TTT CTT CAA CAC CG	Sequencing
	NL7	V	5.6	55	GGA AAA TAA CGA AGA GTT GG	AAC ACG AGC TTT ACA CTT GC	Sequencing
TT19 AT5G17220	TT19C	V	5.65	55	AAT GCC AAC ACA TCT ACT CTC	GGA GCT AAT TCA GTG ACC AG	Sequencing
HUA2 AT5G23150	5-7785924	V	7.78	55	GCC TCA TCT CTC TCT GTC TG	CCA AAA AGA AAA ACA CTT GG	Sequencing
HUA2 AT5G23150	5-7786865	V	7.78	55	TAG ACC TCG GTT ATC AGG TG	TCC CAG TAG AAT CAG ATC CA	Sequencing
	5-7787778	V	7.78	55	CTC ACG TGC TGC TAA AAA G	GCG ACA TAC ACA CAA AAA GA	Sequencing
	5-7788700	V	7.78	55	GAG GAC TTG TCT GCT GCT AT	AGA GGA AAG CTC ACC TTT TT	Sequencing
	5-7789570	V	7.78	55	GTG AAT CTC ACT TCC ATC GT	TGA ACT CGG TTG ATT TTT CT	Sequencing
	hua2-5n	V	7.78	58	GTA GCA CCT TCC TCT TTC TGT	TGG ACA ATT TGA TCA CCC TG	Sequencing
	5-7791127	V	7.78	55	TTT AAA CCC CCT GGA CAT	AAA GCC TGC AGT AGT TTT TG	Sequencing
TTG1 AT5G24520	TTG1-UTR1	V	8.37	55	AGA GTG GAT ATG GTG AGT CG	TTC ATG TCA TCC TAG TGA TCC	Sequencing

TTG1 AT5G24520	TTG1-EX2	V	8.37	55	GGA GAG AAG GAG GAC TGA AC	ATA ATT CAG CTC CAG ATT CG	Sequencing
	TTG1-EX1	V	8.37	55	AGA GCC TGT GTA TCA TCA CC	AAA CGT CTC GGA ACT TGT AG	Sequencing
	TTG1-INT1	V	8.37	55	CCA ATA ATC AAT GCT TCC TC	TAT GGC TAC GAT TTT GAT GG	Sequencing
	TTG1-INT2	V	8.37	55	ATG TTT CTG AGG GCA TAC AG	AAT CAG TTG CAG TGG TCT TC	Sequencing
	TTG1-UTR2	V	8.37	55	TAC TCA CCA AAT TCC CAT TG	TTG TTT ACT GCT GAC TGT GC	Sequencing
CHS AT5G13930	CHS	V	4.48	60	TCA GGC GGA GTA TCC TGA CTA	CGT TTC CGA ATT GTC GAC TT	qRT-PCR
TTG1A AT5G24520.3	TTG1A	V	8.37	60	GTC TTC TTC GCA GCC TGA TT	AAA CCA GCA TGA AGT TTC CAA	qRT-PCR
TTG1B AT5G24520.2	TTG1A	V	8.37	60	TTG GAT TGG TAT TGC TTT TGC	CGG GTC TAT GCT AAA ATC CTT ATT	qRT-PCR
TT2 AT5G35550	TT2	V	13.72	60	AAG GCA AAT GGA GCA CTC TC	CTA CAG CTT TTG CCA CAC CTC	qRT-PCR
DFR AT5G42800	DFR	V	17.16	60	AAC GGA TGT GAC GGT GTT TT	TCC ATT CAC TGT CGG CTT TA	qRT-PCR
TRY AT5G53200	TRY	V	21.58	60	ACA GAC TTG TCG GTG ATA GGT G	CTA TCT CCT CTG GTT GTC TTC CA	qRT-PCR

Acknowledgements

First of all I would like to express my gratitude to Prof. Dr. Maarten Koornneef for giving me the opportunity to work in his department at the MPIPZ in Cologne and fruitful discussions.

Additionally, I would like to extend thanks to Prof. Dr. Ute Höcker of the Universität zu Köln for her evaluation of this PhD thesis.
Many thanks to the head of the thesis committee, Prof. Dr. Wolfgang Werr.

To my group leader and supervisor, Dr. Matthieu Reymond, for offering me this interesting topic for the PhD thesis in his group, I am grateful and would like to express my appreciation. Thanks for his good guidance and kind advice during the PhD thesis, in the last year even from far distance.

I would also like to thank Dr. Richard Reinhardt and the Max Planck Genome Centre Cologne for sequencing the Eri-1 accession.
Thanks to Dr. Jia (Nina) Ding, the bioinformaticion of the Koornneef department for helping in the analysis of the sequencing data of Eri-1 accession.

Many thanks, also, to the other people of my group: Dr. Ruben Alcázar, Dr. Aina Prinzenberg and Dr. Inga Schmalenbach. It was a real pleasure to work with you and I appreciated the helpful discussions. Thanks also to previous members of the AG Reymond, which left before finishing my PhD thesis: Dr. Hugues Barbier, Dr. Anna Ihnatowicz, Dr. Bjorn Pieper, Dr. Samija Amar, Dr. Navot Galpaz and Barbara Eilts and Nele Kaul. It was very nice working with you in the same group. Thanks for the nice atmosphere.

Special thanks go to Amanda Davis for the great help in correcting my English.

Thanks to Regina Gentges for taking care of the progeny of my plants. Thanks also to Birgit Thron for help with administrative issues. I also thank Britta Hoffmann for the library support and the people in the mechanical, electrical and climate workshop for the help with the high light lamps.

Thanks to the IMPRS and the Max-Planck Society for funding of this project.

Thanks my friend and flatmate Rebekka Pilz for taking care of the cats, when I am busy in the lab or writing.

I would like to take this opportunity to express my heartfelt thanks and appreciation to my parents, Wolfgang and Annemarie, for their absolute support and for understanding that time is sometimes very limited.

A very deep and special gratitude goes to my love, Jens Maintz, who always gives me huge motivation and who supported me a lot in the end phase of this thesis.

Erklärung

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

Teilpublikation:
Ghandilyan, A., Ilk, N., Hanhart, C., Mbengue, M., Barboza, L., Schat, H., Koornneef, M., El-Lithy, M., Vreugdenhil, D., Reymond, M., u. a. (2009). A strong effect of growth medium and organ type on the identification of QTLs for phytate and mineral concentrations in three Arabidopsis thaliana RIL populations.
J. Exp. Bot, **60**, 1409-1425.

Nadine Ilk
Köln, April 2012