Genes that underlie natural variation in growth rate and flowering time in local accessions of *Arabidopsis thaliana*

Zafar Iqbal Malik

Thesis submitted for the degree of

Doctor of Philosophy

University of Edinburgh

2014

Abstract

Growth rate and flowering time are agriculturally important traits that are linked to fitness, productivity and reproductive success of plants. To study the genetic basis for natural variation in growth rate and flowering time between local accessions of *Arabidopsis thaliana*, hybrids were produced between fast growing / late flowering and slow growing / early flowering parents. F3 and F5 hybrid families were grown under a range of conditions – under a constant controlled environment, outside over the winter and outside in spring and early summer. Growth rates were estimated from repeated images of rosettes. Flowering time, as number of leaves to flower, was also recorded both in control and natural conditions for F5 lines. Damage by slugs and stress-induced production of anthocyanin pigments were also recorded for plants grown outside. Broad-sense heritability estimates were higher for F5 families than F3, in which more loci will segregate, and ranged from 48% to 89%. No significant correlation between growth rates under different environments was observed in most cases for F3 populations, however significant correlations were detected for F5 families outside and under controlled conditions, suggesting that same genes can affect growth rate in more than one environment. The genotypes of F3 families were determined at thirty-nine SSLP (simple sequence length polymorphism) loci and used in regression with phenotype data to search for quantitative trait loci (QTL). Significant QTLs were detected in F3 families for growth rate, flowering time and anthocyanin production, but not for herbivore damage. To confirm QTL detected in the F3 and to detect additional loci, bulk segregant analysis was carried out in F5 families grown under different conditions. Potentially linked markers were tested further in individual F5 plants and QTL mapped on a finer scale in

i

F5 families that remained heterozygous for candidate regions. *VIP5* and *LDL1* were selected as potential candidate genes for flowering time variation. These genes were sequenced for two parental alleles. A transposon insertion and 5' UTR deletion were found in the *LDL1* allele from the late flowering parent and SNPs (single nucleotide polymorphisms) were observed throughout the gene. However both alleles appeared to be expressed at similar levels. Transgenic lines have been produced carrying the *LDL1* allele from the early flowering parent (4D1) in the background of the later flowering parent (11C1). This work is on-going and will hopefully reveal whether *LDL1* underlies differences in flowering behaviour seen between 11C1 and 4D1.

Lay summary

This thesis demonstrates that there is an extensive genetically-determined variation in growth rate and flowering time between two local accessions of *Arabidopsis thaliana*. To understand the genetic basis for this variation, hybrids between the accessions were grown in a range of natural and controlled conditions. In hybrids, heritability of growth rate and flowering time was lower in autumn in natural conditions, which might reflect longer exposure of plants to variable environments. Correlation between the growth rates of hybrid lines in different conditions suggests that similar genes might influence growth rate in different conditions. Anthocyanin production over winter showed high heritability and a significant correlation with growth rate, suggesting a common genetic basis. Similarly, susceptibility to herbivory also correlated with growth rate. QTL mapping was undertaken in F3 and F5 hybrid populations. This identified significant QTLs for flowering time variation, growth rate and anthocyanin production. Two marker loci, *CIW1* on chromosome 1, and *CIW4* on 3, showed significant association to growth rate in F3, F4 and F5 populations and *CIW1* to flowering time in some segregating F5 families. Two candidate genes for flowering variation - *LDL1* and *VIP5* were investigated further. These genes were sequenced for the parental lines. Besides SNPs in both genes, a 24 bp deletion at the start of *LDL1* gene and 882 bp insertion of a transposon in the 5'UTR were found in the allele from the late flowering parent. Although both alleles were expressed at a similar level, the 5'-RACE PCR showed that the late flowering parent produced only a shortened transcript, suggesting that it might carry an *ldl1* loss-of-function mutation. This could account for the later flowering of this parent and possibly also for its faster growth rate.

Declaration

I hereby declare that I am the author of this thesis. The work presented is my own original work and I have not previously submitted this work in its entirety or in part, for any other degree or professional qualification.

_____

Zafar Iqbal Malik

Date:_____

Acknowledgements

Table of contents

List of figures

List of tables

Abbreviations

~	Approximately
A* after	Area after
A before	Area before
A obs.	Observed area
bp	base pairs
^{o}C	Centigrade scale
c	y intercept of the straight line graph
CIM	Composite Interval Mapping
cM	Centimorgan
CTAB	Hexadecyltrimethyl Ammonium Bromide
Conc.	Concentration
DEPC	Diethylpyrocarbonate
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
e.g.	For example
FAM	Fluorescein amidite
Fig	Figure
GH	Green house
GR	Growth room
HCl	Hydrochloric acid
Hz	Hertz
Hrs.	Hours
IT	Image Tool

LB	Lysogeny broth (Luria-Bertani broth)
LDL1	LSD1-like-1
LOD	Logarithm of odds
LSD1	Lysine-specific demethylase -1
m	Meter
m	Gradient of the straight line graph
M	Molar or molarity
mg	Milligram
min	Minute
ml	Millilitre
mm	Millimetre
μM	Micro molar
mM	Milli molar
M	Mole
NaOAc	Sodium acetate
NaCl	Sodium Chloride
OS	Outside / natural conditions
P	Phosphorus
QTLs	Quantitative Trait Loci
RACE	Rapid amplification of cDNA ends
rpm	Revolutions per minute
RT-PCR	Reverse transcription polymerase chain reaction
SDS	Sodium Dodecyl Sulphate
Sec	Seconds
SNPs	Single nucleotide polymorphisms

SSLPs Simple Sequence Length Polymorphism

Ta Annealing temperature

TE A buffer derived its name from its components Tris and EDTA

Tris Tris (hydroxymethyl) amino methane

μl Micro litre

UTR Un-transcribed region

UV Ultraviolet

VIP5 Vernalization independence

4 (4D1) Slow growing and early flowering genotype

11 (11C1) Fast growing and late flowering genotype

1.0 INTRODUCTION

Arabidopsis is a major model plant which is suited for genetic and molecular studies. The completely known genomic sequence and data for gene disruptions make it a unique research resource. It makes it possible to analyse the function of individual genes or group of genes by studying the phenotypes of mutants. This study of the function of individual genes has increased our knowledge but still there is a need to study the genes, gene combinations and gene products that affect complex traits.

1.1 Growth and growth rate

Growth is usually defined as an irreversible increase in the size of a cell, organ or whole organism. In plants, at early stages of development, it is often accompanied by cell division, although the later growth of many organs, including leaves, can involve only cell expansion (Jackson, 1980). The increase in size of an organism or part of an organism per unit time is referred to as the growth rate. Growth of an organism may stop at maturity, as in case for most mammals, or it may continue throughout life, as in some plants. Growth is highly regulated and coordinated throughout plant development. Two levels of growth coordination can be distinguished in plants, one is the final size (Paul and Foyer, 2001), the second being the specific patterns of organ formation (Beemster et al., 2005). Growth rate is therefore a complex trait in plants that is affected by variation in cell division, cell expansion, rate of photosynthesis and organogenesis (Zhang et al., 2012) and it occurs at different levels ranging from cells to the whole body. Leaf area expansion, a proxy of leaf growth that can be measured using a non-invasive image analysis approach (Arvidsson et al., 2011; El-Lithy et al.,

2004; Granier et al., 2006), was shown to be controlled by many significant genetic factors (El-Lithy et al., 2004; Massonnet et al., 2010; Tisne et al., 2008). Growth rate is also an environmentally sensitive trait interconnecting cell biology, organogenesis, and physiology.

1.1.1 Phenotyping aims

The basic aim of phenotyping in plants is to quantify the different morphological traits (Walter et al., 2012) to investigate basic principles involved or in selection of specific genotypes in plant breeding programmes. Automation of this technology is currently improving day by day, for example for measuring the total leaf area of smaller plants like *Arabidopsis* (Leister et al., 1999) or to measure a number of traits simultaneously and repeatedly at high throughput (Rajendran et al., 2009). At the same time, field monitoring and imaging technologies have also been improved to provide meaningful information about plants phenotypes in the field (Montes et al., 2007), which is important in breeding programmes. Recording of a set of environmental parameters in an experimental period is also required to analyse genotype $\times$ environment interactions.

1.1.2 Tools for phenotyping

Various strategies have been adopted to measure the morphological traits of plants. The most simple and traditional ones are the use of manual devices to take measurements of traits. Reliability in these cases is always questionable and measurements may cause injuries and stress to the plants. This is particularly important when studying traits that involve changes over time, for example growth rates. To quantify morphological traits with more accuracy, automatic phenotyping

approaches are needed. One of the modern technologies, ubiquitous sensor networks (USN) used fixed and rotary sensor-based measurement (Suk et al., 2011). Automated techniques have been applied to measure different traits in *Arabidopsis*, for example, growth of the root (Miller et al., 2010), hypocotyl (Cole et al., 2011) or seedling (Walter et al., 2007), and pathogen resistance (Berger et al., 2007). The uses of digital images of rosette area to estimate growth rates have been used successfully in *Arabidopsis* and in many crop studies (Campillo et al., 2010).

1.2 Plant growth parameters

Growth parameters are the characteristics by which the growth and growth rate of plants can be measured. These include rate of increase in plant height or shoot length, girth, number of leaves, leaf chlorophyll, leaf colour, fresh and dry weight, and rosette area of leaves (Leister et al., 1999; Li et al., 1998). The rate of leaf production as a growth rate measurement has also been used in small herbaceous plants. Fresh and dry weights of plants can also be measured (Wood and Roper, 2000), however these has the disadvantage that repeated measurements cannot be made on the same plants, introducing another source of variance in estimates.

1.2.1 Plant height and shoot length in *Arabidopsis*

Plant height is an important component of plant architecture that is highly correlated with biomass yield in many species. Biomass and height are significantly and positively related with each other in maize, *Zea mays L.* and sorghum (Murray et al., 2008). Plant height is also positively correlated with flowering time in sorghum (Ritter et al., 2008). Plant height is often used to estimate the growth of trees. Variable techniques are used to measure the height of trees depending upon number of

factors like species, location and habits etc. Various factors have been found to increase the rate of increase in tree height, like increasing nitrogen and phosphorus supplies by fertilizer application (Assuero et al., 2004; Lynch et al., 1991). The ratio between root and shoot growth is also key parameter in crop production, life histories and responses to unfavourable conditions (Poorter et al., 2005). For example, drought tends to increase root-to-shoot biomass across a range of species.

Significant variation was observed between *Arabidopsis thaliana* genotypes in branch number and shoot architecture (Ungerer et al., 2002). There are also evidences of secondary growth in *Arabidopsis* shoot, hypocotyl and root (Ko and Han, 2004; Melzer et al., 2008; Sibout et al., 2008), though these aspects of growth remain poorly understood.

1.2.2 Measuring the rosette areas of plants

Measuring fresh and dry weight of plants as growth rate, are destructive methods and need a large number of plants to obtain accurate estimates. A non-destructive approach will always be preferable, e.g. using image analysis. Plant growth rate analysis was done successfully by measuring the area of whole plants (Motooka S, 1991). For *Arabidopsis*, which in its vegetative phase grows as a flat rosette with limited leaf overlap, the use of digital video and image analysis has been effective in estimating plant growth rate non-destructively, even during early developmental stages (Leister et al., 1999). Digital image analysis has become an important tool in biological research over scales ranging from satellite images to micrographs (Nilsson, 1995). For measuring root or shoot length, areas of leaves or whole plants, images are taken at definite time points (e.g. daily or weekly, depending on the rate of growth) and converted to binary images that distinguish the plant from its background. The

4

area, shape or length of the plant or its organs can then be analysed quantitatively from these images. There are number of companies and research groups that have set up automated platforms to calculate root elongation and architecture (Armengaud et al., 2009; Hund et al., 2009; Nagel et al., 2009), root gravitropic curvature (Miller et al., 2007), the projected area of single leaf (Granier et al., 2006) or total leaf area (El-Lithy et al., 2004; Granier et al., 2006; Leister et al., 1999; Walter et al., 2012) and shape parameters for leaves or rosettes (Jansen et al., 2009).

1.2.3 Single time point rosette area as proxy for growth rate

Growth can be measured in a number of ways. Absolute growth usually refers to the rate of change in size over a particular period, for example the final height of a plant divided by the time taken to reach that height. However, growth of many organs is not linear with time and often approaches exponential during periods of fastest growth. Therefore the concept of relative growth rate (RGR) is often used. This measures the rate of increase in size or mass per unit of size or mass already present (Hoffmann, 2002) . For example,

$RGR = \ln(M2) - \ln(M1) / (t2-t1)$

Where $\ln(M1)$ and $\ln(M2)$ are the natural logs of the plants' dry masses at times t1 and t2, respectively. RGR is therefore equivalent to the slope of a plot of $\ln(M)$ against time.

However, it is important to acknowledge a potential disadvantage of RGR; it involves at least two estimates of area, each with an associated error (Hoffman & Poorer 2002). This could potentially lead to a less accurate estimate of genetically determined growth rate than an estimate made from a single measurement. J. Atkinson showed (PhD thesis submitted to The University of Edinburgh, 2006) that rosette area was a

good proxy for mass as it was directly proportional to fresh weight with an r^2 value of

0.89 (Figure 1).
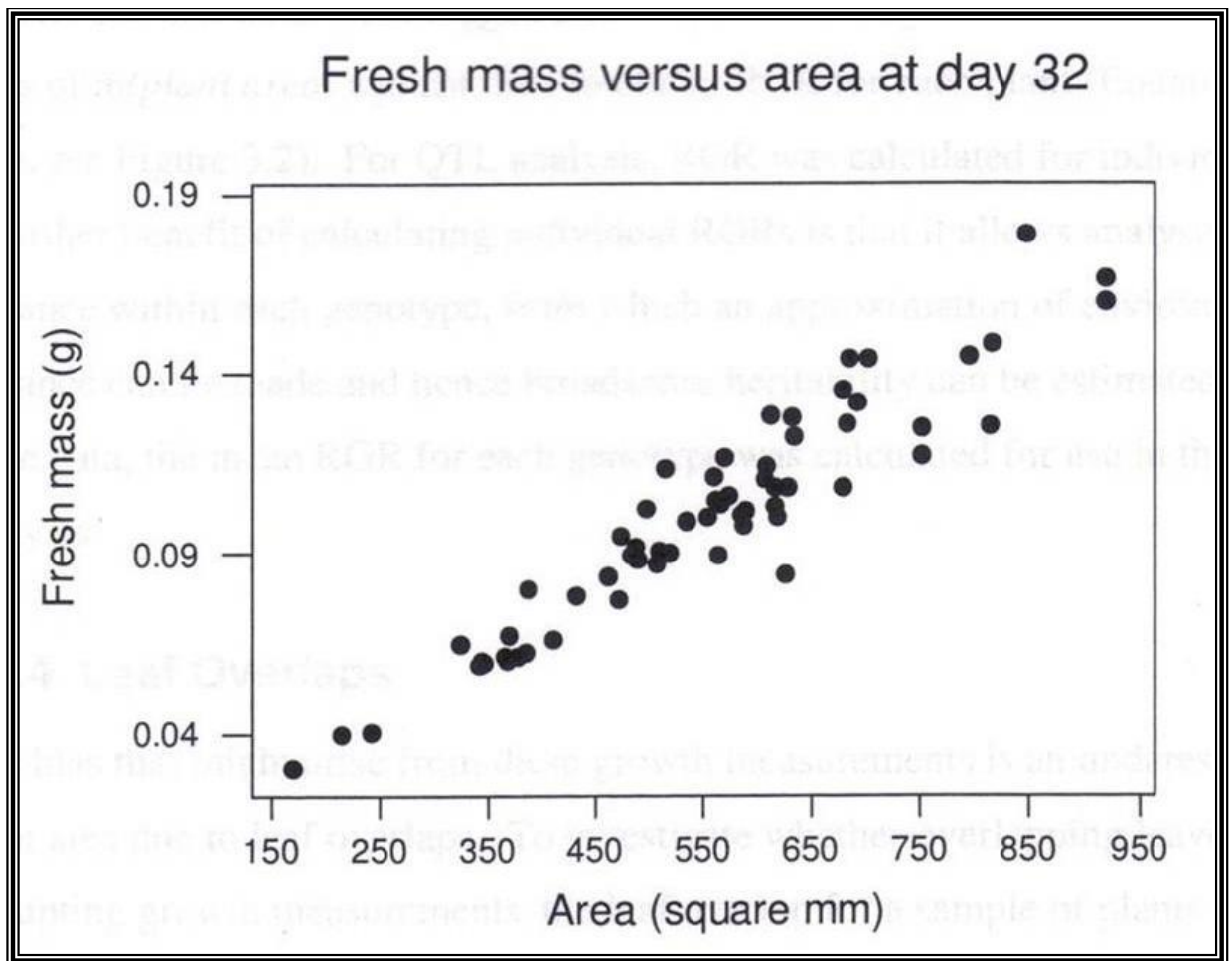
Figure 1: J. Atkinson recorded regression of fresh mass against plant rosette area at 32days.

Plant rosette area is significantly correlated with fresh mass ($r2 = 0.89$) and dry mass ($r2 = 0.96$). 119 plants were used in the experiment (J. Atkinson, PhD thesis to The University of Edinburgh, 2006).

Later stage rosette areas increase in some genotypes did not reflect earlier stage growth rates (Poay Lim, PhD thesis to The University of Edinburgh, 2013). It is highly desirable to measure rosette areas of plants at an early stage of development with minimal effects of other factors on growth like bolting or degree of leaf overlap. It is always necessary to set a time point of area measurement where plants were grown to size that allowed more accurate estimates with minimal leaf overlapping. Poay Lim (PhD thesis to The University of Edinburgh, 2013) showed that different genotypes grow at different rates and continuous to increase with time under uniform conditions before flowering (Figure 2). Rosette areas of genotypes (if germination starts at same point) were compared within an experiment as growth rate, so it is relative growth rate not absolute.
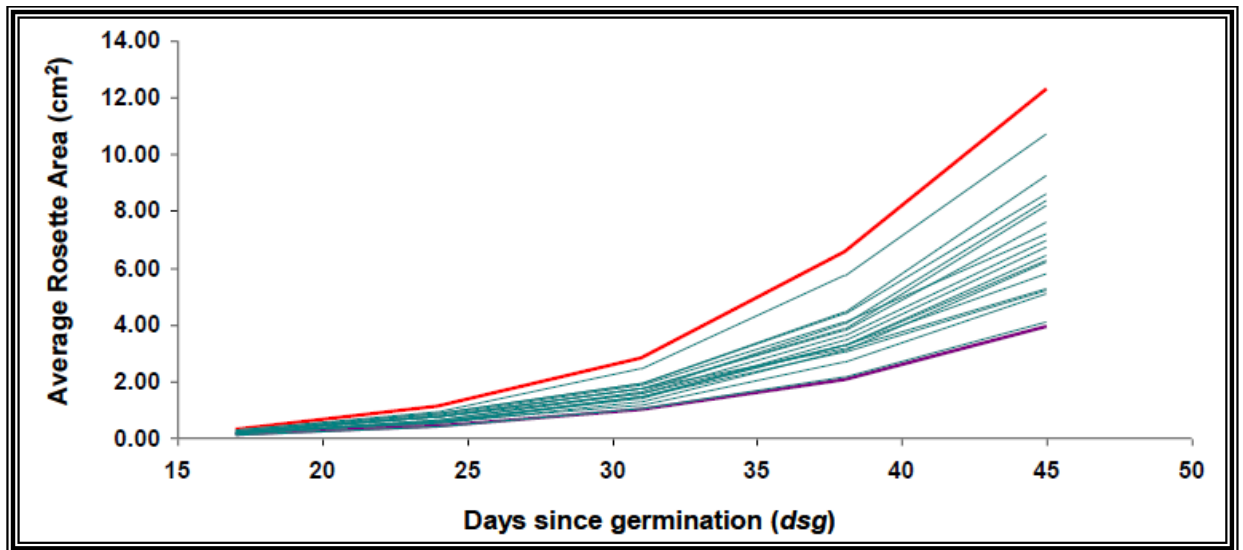
Figure 2: Poay Lim (PhD thesis to The University of Edinburgh, 2013) findings of relationship of average rosette areas of plants to days since germination.

Rosette areas of genotypes were compared within an experiment as growth rate, so it is relative growth rate not absolute.

1.3 Natural variation and evolution

Intraspecific natural variation may be broadly defined as the phenotypic variation within-species that is caused by spontaneously arising mutations. This variation might be maintained by artificial or natural selection or might accumulate in the absence of selection (under neutrality or drift) (Alonso-Blanco et al., 1999). Phenotypic variation which is due to heritable genetic variation is a fundamental prerequisite for evolution by natural or artificial selection. It is the living organism as a whole that contributes (or not) to the next generation, so natural selection affects the genetic structure of a population indirectly via the phenotypes. No genetic variation means no heritable phenotypic variation, no adaptation or drift and hence no evolution.

The interaction between genotype and phenotype has often been conceptualized by the following relationship:

Genotype + environment → phenotype

Or, to account for variation which is not caused by genotype or environmental conditions as:

Genotype + environment + random-variation → phenotype

The phenotype variation, whether in wild or cultivated plants, may be discontinuous if it affected by very few genes that have large effects on the phenotype, relative to the environment. For example, resistance to a particular genotype of a pathogen can be determined by a single resistance (R) gene. However, the majority of traits are determined by multiple genes, which have small effects relative to each other or to variation cause by the environment. The trait therefore typically shows continuous

10

variation within a population and, because it cannot easily be categorised, is usually measured quantitatively. Such traits are therefore referred to as quantitative traits and the underlying genes as quantitative trait loci or QTL (Buescher et al., 2010a; Robertson A 1967).

1.4 Quantitative traits and quantitative trait loci

Although a QTL represents one or more genes within a particular chromosome region, very few QTL have been localised to a gene. Therefore QTL is usually taken to mean a chromosome region that affects the trait value. QTL analysis is a statistical method that links two types of information - phenotypic data (trait measurement) and genotypic data (Kearsey and Farquhar, 1998). Molecular markers are used for genotyping because they are produced at high densities and are unlikely to affect the trait of interest. Several types of markers are commonly used in genotyping including single nucleotide polymorphisms (SNPs), simple sequence repeats (SSRs or microsatellites), restriction fragment length polymorphisms (RFLPs) and transposable element positions (Gupta and Rustgi, 2004). To carry out the QTL analysis, the parental strains are crossed, obtaining F1 individuals which are then selfed or crossed using one of a number of different schemes to generate a segregating population (Darvasi, 1998). Finally, phenotypes and genotypes of the segregating population (e.g., F2, F3 or back-cross generation) are scored. Markers that are genetically linked to a QTL influencing the trait of interest should show a correlation between genotype and phenotype values, whereas unlinked markers will not show significant association with phenotype. It is possible to account for fixed-effects (e.g., known environmental differences or sex) in the analysis. QTL may also be used for traits in which phenotypes are affected by interactions – for example between QTL (epistasis),

between QTL and the environment (by-environment interactions) or between QTL and the sex of the individual (genotype-by-sex), although these might require a specialist experimental design or the use of more individuals to detect significant interactions.

When related traits are found to have QTL that map to similar positions, the variation for these traits might be controlled by the same gene - in genetic terminology pleiotropic. One example of a pleiotropic QTL is the *Arabidopsis ERECTA* gene, which encodes a receptor kinase. Because a mutant allele is present in the *Landsberg erecta* parent of many QTL mapping populations, it has been found to affect many different quantitative traits including shade-avoidance (Patel et al., 2013), resistance to bacterial and fungal pathogens (Godiard et al., 2003; Llorente et al., 2005) and transpiration rate (Masle et al., 2005) in addition to plant morphology.

In practice the ability to detect a QTL and to map it accurately depends on a number of factors. These include the heritability of the trait (the extent to which the phenotype is determined by genotype), the magnitude of the effect of the QTL, the number of recombinant chromosomes being analysed, which is related to the number of hybrid individuals used in mapping, and the density of molecular markers.

1.4.1 QTL analysis in plants

Development of molecular markers that allow cost effective genotyping of mapping populations has allowed QTL analysis to be applied to a wide range of plant species. Many of these studies have been carried out on crop plants because identification of molecular markers that are linked to desirable traits allows the markers to be used to select the traits in breeding programs. For example, SSRs markers have been used to

detect QTLs for fruit size and quality traits in apricot (Ruiz D, 2010) and different types of populations used to map QTLs for fruit weight, firmness, flavour and other traits in tomato (Bertin et al., 2009; Causse et al., 2002) and grapes (Mejía et al., 2007) or kernel shape and size in barley (*Hordeum vulgare* L.).

1.4.2 Natural variation in plant growth rate

Arabidopsis thaliana is predominantly a selfing species, and therefore most plants collected in nature represent inbred lines. These wild homozygous lines are commonly referred to as ecotypes. However the use of this word in *Arabidopsis* does not strictly conform to its ecological definition – a genotype that is adapted to specific local conditions. Because the ecological meaning has been lost, the term accession, as it is often used in germplasm collections, is becoming more commonly used to refer to a sample collected at a specific location (Alonso-Blanco and Koornneef, 2000). Genetic variation within populations of *Arabidopsis*, and especially among populations, has been studied extensively using a variety of tools. *Arabidopsis* has been used extensively for QTL studies. Its complete genome sequence has allowed development of a large number of genetic markers at known positions in chromosomes. Its short generation time allows rapid generation of segregating mapping populations – recombinant inbred lines (RILs), near-isogenic lines (NILs) etc. and its small size allows many plants to be grown in a limited space. In addition, the functions of many *Arabidopsis* genes are known from mutations, allowing identification of candidate genes for QTL. Many of these studies have used genotyped RILs that are available as a public resource. QTL studies have examined a large number of traits in *Arabidopsis*, including flowering time, seed germination and plant architecture etc.

The extensive natural variation that occurs in *Arabidopsis* is being exploited increasingly as a source of genetic variation for the analysis of important adaptive traits, e.g. flowering time, seed size, seed dormancy, pathogen resistance, and tolerance to abiotic stresses etc. (Alonso-Blanco and Koornneef, 2000; Koornneef et al., 2004). Recombinant inbred lines (RILs) provide an immortal population, as each individual is practically homozygous, and large numbers of genetically identical individuals can be obtained, allowing repeated measurements of various traits in different conditions (Alonso-Blanco and Koornneef, 2000; Doerge, 2002)

Plants raised from seeds collected in nature have been analysed phenotypically for numerous characteristics including morphological traits and flowering time; and genotyped with a range of different markers like microsatellites (Vander Zwan et al., 2000) restriction fragment length polymorphisms, mainly analysed as cleaved amplified polymorphic sequences (CAPS) markers (Ullrich et al., 1997), amplified fragment length polymorphisms (AFLPs) (Sharbel et al., 2000) and extensive DNA sequencing (Schmid et al., 2003). Currently 1001 accessions of *A. thaliana* are being resequenced (www.1001.org). Most studies have focused on detecting polymorphisms in single copy nuclear sequences. However, variation has also been studied in including mtDNA (Ullrich et al., 1997), and in repeated sequences, including centromere repeats (Hall et al., 2003) and transposable elements (Frank et al., 1998). Cytogenetic polymorphisms have also been found, but these have been studied only in a limited number of accessions (Koornneef et al., 2004). From all these analyses, we can conclude that within populations of *Arabidopsis,* polymorphisms are relatively frequent. In initial analyses, the relatively high genetic variation found among *Arabidopsis* accessions using genome-wide markers such as

AFLP, CAPS and microsatellites (Ullrich et al., 1997) shows that in general there was no association (or weak association) between geographical origin and genetic distance. Genetic isolation-by-distance (an inverse correlation between geographic distance and sharing of polymorphisms) was found by analysing 79 AFLP markers in a worldwide collection of 142 accessions, with a major representation from central Europe (Sharbel et al., 2000). These authors suggested that after the last glaciation *Arabidopsis* colonized Central and Northern Europe from Asia and from Mediterranean Pleistocene refugia. More recent studies involving at large number of SNPs and world-wide accessions, however, have identified a correlation between genetic distance over a range of geographic scales (Nordborg et al., 2005; Pico et al., 2008; Platt et al., 2010). More than 25 genes have been systematically sequenced and compared in various accessions and the number of genes for DNA sequence comparisons among accessions is increasing rapidly (Tian et al., 2002). These include mainly floral and meristem developmental genes, pathogen resistance and defence genes, and genes encoding metabolic enzymes. Sequence analyses of individual genes revealed nucleotide diversity values ranging from 0.0006 for *ATTI* to 0.0558 for *CLV2* with an average value of 0.006. Similar levels of variation have been estimated by analysing 606 sequence tagged sites (STS) in 12 accessions, where the mean sequence divergence to the Col accession was 0.68 (Clauss and Mitchell-Olds, 2003).

The genetic architecture of growth rate and size variation has also been studied in *Arabidopsis*. Several QTLs were detected for shoot growth variation by using Bur-0 × Col-0 RILs (Vlad et al., 2010). By using a recombinant inbred line population derived from L*er* and Cvi, quantitative trait loci (QTLs) were mapped that affected 12 life history traits related to seed size, fruit size, seed number, and plant resources

(Alonso-Blanco et al., 1999). Quantitative trait locus (QTL) analysis, using 114 (F9 generation) recombinant inbred lines derived from the cross between *Ler* and Shahdara, revealed QTLs for seed weight, plant area, dry weight, relative growth rate, chlorophyll fluorescence and flowering time. Growth traits (plant area, dry weight, and relative growth rate) co-located at five genomic regions. At the bottom of chromosome 5, co-location was found of QTLs for leaf area, leaf initiation speed, specific leaf area and chlorophyll fluorescence, indicating that this locus might be involved in leaf development. No consistent correlation between growth traits and flowering time was observed despite some co-locations. Some of the QTLs detected for flowering time overlapped with loci detected in other recombinant inbred line populations, but also new loci were identified (El-Lithy et al., 2004). This suggests that variation in growth rate can involve different genes in different accessions. Phenotypic variation for shoot growth in the Bur-0 × Col-0 RIL population was mapped to several QTLs. Using a fine-mapping strategy, one of the QTL was mapped to At4g30720, which encodes a new chloroplast-located protein essential to ensure a correct electron flow through the photosynthetic chain and, hence, photosynthesis efficiency and normal growth (Vlad et al., 2010).

Many environmental and non- environmental factors influence these growth parameters (Massonnet et al., 2010; Pereyra-Irujo et al., 2008), thus affecting plant growth.

Several QTL studies have examined the control of root growth rate. QTLs were mapped in the Bay-0 x Sha RIL population for effects on the growth of roots. One locus was found to be responsible for approximately 80% of the variance of the observed difference in root length in the RIL population. This gene, named *BREVIS*

16

RADIX (*BRX*), controls the extent of cell proliferation and elongation in the growth zone of the root tip. *BRX* was isolated by positional cloning and found to encode a member of a small, plant specific family of proteins. Analysis of *Arabidopsis* mutants lacking activity of the other genes suggested that *BRX* is the only gene of this family with a role in root development (Mouchel et al., 2004). An additional 12 QTLs for primary root length, lateral root number and density and for total length of the lateral root system were mapped and one epistatic interaction between QTLs was detected. Sha QTL alleles always increased the length of the lateral roots which may be taken as an adaptation to its very dry natural environment in Tajikistan (Loudet et al., 2005). The same population was used to map genes involved in the root growth response to low phosphate. One of these QTL, *LRPI*, explains 52% of variance of primary root length changes in response to phosphate. A single QTL was also detected for primary root cell elongation in response to low P value (Reymond et al., 2006).

Differences in the accumulation of elements have been found between accessions and over a hundred QTLs for elemental accumulation estimated in RIL populations. A strong effect of plant environment was observed between elements and QTLs controlling elemental accumulation (Buescher et al., 2010b).

1.4.3 Usefulness of RILs in QTLs study

Recombinant Inbred Lines (RILs) are very useful in QTL analysis as each represents a unique combination of parental genotypes. Replicates of each genotype can therefore be grown in the same condition, to increase the accuracy of estimates of genetically-determined phenotype values and reduce the effects of non-genetic variation. They can also be used for the analysis of traits in many environments. During the production of RILs, there are additional opportunities for recombination during the

selfing generations, and therefore a chromosome in a RIL population will typically carry twice as many recombinations as a chromosome in an F2 population, though each chromosome is likely to be homozygous in a RIL. A further improvement is in the advanced intercross approach, in which individuals from the F1 or F2 population are randomly intercrossed, thus increasing the opportunity of recombination before genotypes are fixed by selfing (Darvasi and Soller, 1995). It is also possible to incorporate more than two parental genomes in the population by inter-crossing F1 hybrids from different parents to generate multi-parent advanced genetic inter-cross (MAGIC) lines. Several *A. thaliana* RIL populations have been used for QTL mapping and subsequent molecular identification of the responsible genes. However, only relatively few large populations have been densely genotyped, limiting the resolution of QTL maps (Rosloski et al., 2010). The disadvantages of RILs are that they do not allow the dominance of QTL alleles to be examined, because they consist only of homozygous genotypes and also they have not been subjected to selection. Dominance relationships can be used to infer whether a QTL allele might involve a reduction in activity including a loss of function.

1.5 Flowering

Flowering is a crucial event in the life of plants which is linked to their reproductive success and fitness. Flowering time is controlled by hundreds of different genes in different plant species. Thus so-called flowering-time genes have been placed in a number of genetically defined pathways that integrate external stimuli such as photoperiod, ambient temperature, or prolonged exposure to cold, with endogenous signals including phytohormones and plant age (Amasino and Michaels, 2010; Bäurle and Dean, 2006; Greenup et al., 2009; Kobayashi and Weigel, 2007; Turck et al.,

2008). These pathways promote, repress or initiate flowering in quantitative manner (Coupland, 1995; Levy and Dean, 1998; Simpson and Dean, 2002). The most important environmental factors that influence the initiation of flowering are day length (photoperiod), ambient temperature and prolonged exposure to cold (vernalization). These factors not only change with seasons but also with geographical region

1.5.1 Genetic control of flowering in *Arabidopsis*

Studies have shown that flowering in *Arabidopsis* is regulated by range of internal and external stimuli. Flowering time genes identified so far have been placed into four main pathways: the autonomous, vernalization, photoperiod and gibberellic acid (GA) pathways. The vernalization and autonomous pathways control activity of the flowering repressor encoded by the *FLOWERING LOCUS C* (*FLC*) gene, so we categorise flowering time pathways in *Arabidopsis* into *FLC*-dependent and *FLC*-independent pathways (i.e., GA and photoperiod pathways).

1.5.2 *FLOWERING LOCUS C* (*FLC*)

FLOWERING LOCUS C (*FLC*) is a MADS-box gene that encodes a well characterized floral repressor. Related MADS box genes, from *MADS AFFECTING FLOWERING 1* (*MAF1* or *FLM*), *MAF2* and *MAF5* have been found to have similar roles in flowering (Ratcliffe et al., 2003; Ratcliffe et al., 2001).

When plants are exposed to cold, *FLC* expression is reduced. If this reduced level of expression is maintained, there will be early flowering (Sheldon et al., 2000). Flowering in *Arabidopsis* is also regulated through transcriptional and epigenetic control of *FLC* by the autonomous pathway. The first intron of *FLC* is 3.5Kb long

which makes about half of the whole gene (Hanada et al., 2007). Human genes having long introns were found to have more conserved sequences than average which indicated that they might be involved in regulation (Sironi et al., 2005).

1.5.2.1 Regulation of *FLC* by FRIGIDA (FRI)

FLC expression is up-regulated by the FRI protein that consists of double coiled domains and has 609 amino acids. It acts as a complex consisting of FRI, FRL1 and FRL2 (FRIGIDA LIKE1 and 2) and the C2H2 zinc-finger protein SUPPRESSOR OF FRIGIDA4 (SUF4). Mutation in any of the genes encoding members of this complex results in lower *FLC* expression. The complex activates *FLC* transcription by direct binding to its 5' promoter region (Sheldon et al., 2002).

There are also other components that are involved FRI-mediated up-regulation of *FLC* that are similar to SWR1 and Paf1C in yeast. SWR1-like members consist of PHOTOPERIOD-INDEPENDENT EARLY FLOWERING 1 (PIE1), ACTIN RELATED PROTEINS 6 (ARP6) and SERRATED LEAVES AND EARLY FLOWERING (SEF) and Paf1 C consists of VIP4 and VIP5 (VERNALIZATION INDEPENDENCE4 and 5) and ELF7 and ELF8 (EARLY FLOWERING7 and 8). All these genes have a more general role than FRI in gene expression regulation (Choi et al., 2007; March-Diaz et al., 2007).

FLX (FLC EXPRESSOR), a leucine zipper-containing protein, has also been found to regulate *FLC*. Plants carrying *flx* mutation have reduced levels of *FLC* expression, while other regulators of *FLC* remain unaffected (Andersson et al., 2008). The effect of the *flx* mutation is reduced if there is also a mutation in the autonomous pathway consistent with FLX acting together with FRI. However, FLX also promotes the expression of *FLC*-like *MAF1* and *MAF2* genes, unlike FRI (Andersson et al., 2008).

1.5.2.2 The Autonomous pathway

Components of the autonomous pathway control flowering via RNA processing and epigenetic regulation of *FLC*. The known autonomous pathway components are: FCA, FY, FPA, FVE, FLD and FLK (FLOWERING LOCUS CA, Y, PA, VE, LD and LK) and LUMINIDEPENDENS (LD). FLD and FVE regulate *FLC* via histone deacetylation, independently of all other components in the pathway (Ausin et al., 2004; He et al., 2003). FCA and FY are RNA binding proteins and their interaction helps in determination of polyadenylation site selection in antisense transcripts from *FLC*, so mutation in *FCA* and *FY* produce more transcripts that are antisense to the *FLC* promoter and this might enhance *FLC* expression (Hornyik et al., 2010; Quesada et al., 2003; Simpson, 2004). However, activity of both these proteins is not limited to *FLC*. FPA and FLK are also RNA binding proteins that work independently from each other and from FCA and FY (Lim et al., 2004; Schomburg et al., 2001a). FPA has a similar effect on polyadenylation site selection to FCA and FY, though it functions independently of them (Hornyik et al., 2010). *LD* is known to encode a homeodomain protein. Although homeodomains are typically associated with binding to DNA, there is evidence that they can bind to RNA and this would be consistent with regulation of *FLC* by different non-coding RNAs (Dubnau and Struhl, 1996; Lee et al., 1994a; Rivera-Pomar et al., 1996).

Components of the autonomous pathway are widely conserved in flowering plants whereas *FLC* is not; as yet orthologs of *FLC* have been identified only in the Brassicaceae. This suggests that the autonomous pathway might have more general role in regulation of gene expression than just flowering.

1.5.2.3 The vernalization pathway

The process by which flowering is promoted as plants sense exposure to the cold temperatures of winter, vernalization, is one of the systems plants have developed to sense environmental cues and modify their growth and development accordingly. Vernalization is an adaptive trait that helps to prevent flowering before winter and permits flowering in the favourable conditions of spring. It has been known for many years that a vernalized plant can maintain the memory of this vernalization mitotically under warmer conditions and that an unvernalized state is reset during development of the embryo (Sheldon et al., 2008). The autonomous and vernalization pathways functions to suppress *FLC* but there is evidence that some vernalization responses are independent of *FLC*. For example, *MADS AFFECTING FLOWERING2* (*MAF2*) was shown to be involved in preventing flowering after short periods of cold - *maf2* mutants showed the same response to 10 days of cold as wild type plants after 85 days of cold (Ratcliffe et al., 2003) .

Additionally ectopic expression of the *SOC1* homolog *AGAMOUS LIKE 19* (*AGL19*) has been shown to result in rapid induction of flowering. *AGL19* expression increases in response to cold independently of *VERNALIZATION INSENSITIVE 3* (*VIN3*), which is necessary for the repression of *FLC* by cold, as *agl19 vin3* double mutants have a more impaired vernalization response than either of the single mutants. In contrast to *FLC*, the response of *AGL19* to vernalization is transient and is reversed once plants are returned to the warm (Schonrock et al., 2006).

Several studies have shown that vernalization regulates flowering time and *FLC* expression in a dose-dependent manner; longer cold periods giving lower *FLC* mRNA levels and shorter times to flowering (Michaels and Amasino, 1999a; Sheldon et al.,

2000). However there is also evidence to suggest that it is not the absolute amount by which cold initially reduces *FLC* expression that is crucial for the effect on flowering time but the stability of that repression if the plant is returned to warm conditions (Shindo et al., 2006). In natural Swedish accessions of *Arabidopsis* neither the initial unvernalized level of *FLC* nor the decrease in its RNA levels in response to cold correlated with flowering times. Instead a correlation was found with the length of vernalization required for stable repression of *FLC* so that expression did not increase again if plants were returned to the warm. Three genes are known to be involved in vernalization pathway: *VERNALIZATION1*, *2* and *3* (*VRN1 VRN2* and *VIN3*). *VIN3* encodes a protein containing a plant homeodomain (PHD) and fibronectrin type III repeats, usually involved in protein-protein interactions. It is induced in response to cold and required to establish cold-induced epigenetic silencing of *FLC*. *VRN1* encodes a Myb-related protein and *VRN2* a component of the Polycomb repressive complex 2 (PRC2), which acts by methylating lysine residues of histone H3 in chromatin, which is thought to lead to stable repression (Gendall et al., 2001; Levy et al., 2002).

1.5.3.4 Histone Deacetylase (HDAC) co-repressor complex

A number of nueronal genes are silenced in non-nueronal tissues during differentiation in mammalian development. A number of these genes contain a 23 bp regulatory element called repressor element-1 (RE1) (Ballas et al., 2001; Maue et al., 1990). This RE1 is bound by RE1-silencing transcription factor (REST) directly (Chong et al., 1995).

The components of the complex provide enzymatic activity that represses the expression of target neuronal genes by modifying histones. The enzymatic activities

of complex include (1) histone deacetylases (HDACs) 1 and 2 (Andres et al., 1999); (2) LSD1, which contains two domains – a SWIRM domain that helps in anchoring the tail of the histone H3 molecule (Tochio et al., 2006), and a PAO (polyamine oxidase)-like domain that acts as a histone H3 lysine demethylase (Shi et al., 2004) and (3) G9a, a SET domain protein with histone methyltransferase (HMT) activity (Roopra et al., 2004; Tachibana et al., 2001). Besides these major components, other proteins, e.g., REST-associated mSin3, methyl-DNA binding factor MeCP$_2$, are parts of the co-repressor complex (Ballas and Mandel, 2005; Grimes et al., 2000).

In Arabidopsis, FLOWERING LOCUS D (FLD) is a homolog of LSD1 that down regulates *FLC* via histone hypo-methylation. However, the role of FLD in repression of *FLC* is not fully understood. Histone modifications are well characterised in *Arabidopsis* e.g. its chromatin contains histone H3 methylated at lysine (K) residues 4, 9, 27 and 36, with H3K4/H3K36 and H3K9/H3K27 representing active and inactive chromatin, respectively (Zhao et al., 2005). Such modifications play an important role in regulation of *FLC* expression e.g. histone H3 trimethylation at Lys-4 (H3K4me3) and histone acetylation are associated with active *FLC* transcription, whereas histone deacetylation and histone H3 methylation at Lys-9 (H3K9) and Lys-27 (H3K27) are associated with *FLC* repression (He and Amasino, 2005b).

Similarly, vernalization leads to repressive histone modifications of *FLC* chromatin, including deacetylation, and increased methylation of H3K9 and H3K27 (Bastow et al., 2004; Jean Finnegan et al., 2005; Sung et al., 2006). Activation of *FLC* expression and the corresponding increase in H3K4 trimethylation require the PAF1 (for RNA Polymerase II–Associated Factor1)–like complex (He et al., 2004). H3K4 methylation, which is associated with actively transcribed genes, plays an important

role in regulating transcription (Martin and Zhang, 2005). In yeast (*Saccharomyces cerevisiae*), trimethylated H3K4 is associated with active euchromatic genes (Santos-Rosa et al., 2002), whereas H3K4 dimethylation (H3K4me2) occurs in both inactive and active euchromatic genes (Ng et al., 2003). Similar to its association in *S. cerevisiae*, H3K4 trimethylation is associated with active transcribed genes in multicellular eukaryotes; however, in contrast with yeast, H3K4 dimethylation is also associated with active genes in multicellular eukaryotes (Schneider et al., 2004). It has also been shown that in *Arabidopsis*, elevated levels of H3K4me2 and H3K4me3 are associated with active genes and that these modifications occur in 5′ promoters and coding regions but are absent from nontranscribed intergenic regions (Alvarez-Venegas and Avramova, 2005).

Histone H3K4 methylation is dynamically regulated by histone methylases and demethylases (Martin and Zhang, 2005). A component of transcriptional corepressor complexes, Lysine-Specific Demethylase1, has been shown to demethylate H3K4 and repress target gene expression in mammalian cells (Shi et al., 2004). Human LSD1 specifically demethylates monomethyl and dimethyl H3K4 (Forneris et al., 2005; Lee et al., 2005; Shi et al., 2004) and, when complexed with an androgen receptor, also destabilizes dimethyl H3K9 (H3K9me2)(Metzger et al., 2005). LSD1 is an integral component of several mammalian histone deacetylase (HDAC) corepressor complexes (Hakimi et al., 2002; Humphrey et al., 2001) in which HDACs and LSD1 may act together to remove activating acetyl and methyl histone modifications (Lee et al., 2006; Shi et al., 2005). Consistent with this model, in one such complex (the BRAF-HDAC complex), the enzymatic activities of HDACs and LSD1 are closely

linked, as HDAC inhibitors diminish histone demethylation activity and the lower LSD1 activity decreases the deacetylation activity of this complex (Lee et al., 2006).

1.6 Natural variation in flowering in cereals crops

Molecular studies of natural variation in flowering time in barley identified 25 different QTL with major effects (Alonso-Blanco et al., 2009). One to four segregating QTL were detected in each of the eight different environments including growth chambers and natural conditions. Functional diversity of CO-like (*CONSTANS*) and *FT*-like genes was also found in barley that correlate with quantitative responses to photoperiod (Kikuchi et al., 2009). Three *VERNALIZATION* (*VRN*) loci have been detected in wheat and barley that affect the flowering in response to low temperature. Three major loci controlling the vernalization response in both wheat and barley map to collinear locations in their respective genomes, suggesting that they represent orthologous genes (Dubcovsky et al., 1998; Karsai et al., 2005; Laurie et al., 1995).

In *Brachypodium*, a single orthologue of the rice flowering time gene *Hd3a* (an *FT*-like gene) was detected (Hasterok et al., 2006). When *Brachypodium* was transformed with floral repressor, *TERMINAL FLOWER 1* (*TF1*) from *Arabidopsis* or ryegrass, which antagonises FT activity, it repressed flowering (Jensen et al., 2001; Olsen et al., 2006; Shannon and Meeks-Wagner, 1991), suggesting conservation of the mechanisms controlling flowering within grasses Little is known regarding effect of variation in photoperiod or day-length on flowering time of sorghum (Menz et al., 2002).
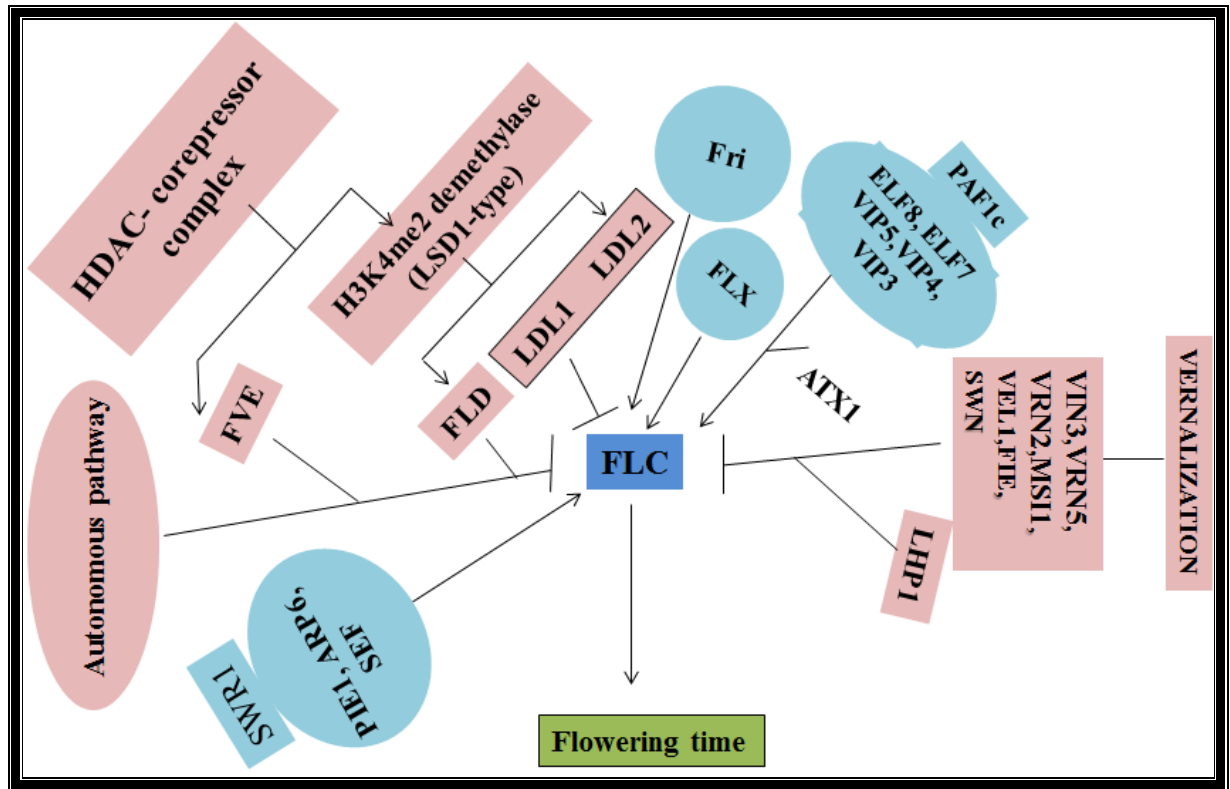
Figure 3: The transition from vegetative to flowering time is controlled by regulatory networks in Arabidopsis.

This figure shows FLC-dependent flowering regulatory networks in which Fri and PAF1c upregulates *FLC* (shown in light blue colour) whereas autonomous pathway, vernalization and HDAC corepressor are down regulators of *FLC* (shown in pink colour).

1.7 Natural variation in flowering in Brassicaceae

Genes have been identified for flowering time in response to vernalization in *Brassica napus* (Ferreira et al., 1995) in populations produced by crosses of annual and biennial oilseed cultivars. Like *Arabidopsis*, other species in the family of Brassicaceae respond to vernalization and this appears to involve conserved roles for *FLC*-like and *FRI*-like genes (Osborn et al., 1997). In fact, several *FLC* homologs have been isolated from *Brassica* species, such as *B. napus* (Tadege et al., 2001) and doubled haploid lines of *Brassica rapa* and *Brassica oleracea* (Schranz et al., 2002). Moreover, genetic manipulation of *FLC* expression has been proven to modify flowering time in both *Arabidopsis* (Michaels and Amasino, 1999a; Sheldon et al., 1999b) and *B. napus* (Tadege et al., 2001). Constitutive expression of *Arabidopsis thaliana FLC* in an early-flowering *Brassica napus* cultivar delayed flowering by 2–6 weeks (Tadege et al., 2001). *BrFLC1, BrFLC2,* and *BrFLC5*, which are *B. rapa* homologs of *FLC*, co-segregate with QTL that determine flowering time in late-flowering ecotypes of *B. rapa* (Schranz et al., 2002). Five *FLC*-related homologs (*BnFLC1-5*) isolated from *B. napus* delayed flowering significantly when expressed in *Arabidopsis*. Two homologs of *FLC*, *BoFLC3-2*, and *BoFLC4-1*, have been isolated from cabbage and expression patterns were analysed by using reporter gene in *Arabidopsis*. These data showed the molecular mechanisms that repress *FLC* expression in the vernalization response can also act on *BoFLC* genes (Lin et al., 2005). These results showed that *FLC* homologs of *Brassica* species likely act in a similar way to *FLC* in *Arabidopsis* and play a central role as repressors of flowering.

1.8 Natural genetic variation in flowering time in *Arabidopsis*

The transition timing from the vegetative to reproductive phase of a plant (called flowering time) is crucial for reproductive success of plants, for example to ensure that flowers are produced in a favourable environment and in synchrony with pollinators and potential mates. Mutant screens have identified around 100 genes that affect in the different flowering time pathways in *Arabidopsis* (Koornneef et al., 1998; Koornneef et al., 2004; Mouradov et al., 2002). However *Arabidopsis* accessions show a large amount of heritable variation for flowering time. Some of the accessions flower in few weeks whereas others do not flower for months unless vernalized. *Arabidopsis* shows two extreme life cycles, summer and winter annuals, in which winter annuals germinate one year, overwinter as vegetative rosettes and flower the following year (Grennan, 2006; Johanson et al., 2000). They are typically late flowering and responsive to vernalization. Summer annuals, in contrast, germinate and flower in the same year and typically are quicker flowering and respond less to vernalization or show no response. Natural accessions also differ in their responses to photoperiod and vernalization (Balasubramanian and Weigel, 2006; Henderson et al., 2003). Flowering time variation is usually a multigenic trait and QTLs analysis is required to identify and map corresponding QTLs. A minimum number of 14 QTLs for flowering time were identified in *Arabidopsis* accessions in environments differing for photoperiod and vernalization treatment (Alonso-Blanco et al., 1998; Clarke et al., 1995; Jansen et al., 1995; Loudet et al., 2002; Stratton, 1998; Weinig et al., 2002). Comparing the effects and position of QTLs with known flowering genes has enabled the identification of candidate genes for most of the QTLs (Alonso-Blanco et al., 1998; Loudet et al., 2002; Ungerer et al., 2002).

Many summer-annual accessions like Columbia (Col), Wassilewskija (Ws), and Landsberg *erecta (*L*er)*, are early flowering without vernalization as they carry loss-of-function mutations in *FRI* (Johanson et al., 2000). Low *FLC* levels after vernalization in the presence of *FRI* show that vernalization has the ability to supersede the ability of *FRI* to activate *FLC* (Michaels and Amasino, 1999a; Sheldon et al., 1999b).

In both mutant strains and natural accessions, there is a broad correlation of *FLC* expression levels and flowering time such that when *FLC* level is low, flowering occurs early and *vice versa* (Gazzani et al., 2003; Michaels and Amasino, 1999a; Michaels et al., 2003; Sheldon et al., 1999b; Sheldon et al., 2000). However, FLC expression variation was also observed among late-flowering accessions (Werner et al., 2005a). Genetic analyses also demonstrate that the difference in flowering behaviour between late- and early-flowering ecotypes is due to allelic variation at one or both of two loci, *FRIGIDA* (*FRI*) and *FLOWERING LOCUS C* (*FLC*) (Burn et al., 1993; Clarke and Dean, 1994; Koornneef et al., 1994; Lee et al., 1994b). Vernalization causes epigenetic silencing of the *FLC* locus and early flowering. Independently of vernalization, *FLC* is also negatively regulated by the autonomous pathway, which was thought to act independently of the environmental cues. However, it has been found that this pathway may also mediate response to ambient growth temperature (Blazquez et al., 2003).

1.9 Summary of previous work in Hudson lab

J. Atkinson (PhD thesis to The University of Edinburgh, 2006) finding that rosette area is significantly associated to dry mass (Figure 1) has been referred in my thesis. Poay Lim (PhD thesis to The University of Edinburgh, 2013) studied natural variation

in growth rate, flowering time and seed production in local accessions of *A. thaliana*. Her findings that different plant genotypes respond variably to seasons (Figure 7), variation in growth rate within and between families in control environment (Figure 6), rosette area regression to days to germination (Figure 2) and table of sites of collections of wild plants (Table 1) was taken from her thesis. However, no direct data collected by J. Atkinson and Poay Lim was used in my thesis.

Hayley Mc Culloch (PhD thesis to The University of Edinburgh, 2011) worked on the genetic basis of natural variation in flowering time in local accessions of *A. thaliana*. She grew and collected flowering time data of F3 plants both from field and growth room conditions. This data was used with her permission in QTL mapping of flowering time of F3 plants in my project. No other work or her data was used in my experiments.

2.0 MATERIALS AND METHODS

2.1 Seeds sterilization and growth conditions of plants

Seeds were stratified at $4^{o}C$ in 0.1% agar for 48 to 72 hours to break dormancy and promote germination. Plants were grown using pre-mixed compost (Levington F2) without any added insecticides. Pots were filled with compost and compressed very lightly to give a firm bed. Seeds were sown on the surface of the compost with a pipette. Trays were covered with transparent sheets until the seeds had germinated. The plants were then watered from below when the surface of the compost started drying. Thinning was carried out to retain only one plant per pot. For populations grown outside, one set of plants was transferred to a garden approximately after one week under natural condition whereas the other set was kept in a greenhouse maintained at $16^{o}C$-$21^{o}C$ with 16 hrs photoperiod and with light levels of ~150umol/m^2/s or in a growth room maintained at $21^{o}C$ during the day and night with 16 hrs photoperiod and light levels of ~100 μmol/m^2/s and ~65% relative humidity during the day and ~50% relative humidity at night. Photographs were taken on a weekly basis and were analysed with Adobe Photoshop and Image Tool to calculate rosette areas of plants.

2.2 Adobe Photoshop

Images were taken from above trays of plants, each with a size marker of known area for calibration (a 10p coin, which has an area of 471 mm^2). Images were prepared for analysis with Adobe Photoshop (Fig. 4a). First the resolution was set to 72 pixels/inch. A green sample was selected from a rosette with the eye dropper tool and

also the size marker was selected with the magic wand tool. Then, the size marker was deleted and filled its place with the green colour using the paint-bucket tool. After this, the middle of a typical, large rosette or several leaves if plants were smaller were selected by using the elliptical marquee tool. The next step was to select similar colours and to grow the selection so that it included all the plant tissue. After setting white as the background colour, the plant images were cut, selected the remaining background and deleted, and pasted the rosettes and marker back onto the white background (Fig. 4b). The median blur image function, set to 2 pixel radius, was used to fill holes in the image and to remove small background objects that had been selected with rosettes. The image was then turned into a binary (black and white) image (Fig. 4c). To make a binary image, the adjust hue-saturation function was used to set the lightness to -100 and the image was flattened afterwards (Fig. 4a). These steps were recorded as an action so that they could be applied to other images. To remove remaining background objects, the rosettes were selected, with the magic wand tool and were cut, the background was selected and deleted and pasted the rosette back onto white background. Any unconnected parts of plants were joined using the paintbrush, so that each would be treated as one object by ImageTool. Finally, the image was saved as a jpg file.
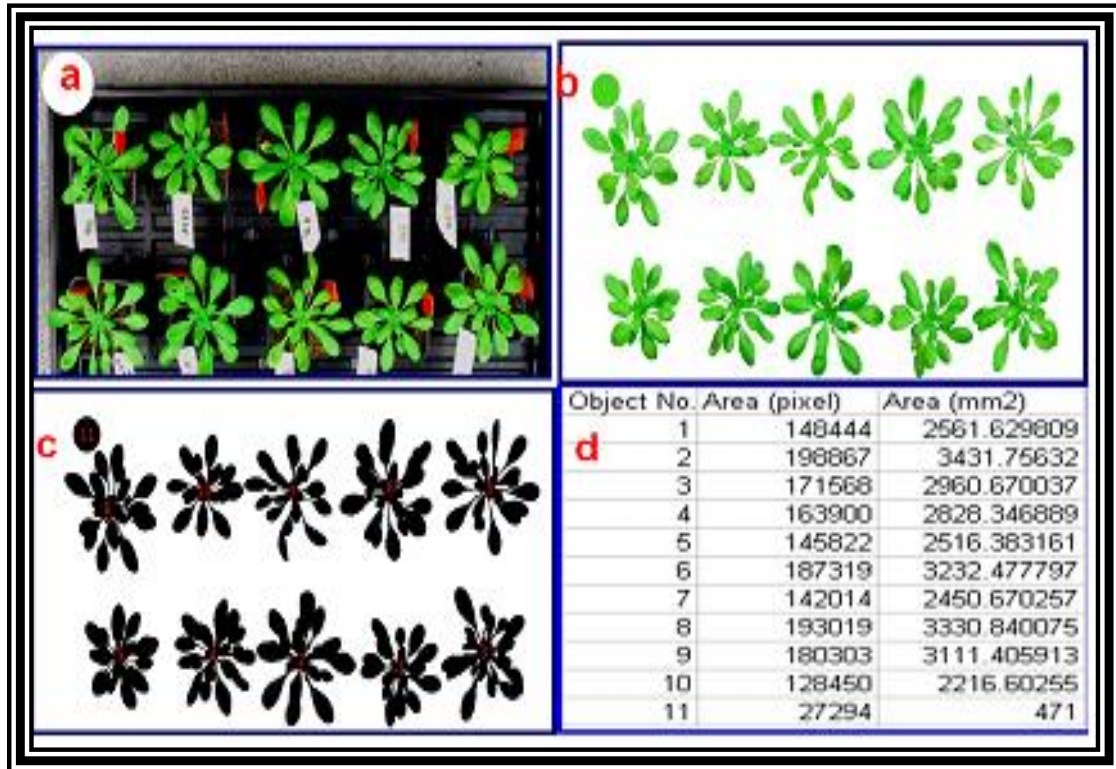
Figure 4: Working with Adobe Photoshop and Image Tool.

Photographs of rosettes and a size standard (a) were cut and pasted onto a white background (b) and converted to binary images (c) in Adobe Photoshop. The binary images were used to calculate rosette areas, using Image Tool (d).

2.3 *Arabidopsis* collection sites in and around Edinburgh

Wild *A. thaliana* is found in different types of habitats in and around Edinburgh, such as dry ground, cliffs, and skeletal soils over rocks and gardens like The Royal Botanic Garden Edinburgh etc. In this study, Andrew Hudson and an Honours student, S. Whithall, sampled thirteen populations of *A. thaliana* in a radius of 5 kilometres from The University of Edinburgh King's Buildings campus at elevation ranging from 62 to 249 meters above sea level (Table 1 from Poay Lim PhD thesis to The University of Edinburgh, 2013). Two to twenty plants were collected from each locality within a 10 m radius and were transferred to green house with long day conditions. Seeds were collected from these plants. One individual was grown from each parent and its seeds collected and coded for identification. For example, 11C1 denotes a particular genotype, "11" represents the site of collection, "C" represents the pot in which plants were originally transferred and "1" represents the first plant of the pot. Each coded stock is also referred as family (Thesis - from Poay Lim, PhD thesis to The University of Edinburgh, 2013).

Site ID	Location	Elevation from sea level	Population size
1	U. of Edinburgh, Kings Buildings, S. of Swann Building	70 m	10% cover
2	U. of Edinburgh, Kings Buildings, N. of Swann Building	70 m	< 1% cover
3	U. of Edinburgh, Kings Buildings, S. of Rutherford Building	67 m	15% cover
4	U. of Edinburgh, Kings Buildings, Forestry plots	62 m	15% cover
5	U. of Edinburgh, Kings Buildings, S. of JCMB	75 m	80% cover
6	U. of Edinburgh, Kings Buildings, W. of CSEC	78 m	30% cover
7	U. of Edinburgh, Kings Buildings, W. of CSEC	79 m	

8	U. of Edinburgh, Kings Buildings, W. of CSEC	79 m	5% cover
9	U. of Edinburgh, Kings Buildings, W. of SAC	79 m	5% cover
10	Loanhead	148 m	5% cover
11	Hillend	249 m	20% cover
12	Liberton	115 m	50% cover
13	Straiton	152 m	50% cover

Table 1: Collection sites detail of *A. thaliana* local accessions from in and around Edinburgh.

2.4 Measurements of rosette areas of plants in Image Tool

UTHSCSA ImageTool (IT) is an image processing and analysis programme (compdent.uthscsa.edu/dig/itdesc.html). For calculation of rosette areas, Image Tool was set to recognise an object as a minimum of 200 black pixels to avoid picking up any speckles as objects. A binary image of plants was opened and the Find Object function used on manual to adjust threshold levels so that rosettes were selected completely. The programme then said how many objects it had found and numbered them. The object outlines and their numbers are shown in red on the image (Fig.4c). Image Tool numbers objects from the bottom upwards in the order that it first encounters them. The areas of each object, in pixels, were written to a table of measurements which could be copied and pasted into an Excel spread sheet. These values were converted to mm^2 using the known area of the marker e.g., a 10p coin has an area of 471.44 mm^2 and so the area of a rosette in mm^2= (area of rosette in pixels /area of coin in pixels) x 471.44 mm^2.

2.5 Flowering time measurements

Five plants per line for an F5 population (89 lines in total) were grown together in a growth room. The number of rosette leaves produced before flowering was counted for each plant. Cotyledons and cauline leaves were not taken into consideration. For each line, the mean value of its five members was calculated. The number of days to flowering was also recorded for each plant. Again, the mean value of the line was calculated.

2.6 Scoring of anthocyanin production

Arabidopsis plants produce anthocyanin when experiencing cold stress and change their colour from green to purple. Winter-grown F3 population changed colour during frosty conditions. Photographs of plants were used to score this colour change on a scale of 1-5, with 1 being the greenest plant and 5 being the most purple, against images of a plant representing each point on the scale.

2.7 Herbivory estimates

When plants are grown in natural conditions, herbivore attack may occur. In our experiment, an F3 population grown outside in the winter experienced herbivory, probably by slugs. Not all plants in the population were attacked. To estimate the amount of tissue eaten, the rosette areas of all plants were calculated before and after the attack. Plants that had not been eaten were used to calculate the mean relative increase in rosette area during the period of the attack by linear regression. The mean relative increase (the slope, m, of the regression of the final area on the initial area) and the y intercept (c) were used to estimate the expected final area of plants had they not been eaten (A*):

Expected areas of plants if they had not been eaten (**A*** after) = (**A** before $\times$ m) + c

Therefore the area of the plants that had been eaten could be estimated by the following formula, where A obs. is the measured area of the plants after attack:

Eaten areas of plants= **A*** after − **A** obs.

(Where A obs, was the observed area after attack)

Percentage eaten areas of plants= (**A*** after − **A** obs.) $\times$100 / **A*** after

2.8 Quantitative Trait Loci (QTLs) analysis

Average values for each trait were regressed onto genotypes to search for Quantitative trait loci (QTLs). QTL analysis was performed with Windows QTL Cartographer software v2.5 (http://statgen.ncsu.edu/qtlcart/WQTLCart.htm), using the option for mapping in inbred lines. The Composite Interval Mapping (CIM) analysis technique (Jansen, 1996b) was used to detect the potential QTLs with 500 permutations and 0.05 significance levels for all traits with 1.0 cM precision.

2.9 Bulk –Segregant analysis

Monogenic qualitative traits are usually analysed by this rapid mapping technique. In this case it was used to test for linkage of markers to growth rate QTL that were segregating in F5 populations. Pools of DNA were made from the eight fastest and eight slowest growing plants of an F5 populations, grown either outside in natural conditions or in growth room conditions. No more than one fast or slow plant was taken from a family (offspring of the same F4 parent), to minimise the effects of allele sharing due to common ancestry. These pools were then genotyped with all of the 19 microsatellite markers which had been found to be polymorphic in the F3 population.

2.10 Heritability estimates

One-way analysis of variance was performed to calculate the broad-sense heritability of each trait in each population. This method is based on Box 9.1 and Box 9.2 of Sokal and Rohlf (Sokal and Rohlf, 1995) (http://udel.edu/~mcdonald/anova.xls and http://udel.edu/~mcdonald/statintro.html). It involved calculating the proportion of the total variance in the trait that occurred between lines (which is expected to have a genetic basis). This is likely to under-estimate the heritability of the trait, because it

assumes that all members of a line are genetically identical and therefore that variance within a line has only a non-genetic (environmental) basis.

2.11 Collection of tissues for DNA extraction

Similar amounts of plant tissue from all five members of a line (about 200 mg from each plant) were placed in 1.5 ml Eppendorf tubes and leaves pushed down with the help of forceps. Two steel ball bearings were added to the tube. Tissues were either used directly for DNA extraction or kept in a -80°C freezer until needed.

2.11.1 Genomic DNA extraction by CTAB (Hexadecyltrimethyl ammonium bromide) method

This is partially modified from the method of (Doyle and Doyle, 1987) for plant DNA extraction. A 24 holes block for a Retch mixer mill was chilled at -80°C. Tubes with frozen tissues were quickly transferred to the block and after the block had been assembled and clamped into the mixer mill, it was shaken for 1 min at maximum frequency (50 Hz). Working quickly, the lids of the block were rotated 180° so that the arrangement of tubes was reversed such that the tubes previously at the front were now are at the back and the blocks shaken again. After disassembling the block and opening the tubes, 1 ml of extraction buffer was added (Table 2). Then the tubes were recapped and the block reassembled and shaken it for 1 min to mix the tissue with the buffer. After, incubating the tubes in a heating block at 65°C for 10-20 min in a fume hood, it was allowed to cool for ~2 min. Each tube was then filled with 500 µl of chloroform and mixed. Tubes were centrifuged at 13000 x g for 5 min. The material in the tubes separated into three distinct layers, a top clear aqueous layer, a middle interface of precipitated proteins etc. and a green organic layer at the bottom. The top

clear aqueous layer was removed to a clean tube and 440 µl of isopropanol added and mixed gently by rocking the tubes end over end. Centrifuging for 5 minutes recovered a grey pellet of nucleic acids. The alcohol was tipped off, spun the tubes again for 1 min to dry the sides of the tubes and removed the rest of the alcohol with a pipette without displacing the nucleic acid pellet. The pellet was dissolved in 100 µl of TE buffer (Table 4) with 1/1000th volume of 10 mg/ml RNase A, and kept overnight to ensure complete dissolution of DNA. DNA was then precipitated by adding 260 µl of a mixture of 10 ml of 100% alcohol and 400 µl of 3M NaOAc solution (pH 7.4) which was mixed completely by rocking. DNA was recovered by centrifugation at maximum speed for 5 min, giving a transparent pellet. After removing the alcohol, the DNA was air dried and dissolved in 50 µl of dH_2O or TE buffer. Finally, the concentration of DNA was measured with a Nanodrop spectrometer and the DNA diluted to ~10 ng/µl.

Stock	Amount	Final concentration
1 M Tris-HCl, pH 8.0	15 ml	100 mM
0.5 M EDTA, pH 8.0	6 ml	20 mM
5 M NaCl	42 ml	1.4 M
CTAB	3 g	2% w/v
dH2O	50 ml	

Final volume adjusted to 150ml with dH2O

Table 2: Recipe for CTAB extraction buffer.

2.11.2 Genomic DNA extraction by Edwards method

The method of (Edwards et al., 1991) was modified for the extraction of genomic DNA. Frozen samples were placed in a mixer-block cooled at -80°C. Tissues were ground at maximum speed for 1 min in the mixer mill, as for the CTAB extraction method. This was repeated after rotating the block by 180°. Extraction buffer (400 µl) was then added and the contents of the tube mixed for 1 min in the mixer mill. The extraction buffer consisted of 200 mM Tris HCl pH 7.4, 250 mM NaCl, 25 mM EDTA and 0.5% SDS (Table 3). Samples were incubated at room temperature for one hour and spun at 13,000 x *g* in a centrifuge for 5 min. Supernatant (300 µl) was transferred to a clean tube and 300 µl of isopropanol was added and the contents mixed. Tubes were centrifuged for 5 min at maximum speed. Liquid was removed and tubes were briefly spun to dry the sides. The remaining liquid was removed with a yellow pipette tip.

The pellet was dissolved in 150 µl of 0.1x TE (10 mM Tris pH 8.0, 0.1 µM EDTA) and 1 µl of the solution was used in each PCR reaction. The remaining solution was stored at -20°C.

Stock	Amount	Final concentration
200 mM	5 ml of Tris HCl (pH7.5)	2 M Tris HCl (pH7.5)
250 mM	2.5 ml NaCl	5 M NaCl
25 mM	2.5 ml of EDTA	0.5 M EDTA
0.5%	2.5 ml of SDS	10% SDS

Adjusted final volume to 50 ml with dH2O

Table 3: Recipe of Edward's extraction buffer solution.

2.11.3 10x TE buffer recipe

This buffer consists of 100 mM Tris-HCl (pH 7.5) and 10 mM EDTA (pH 8.0). The working concentration is 1x or lower. To make 1 litre 10x TE (Tris-EDTA) buffer, the following were mixed,

1	100 ml 1 M Tris-HCl (pH 7.5)
2	20 ml 500 mM EDTA (pH 7.5)
3	880 ml ultrapure water

Table 4: Recipe for 10x TE buffer.

2.12 Simple Sequence Length Polymorphism (SSLP) markers

SSLP markers (microsatellites) were used for genotyping. A total of 137 microsatellite markers were tested for polymorphism between the mapping parents, 11C1 and 4D1, of the hybrids in my experiments. SSLPs markers used for genotyping were described in two main sources; The *Arabidopsis* Information Service (http://arabidopsis.org/servlets/Search?action=new_search&type=marker) and the INRAMSAT database (http://www.inra.fr/internet/Produits/vast/msat.php). The map position of each SSLP marker in cM was estimated from its position in the *Arabidopsis* genome sequence relative to makers for which recombination frequencies had been determined in other mapping populations. For QTL analysis, the first marker on each of the 5 linkage groups was given the position 0 cM. A list of the 39 markers which showed polymorphism for 11C1 and 4D1 with their complete description is given in Table 5.

Locus name	Map position (cM)	Primer sequence	Chr: No.	Annealing temperature (Ta) °C
ACC2	0.00	F: 5'-AAATGTGCAATTGCCTTC-3'	1	49
		R: 5'-AGAAGTTTAGACAGGTAC-3'		49
F21M12	6.64	F: 5'-TTACTTTTTGCCTCTTGTCATTG-3'	1	50
		R: 5'-GGCTTTCTCGAAATCTGTCC-3'		52
NGA392	41.64	F: 5'-GGTGTTAAATGCGGTGTTC-3'	1	55
		R: 5'-TTGAATAATTTGTAGCCATG-3'		50
CIW1	72.00	F: 5'-ACATTTTCTCAATCCTTACTC-3'	1	53
		R: 5'-GAGAGCTTCTTTATTTGTGAT-3'		53
F5114-49495	92.08	F: 5'-CTGCCTGAAATTGTCGAAAC-3'	1	56
		R: 5'-GGCATCACAGTTCTGATTCC-3'		58
NGA692	119.25	F: 5'-AGCGTTTAGCTCAACCCTAGG-3'	1	60
		R: 5'-TTTAGAGAGAGAGAGCGCGG-3'		61
ALT_ NYUP 8H12R	132.00	F: 5'-CACGACGTTGTAAAACGACCACTGAAACCACTTCCCACA-3'	1	68
		R: 5'-GCTTGAGCCAAGTCGAGAGT-3'		54
CIW2	0.00	F: 5'-CCCAAAAGTTAATTATACTGT-3'	2	52
		R: 5'-CCGGGTTAATAATAAATGT-3'		49
CIW3	24.00	F: 5'-GAAACTCAATGAAATCCACTT-3'	2	46
		R: 5'-TGAACTTGTTGTGAGCTTTGA-3'		47
MSAT 2.36	30.00	F: 5'-GATCTGCCTCTTGATCAGC-3'	2	58
		R: 5'-CCAAGAACTCAAAACCGTT-3'		53

PLS3	44.00	F: 5'-TAGTCGTTTCTCTGGTTGTAG-3' R: 5'-TTGCCTGTCGATGTAGATTTGT-3'	2	58 58
CZSOD2	56.94	F: 5'-GAATCTCAATATGTGTCAAC-3' R: 5'-GCATTACTCCGGTGTCGTC-3'	2	52 60
NGA361	63.02	F: 5'-ACATATCAATATATTAAAGTAGC-3' R: 5'-AAAGAGATGAGAATTTGGAC-3'	2	52 52
UBIQUE	82.00	F: 5'-CACGACGTTGTAAAACGACACGACATGGCAGATTTCTCC-3' R: 5'-AGGCAAATGTCCATTTCATTG-3'	2	68 48
NGA172	0.00	F: 5'-CATCCGAATGCCATTGTTC-3' R: 5'-AGCTGCTTCCTTATAGCGTCC-3'	3	49 54
NGA162	20.56	F: 5'-CTCTGTCACTCTTTTCCTCTGG-3' R: 5'-CATGCAATTTGCATCTGAGG-3'	3	55 50
CIW11	43.00	F: 5'-CCCCGAGTTGAGGTATT-3' R: 5'-GAAGAAATTCCTAAAGCATTC-3'	3	47 46
CIW4	76.00	F: 5'-GTTCATTAAACTTGCGTGTGT-3' R: 5'-TACGGTCAGATTGAGTGATTC-3'	3	55 57
FUS6.2	86.41	F: 5'-TTCCTTGATCAGATTTGGTCG-3' R: 5'-TCGTTACACTGGCTTGCTTG-3'	3	57 58
NGA6	90.00	F: 5'-ATGGAGAAGCTTACACTGATC-3' R: 5'-TGGATTTCTTCCTCTCTTCAC-3'	3	57 57
JV30/31	0.00	F: 5'-CATTAAAATCACCGCCAAAAA-3' R: 5'-TTTTGTTACATCGAACCACACA-3'	4	53 56

NGA8	16.56	F: 5'-TGGCTTTCGTTTATAAACATCC-3'	4	56
		R: 5'-GAGGGCAAATCTTTATTTCGG-3'		58
DET1.2	21.44	F: 5'-GGTGAAAATGGAGGAGACG -3'	4	52
		R: 5'-TTCAAACACCAATATCAGGCC-3'		50
CIW6	37.00	F: 5'-CTCGTAGTGCACTTTCATCA-3'	4	56
		R: 5'-CACATGGTTAGGGAAACAATA-3'		55
MSAT4.12	43.00	F: 5'-AAAGGAAGAAGACTGTT-3'	4	52
		R: 5'-AGAAGAAGAAGCGAGATT-3'		49
MSAT4.14	47.00	F: 5'-GACCGTTTCTAGTGCTCACA-3'	4	58
		R: 5'-ACGGAATAAGCGGAGGA-3'		52
CIW7	65.00	F: 5'-AATTTGGAGATTAGCTGGAA-3'	4	46
		R: 5'-CCATGTTGATGATAAGCACAA-3'		48
MSAT4.9	40.00	F: 5'-GAAATCAACGGCTGAG-3'	4	48
		R: 5'-AAGTAATTAAGACGCTGAGA-3'		52
MSAT4.12	43.00	F: 5'-GGAACAAGAACACAGTGAA-3'	4	53
		R: 5'-ATAAATCTAGGCAGGACAAG-3'		54
MSAT4.18	50.00	F: 5'-TGTAAATATCGGCTTCTAAG-3'	4	52
		R: 5'-CTGAAACAAATCGCATTA-3'		47
MSAT4.19	53.00	F: 5'-TGAACTAAAGACTTGATGCC-3'	4	54
		R: 5'-CCAAACGCAAATAGTGTT-3'		49
NGA 1139	73.41	F: 5'-TTTTTCCTTGTGTTGCATTC-3'	4	46
		R: 5'-TAGCCGGATGAGTTGGTACC-3'		54

NGA158	0.00	F: 5'-ACCTGAACCATCCTCCGTC-3' R: 5'-TCATTTTGGCCGACTTAGC-3'	5	60 55
ALT_NGA 151	29.62	F: 5'-CAGTCTAAAAGCGAGAGTATGATG-3' R: 5'- GTTTTGGGAAGTTTTGCTGG-3'	5	54 50
CIW8	42.00	F: 5'-TAGTGAAACCTTTCTCAGAT-3' R: 5'-TTATGTTTTCTTCAATCAGTT-3'	5	46 43
S0262	65.20	F: 5'-ATCATCTGCCCATGGTTTTT-3' R: 5'-TTGCTTTTTGGTTATATTCGGA-3'	5	54 55
CIW9	88.00	F: 5'-CAGACGTATCAAATGACAAATG-3' R: 5'- GACTACTGCTCAAACTATTCGG-3'	5	56 60
NGA129	105.41	F: 5'-CACACTGAAGATGGTCTTGAGG-3' R: 5'-TCAGGAGGAACTAAAGTGAGGG-3'	5	62 62
ALT_ CIW10	115.00	F: 5'-CCACATTTTCCTTCTTTCATA-3' R: 5'- CAACATTTAGCAAATCAACTT-3'	5	46 45
JV61/62	130.00	F: 5'-CGCTTTCCTTGTGTCATTCC-3' R: 5'-AAATGCAAATATTGATGTGTGAAA-3'	5	52 47

Table 5: SSLPs with complete description.

2.12.1 SSLP analysis

SSLPs were amplified from genomic DNA by PCR. For genotyping of the F3 mapping population, the PCR products were fluorescently labelled for analysis in a capillary DNA sequencing machine. Three primers were used in amplification of fluorescently labelled products. One matched the *Arabidopsis* sequence to one side of the polymorphic site. The second matched the *Arabidopsis* sequence to the other side of the polymorphism at its 3' end, but had the sequence of the M13 forward sequencing primer (5'-GTAAAACGACGGCCAGT-3') at its 5' end. The third primer consisted of the M13 sequencing primer labelled with one of four fluorescent dyes at its 5' end (6-FAM, VIC, NED or PET). PCR therefore amplified a fluorescently labelled product and the use of four different dyes allowed the products of different PCR reactions to be analysed together. The PCR reactions were pooled, after being diluted according to the amount of product present for each primer combination, and further diluted 1:10 with ABI Hi-Dye formamide mix and 1:500 dilution of ABI LIZ-500 size standard. Reactions were analysed using ABI 3730 DNA sequencer. Alternatively, the fluorescent M13 primer was omitted and fragments analysed by gel electrophoresis stained with ethidium bromide. Files from the ABI 3730 were analysed using ABI Gene Scan analysis software and then imported into Genographer (www.softpedia.com › Windows › Science / CAD), which displays fluorescent amplification products as bands and determines the size of these products. The SSLP loci that have been analysed are shown in Table 5.

2.12.2 PCR reactions

All amplifications were performed using a GRI DNA Engine or a G-Storm thermocycler using home-made *Taq* polymerase. Each reaction was performed with 1µl or 2 µl of DNA (10-20 ng of DNA) with a mix 19 µl or 18 µl reactions mix respectively (Table 6).

Constituents	Amount	Description
10x buffer	2 µl	Standard 10x buffer
10 mM dNTPs	0.4 µl	Rova lab.
Forward primer (10 µM)	0.4 µl	IDT
Reverse primer (10 µM)	0.4 µl	IDT
Taq polymerase (5 units/µl)	0.1 µl	
H_2O	14.7 µl	Sterilized
TOTAL	18 µl	

Table 6: Components of PCR mixes.

All PCR reactions were carried out with the standard PCR programme shown in Table 6.

Step 1	94$^{\mathrm{o}}$C	2 mins.
Step 2	94$^{\mathrm{o}}$C	30 secs.
Step 3	Ta*	30 secs.
Step 4	72$^{\mathrm{o}}$C	1 min.
Step 5	Go to step 2	34X
Step 6	72$^{\mathrm{o}}$C	10 mins
Step 7	4$^{\mathrm{o}}$C	4 hrs.
Step 8	End	

* the annealing temperature, Ta, was varied according to the primers used.

Table 7: Standard PCR programme.

2.13 Agarose Gel Electrophoresis

2.13.1 Materials

An adequate volume of electrophoresis buffer was prepared (0.5x TBE; Table 8) to fill the electrophoresis tank and to prepare the gel. For the gel, electrophoresis-grade agarose was added to the buffer and dissolved by heating in a microwave. The gels that were used in my experiments contained between 1.0% and 3.0% agarose and 50 ng/ml ethidium bromide. In the next step, the melted agarose was poured into a gel tray and inserted the gel comb, ensuring that no bubbles were present.

1	108 g Tris base
2	55 g boric acid
3	40 ml 0.5 M EDTA (pH 8.0)
4	Autoclave for 20 min

Table 8: 10x TBE Buffer (1 Litre).

2.13.2 Loading and running agarose gels

After the gel had set, the combs were removed and tried not to tear the wells. After placing the gel into the electrophoresis tank, a sufficient amount of 0.5x TBE buffer was added to cover the gel to a depth of about 1 mm. Around 5 µl of each DNA sample was loaded, with 1/10th volume of 10x loading buffer, when needed. Size markers, usually 1 kb (GeneRulerTm) or 100 bp DNA ladder (New England Bio Labs, 500 µg/ml), were loaded in at least one well. Leads were attached in such a way that DNA migrated towards the anode. The voltage was set to give a potential gradient of ~0.5 – 0.8 V/cm (100 V to 150 V for a 20 cm tank) to begin electrophoresis. After 45 minutes to 1 hour, the power supply was turned off. Since ethidium bromide is present in the gel, the DNA could be visualized by placing the gel on a UV light source and photographed directly. The sizes of PCR products were estimated by comparison to fragments of the size standard.

2.14 Cloning, bacterial transformation and sequencing

Regions of the flowering time genes, *LDL1* and *VIP5*, were amplified from parental plants and either sequenced directly, or purified from gels before sequencing or cloning in a plasmid vector. Both genes were initially amplified in three overlapping fragments. PCR reactions were conducted as described in 2.11.2. When more than one product was produced in PCR, fragments were purified by cutting the required band from an agarose gel and extracting the DNA with a QIAquick Gel extraction Kit by QIAGEN. Products were either sequenced directly, or cloned into pJET1.2 plasmid, using a CloneJET™ PCR cloning kit by Fermentas. The kit was used with the modifications shown in Table 9. The reaction components shown in Table 8 were

mixed in a 500 µl PCR tube spun and incubated at 70°C for 5 min in a PRC machine to make PCR products blunt-ended. Tubes were then chilled on ice and 0.5 µl each of the pJET1.2 blunt vector (http://abo.com.pl/pl/p/CloneJET-PCR-Cloning-Kit/14261; Figure 5) and T4 DNA ligase were added. Components were mixed, spun and incubated at room temperature for 10 minutes before being used in transformation of *E. coli*.
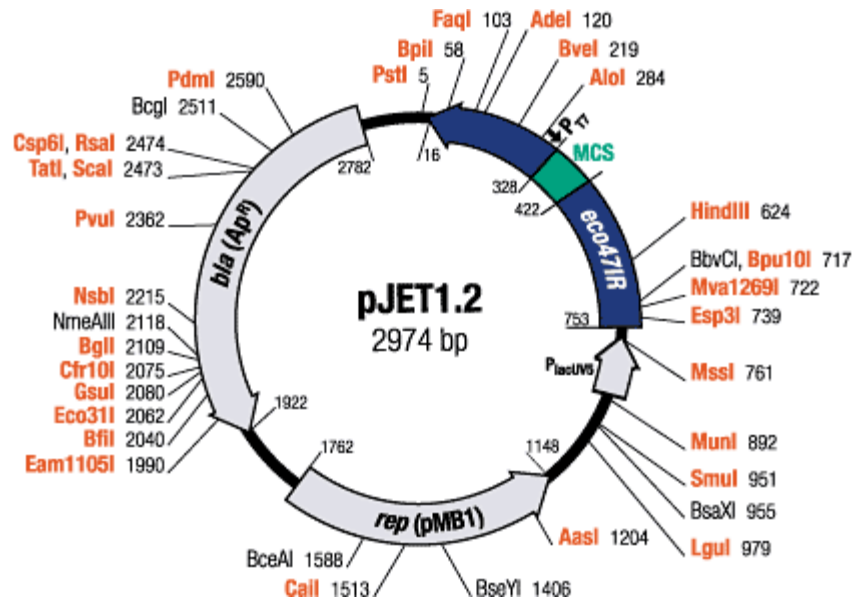
Figure 5: pJET 1.2/ blunt vector map (http://abo.com.pl/pl/p/CloneJET-PCR-Cloning-Kit/14261).

Constituents	Concentration
2x reaction buffer	5 µl
PCR products	0.5 µl
Water	12 µl
DNA blunting enzyme (Proof reading Taq Polymerase)	0.5 µl
	18 µl

Table 9: pJET cloning kit components with their amounts.

2.14.1 Preparation of KCM competent cells

A seed culture of DH5α was grown in 5 ml of LB media. This was kept overnight at $37^{o}C$ with shaking. The following day, 1.5 ml of the culture was used to inoculate 500 ml of LB medium in 2 litre flask. The optical density (OD) of the culture (at 600 nm) was checked after 4 hrs and the culture allowed to grow until it had reached around 0.30. The culture was then placed on ice for 5 min and 50 ml aliquots in chilled Falcon tubes were centrifuged for 15 min at $4^{o}C$ (10,000 x g). The pellet of cells was kept and supernatant was discarded. Transformation and storage solution (TSS) (1 ml) was used to resuspend the cells and aliquots of 100 µl of cells in 1.5 ml Eppendorf tubes were frozen in liquid nitrogen and stored at $-80^{o}C$.

2.14.2 DH5α transformation

DH5α transformation was carried out using the protocol available online at (http://openwetware.org/wiki/Huang:_DH5alpha_transformation), with the following changes. Competent cells were thawed on ice and 100 µl of cells added to 5 µl of ligation in a 1.5 ml Eppendorf tube. Tubes were then incubated on ice for 30 min, heat shocked in a water bath at $42^{o}C$ for 90 seconds and returned to ice for 2 minutes. L-broth (1 ml) was added and cells incubated at $37^{o}C$ for an hour with continuous shaking. The culture was centrifuged at 6000 x g for 5 min. Most (900 µl) of the medium was removed and the remainder used to resuspend the cells. Cells were spread on LB agar plates with ampicillin (200 µg/ml) and incubated at $37^{o}C$ in inverted position overnight.

2.14.3 Bacterial transformation using di-methyl sulphoxide (DMSO) and polyethylene glycol (PEG)

This is a modified version of the protocol described by (Chung et al., 1989) and later published by (Walhout et al., 2000) with modification. Competent cells were allowed to thaw on ice. 100 µl (70 µl H$_2$O, 10 µl ligation and 20 µl of 5X keratinocyte conditioned medium buffer (KCM buffer recipe; Table 10) were added to an Eppendorf tube containing 100 µl of a suspension of thawed competent cells and incubated on ice for 20 minutes. The mixture was then heat-shocked at 37°C for 5 min and chilled again on ice for 5min. LB was heated to 65°C for 5 sec and 800 µl of warm LB was added to the mixture. It was incubated at 37°C for 40 min and 300 µl was then spread on LB agar plates with ampicillin (200 µg/ml). Plates were incubated overnight at 37°C in an inverted position.

Reagents	Conc.
KCl	0.5 M
CaCl$_2$	0.15 M
MgCl$_2$	0.25 M

Table 10: showing components of 5X KCM buffer.

2.13.4 Making of LB-agar plates

Solid LB-agar (250 ml) was heated in microwave to about 100°C to melt it and then allowed to cool to approximately 65°C. Ampicillin was added from a 50 mg/ml stock in water to give a final concentration of 50 µg/ml and mixed. Around 30 ml of agar was poured into each plate in a laminar flow hood. After the agar had set, plates were allowed to dry for 15 minutes in the laminar flow hood by lifting their lids.

2.14.5 PCR reactions with pJET primers

Colony PCR reactions were done to test for the presence of an insert following the standard PCR protocols in Table 11, except that no DNA solution was added as template. The template was provided by touching a colony with a yellow pipette tip and stirring the tip in the PCR mixture. The primers flanked the cloning site in pJET1.2 (for sequence Appendix A and for cloning site Figure 5).

Constituents	Conc.	Description
10X buffer	2 µl	Standard 10X buffer
2.5 mM each dNTPs (or 10 mM)	1.6 µl (0.4 µl)	Rova lab.
pJET forward primer (10 µM)	0.4 µl	pJET cloning kit
pJET reverse primer(10 µM)	0.4 µl	pJET cloning kit
Taq polymerase (5U/µl)	0.1 µl	
H_2O	13.5 µl	Sterilized
	18 µl	

Table 11: Colony PCR reactions with pJET primers.

PCR products were loaded in gels for identification of plasmids carrying inserts of the expected size. For selected plasmids, 1 µl of the PCR product was used in a sequencing reaction with 0.5 µl of (10 µM) primer.

2.14.6 Gel purification of DNA amplicons

Gel purification was done with a QIAquick Gel Extraction Kit. The required DNA fragment was excised from the agarose gel with clean and sharp scalpel under UV light. The gel slice was weighed in an Eppendorf tube and 6 volumes of Buffer QG were added to 1 gel volume. Tubes were incubated at 50°C with occasional vortex mixing until the gel slice has completely dissolved. Then, 10 µl of 3M sodium acetate (pH 5.0) was added and mixed. The colour of the mixture turned yellow. One gel volume of isopropanol was added to the sample and mixed. The mixture was loaded into a QIAquick spin column in a 2 ml collection tube. The column was centrifuged for 1 minute at 13,000 x *g* to force the liquid through it. The flow-through was discarded and the column placed back into the same tube. Since the purification was being carried out for sequencing, 0.5 ml of Buffer QG was added to the column and centrifuged it for 1 minute to wash it. The flow through was again discarded and the QIAquick column placed back into the same tube. A second wash of 0.75 ml buffer PE was added to the column and the column was allowed to stand for 2-5 minutes before being centrifuged. The flow through was discarded again and the column placed back into the same tube and centrifuged once more to remove the residual wash buffer. Then, the QIAquick column was placed into a clean 1.5 ml Eppendorf tube and 50 µl buffers EB (10 mM Tris HCl, pH 8.5) was added to the centre of the column. The eluted DNA in solution was collected by centrifuging the column and then the flow through were concentrated in a rotary evaporator, at medium heat, until only 10 µl of liquid was left. Purified DNA fragments were cloned into pJET1.2 as described in 2.14.2-6. Alternatively, purified DNA was used directly for sequencing.

2.14.7 DNA Sequencing

PCR products were either sequenced directly or after gel purification (as described in 2.13.5). Template (1 µl of PCR product) and primers (0.5 µl of 10 µM primers) were mixed in a 0.2 ml PCR tubes and sent to The GenePool Genomics and Bioinformatics Facility, Ashworth laboratories, the University Of Edinburgh. Sanger sequencing was carried out by the GenePool using the BigDye-terminator V3.1. Sequence files were returned in .ABI format and were aligned with Seqman II software (Lasergene Seqman, DNAStar).

2.15 Reverse transcription polymerase chain reaction (RT-PCR) for gene expression analysis

Expression analysis of *LDL1* and *VIP5* genes was done by RT-PCR.

2.15.1 Trizol method of RNA extraction

RNA extraction was done by a modification of the Trizol method described by (Chomczynski and Mackey, 1995) followed by purification on silica spun columns. Less than 100 mg of tissues, either fresh or frozen, were grounds in Trizol (1 ml /100 mg of tissue) in mortar and pestle in the fume hood. Blue pipette tips with the ends cut off were used to transfer ground tissue to Eppendorf tubes. These were centrifuged at 13,000 x g for 10 min at 4°C. The supernatant was taken and 0.2 volume of CHCl$_3$ added vortexed and spun for 15 min. at 4°C. The supernatant was transferred to a clean tube and 100% ethanol was added to give a final concentration of 35% ethanol (i.e. for 500 µl supernatant 270 µl 100% ethanol was added). The components were mixed well and up to 700 µl added to a spun column (Invitrogen Purelink RNA column). The column was spun at 12,000 x g for 15 sec. at room

68

temperature. The flow-through was discarded and more sample was loaded until all the sample has been through the column. The flow-through was discarded and 700 µl of wash buffer I was added and spun. The flow-through and collection tube were then discarded. The column was put in a new collection tube and 500 µl of wash buffer II (with ethanol) was added, spun through the column and discarded. This step with wash buffer II was repeated and the wash solution was again discarded. The column was spun for 1 min to remove any remaining wash buffer. It was transferred to new collection tube, 50 µl of RNase-free water was added to the column and it was spun for 1 min to elute the RNA. The RNA sample was put on ice and the column was discarded. An aliquot of the RNA solution (5 µl) was taken to estimate its concentration with a Nanodrop spectrophotometer and the rest was kept frozen at -80°C or in liquid nitrogen. The concentration was measured in a 2 µl aliquot using the Nanodrop. A_{260} and A_{280} absorbance readings were also taken. A 1 µg/µl solution of pure RNA should have an A_{260} of 25 and an A_{260}/A_{280} ratio close to 2.0 in water.

Because RNA is very susceptible to digestion by RNase, which is present on skin etc. and can survive even after autoclaving, several precautionary measures were taken. Tubes and pipette tips were used from newly opened bags and handled with gloves. Finally, RNA was dissolved in Diethylpyrocarbonate (DEPC)-treated water (RNase-free).

2.15.2 cDNA synthesis

PCR tubes with caps and Eppendorf tubes were taken from newly opened bags, wrapped in Saran Wrap and irradiated on UV transilliminator for 1 min to destroy any RNAse contamination.

2.15.3 RT buffer, dNTPs and Oligo dT primer mix

RT buffer, dNTPs and oligo dT primer were thawed. PCR tubes were on ice and contains the following reaction mix

Components	Volume
RNA (1 μg total)	1.0 μl
Oligo dT primer (100 μM)	1.0 μl
H$_2$O (RNase-free water)	3.0 μl

Table 12: showing RT buffer, dNTPs and Oligo dT primer mix.

Step 1	85°C	5 mins.
Step 2	4°C	On hold
Step 3	25°C	5 min.
Step 4	42°C	60 min.
Step 5	70°C	10 min.
Step 6	4°C	On hold

Table 13: showing RT-PCR programme.

Components	Volume
5X buffer	8.0 μl
dNTPs (10 mM each)	2.0 μl
H$_2$O (RNase-free water)	17.0 μl
RNase inhibitor (40U/μl)	1.0 μl
AMV RT (10U/μl)	3.0 μl

Table 14: showing RT cocktail mix.

15 μl of cocktail mix was added to each of PCR tubes, mixed, spun again and the PCR programme continued. When the PCR programme finished, 30 μl of H$_2$O was added to each tube and stored at -20$^\text{o}$C before PCR.

Constituents	Volume	Description
10X buffer	1.0 µl	Standard 10X buffer
2.5 mM of each dNTPs	0.2 µl	Rova lab.
Forward primer (10 µM)	0.2 µl	IDT
Reverse primer (10µM)	0.2 µl	IDT
Taq polymerase (5U/µl)	0.1 µl	
cDNA	3.0	
H$_2$O	5.3 µl	RNase-free water
TOTAL	10 µl	

Table 15: showing protocols for RT-PCR.

Step 1	94 ^{o}C	1 min.
Step 2	94 ^{o}C	10 secs.
Step 3	Ta ^{o}C	15 secs.
Step 4	72 ^{o}C	35 sec.
Step 5	Go to step 2	20X
Step 6	4 ^{o}C	On hold
Step 7	Go to step 2	25X
Step 8	4 ^{o}C	On hold
Step 9	Go to step 2	30X
Step 10	72 ^{o}C	10 min.
Step 11	End	

Table 16: showing RT-PCR programme

2.16 5'-RACE (rapid amplification of cDNA ends)

When reverse transcriptase reaches the 5' end of the mRNA, it adds oligo (dC) to the 3' end of the cDNA (manganese ions are essential for this activity) (Goldschmidt et al., 2006). The Smart II oligo anneals to this GGG extension, providing a template for reverse transcriptase to continue cDNA synthesis, incorporating the complement of the Smart II oligo at the end of cDNA.

Amplification is carried out with a mixture of UPML and UPML. The L "long" primer anneals to most of the Smart II sequence and S "Short" primer to its 5' part. A mixture of these primers is used in a first-round of amplification with a gene-specific primer. If necessary nested PCR was carried out with NUP (nested upstream primer) and a gene-specific primer closer to the 5' end of transcript.

Name	Sequence	Tm (^{0}C)
SMARTII	5'-AAGCAGTGGTAACAACGCAGAGTACGCGGG-3'	62
NUP	5'-AAGCAGTGGTAACAACGCAGAGT-3'	62
UPML	5'-CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT-3'	62
UPMS	5'-CTAATACGACTCACTATAGGGC-3'	62
GSP1	5'-GAGGAGGAGGAAGAGGAAGAGGCTGA-3'	62
GSP2	5'-CTTCTTCAATATTTTGCTCCGATTGG-3'	62

Table 17: showing primers for 5' RACE.

Components	Volume
Oligo dT (30 μM)	1.0 μl
dNTPs (10 mM each)	1.0 μl
Smart II oligo (30 μM)	1.0 μl
RNA (1 μg total RNA)	1.0 μl
H_2O	11.5 μl
	14.5 μl

Table 18: showing reaction mix for 5' RACE.

Components	Volume
5X RT buffer	4.0 μl
0.1M DTT	2.0 μl
BSA (10 mg/ml)	0.5 μl
$MnCl_2$ (80 mM)	0.5 μl
RNase inhibitor (40U/μl)	0.5 μl
Superscript II RT (200 U/ μl)	1.0 μl
TOTAL	8.5 μl

Table 19: showing cocktail mix.

This mix was always kept on ice.

Temp.(^{0}C)	Time
80	3 min.
4	On hold
25	3 min
42	60 min.
4	On hold

Table 20: showing PCR programme for 5' RACE.

Components	Volume
10X yellow PCR buffer	2.0 µl
dNTPs (10 mM)	0.4 µl
UP primer mix[1]	0.4 µl
GSP1(10 µM)[2]	0.4 µl
cDNA template	2.0 µl
Taq (5U/µl)	0.1 µl
H$_2$O	14.7 µl
TOTAL	20.0

Table 21: showing constituents of first round PCR with cDNA.

[1]: This is a mixture of 10 µl of 10 µM UPMS, 2 µl of 10 µM UPML and 88 µl H$_2$O.

[2]: More downstream of gene-specific primers.

A normal PCR programme, as described in 2.12.3, was used with annealing temperature of 62^0C. Because transcription can be initiated at different positions, slightly fuzzy bands were expected. Because there were number of visible bands, nested PCR was done to increase specificity. First-round PCR was diluted (~1:100 with water) and 1 µl of this dilution used in PCR with GSP2 and NUP (at usual concentrations).

2.16.1 Purification of cDNA free of Oligos using Qiaquick gel extraction kit

Purification of the cDNA was done with a Qiaquick Gel Extraction Kit to remove oligonucleotides that might interfere with amplification. Buffer QG (60 µl from the gel extraction kit) was added to each RT reaction and mixed well. Isopropanol (20 µl) was added to each reaction and again mixed well. The liquid was loaded onto a spun column in a collection tube and spun for 1 min at 17, 0000 x g. The flow-through was reloaded and again spun and discarded. PE (0.75 ml with added ethanol) was added to column to give it a wash. It was left for 2 min and then spun and the liquid discarded. This washing step was repeated once more. The column was put in a clean Eppendorf tube with its lid cut off and 50 µl H_2O was added and kept for 1 min. and spun. The column was discarded and flow through was stored at -20^{0}C.

2.17 Gateway Cloning

2.17.1 To Create a Gateway Entry Clone

The *LDL1* allele from parent 4D1 (genomic DNA) was amplified by PCR and transferred to a T-DNA binary vector using Gateway recombination. This involved three steps,

Step One – To produce PCR product

The *LDL1* allele was amplified by PCR as described in 2.12.2 (Table 6), except that Q5 proof-reading polymerase (New England Bio labs) was used to reduce the chance of PCR-generated mutations.

Step Two – To perform the TOPO Cloning Reaction

The PCR product was cloned in the plasmid vector pENTR dTOPO (http://www.lifetechnologies.com/order/catalog/product/K240020). This uses topoisomerase, rather than ligase, to join the vector and PCR product. The components were mixed gently and incubated for 5 minutes at room temperature and then kept on ice while proceeding to next step.

Step Three - Transformation of *E. coli*

Competent *E. coli* cells (section 2.14.2) were thawed and mixed with 2 µl of the TOPO cloning reaction. The mix was incubated on ice for around 20 minutes. Cells were then heat shocked for 30 sec. at 42°C without shaking and tubes transferred to ice immediately. Super Optimal broth with Catabolite repression (SOC) medium (1 ml) was added and cells incubated at 37°C for 1 hour with shaking. The bacterial culture was spread on LB agar plates containing 100 µg/ml kanamycin sulphate and

incubated overnight at 37°C. Colonies were assessed for the presence of the correct insert by colony PCR (section 2.14.5) using *LDL1* primers. An overnight culture of a positive clone was grown in 5 ml of LB with 100 μM kanamycin and the plasmid extracted using a Qiagen Qiaprep spin mini kit. The insert of the plasmid was sequenced with the same primers used to sequence PCR products from genomic DNA, to validate the absence of PCR-generated mutations.

2.17.2 LR Clonase Reaction

The cloning site in the pENTR D-TOPO vector is flanked by *attL* recombination sites (*attL1* and *attL2*). This allows the insert to be transferred to a Gateway vector carrying *attR* sequences by recombination with a commercial mixture of recombinase enzymes (LR Clonase, Invitrogen). The following components were added to 1.5 ml Eppendorf tube at room temperature and mixed.

1	pENTR/D-TOPO donor plasmid (~150 ng/μl)	1 μl
2	pGWB1 acceptor plasmid (150 ng/μl)	1 μl
3	TE buffer, pH 8.0	8 μl
4	Total	10 μl

The LR Clonase™ II enzyme was thawed on ice for about 2 minutes and vortexed briefly twice to mix (2 seconds each time). 2 μl of the enzyme mix was then added to the above reaction mix, vortexed twice and micro-centrifuged. The reaction was incubated for 1 hour at 25°C in a PCR machine. 1 μl of the Proteinase K solution

supplied with the Clonase was then added to the mix to terminate the reaction. The sample was mixed and incubated at 37°C for 10 minutes.

2.17.3 Transformation

1 µl of each LR reaction was transformed into 50 µl of competent DH5α cells and 100 µl of each transformation were spread onto selective plates containing 100 µg/ml kanamycin and 50 µg/ml hygromycin (pGWB1 encodes resistance to both antibiotics, but the pENTR dTOPO donor also encodes kanamycin resistance). Colonies were screened for the insert by PCR and the recombinant plasmid was extracted and transformed into *Agrobacterium* with the freeze-thaw method. No control was used in this case.

Reagent	Volume
Fresh PCR product	3 μl
Salt Solution	1 μl
Dilute Salt Solution	--
Sterile Water	5 μl to a final volume of 5 μl
pENTR/D-TOPO Vector	1 μl
Total volume	10 μl

Table 22: Constituents of TOPO Cloning Reaction.

2.17.4 *Agrobacterium* strain GV3101:

This strain carries resistant to gentamycin, encoded by the binary helper plasmid, and rifampicin encoded in the bacterial chromosome, therefore, 50 µg/ml Gentamycin and 10 µg/ ml rifampicin were added to LB broth or agar for selection. GV3101 is sensitive to kanamycin, so is a good strain for use with binary vectors that confer kanamycin resistance (or chloramphenicol resistance) in bacteria. The binary helper plasmid consists of a disarmed Ti plasmid that possesses the virulence genes needed for T-DNA transfer but has no functional T-DNA region of its own. The strain was grown at 28-30°C and stored as glycerol stock (800 ml of fresh overnight culture + 200 µl sterile 80% glycerol) at -80°C.

2.18 Transformation of *A. tumefaciens* with plasmid DNA (binary vector system)

A freeze/thaw shock transformation method was used for transformation of GV3101, thus eliminating the need for the old genetic method of transferring a plasmid maintained in *E. coli* by tri-parental mating.

2.18.1 Freeze/thaw shock transformation

A single colony of GV3101 was picked and was and inoculated into 3 ml of LB with gentamycin and rifampicin in a 15 ml snap-cap tube. It was grown at 30°C overnight in a shaking incubator. 50 ml of LB with antibiotics was inoculated in a 250 ml flask with 0.5 ml (1/100 volume) of the overnight culture and grow at 30°C for around 4 hours until it reached mid-log phase (an OD600 between 0.5 and 1.0). The culture was chilled for 5-10 min on ice and centrifuged at 3000 x g for 5 min at 4°C in

chilled, sterile centrifuge tubes. The supernatant was discarded and the tube drained tube by keeping them inverted for 30-60 seconds. The bacteria were resuspended in 1 ml of ice cold 20 mM $CaCl_2$ and 0.1 ml of bacterial suspension was dispensed into pre-chilled 1.5 ml Eppendorf tubes on ice, frozen in liquid nitrogen and stored at -80°C. For transformation, cell suspensions were allowed to thaw on ice; around 1 μg of plasmid DNA was added and mixed gently. Tubes were then frozen in liquid nitrogen and thawed in a water bath at 37°C for 5 min. One ml of LB was added to each tube, and the suspension transferred it to a 15 ml snap-cap tube and incubated for ~2 hours in a shaking incubator at 30°C. The mix was then transferred into a 1.5 ml microfuge tube and spun at 13,000 xg for 5 minutes to pellet cells. Most of the supernatant was discarded, the pellet resuspended in the remainder and spread onto LB plates with antibiotic selection (100 μg/ml kanamycin, 50 μg/ml gentamycin). Plates were incubated at 30°C and transformed colonies became visible on the second day of incubation. The presence of the correct T-DNA was confirmed by colony PCR.

2.18.2 Floral Dip of *Arabidopsis* transformation

This method was adapted from (Clough and Bent, 1998). Plants were dipped once the first flowers had started to open. Three days prior to plant transformation, 5-ml of liquid culture of *Agrobacterium* carrying a binary vector was inoculated and incubated at 28ºC with vigorous agitation. LB medium was used containing antibiotics that select both the Ti and the T-DNA plasmids (gentamycin and kanamycin, respectively). After 2 days, 200 ml of LB medium was inoculated with 1 ml of the pre-culture and incubated again with vigorous agitation for an additional 24 hours at 28ºC. *Agrobacterium* was pelleted by centrifuging at 6 000 rpm in a GSA rotor and

the cell pellet resuspended in 400 ml of 5% sucrose solution with 0.1% Silwet. The suspension was transferred to the lid from a box of disposable pipette tips and plants were inverted to dip the inflorescence shoots into the suspension and allowed to soak for ~30 seconds. After dipping, the pots were laid on their sides and covered with a transparent propagator for the 24 hours. After 24 hours, the covers were removed and the plants stood upright. Plants were dipped again after 1 week. Seeds were collected about 3 weeks later. No selection of transformed plants was done due to shortage of time.

3.0 RESULTS AND DISCUSSION

3.1 Growth rate variation, anthocyanin synthesis and herbivory in plants of F3 and F5 populations.

3.2 Introduction

Heritable variation had been found in the growth rate and flowering times of local accessions of *Arabidopsis* (Figure 6). These accessions had been collected from in and around Edinburgh. Evidence had also been found that some local genotypes grew faster, relative to the genotypes as a whole, in spring compared to autumn (Figure 7), suggesting that variation in local populations might be maintained by adaptation to factors that varied between season. However some genotypes remained fast or slow growing in both seasons. This suggested that the local accessions might differ for genes that affect growth rate in both spring and autumn or that affect growth more in one season than the other. A major aim of this chapter was to study the inheritance of growth rate variation in a hybrid population between local accessions that responded differently to season and to test how different genes might respond to season by growing recombinant offspring under different field conditions. Plants grown in the field were also seen to produce different levels of anthocyanin pigment in response to freezing temperature and to show different levels of herbivore damage. The heritability of these traits was also examined and tested for correlation to growth rate and flowering time. Different genotypes also flowered at different times in the field and responded differently to vernalization (the effects of low temperature on

flowering), so could flower later than others in one season but before others in another season.

Growth of plants responds to both external and internal factors and final plant size results from integration of both environmental and genetic factors (Granier and Tardieu, 2009). Natural genetic variation has been reported for different growth rate parameters that vary from few to many times (Koornneef et al., 2004). Genetic variation in growth rate, measured using digital images of rosette area has also been described in *Arabidopsis* (Leister et al., 1999). High growth rate has been reported to be positively correlated with the fitness of plants (Milla et al., 2009). Correlations have also been found between growth rate parameters and other aspects of plant behaviour. These include positive correlations between stable carbon isotope ratios and flowering time (McKay et al., 2003). Genetic variation has also been assessed in *Arabidopsis* grown in variable environments differing mainly in mineral content like phosphorus and nitrogen (Loudet et al., 2003a; Rauh et al., 2002). Because *Arabidopsis* is a member of the Brassicaceae, an economically important family, genes involved in growth rate variation in *Arabidopsis* may be relevant to crop breeding programmes.

The ability of a plant to maintain fitness despite herbivore damage is expected to change during the life cycle of a plant as it is linked to growth and resources available for growth (Tucker and Avila-Sakar, 2010). Studying correlations between herbivory and growth rate and flowering time might reveal interactions between plant development and herbivory.

Attraction of pollinators and seed dispersers are major functions of coloured anthocyanin in plants. Anthocyanins also provide protection against UV (Sarma and

Sharma, 1999) and cold temperature (Christie et al., 1994). They are also produced in response to microbial attack (Lorenc-Kukula et al., 2005) and drought (Castellarin et al., 2007). Scoring anthocyanin productions in response to cold stress is a part of this study.

3.3 Growth rate variations in control conditions
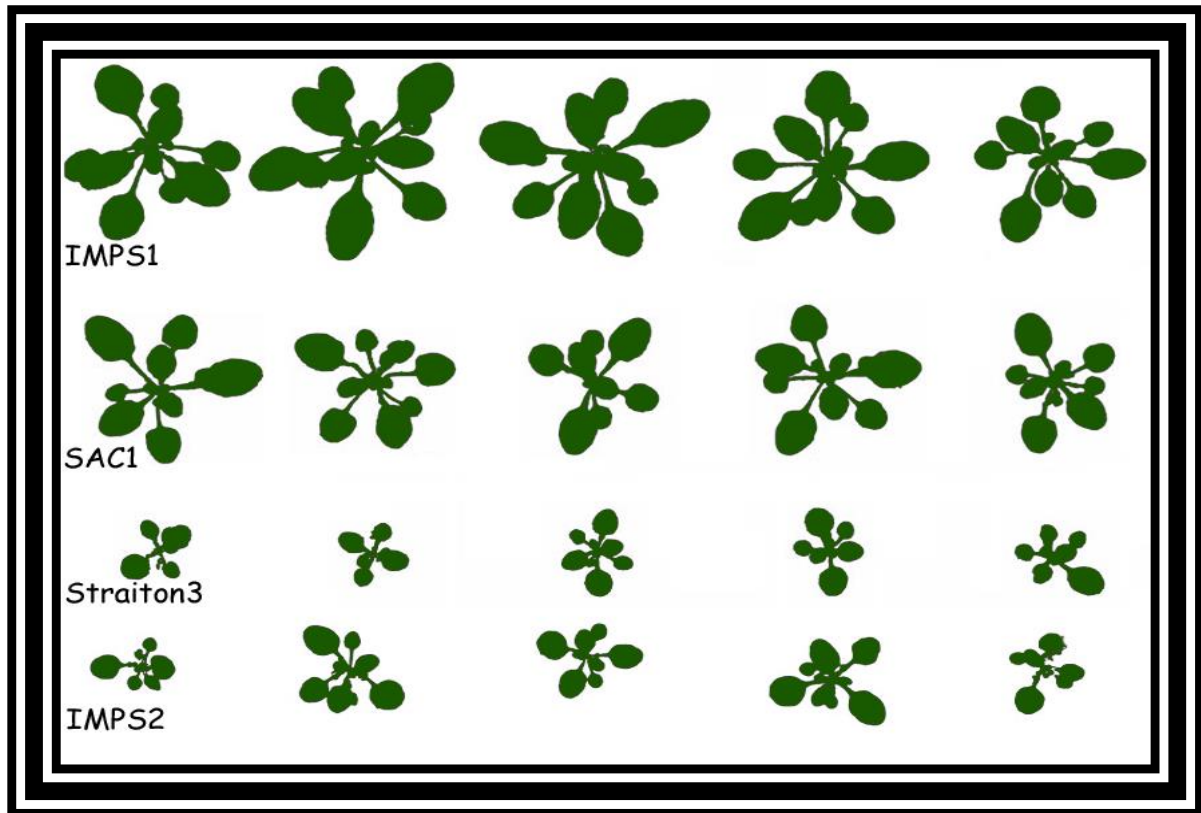
Figure 6: Growth rate variations within and between different lines in local accessions of *Arabidopsis thaliana*.

Variation in growth rate was observed not only in the plants of different lines (due to difference in environmental conditions and genotype) but also in the plants of the same line (due to non-genetic variation, including environmental conditions). Most of the variation in this case occurred within lines, showing that the trait is highly heritable. Images from (Lim and Hudson, 2013).
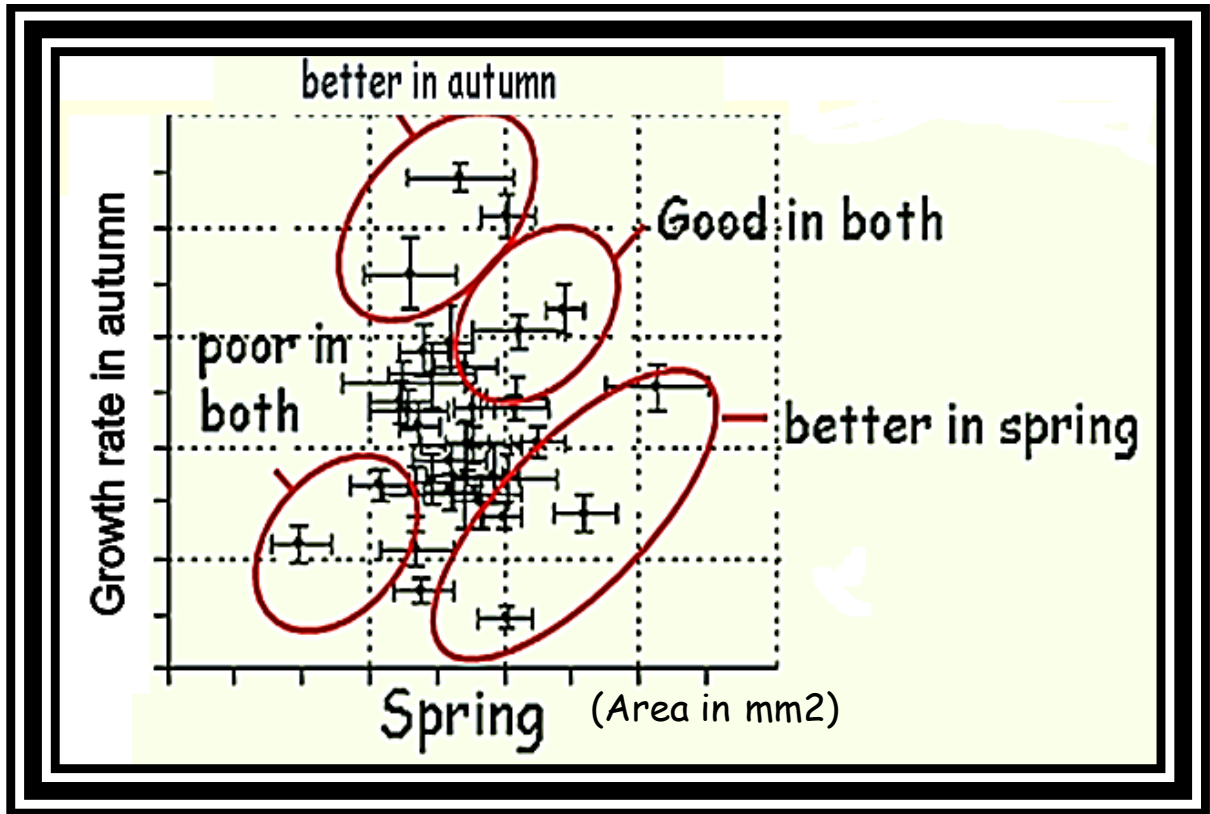
Figure 7: Growth of plants in different seasons.

Each point represents the mean growth rate of one genotype (± SEM) under autumn compared to spring conditions. Different genotype responded differently when they were grown in spring or autumn. Some showed better growth, relative to the rest of the plants, in both spring and autumn whereas some showed poor growth in both seasons. Others performed better in spring or in autumn. (Figure from (Lim and Hudson, 2013).

3.4 Mapping of growth rate genes

A fast growing accession of *Arabidopsis thaliana,* 11C1 was crossed to slow growing 4D1 as female to get F1 progeny and F1 lines were allowed to self-pollinate for a further four generations to get F5 families. F3 and F5 plants were used in my project (Figure 8). Because members of an F3 or F5 family are more genetically similar to each other than they are to other families in the same generation, they can be used to replicate similar genotypes either under the same conditions, to give a better estimate of the mean phenotype (e.g., growth rate) for the genotype, or under different conditions, to examine the responses of genotypes to the environment.

Earlier results had indicated that 11C1 is faster growing both under long day and short day conditions than the mother, 4D1. To map growth rate genes, the F3 generation were grown in a greenhouse (controlled conditions) and outside (natural conditions) in both spring and autumn. Similarly, F5 plants were grown in a growth room (controlled conditions) and outside (natural conditions) in both seasons. F3 families were used in initial QTL analysis whereas F5 lines were used in bulk-segregant analysis. Pedigrees of both parents and offspring used in the experiments are shown in Figure 8.

P   **Fast growing 11C1 (♂) × Slow growing 4D1 (♀)**

⇓

F1   **(1)**, (2), (3), (4,) (5), (6), (7), (8), (9), (10) etc.

⇓

F2   (1-1), (1-2), (1-3), (1-4), (1-5)

⇓

F3   (11-1), (11-2), (11-3), (11-4), (11-5)
      (Offspring used in QTL analysis)

⇓

F4   **(111**-1), (111-2, (111-3), (111-4), (111-5)

⇓

F5   (Offspring used in bulk-segregant analysis)
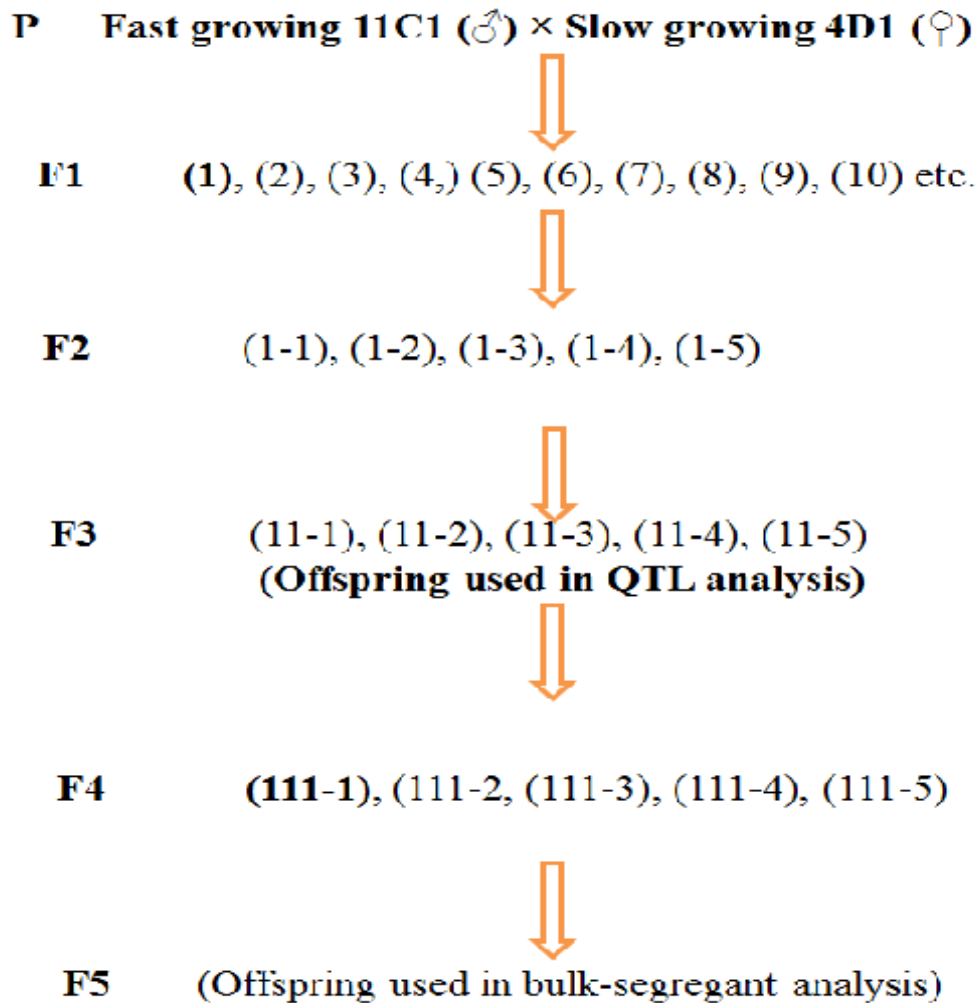
Figure 8: Pedigree of parent plants and offspring used in experiments.

This figure shows production of populations from F1 to F5. A single plant from each family was allowed to self-pollinate in the F2 generation. The seeds of more than one member of each F3 family were harvested separately to grow F4 families; therefore some families in the F4 and F5 generations are more closely related to each other than to the rest. Each F5 family was produced from the bulked seeds of five members of an F4 family. Therefore the frequency of F5 families carrying both parental alleles is usually lower than 50%.

3.4.1 Estimation of plant growth in a greenhouse (controlled conditions) of F3 population

In the green house, 71 F3 hybrid lines were grown in autumn (GH1) and 82 lines in spring (GH2) with five members of each line. Plants of each line were randomised in different trays for observation to minimise the effects of environmental variation within the greenhouse on the mean growth rate of each line. The two parent lines were also grown. Plant rosette areas were estimated with Adobe Photoshop and Image Tool in mm^2. Area was used as a proxy for growth rate because it had been shown to be more heritable than estimates of relative growth rate made from two or more rosette area estimates and because the increase in rosette area with time is closer to linear than the exponential rate assumed in calculation of relative growth rate (Poay Lim, personal communication). In GH1 conditions, plants showed transgressive segregation for growth rate - some plants showed a growth rate that was higher than the 11C1 parents; others grew slower than the 4D1 parent (Figure 9a). This suggests that the parents carry different combinations of alleles that increase or decrease growth in autumn conditions, so that some of their offspring can inherit more increasing or decreasing alleles than were present in one of their parents. In spring, however, 11C1 grew faster than all the F3 lines (Figure 9b), suggesting that 11C1 carries mainly genes that, under spring conditions, increase growth.

3.4.2 Estimation of plant growth in outside garden (natural conditions) of F3 population.

A set of 87 F3 lines was grown outside in autumn and another set of 80 lines in spring with five members of each line. Plants were grown in standard compost in pots, to

minimise any differences in substrate, and pots were randomised as in the greenhouse. A near-normal distribution of plant areas was again seen (Figure 9c and 9d). Transgressive segregation was again observed, suggesting that both parents may contain fast and slow alleles of growth rate genes.

3.4.3 Estimation of plant growth in an outside garden (natural conditions) of F5 population.

Eighty-nine lines were grown in the outside garden in spring (OS3) and autumn (OS4) with five members of each line. The populations showed transgressive segregation for growth rate (Figure 10a and 10b). This suggests that the parents carry different combinations of alleles that increase or decrease growth in outside autumn and spring conditions.

3.4.4 Estimation of plant growth rates in a growth room (control conditions) for the F5 population.

The 89 F5 lines were grown in spring (GR1) and autumn (GR2) in a growth-room with five members of each line. In GR1 and GR2 plants showed transgressive segregation for growth rate. Some plants showed a growth rate that was higher than the 11C1 parents; others grew slower than the 4D1 parent (Figure 10c and 10d). This suggests that the parents carry different combinations of alleles that increase or decrease growth in control conditions of growth room in autumn and spring.

Though transgressive segregation was observed in F3 and F5 populations in all conditions there were differences between the sets of plants (Figure 9 and 10). For the F3 population grown in the green house both in autumn and spring and in natural conditions in spring most of hybrid lines are skewed to lower values than the parental

lines, compared to natural conditions in autumn where most of the hybrids are in between the parental lines (Figure 9). This suggests that alleles of the genes that underlie differences in natural conditions (autumn) are mostly distributed so that the fast parent has the increasing alleles and the slow parent the decreasing alleles. For the genes that determine growth of F3 population in other conditions, the alleles are more mixed in their distribution between parents. However, in the F5 population the parents grow faster than most of hybrids in natural conditions in spring compared to autumn and in growth room in both seasons where most of the hybrids are in between parental lines (Figure 10). These results suggest that genes that underlie variation in growth rate in the growth room (both in autumn and spring) and in autumn when grown in natural conditions have a more mixed distribution between parents than the genes expressed in natural conditions in spring.
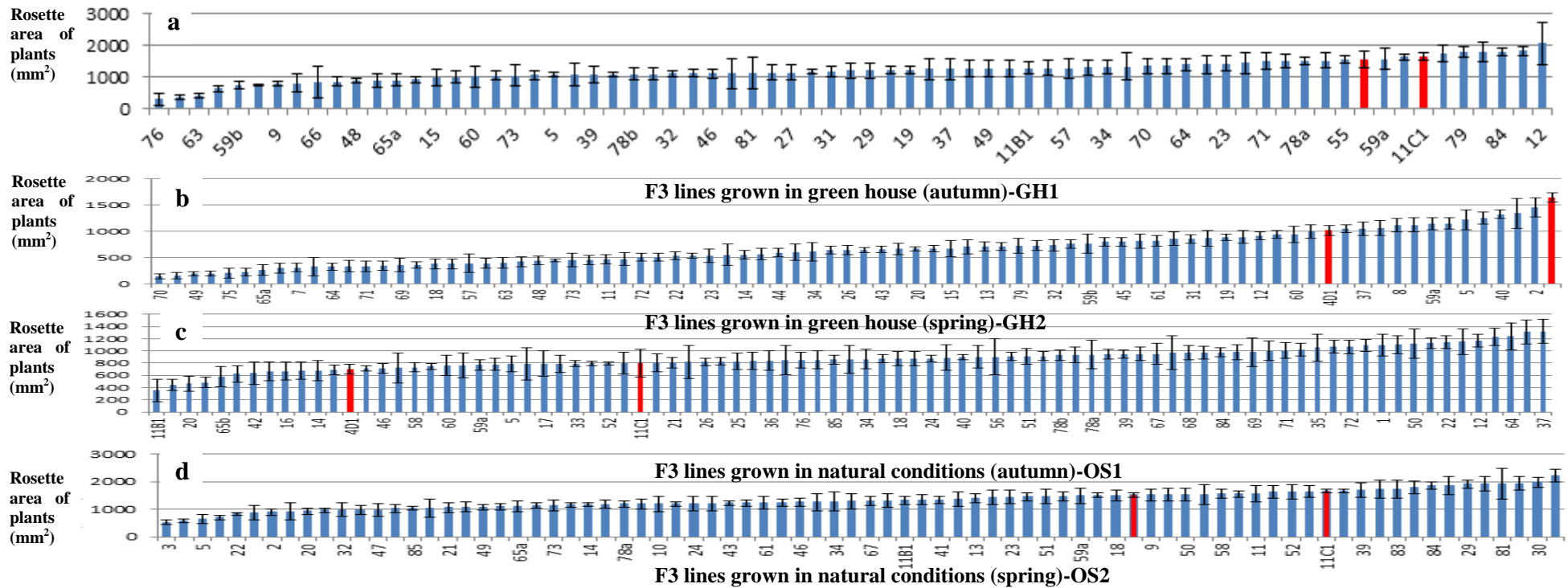
Figure 9: Variation in growth rate in (a) green house (autumn) and (b) green house (spring) and (c) natural conditions (autumn) and (d) natural conditions (spring).

These graphs showed the distribution of mean rosette areas (mm^2) with ($\pm$ SEM) of 5 plants of each F3 line (blue bars) including parental lines in the 6th week after sowing (parental lines are shown in red bars). Transgressive segregation was observed in F3 lines in both autumn and spring seasons in the green house and natural conditions, suggesting that parental lines contain both fast and slow alleles of growth rate genes.
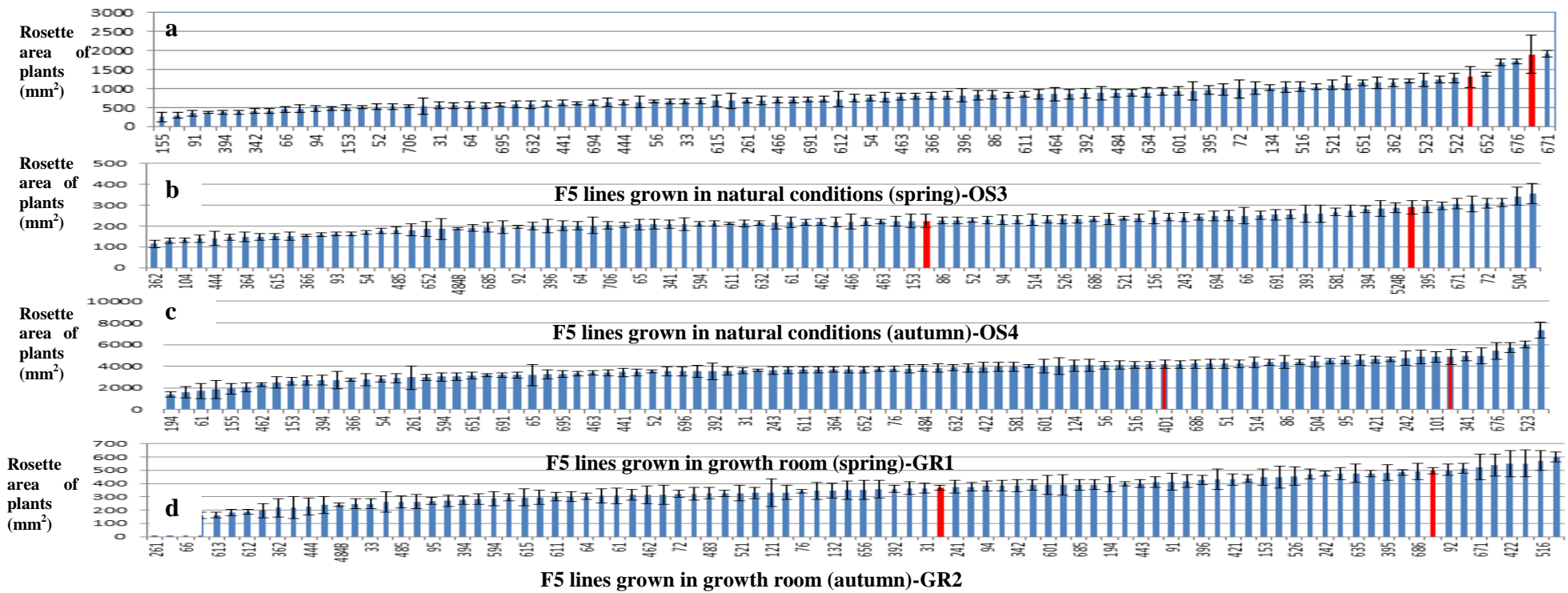
Figure 10: Growth rate variation in plants of F5 population with (± SEM) grown in (a) natural conditions (spring) and (b) natural conditions (autumn) and (c) growth room (spring) and (b) growth room (autumn).

These variations in growth rate of F5 lines (blue bars) in growth room and natural conditions in both seasons suggest that parental lines (shown in red bars) contain mixtures of fast and slow alleles of growth rate genes.

98

3.4.5 Growth rate variation within an experiment

F3 lines growing in the green house in autumn were analysed for area after weeks 6 and 7. The mean area measurements for families on these two dates were correlated, but only loosely (Figure 11). These results suggest either that there is a significant error in estimating rosette area at one or both time-points and that growth rate is not constant for all lines during the experiment.
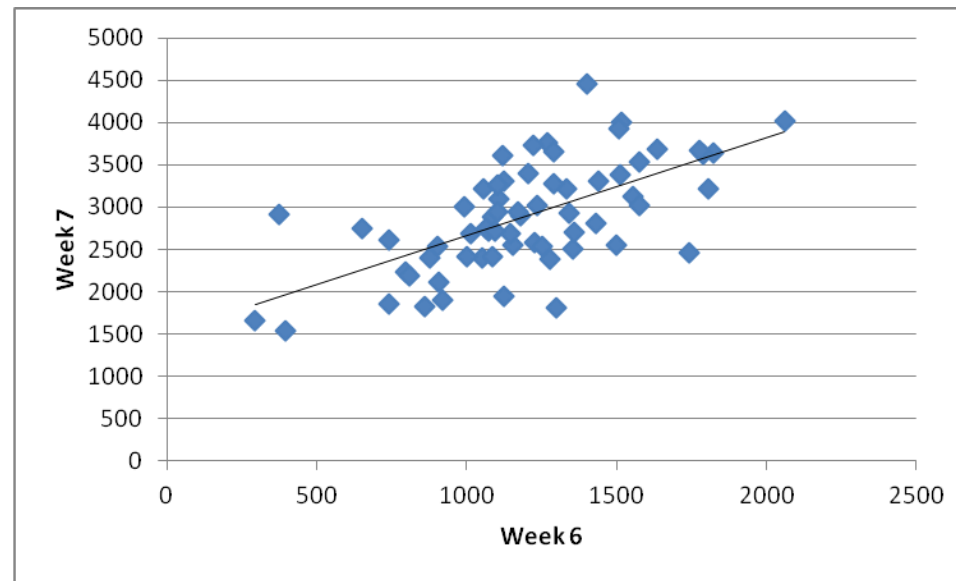
Figure 11: Comparison rosette areas for F3 families in the green house in week 6 and 7.

Although areas at week 7 and week 6 are correlated, there is considerable variation around the regression line with p (uncorr) = 2×10^{-8} and r= 0.63. Values are mean rosette areas of 5 F3 plants.

3.5 Heritability

If members of the same lines are genetically identical, the differences in growth within a line will only be due to non-genetic differences ("environment") while variation between lines will be due to genetics and environment (Figure 6). The broad-sense heritability of a character can therefore be estimated as the proportion of the total variance that occurs between lines. However, although members of an F3 line are more closely related to each other than to other F3 lines, they are not genetically identical - 50% of families on average will segregate for each QTL (Figure 12). Although heterozygosity will be lower in the F5 lines, 25% of families will have a mixture of genotypes at each QTL locus (each family consists of the bulked offspring of five F3 parents). Therefore, some of the variations within each line will be due to genetics, rather than environment. Therefore, estimates of heritability (H^2) values from F3 and F5 will probably be the underestimates and the estimates should be lower for the F3 than the F5. Estimates of between-family variance also do not take into account any environmental effects on the seeds, because the seeds of a line were produced on the same parent at the same time. This might increase the estimate of heritability. However, previous studies have found that these maternal effects are low (Mousseau and Fox, 1998).
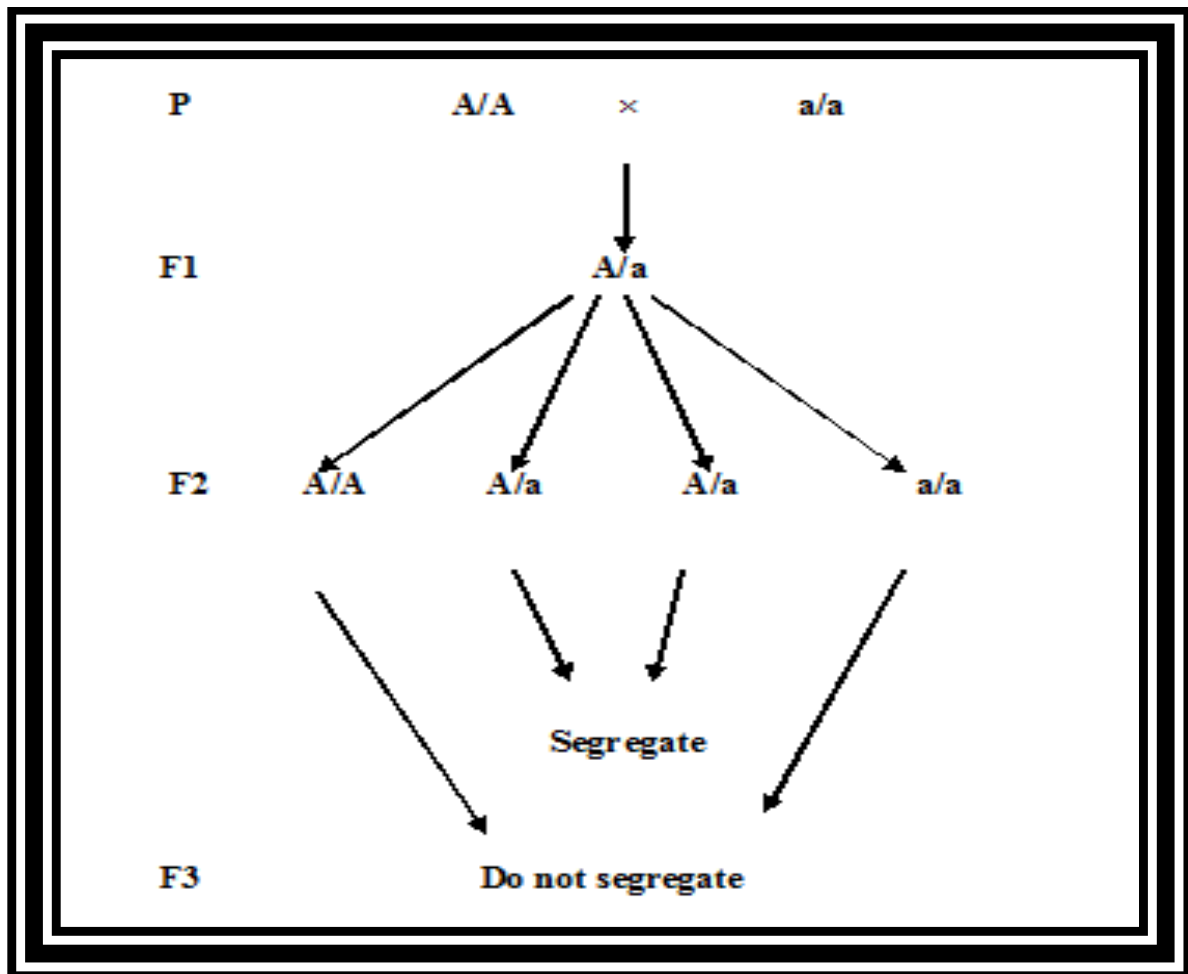
Figure 12: 50% families do not segregate in F3 (only for one locus)

Calculation of heritability is done with following formula,

Heritability (H^2) = Variance between plant lines / Total variance

Population and seasons	Natural (outside) conditions	Green house conditions	Growth room conditions
F3 autumn	56%	54%	--
F3 spring	72%	86%	--
F5 spring	86%	--	80%
F5 autumn	80%	--	80%

Table 23: Heritability estimates for F3 and F5 populations in different environmental conditions and seasons.

These values are conservative as not all lines in F3 and F5 populations are genetically identical, as assumed while calculating these values.

Heritability of growth rate for the F3 generation in the green house was estimated as 54% in autumn, and 86% in spring. It was similar outside in autumn (56%) and lowers than in the greenhouse (72%) in spring (Table 23). In general higher heritability was observed in spring compared to autumn. One possible explanation for this is that autumn plants grew for a longer period over winter and therefore had more opportunity to be affected differently by environmental factors. For the F5 population heritability estimates ranged from 80% to 86% (Table 23). These were generally higher than for the F3 population, as expected because of the lower frequency of segregating families in this generation. Again, estimates for populations grown outside were higher in spring, presumably because they have shorter life cycles in spring. This is supported by the similarity of the values for the growth room (controlled conditions) in both seasons. These heritability values always give minimum estimates as we assume that all members of same line are genetically identical.

3.6 Growth rate correlations

If the same genes determine growth rate variation under different conditions, then the growth rates of families under different conditions should be correlated. Growth rate correlations between lines of F3 and F5 populations grown under different environmental conditions were tested.

3.6.1 Growth rate correlations between lines of F3 population

Correlations were examined between the growth rates of F3 lines grown in green house and natural conditions in autumn and spring.

TRAITS	GH1		GH2		OS1	
Green house autumn (GH1)	**p**	**r**				
Green house spring (GH2)	0.70	0.04	**p**	**r**		
Natural conditions autumn (OS1)	0.07	0.21	0.96	0.004	**p**	**r**
Natural conditions spring (OS2)	0.01	0.28	0.24	0.13	0.07	0.12

Table 24:   Growth rate correlations between F3 lines in different environmental conditions and seasons.

Mean area values for each family were used in each condition of growth. Non-significant correlations suggest the involvement different genes in growth in different conditions or high non-genetic components of variation.

No significant correlation was observed between GH1 and GH2 [p (uncorr) = 0.70 and r=0.04], GH2 and OS2 [p (uncorr) = 0.24 and r = 0.13], and GH2 and OS1 [p (uncorr) = 0.96 and r = 0.004] (Table 20). Significant but low correlations were seen between (OS1 and OS2 [p (uncorr) = 0.0.07 and r =0.12], GH1 and OS1 [p (uncorr) = 0.07 and r = 0.21] and GH1 and OS2 [p (uncorr) = 0.01 and r =0.28] (Table 24). This suggests either that different genes are involved in growth of plants in different seasons and also in different environmental conditions. Alternatively, the environmental component of variance might be high enough to hide any genetically determined correlation. When the same plants were compared at two different times OS1 (November) and OS1 (March), correlations were significant and higher (p (uncorr) = 2.1E-06 and r= 0.48). Although it is not surprising that plants that were large in autumn remained large in spring, and *vice versa*, the relatively low level of correlation suggests that environmental variance is high.

The growth rate values for the F3 lines were used together with their genotypes to test for growth rate QTL under the different conditions.

3.6.2 Growth rate correlations between lines of F5 population

Correlations were also examined between the growth rates of F5 lines grown in natural and growth room conditions in both autumn and spring. If different genes are involved in growth rate in different conditions, then the fastest growing plant in one condition might not be the fastest growing in other conditions and *vice versa*.

TRAITS	GR1					
Growth room spring (GR1)	p	r	GR2			
Growth room autumn (GR2)	0.003	0.30	p	r	OS3	
Natural conditions spring (OS3)	2.6×10^{-9}	0.58	0.07	0.19	p	r
Natural conditions autumn (OS4)	0.003	0.30	3.5×10^{-5}	0.42	0.001	0.33

Table 25: Growth rate correlations between F5 lines in different environmental conditions and seasons.

Mean values for family rosette area were used. High correlation values suggest the involvement of similar genes in growth rate in different conditions.

Relatively high correlations (r>0.30) were seen in all cases except OS3 and GH1 (r=0.19; Table 25). These suggest that the same genes might affect growth rate in different environmental conditions in autumn and spring. Correlations are generally higher in the F5 than in the F3. This can be explained that there will be more homozygous families for growth rate QTL in F5 population compared to the F3. These growth rate values of F5 lines were used in the bulk-segregant analysis described later.

3.6.3 Growth rate and flowering time correlations of F5 population

Correlations were also examined between growth rates and flowering time of F5 lines grown in growth room and natural conditions in autumn and spring (Table 26). Significant correlations were observed between growth rates and flowering time of lines grown in natural conditions and between lines grown in natural conditions and growth room (Table 26; r>0.32) suggesting the involvement of similar genes in growth rate and flowering time in different conditions. However, no significant correlation was found between growth rate and flowering time in other sets of plants in the F5 population (Table 26) which suggest the involvement of different genes in growth rate and flowering time in growth room and natural conditions. Similarly, no significant correlations were found between flowering times of F5 plants grown in natural conditions and growth room in autumn (data not shown).

Traits	Natural conditions autumn (OS4)-flowering time		Growth room autumn (GR2)-flowering time	
	p	r		
Natural conditions autumn (OS4)-growth rate	0.0001	0.39	p	r
Natural conditions spring (OS3)-growth rate	0.59	0.05	0.34	-0.1
Growth room autumn (GR2)-growth rate	0.001	0.32	0.33	0.10
Growth room spring (GR1)-growth rate	0.21	0.13	0.59	0.05

Table 26: Growth rate and flowering time correlations between F5 lines in different environmental conditions in autumn.

Mean values for family rosette areas were used as an estimate of growth rate and mean family values for the number of leaves were used as flowering time estimates. High correlation values suggest the involvement of similar genes in growth rate and flowering time in different conditions.

3.7 Anthocyanin production

Plant leaves change their colour in winter mainly due to the anthocyanin production, which can therefore be considered as marker for cold stress. Some members of the set of 87 F3 lines that were grown outside from October produced anthocyanin during the winter. Anthocyanin production was scored at the end of January on a scale of 1-5 from green (Fig. 13c) to purple (Fig. 13d) by comparing images of plants. A heritability value of 48% was found for this trait. This trait data was also used in QTL analysis, as described later. A significant correlation was found between anthocyanin production and growth rate of plants in natural condition in autumn (Table 27) suggesting that faster growing plants tend to produce more anthocyanin.

Trait	Green house (autumn)		Green house (spring)		Natural conditions (autumn)		Natural conditions (spring)		Flowering time-natural conditions-spring		Flowering time-growth room-spring		Anthocyanin in autumn	
	p	r	p	r	p	r	p	r	p	r	p	r	p	r
Anthocyanin production	0.17	0.16	0.28	0.13	0.02	0.27	0.50	0.08	0.64	-0.10	0.39	0.19		
Herbivory	0.40	0.10	0.30	0.12	0.05	0.23	0.88	-0.01	0.006	-0.57	0.74	-0.07	0.31	0.23

Table 27: Anthocyanin synthesis and herbivory correlations with growth rates and flowering times of F3 lines.

Significant correlation between anthocyanin production and growth rate suggest that fast growing plants tends to produce more anthocyanin. Negative significant correlation between herbivory and flowering time suggests that early flowering plants are more susceptible to herbivory.

3.8 Herbivory

Between October and January, herbivores, probably slugs, attacked and damaged plants in the field (Figure 13a, b). Most of the plants were eaten either partially or totally. The percentage area of each plant eaten was estimate as the difference in the predicted area in January (estimated by regression of January and October rosette areas for uneaten plants – see Materials & Methods) and the area remaining. A minimum heritability estimate of 48% was found for this trait. The data for this trait are used later in QTL analysis. Plants were randomly placed in the garden and there were no protective measures against slugs. However, herbivory may rely on the other factors like positioning of plants as well. A significant correlation was found between herbivory of families and their growth rates (Table 27), suggesting that faster growing plants are more susceptible or perhaps more attractive to herbivory. A significant negative correlation was also found between herbivory and flowering time (Table 27), suggest that earlier flowering plants are more susceptible to herbivory.
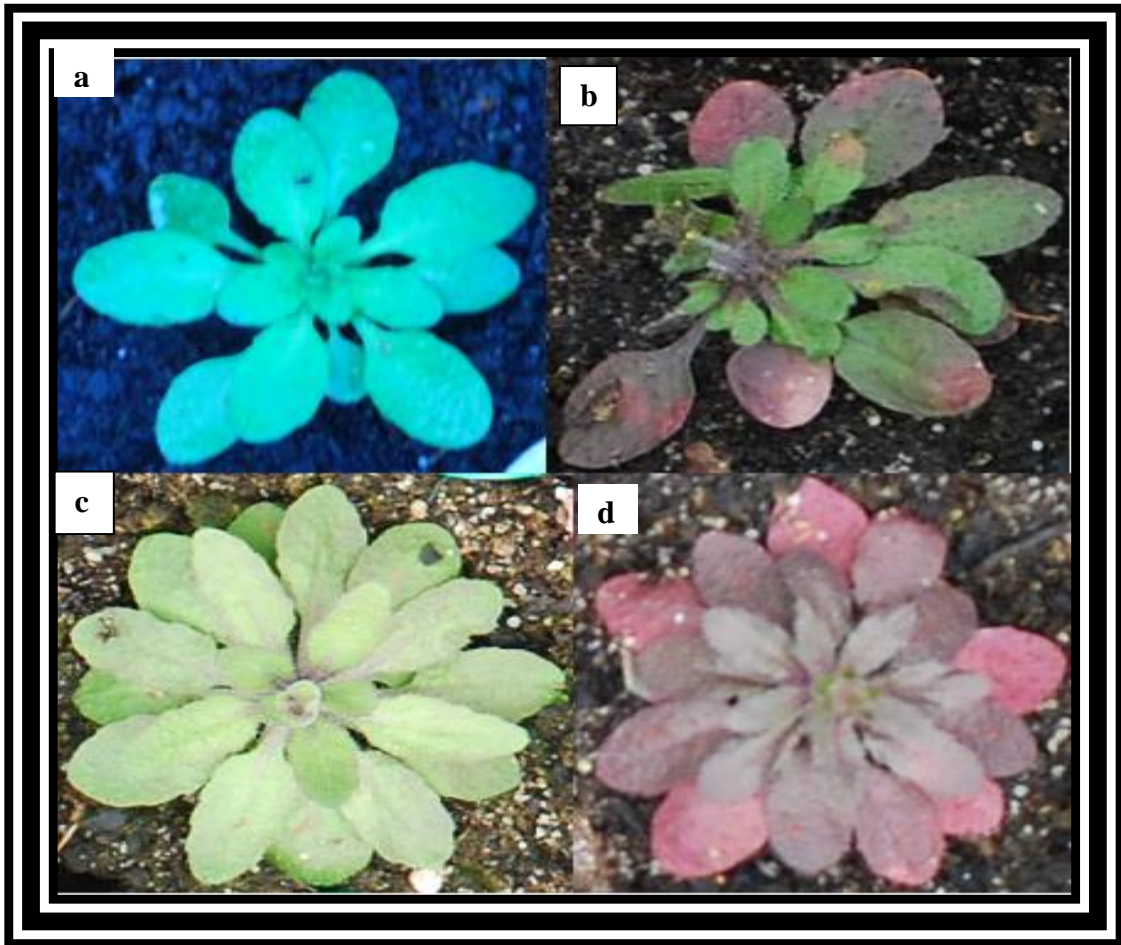
Figure 13: Herbivory and anthocyanin production

a) shows a plant in October that was partly eaten by January (b). (c) is an example of a plant

with the lowest level of anthocyanin in January and (d) a plant with the highest level.

3.9 Discussion

Relatively high heritability estimates for growth rate of either F3 or F5 populations showed that much of the variation, in either controlled or natural conditions, is genetically determined.

Higher heritability in growth rates was observed in spring than autumn in both F3 and F5 populations, in both natural conditions and in more uniform environmental conditions of the greenhouse and growth room. This trend of higher heritability of growth rate variation in spring than autumn continued with same pattern at different time points of plants growth. There is evidence that spring phenology has a strong genetic component in perennial plants (Pellis et al., 2004; Tsarouhas et al., 2003; Weih, 2009) and this might affect productivity, because it enables the plants to use the conditions that are more favourable during spring and early summer than in autumn. Autumn phenology is closely linked to frost in many plants and a correlation between frost resistance and low growth or growth cessation has been confirmed in perennial plants (Ögren, 1999). The nitrogen contents of plants are also positively related to plants survival in winter and shoot growth in spring (Chapin et al., 1990). There is also evidence that *Arabidopsis* shows inherently high growth rate when grown in favourable conditions (Grime and Hunt, 1975; Poorter and Remkes, 1990).

In the F3 population, growth rate of plants grown in the green house in autumn were significantly correlated with plants grown in natural conditions in autumn and spring. Similarly, plants grown in autumn and spring in natural conditions were also significantly correlated with each other. These results support the idea of involvement of similar genes in plants growth in different conditions. No significant

114

correlation in growth rate in other plants sets of F3 generation suggest that there might be different genetic factors involved in growth in different conditions. Significant correlations in growth rate were observed for all sets of plants grown in both natural conditions and the growth room in autumn and spring for F5 populations. This again suggests that similar genetic factors might be involved in growth rate in different conditions.

High broad-sense heritability estimate of anthocyanin production shows that it is also genetically determined trait. Anthocyanin production was significantly correlated with the growth rate of plants outside in winter. This means that fast growing plants tends to produce more anthocyanin and *vice versa.* This also suggests the involvement of similar genes in growth rate and anthocyanin production. One possibility is that the same genes control anthocyanin production and growth in parallel. So, for example, plants that are better adapted to low temperature respond by maintaining their growth and producing more anthocyanin. Alternatively, the underlying genes might affect anthocyanin production, which then affects growth or *vice versa*. This hypothesis could be tested by studying the effects of anthocyanin mutations on growth rate or the effects of isolated growth rate QTL on anthocyanin production.

The results also showed that susceptibility to herbivory is also heritable trait. A variety of herbivores use plants as food and different plants have different methods to tackle these attacks and survive. At the individual level, these herbivores have negative effect on plant growth, reproduction and survival (Crawley, 2009). Plants use both direct and indirect strategies of defence to herbivores (Snoeren et al., 2010). Indirect defence in plants has been studied extensively (Arimura et al., 2005; Dicke

115

et al., 2009). The epicuticular wax loads, trichomes or glucosinolate levels are the direct defence methods for plants and are proved to be natural variation in these traits in *Arabidopsis thaliana* (Reymond et al., 2004). These direct methods of defence might explain the observed variation in leaf consumption of *Arabidopsis* by insect herbivores (Mewis et al., 2005). Genetic variation has also been reported for resistance to herbivores (Broekgaarden et al., 2008; Kusnierczyk et al., 2007; Steppuhn et al., 2008).

Herbivory was significantly correlated with growth rate (faster growing plants tended to be more susceptible) and it showed a significant negative correlation with flowering time (earlier flowering plants were more susceptible). Again this correlation can be explained if the same genes (linked genes) affect more than one trait. It is possible that either traits respond to the same gene, or that variation in growth rate or flowering time affect herbivory, or *vice versa*. For example, the transition to flowering might involve an increase in compounds that are attractive to herbivores or a decrease in protective compounds. Similarly, there could be a cost to producing protective compounds that affects growth, so plants that are more resistant grow more slowly.

The cost-of-resistance hypothesis holds for many cultivated plants where slow growing species have more constitutive resistance than fast-growing species, but not for wild plant species (Kempel et al., 2011). If resistance is provided by low cost compounds then the trade-off between growth and constitutive resistance does not exist (Valverde et al., 2003). Lower concentrations of compounds are accumulated in fast-growing species than slow growing species which tend to accumulate compounds that reduce the digestibility of the plants (Poorter, 1992). My results are

also in agreement with the study that early bolting plants are more susceptible to herbivores but that susceptibility does not necessarily reduce the plants fitness (Weinig et al., 2003).

Because relatively high broad-sense heritability estimates for variation in growth rate in different environmental conditions, anthocyanin production and herbivory suggest that these are mainly genetically determined traits. I focus my attention on dissecting the genetic basis of these traits by QTL mapping in the experiments that follow.

4.0 QTL analysis of growth variations, anthocyanin production and herbivory

4.1 Introduction

To identify the number and positions of loci that account for natural genetic variation in growth, anthocyanin production and herbivory in two local accessions of *Arabidopsis thaliana* are the main aims of this chapter. Little is known about the genetic basis of natural variation in plant growth, despite known variations in growth related traits in plants. QTL analysis of rate of leaf production in vegetative development indicated that the differences between L*er* and Fei-0 accessions are caused by 10 detectable loci. Comparative study of these hybrid also showed that most of the genomic regions involved in flowering time and total leaf number are also involved in the rate of leaf production (Mendez-Vigo et al., 2010). In (L*er* × Cvi) population of RILs, eight QTLs that control height were detected on chromosomes 2, 3 and 5 and 19 QTLs for leaf characters, five of them on chromosome 5, were also found (Bandaranayake et al., 2004). Similarly, QTLs for plant height were also detected on chromosomes 2 and 5 in a L*er* x Col population (Ungerer et al., 2002). QTLs for dry weight and rosette area were also found around the *ERECTA* locus in the populations having L*er* as one of the parents (Alonso-Blanco et al., 1999; Loudet et al., 2003a; Loudet et al., 2003b; Weinig et al., 2002). This can be explained by the fact that L*er* carries an *erecta* loss-of-function mutation that affects many aspects of plant development. QTL mapping experiments have also been carried out for metabolic and biochemical components of plants growth. A major QTL was detected for sugar contents in mature seeds of Ler × Cvi RIL

population (Bentsink et al., 2000). QTLs were also mapped for enzymes activities in primary and secondary metabolism in leaves of the L*er* × Col RIL population (Mitchell-Olds and Pedersen, 1998). Loci have identified in many populations for leaf and seeds phosphate contents (Bentsink et al., 2003; Loudet et al., 2003a) and for growth rate parameters in environments having different nitrogen sources (Loudet et al., 2003b; Rauh et al., 2002).

Flavonoids and phenylpropanoid metabolites, found in the plant kingdom including *Arabidopsis thaliana,* have an important role in the life of plants by providing blue, red and purple pigments. These protect plants against UV radiation, attract pollinators and other beneficial organisms, or also have a role in plant–microbe interactions (Buer et al., 2010). Plants also have specific responses to environmental stresses which helped them to survive in adverse environments (Chinnusamy et al., 2004), for example when plants are exposed to cold, it triggers well regulated cold-acclimation processes. This involves increased expression of a set of cold-regulation genes that are likely to mitigate cold stress, but their main functions are largely unknown (Christie et al., 1994). In contrast the anthocyanin biosynthetic pathway, and its regulation is well described in plants including *Arabidopsis thaliana* (Pelletier et al., 1997) and anthocyanin biosynthesis is increased in response to several stresses. However, little is known about natural variation of anthocyanin production in *Arabidopsis*, or in responses to cold in natural environments.

In plants, defence against herbivores is achieved through tolerance and resistance (Strauss and Agrawal, 1999). The ability to tolerate damage mainly depends on meristems and available resources to growth and reproduction (Strauss and Agrawal, 1999), while resistance can involve a range of deterrent, both chemical and physical.

The relationship between growth and resistance or tolerance to herbivores is complex. For example growth enables plants to get more resources which they can invest in resistance (Boege and Marquis, 2005; Val and Dirzo, 2003). On the other hand, defence mechanisms may require an investment that could otherwise be used for growth. Resistance genes increase plant fitness when a pathogen is present but can be a disadvantage when the pathogen is not present (Todesco et al., 2010). Studying the genetic basis of variation in herbivory in local accession will provide more information about the genomic regions involved in this trait.

4.2 Genotyping the F3 populations with Simple Sequence Length Polymorphisms (SSLPs)

A total of 137 microsatellite markers were tested for polymorphism between the parents, 11C1 and 4D1, of the hybrids in my experiment. Thirty-nine were found to be polymorphic between the parental lines. The map position of each SSLP marker in cM was estimated from its position in the *Arabidopsis* genome sequence relative to markers for which recombination frequencies had been determined in other mapping populations (Figure 14).

4.2.1 Genotype scoring on agarose gels and using an ABI 3730 sequencer

SSLPs were amplified by PCR and the products were analysed either in agarose gels (Figure 15a) or with an ABI 3730 sequencer (Figure 15). All the loci had co-dominant alleles, so if an F3 hybrid produced a single band, it was homozygous for the allele from either 4D1 or 11C1 and was scored as "a" or "b" respectively, and if there were two bands, it was heterozygous and was scored as "h" (Figure 15a). For products separated on the sequencer, a third primer was used, which consisted of the

120

M13 sequencing primer labelled with one of four fluorescent dyes at its 5' end (6-FAM, VIC, NED or PET). PCR therefore amplified a fluorescently labelled product and the use of four different dyes allowed the products of different PCR reactions to be analysed together. Genotype screening results were in four colours and were scored as "a" or "b" if they were homozygous for the 4D1 or 11C1 alleles and were scored "h" if they had both bands (Figure 15b).

4.3 QTL mapping of growth variations, anthocyanin production and herbivory of F3 population

To detect the genetic regions that are involved in growth rate variation of plants grown in the greenhouse or natural conditions in autumn or spring, Composite Interval Mapping (CIM) QTL analysis (Jansen, 1996a; Zeng, 1994) was performed for the F3 population. QTL mapping was also carried out to dissect the genetic basis of anthocyanin production in winter in natural conditions and herbivory in autumn in natural conditions.

Because plant growth rate is very sensitive to environmental variation, which might reduce the genetically-determined component of variation, tests were carried out to determine the best age at which to estimate growth rate for QTL analysis. Rosette areas of randomly selected lines were quantified at different time points and the broad-sense heritability calculated. As an example heritability values estimated from five lines grown in the growth room was lower in week 5 (60%) compared to week 6 (66%). Therefore phenotypic data for growth rate (rosette area) was estimated for all line at week 6 and included in QTL analysis (Appendix E). Growth data for plants grown in the greenhouse and natural conditions in autumn and spring were also used.

However, anthocyanin production and herbivory (percentage plant area eaten) were quantified only for the plants grown outside from an autumn sowing and were measured in winter (January), because only this set of plants showed a significant colour change or herbivore damage. Five plants of each F3 family were grown in each condition and the mean value of each phenotypic trait for each family was used in QTL analysis. A bulk of each F3 family was genotyped, so the experiments were equivalent to mapping in an F2, except that F3 lines have an advantage of a better estimate of the genetically-determined phenotype value. With these data sets, QTL analysis was performed for each trait in each condition. Significant QTLs for growth rate were detected for plants grown in natural conditions or the greenhouse in autumn (Figure 16). The significant QTLs were at the bottom of chromosomes 1 and 5 for plants grown in natural conditions and at the top of chromosome 1 for plants grown in the greenhouse (Figure 16). No significant QTLs for growth variations were detected for plants grown in natural conditions (Figure 15) or in the greenhouse in spring (Figure 17). Significant QTL for anthocyanin production was detected at the top of chromosome 2 (Figure 16) but no significant QTL was found for herbivory (Figure 18).
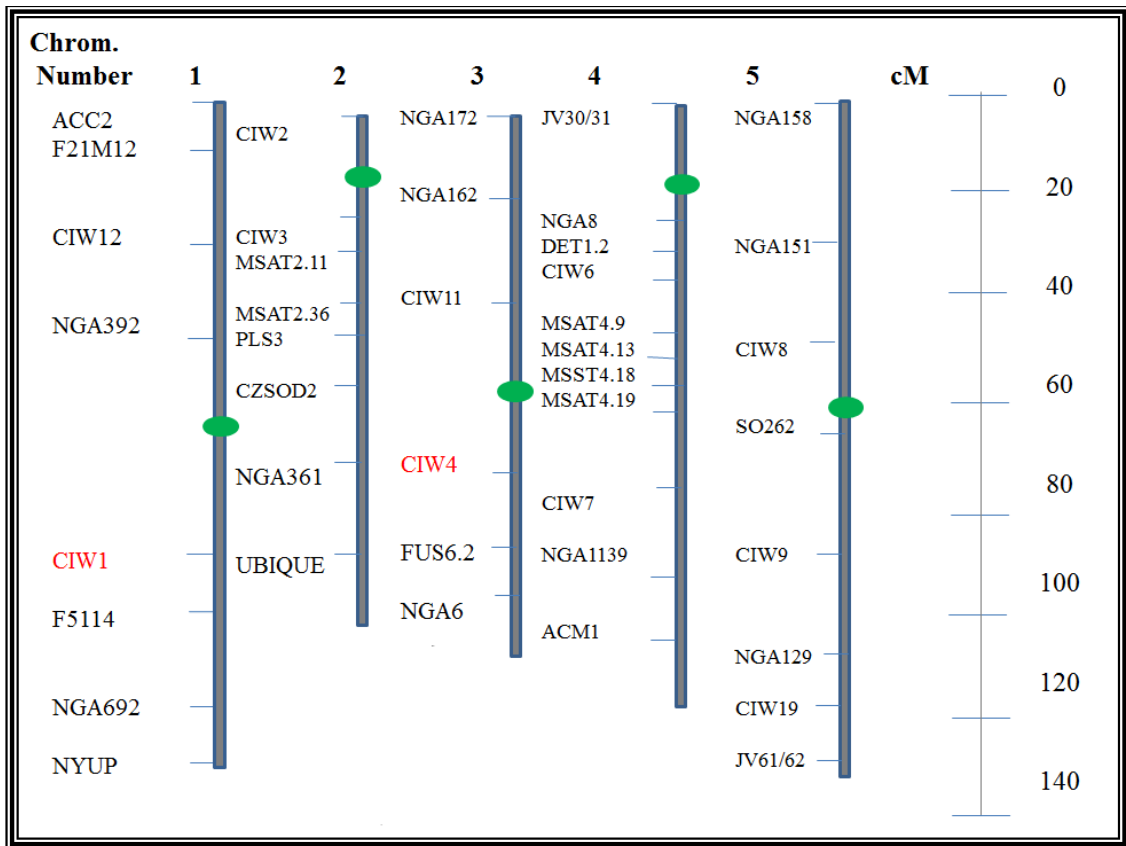
Figure 14: Genetic linkage map of *Arabidopsis thaliana*.

Positions of microsatellite markers used in genotyping the F3 and map distances are shown in cM.

Potential loci linked to growth rate and flowering time are shown in red colour.
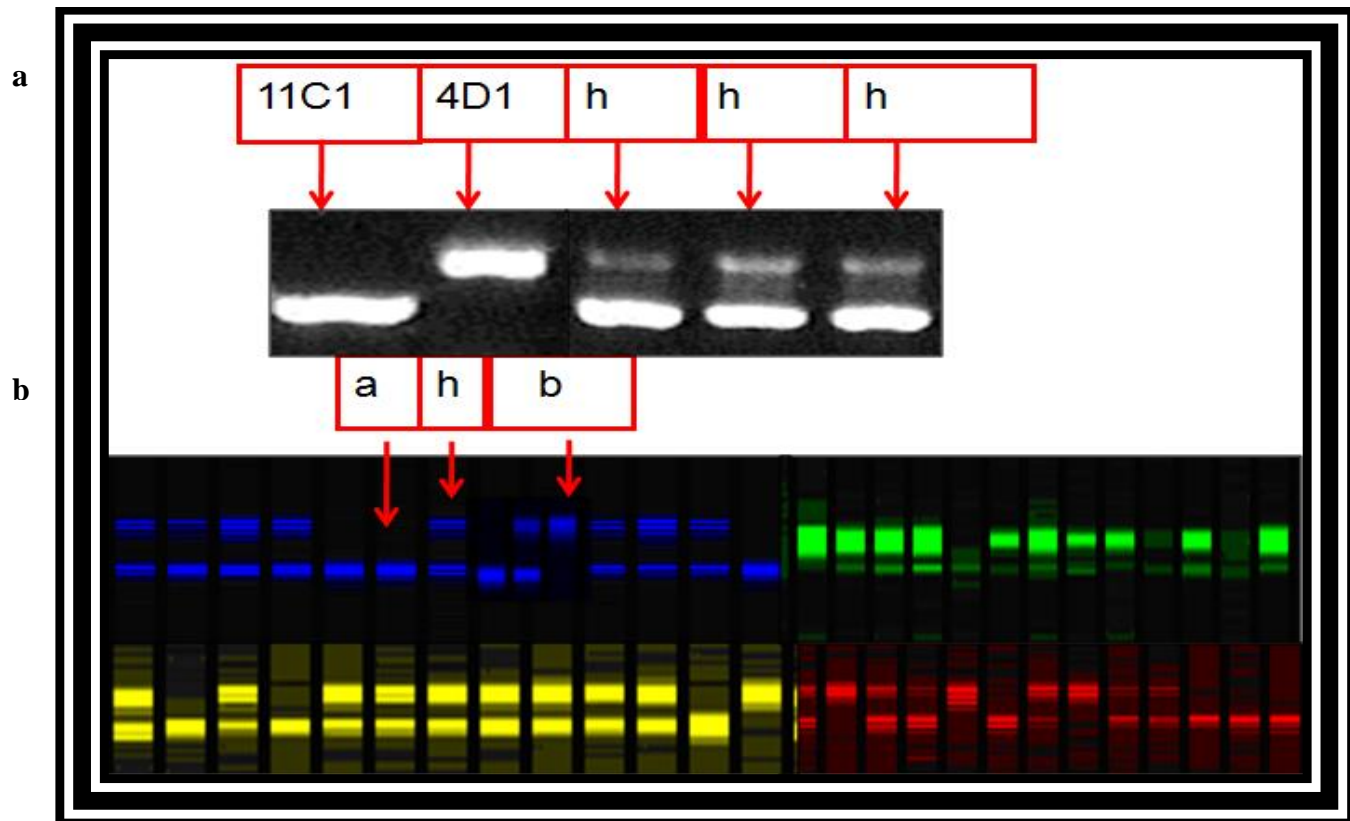
Figure 15: Examples of SSLP genotyping.

a) Fragments separated in an agarose gel. 11C1 and 4D1 show the mapping parents. The double bands of three F3 families had both parental alleles and were scored as heterozygotes ("h"). Single bands were scored as either "a" or "b" as they were homozygous for either the 4D1 or 11C1 allele, respectively.

b) Shows a virtual gel, reconstructed from trace files produced by the ABI3730 sequencer. Each colour shows a different SSLP locus. Three representative genotypes are shown for the blue locus.

Figure 16: Composite Interval Mapping of QTL for growth variation in the F3 population in autumn

(a) Plants grown outside in natural conditions. The x-axis shows the distance along the five *Arabidopsis* chromosomes, as in Figure 14, with the genotyped marker loci. The y-axis shows the log-of-odds (LOD) score for the probability of a QTL at each chromosome position. The horizontal line at 2.5 LOD is an arbitrary threshold for significance of a QTL as a conventional value used. Significant QTLs are detected at the bottom of chromosomes 1 and 5. In plants grown in the greenhouse (b) a significant QTL was detected at the top of chromosome 1.

Figure 17: QTL mapping for plant growth variation in the F3 population in spring.

No QTL were detected at a significant level of 2.5 LOD under either natural conditions (a) or in the greenhouse (b).
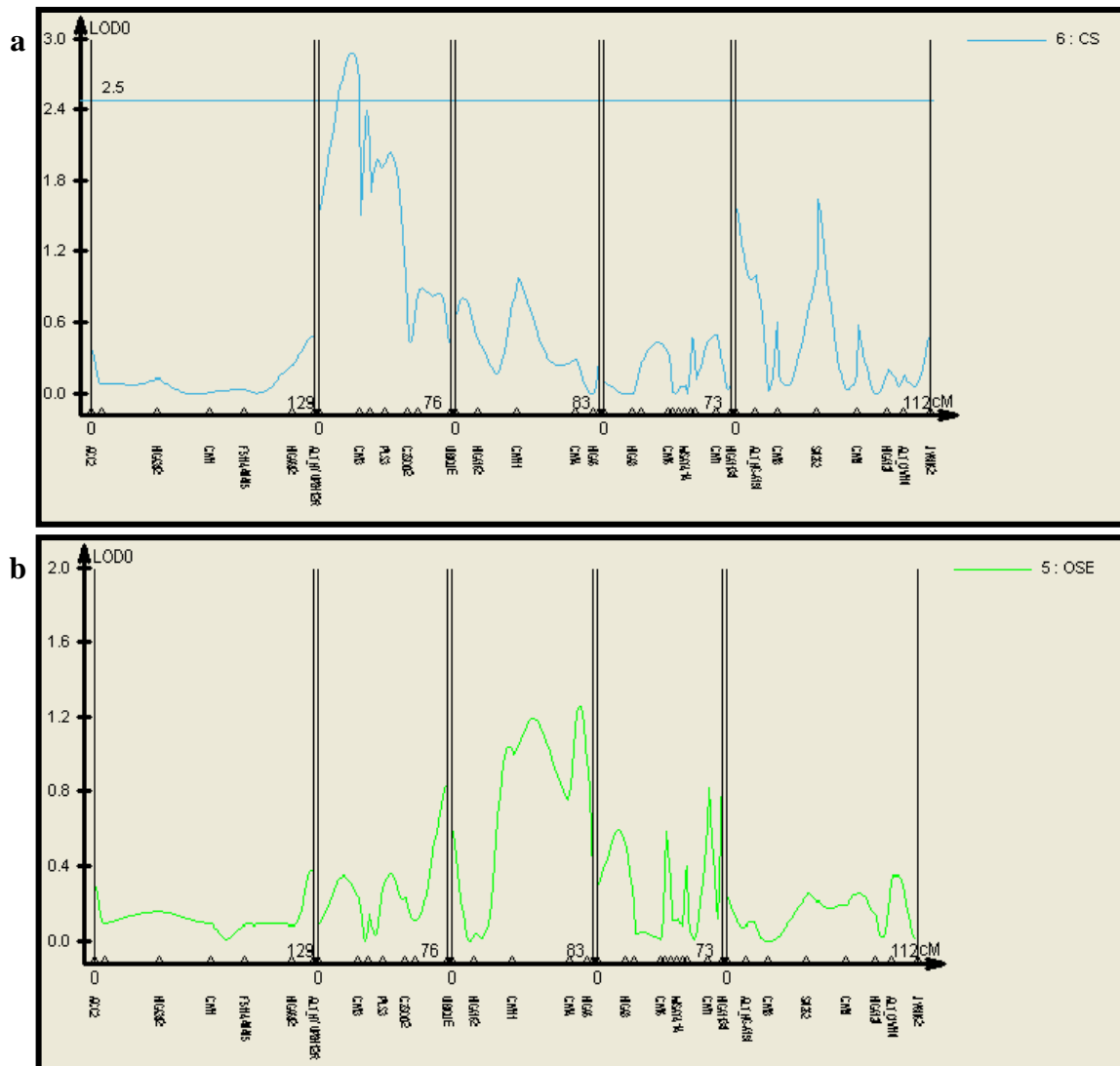
Figure 18: QTL mapping for anthocyanin production and herbivory in F3 plants in winter.

A significant QTL was detected at the top of chromosome 2 for anthocyanin production (a) but no significant QTL was found for herbivory (b).

Phenotypic traits in QTL analysis included plants growth rate variation in the green house and natural conditions in both seasons (autumn and spring) and herbivory and anthocyanin production in autumn. Significant QTLs were detected for growth rate for the autumn plants grown in both conditions but not for spring plants. A QTL for anthocyanin production was also detected at the top chromosome 2. However, relatively few QTLs were detected in the analysis. This could be because heritability was low (which is likely, given that many of the families will still segregate), population size was relatively small, or because there are many genes involved and each has a small effect. Though significant correlations were found in phenotypic traits between different sets of plants in the F3, suggesting that the same genes might underlie variation in different traits, none of the significant QTL for different traits map to the same position.

This experiment was therefore extended to the F5 population where comparatively fewer families will segregate and heritability estimates are higher..

4.4 Bulk-segregant analysis for F5 population

Bulk-segregant analysis was used to test for linkage of markers to growth rate QTLs in the F5 population. Pools of the 9 fastest and 8 slowest growing plants in the growth room or in natural conditions were genotyped with all 39 SSLP markers that had been found polymorphic in the F3 population. Only plants from different F3 parents were used in the same pool, to prevent over-representation of alleles due to shared grandparents. If a marker is linked to a growth rate QTL, its alleles are expected to be biased in their abundance between the fast and slow pools. Some markers showed a bias between fast and slow pools in both conditions (e.g., *CIW4*;

Figure 19), as expected of markers linked to QTL that controlled growth in both. Some were biased in one condition but not in other, like *ACM1* (Figure 19b) whereas most did not show an obvious bias in either condition, like *JV61/62* (Figure 19a). I focused on markers which were linked to potential growth rate QTL detected in the F3 population and were also biased in F5 bulk-segregant analysis. These markers were *F21M12*, *CIW1* and *F5114-49495* on chromosome 1, *CIW11* and *CIW4* on chromosome 3 and *CIW7* on chromosome 4.

Loci name	Chrom. number	GR(F)	GR(S)	OS(F)	OS(S)
CIW12	1	×	×	×	×
NGA392	1	×	×	×	×
CIW1	1	**/**(11)	**/**(4)	**/**(4)	×
CIW2	2	×	×	**/**(4)	×
MSAT2.11	2	×	×	×	×
UBIQUE	2	×	×	×	×
NGA172	3	×	×	×	×
NGA162	3	×	×	×	×
CIW11	3	**/**(4)	×	×	×
CIW4	3	**/**(11)	**/**(4)	**/**(11)	**/**(4)
CIW7	4	×	**/**(11)	×	**/**(11)
MSAT4.9	4	**/**(11)	×	**/**(4)	×
MSAT4.12	4	**/**(4)	**/**(4)	**/**(11)	×
MSAT4.18	4	×	×	×	×
MSAT4.19	4	×	×	×	×
NGA1139	4	×	**/**(4)	×	**/**(11)
NGA158	5	×	×	×	×
NGA151	5	×	×	×	×
NGA129	5	**/**(4)	×	×	×

Table 28: shows a list of SSLPs markers which are biased and unbiased in one or the other conditions.

Bulk-segregant analysis showed that some markers are unbiased (shown in crosses) and some are biased in one or the other condition (shown in red ticks) suggesting linkage with the growth increasing or decreasing allele in outside (OS) and growth room (GR) conditions in spring (S) and autumn (A) (shown red in brackets as 4 or 11).

4.5 Further testing of marker linkage to growth rate QTL

The markers that showed biased segregation in the F5, suggesting linkage to growth rate QTL, were tested for possible associations with growth rate variation in the F3 populations (Table 29, 30). Linkage of these markers to growth rate was also studied for the F5 population grown in the growth room and natural conditions both in autumn and spring, using genotypes from their F3 parents. No significant association between genotype and growth rate was found for any of the markers in the F3 population under any of the conditions (Tables 29, 30). This is consistent with the lower heritability estimates for growth rate in this generation and the inability to detect significant QTLs for growth rate with genome-wide scans. In the F5 generation, *CIW1* (Chromosome 1) was significantly associated to growth rate in natural conditions in spring (Table 31) and in growth room in autumn with 11 as growth increasing allele (Table 32). However, *CIW4* (Chromosome 3) was significantly associated with the growth of F5 families in natural conditions both in autumn and spring (Table 31) and also in spring in the growth room (Table 32), with the 4 allele increasing growth in all cases. These results were unexpected because conditions in the growth room were expected to be similar regardless of season.
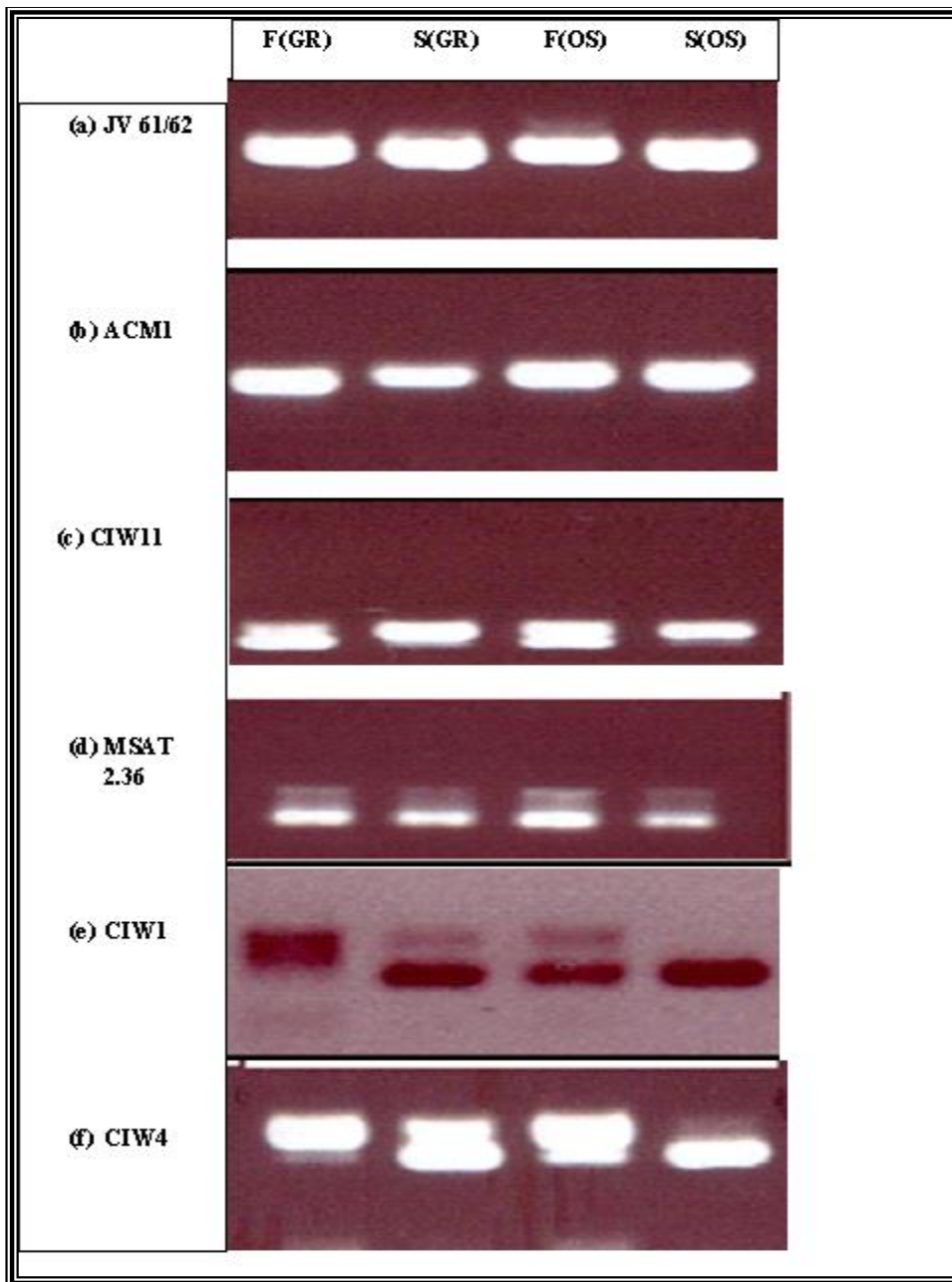
Figure 19: Bulk-segregant analysis suggesting linkage of markers to growth variation QTL segregating in the F5 generation.

For each marker, the first two lanes represent pools of fast and slow growing plants from the growth room, respectively, and the next two lanes fast and slow pools from natural conditions. This figure has examples of markers showing biased distribution of allele between fast and slow pools in one or other condition in F5 generation. For example *CIW11 (c) MSAT 2.36 (d), CIW4* (f) and *CIW1* (e) are biased under both conditions whereas JV 61/62 only in F (OS), ACM1 is unbiased.

132

Marker	Chromosome number	*p* value (natural conditions autumn)	*p* value- (greenhouse autumn)	Comparison of genotypes (natural conditions autumn)			Comparison of genotypes (greenhouse autumn)		
				4/4	4/11	11/11	4	4/11	11
F21M12	1	0.20	0.41	804	877	923	1045	1057	1275
CIW1	1	0.19	0.64	843	917	833	1185	1231	1164
F5114-49495	1	0.44	0.75	919	864	869	1144	1144	1173
CIW11	3	0.45	0.92	911	864	842	1167	1176	1194
CIW4	3	0.11	0.87	920	883	819	1182	1168	1167
CIW7	4	0.95	0.09	875	882	868	1270	1260	1065

Table 29: Markers tested for association to growth rate variation in F3 plants in autumn

Plants were genotyped with the markers and tested for association with growth rate by comparing the mean areas of the two homozygous classes to obtain and estimate of the probability that the genotypes and traits are uncorrelated, *p*, using a Student's t-test. Mean values for rosette area in mm^2 are given for the three genotypes at each locus, measured on the same day for each set of plants.

Marker	Chromosome number	p value (natural conditions spring)	p value (greenhouse spring)	Comparison of genotypes (natural conditions spring)			Comparison of genotypes (greenhouse spring)		
				4/4	4/11	11/11	4/4	4/11	11/11
F21M12	1	0.12	0.95	1256	704	1369	675	696	669
CIW1	1	0.46	0.96	1311	1352	1228	695	667	668
F5114-49495	1	0.13	0.76	1347	1371	1163	670	668	733
CIW11	3	0.23	0.16	1194	1368	1314	579	744	621
CIW4	3	0.83	0.48	1304	1271	1231	721	628	745
CIW7	4	0.85	0.33	1275	1336	1311	584	708	715

Table 30: Markers tested for association to growth rate variation in F3 plants in spring

Plants were genotyped with the markers and tested for association with growth rate by comparing the mean areas of the two homozygous classes to obtain and estimate of the probability that the genotypes and traits are uncorrelated, *p*. Mean values for rosette area in mm^2 are given for the three genotypes at each locus, measured on the same day for each set of plants.

Marker	Chromosome number	p value (natural conditions spring)	p value (natural conditions autumn)	Comparison of genotypes (natural conditions spring)			Comparison of genotypes (natural conditions autumn)		
				4/4	4/11	11/11	4/4	4/11	11/11
F21M12	1	0.85	0.47	805	718	715	240	228	214
CIW1	1	0.05	0.09	839	720	993	260	226	214
F5114-49495	1	0.99	0.08	792	771	776	243	290	213
CIW11	3	0.19	0.71	878	636	692	233	217	217
CIW4	3	0.001	0.035	985	712	592	250	188	205
CIW7	4	0.049	0.53	678	774	976	217	248	229

Table 31: Markers tested for association to growth rate variation in F5 plants under natural conditions in autumn and spring

Plants were genotyped with the markers and tested for association with growth rate by comparing the mean areas of the two homozygous classes to obtain and estimate of the probability that the genotypes and traits are uncorrelated, p. Mean values for rosette area in mm^2 are given for the three genotypes at each locus, measured on the same day for each set of plants.

Marker	Chromosome number	*p* value (growth room spring)	*p* value (growth room autumn)	Comparison of genotypes (growth room spring)			Comparison of genotypes (growth room autumn)		
				4/4	4/11	11/11	4/4	4/11	11/11
F21M12	1	0.83	0.33	3553	3512	3738	314	388	316
CIW1	1	0.95	0.02	3700	3524	3941	286	394	289
F5114-49495	1	0.54	0.13	3953	3612	3293	278	439	341
CIW11	3	0.35	0.58	3883	3495	3369	331	267	343
CIW4	3	0.005	0.53	4181	2868	3079	355	314	310
CIW7	4	0.06	0.11	3338	3599	4224	308	294	392

Table 32: Markers tested for association to growth rate variation in F5 plants in the growth room in autumn and spring

Plants were genotyped with the markers and tested for association with growth rate by comparing the mean areas of the two homozygous classes to obtain and estimate of the probability that the genotypes and traits are uncorrelated, *p*. Mean values for rosette area in mm^2 are given for the three genotype classes, measured on the same day for each set of plants.

4.6 Testing QTL in F4 families

Analysis of F3 and F5 populations suggested that QTLs for growth variation that were expressed under different conditions were present in Chromosome 1 and Chromosome 3, linked to *CIW1* and *CIW4*, respectively. To validate these QTLs, I randomly selected three F3 families – one that was heterozygous at *CIW1* and *CIW4* and two that were heterozygous at either *CIW1* or *CIW4* (Table 33). Twenty-four plants of each line were grown in the greenhouse and were genotyped at the segregating loci. Rosette areas of the homozygous genotypes were compared with a *t*-test. However, only family 18 showed an association, where plants homozygous for the *CIW4* alleles from 4D1 grew significantly faster ($p=0.1$) (Table 33). Failure to detect significant associations between growth rate and both *CIW1* and *CIW4* genotypes in the other F3 families could be because a small number of plants were used. Another explanation is that recombination had occurred between the marker and the linked QTL, so that the QTL alleles were no longer segregating in some of the families. Alternatively, the effects of the QTLs could be dependent on other genes that were not present in all the families (epistasis).

4.7 Fine mapping of growth rate QTLs in F5 families

To further test for linkage between marker genes and growth rate, larger families of 72 plants were grown in the growth room from F4 parents that were heterozygous at *CIW1* (three families) or *CIW4* (two families). F5 plants were genotyped at either *CIW1* or *CIW4* and their rosette areas estimated. To examine whether differences in rosette area reflected differences in leaf size, leaf number or both, the number of leaves >3 mm in length were also counted at the same time. Association of markers

137

to trait were examined with a *t*-test to compare homozygous marker genotypes and obtain a *p*-value (Table 34). For leaf number the only significant association was detected in family 54 in which the *CIW4* allele from 4D1 significantly increased leaf number ($p=0.05$) (Table 34). However, no significant effect on rosette area was detected, suggesting that plants that produce more leaves produce smaller leaves. No other significant associations with *CIW4* genotype were detected. Rosette area was significantly associated with *CIW1* genotype in family 671 ($p=0.04$). However in this case the allele from 11C1 is linked to increased plant growth, consistent with our previous results where the 11C1 allele of *CIW1* had been associated with faster growth in growth room conditions in autumn.

CIW1 and *CIW4* did not show any association to growth rate variation of F3 plants grown in natural or green house conditions in autumn and spring but were significantly associated with growth rate in F4 and F5 populations and in some of the larger F5 families grown in control conditions of the green house or growth rooms. However, these markers were not consistently associated to growth rate in all conditions. For example, *CIW1* was significantly associated in natural conditions in spring but not in autumn whereas *CIW4* was significantly associated to growth rate in both seasons in natural conditions. In the growth room, *CIW1* was significantly associated to growth rate in autumn but not in spring whereas *CIW4* showed significant association in spring but not in autumn.

However, when linkage to growth rate QTL was detected, the QTL were consistent in the directions of their effects. For the QTL linked to *CIW1* the allele from 11C1 increased growth, while for the QTL linked to *CIW4* the allele from 11C1 slowed growth.

F3 family number	Genotype at *CIW1*	Genotype at *CIW4*	Association of *CIW1* with growth rate in F4 progeny				Association of *CIW4* with growth rate in F4 progeny			
			4/4	4/11	11/11	*p*	4/4	4/11	11/11	*p*
6	4/11	11/11	748	734	816	0.675	-		-	-
13	4/11	4/11	185	213	227	0.604	175	206	220	0.585
18	4/4	4/11	-		-	-	614	602	475	0.1

Table 33: Testing association of *CIW1* and *CIW4* genotypes to growth rate variations in F4 populations

The mean areas of the homozygous genotypes (in mm^2) are shown for F4 families segregating at either *CIW1*, *CIW4* or both loci and compared with a *t*-test to estimate the probability (p) that they are not the same. Only family 18, segregating for *CIW4*, showed an association to growth rate, with the 4D1 allele increasing growth.

Family number	Marker	Trait	_p_-value	Mean values for genotypes		
				4/4	**4/11**	**11/11**
54	_CIW4_	Area in mm^2	0.54	1037	1105	1083
		No. of leaves	0.05	11.78	12.4	11.76
261	_CIW4_	Area in mm^2	0.41	909	1026	1012
		No. of leaves	0.49	13.7	14.2	14.3
521	_CIW1_	Area in mm^2	0.41	277	250	257
		No. of leaves	0.53	8.13	7.9	8.23
594	_CIW1_	Area in mm^2	0.80	276	271	265
		No. of leaves	0.29	8.6	8.6	9.0
671	_CIW1_	Area in mm^2	0.04	246	298	289
		No. of leaves	0.86	9.3	9.0	9.1

Table 34: Testing association of _CIW1_ and _CIW4_ genotypes to growth rate and leaf production rate in F5 families

Members of F5 families segregating at either _CIW1_ or _CIW4_ were genotyped and the mean number of leaves or rosette areas compared for the two homozygous genotypes.

4.8 Discussion

The work described in this chapter aimed to map QTL for variation in growth rate between two local accessions of *Arabidopsis* − an important life-history trait in plants. It also examined the genetic basis for resistance to herbivory and anthocyanin production in response to cold.

Significant QTLs for variation in growth rate were mapped at the bottom and top of chromosomes 1 and at the top of chromosome 5. However, they were not detected under all conditions, suggesting that they might be involved in the responses of growth to different environments. Other QTL involved in growth have been detected at similar locations in other mapping populations. The characteristic of QTL at the top of chromosome 1 that affects growth rate in greenhouse in autumn might be the same as DM10.1 described by (Loudet et al., 2003b) in the Bay × Sha population in which the Sha allele has a negative effect. Probably, the most interesting QTL is found at the bottom of chromosome 5. This is the region where QTLs for flowering time, growth rate, and leaf initiation speed, chlorophyll fluorescence and specific leaf area were reported (Loudet et al., 2003b).

Bulk-segregant analysis using F5 families was done with 19 polymorphic markers and pools of fast and slow growing plants from either growth room or natural conditions in spring. This suggested the regions containing *F21M12*, *CIW1* and *F5114-49495* on chromosome 1, *CIW11* and *CIW4* on chromosome 3 and *CIW7* on chromosome 4 as candidates for growth rate QTL.

However, these studies highlighted a number of problems with mapping potential growth rate QTL. Firstly, the effects of putative QTL could not always be detected,

even under conditions that were assumed to be consistent in the growth room. This could be because there were differences in the environment or because the effects of QTL were dependent on epistatic interactions with other loci. Secondly, the effects of the putative QTL were small, relative to the non-genetic variance in growth rate, so it was not possible to assume the QTL genotype of a plant from its growth rate with any certainty. Both these factors would make fine-mapping growth rate QTL very difficult. For example, it would be necessary to reduce variability between experiments and measure a number of progeny to estimate a genotype's mean growth rate. For these reasons, fine-mapping of growth rate QTLs was not pursued.

5.0 Flowering time variations, QTL analysis and fine mapping.

5.1 Introduction

The aim of this chapter is to assess the genetic basis for flowering time variation in the mapping population and how it responds to different environmental conditions when grown outside.

The timing of the transition from the vegetative to reproductive phase of a plant (called flowering time) is crucial for reproductive success in any environment. It has also been demonstrated in *Arabidopsis* and other species that flowering time varies in response to environment and within populations (Korves et al., 2007). Different genotypes also flowered at different times in the field and responded differently to vernalization (the effects of low temperature on flowering), so could flower later than others in one season but before others in another season.

A number of genes have been found to underlie natural variation in flowering behaviour in *Arabidopsis*, including *FRIGIDA* (*FRI*). *fri* loss-of-function mutations flower early, even without vernalization, because they are unable to maintain high levels of the floral repressor *FLC* (Michaels and Amasino, 1999a; Sheldon et al., 1999b). Winter annuals (genotypes with a long life cycle) usually have functional alleles of both *FRI* and *FLC* whereas summer annuals (rapid cycling accessions) have mutations in *FRI* or *FLC* or both (Gazzani et al., 2003; Johanson et al., 2000; Michaels et al., 2003). This rapid-cycling behaviour has been suggested to be advantageous in disturbed habitats, such as cultivated fields (Keddy, 2007). Interestingly, the 11C1 parent of the mapping population had been found to carry a *fri* loss-of-function mutation (McCulloch and Hudson, 2011) but to flower late in the

absence of vernalization. This suggests that it carries other alleles that suppress the effects of the *fri* mutation. It had also been found to express *FLC* at relatively high levels, suggesting that the suppressor mutations might lead to increased expression of *FLC* independently of *FRI*. Therefore an additional aim of this work was to identify a potential suppressor of *fri* that could account for the late flowering of the 11C1 mapping parent.

5.2 Mapping flowering time genes

Flowering time data for F3 families in the growth room and in natural conditions in autumn were collected by Hayley McCulloch. Broad-sense heritability estimates were high - 80% in the growth room and 70% in natural conditions. This flowering time data was used in QTLs analysis (McCulloch and Hudson, 2011) and detail is given in appendix D .

F5 plants were also grown in a growth room (control conditions) and outside (natural conditions) in autumn and flowering time measured as the number of rosette leaves produced. In each case 89 F5 families were grown with five members of each line. Plants were randomised in different trays for observations. Again heritability estimates were high for the F5 − 89% under the controlled conditions of the growth room and 71% outside. The lower value outside is likely to reflect environmental variation experienced by the plants. The detail of data is given in appendix D.

The families showed a nearly normal distribution of leaf numbers at flowering. However, both 11C1 and 4D1 flowered sooner than the majority of the F5 lines (Figure 20). This skew in flowering time is not easily explained. It suggests that interactions between genes from both parents delay flowering.

In the outside garden, 89 plant lines (OS4) of F5 population with 5 members of each line were grown in autumn. Again a normal distribution of flowering times was observed, although in this case, the 4D1 parent was among the earliest flowering and the 11C1 among the late flowering genotypes. This suggests that the parents carry mainly alleles that either increase or decrease flowering time under the autumn conditions.

5.3 Flowering time variation within an experiment

Some F5 lines are early flowering in both conditions, some late in both and some early in one but late in the other. For example, lines like 92 were late flowering in both growth room and natural conditions but 194 and 76 are early flowering in one condition but late flowering in the other. However, some lines having shared F4 parents also behave differently. Lines like 523 and 524 are both offspring of F3 parent 52. but 523 is early flowering whereas 524 is late flowering in natural conditions suggesting the involvement of different genes in different conditions in flowering time (Figure 21). These results suggest that the difference might be due to sensitivity to vernalization or to day length, because plants in the field, but not the growth room, experience vernalization over winter and seasonal fluctuations in day length.

Number of leaves to flower (N)
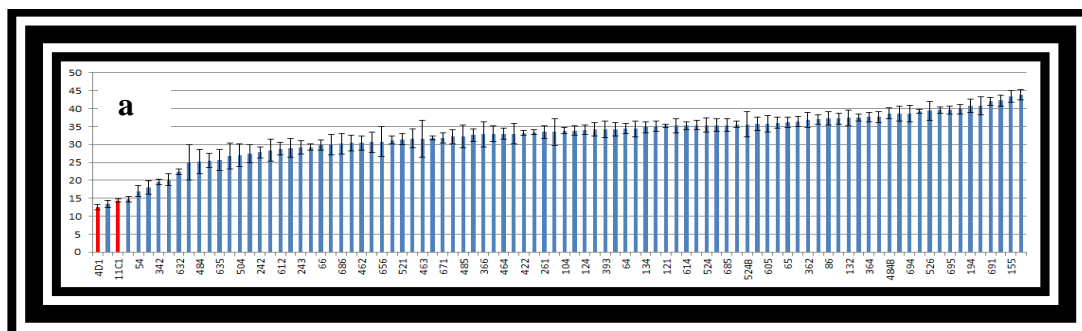
F5 lines grown in growth room (autumn)-GR2

Number of leaves to flower (N)
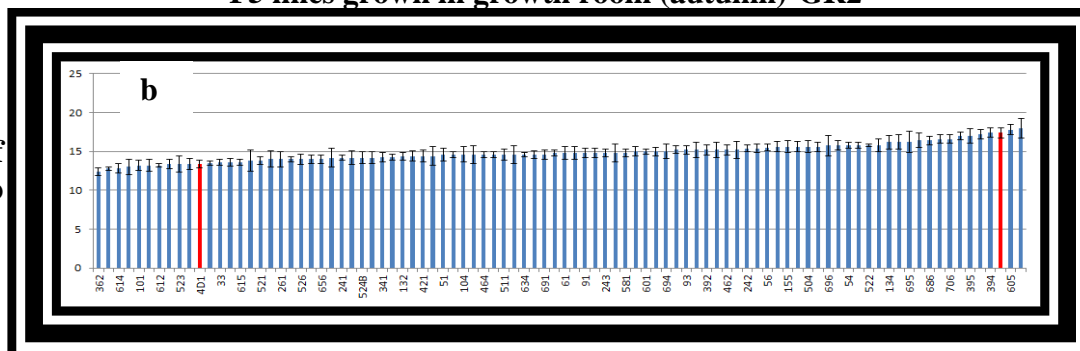
F5 lines grown in natural conditions (autumn)-OS4

Figure 20: Flowering time variation in plants of F5 populations grown in (a) a growth room (autumn) and (b) natural conditions (autumn).

These graphs showed the distribution of mean number of leaves to flower (± SEM) for 5 plants of each F5 line (blue bars) including parental lines (red bars).
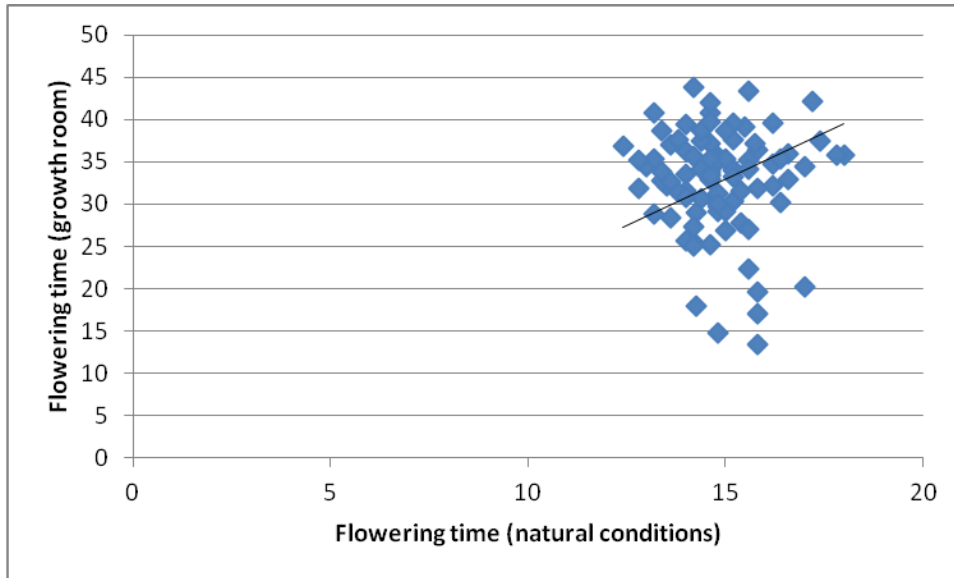
Figure 21: Relationship between flowering time of F5 families in natural and growth room conditions in autumn.

No significant correlation in flowering time is seen between the two conditions.

5.4 QTL mapping for flowering time

To dissect the genetic basis of flowering time variation, Composite Interval Mapping (CIM) analysis (Jansen, 1996a; Zeng, 1994) was performed for plants of F3 population, using the genotypes for 39 SSLP markers, as for growth rate. Significant QTLs were detected for plants in the growth room but not in natural conditions (Figure 22). The significant QTL were in the lower part of Chromosome 1 and at the bottom of Chromosome 5. A candidate QTL at the bottom of Chromosome 5 had already been detected by Hayley McCulloch when testing for association between genes that function in the vernalization pathway and sensitivity to vernalization in this hybrid family. This QTL was therefore not investigated further here. However, the potential QTL in Chromosome 1 had not been identified before; therefore this locus had been focused further. Although 11C1 carries a *fri* loss-of-function mutation no QTL was detected in the vicinity of *FRI* (at the top of Chromosome 4). This suggests that either the *FRI* allele from 4D1 is also not functional, although it is transcribed at wild-type levels to produce RNA encoding an apparently wild-type FRI protein (McCulloch and Hudson, 2011), or that the effects of the *fri* mutation from 11C1 are not detected because of the segregation of one or more suppressors.

5.5 Analysis for F5 plants with selected markers on Chromosomes 1

To further investigate the possibility of a flowering time QTL in chromosome 1, a few markers were selected closer to the potential QTL for single marker analysis in the F5 population (Table 35).
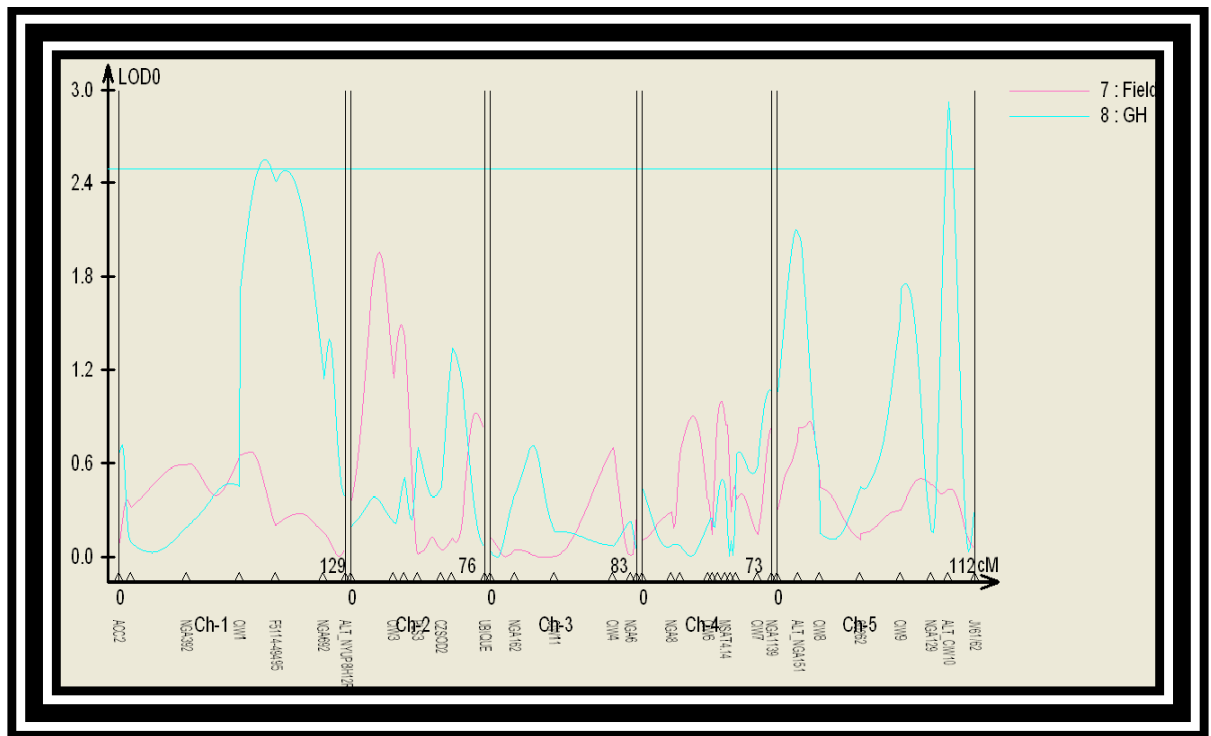
Figure 22: QTL mapping for flowering time in plants for F3 population in natural conditions and the growth room.

Failure to detect significant QTL under natural conditions might reflect lower heritability under field conditions. Alternatively the effects of QTL detected in the growth room might depend on environmental factors that varied between field and growth room. The x-axis shows the five *Arabidopsis* chromosomes and the y-axis the LOD score for the likelihood of a QTL. The trace for plants in the growth room is cyan and for plants in the field in pink.

Association of these markers to flowering time was tested by comparing the mean flowering time of the different homozygous genotypes in the F3 and F5 families with *t*-tests. All those markers that have significant associations to flowering time showed that the 11C1 allele carries an allele that increases time to flower (Table 35). Markers *CIW1*, *F5114-49495* and *NGA692* were also significantly associated to flowering time in F3 plants in the growth room, with the 11C1 allele delaying flowering. None of these markers were significantly linked to flowering time of F3 plants grown in natural conditions. *CIW1* was also significantly associated (*p*=0.005) with the flowering times of F5 plants in growth room conditions, with the 11 allele again delaying flowering. This supported the idea that there is a QTL for flowering time variation in Chromosome 1 and that *CIW1* is the most closely-linked of the markers that were tested.

5.6 The fine mapping of flowering time QTLs with F5 plants

Seventy-two F5 offspring from each of three F4 families that segregated at *CIW1* were grown in the growth room. The number of vegetative leaves produced before flowing and numbers of days to flower were recorded for these plants. Association of markers to traits were assessed with *t*-tests comparing the mean trait values for the homozygous genotypes (Table 36). Both the number of leaves and the number of days to flowering were significantly associated with *CIW1* genotype in all families except family 594, where the difference in the number of leaves was not significantly different (*p*=0.09) (Table 36). In all cases, the 11C1 allele delayed flowering. These results confirm the presence of a QTL for flowering time variation in this region. Because the allele from 11C1 delays flowering, the QTL is also a candidate for the suppressor of the *fri* mutation present in the 11C1 parent.

150

Potential candidate markers	F3 (GR)-autumn *p*- value	11/11	11/4	4/4	F3 (OS)-autumn *p*-value	11/11	11/4	4/4	F5 (GR) autumn *p*-value	11/11	11/4	4/4	F5 (OS)-autumn *p*-value	11/11	11/4	4/4
NGA392	0.09	27	37	30	0.35	14	11	16	--	--		--	--	--		--
CIW1	0.01	36	37	25	0.78	29.9	29	29.5	0.005	36	33	28	0.25	14.3	15	14.8
F5114-49495	0.01	36	36	27	0.70	29.5	28	29.8	0.13	34	37	30	0.76	14.7	14.3	14.5
NGA692	0.03	33	37	26	0.79	30	29	28	--	--		--	--	--		--

Table 35: Testing associations and allelic effects between Chromosome 1 genotypes and flowering time under natural and controlled conditions.

Mean values of leaves produced before flowering are shown for homozygotes and the results of a *t*-test (probability that the means are the same) are given as the *p* value.

Plant lines	Trait	11/11	11/4	4/4	p-value
521 (spring)	Number of leaves to flower	23	19	17	8.5×10^{-9}
	Number of days to flower	41	37	35	8.2×10^{-11}
594 (spring)	Number of leaves to flower	26	24	23	0.09
	Number of days to flower	40	39	37	0.04
671 (spring)	Number of leaves to flower	28	27	25	0.02
	Number of days to flower	48	46	43	0.0006

Table 36: Testing association of *CIW1* genotypes to flowering time in F5 families

Mean values for the three genotypes are given, together with the p value that to homozygous classes are the same.

5.7 Discussion

Flowering time variation, an important trait for plants fitness and reproductive success, has been studied and mapped in the local accession of *Arabidopsis thaliana* in this chapter. A wider variation in flowering time was observed in the plants of the F5 population grown in growth room conditions than in natural conditions over winter. The average number of leaves produced by the 10 earliest flowering lines in the growth room ranged from 13.4 to 25.5 and for the 10 latest from 39.2 to 43.8. This compares with 12.4 to 13.4 leaves for early flowering plants and 16.4 to 18 leaves for late flowering plants outside. Although plants grew for longer outside, they spent most of their time in winter with colder conditions, shorter days and lower light intensity than in the growth room. The rate of leaf production of these plants is therefore expected to be lower than in the growth room. The plants outside also experienced a prolonged period of cold (vernalization), which advances flowering in many genotypes. These two factors might explain why plants outside produced fewer leaves before flowering and showed less variation in flowering time. A similar phenomenon was seen when flowering times of ~75 local accessions were compared under natural and controlled conditions (Lim and Hudson, 2013).

QTL mapping in F3 hybrids identified two significant QTLs, one in the lower arm of Chromosomes 1 and at the bottom of Chromosome 5 (Figure 22). The Chromosome 1 QTL was confirmed in F5 populations and suggested to be closest to *CIW1* of the markers that were tested for association. Potential candidate genes for flowering time in this or close to this region are *FLOWERING LOCUS T* (*FT*) (Huang et al., 2011; Li et al., 2010; Salome et al., 2011; Strange et al., 2011), *FLOWERING LOCUS M* (*FLM*) (Caicedo et al., 2009; Rosloski et al., 2010; Schlappi, 2006; Werner et al., 2005b), *LSD1-LIKE 1* (*LDL1*) (Jiang et al., 2007), *VERNALIZATION INDEPENDENCE 5* (*VIP5*) (Oh et al., 2004; Theologis et al.,

153

2000; Zhang et al., 2003; Zhang and van Nocker, 2002). Testing the involvement of two of these genes – *LDL1* and *VIP5* – is the part of next chapter.

6.0 Sequence and expression variation analysis of candidate genes

6.1 Introduction

A QTL for flowering time, linked to marker *CIW1* had been identified at which the allele from 11C1 delayed flowering. Because 11C1 is late flowering, sensitive to vernalization and expressed *FLC*, though it carries a *fri* loss of function mutation, the QTL allele might allow *FLC* expression (McCulloch and Hudson, 2011). Two genes known to have an effect on *FLC* expression − *LDL1* and *VIP5* − map in the region of the QTL, making them candidate genes. The aim of this chapter is to test whether variation in either *LDL1* or *VIP5* could contribute to this phenotypic variation. *LDL1* and *VIP5* sequences were compared from the mapping parents to see whether they are associated with differences in gene expression and flowering time behaviour.

For plants, the timing of flowering initiation has been critical during evolution to maximize reproductive success (Amasino and Michaels, 2010). A key component in regulatory network of flowering time in *Arabidopsis* is FLOWERING LOCUS C (FLC), a MADS box transcription factor that blocks the floral transition (Michaels and Amasino, 1999b; Sheldon et al., 1999a). FRIGIDA (FRI) activates *FLC* expression such that in the absence of vernalization, flowering is delayed (Boss et al., 2004; Sung and Amasino, 2005). Histone modifications play a central role in *FLC* expression level regulation and flowering time control (He and Amasino, 2005a; Sung and Amasino, 2005). Transcriptionally active genes are associated with the dimethylation and trimethylation of histone H3 lysine 4 (H3K4) (Cao et al., 2008; Pien et al., 2008; Schmid et al., 2003). These modifications have been reported at *FLC* chromatin and also regulates flowering by *FLC* expression (He and Amasino, 2005a; Kim et al., 2009; Liu et al., 2010). *FLOWERING LOCUS D* (*FLD*), *LSD1-LIKE 1* (*LDL1*), and *LSD1-LIKE 2* (*LDL2*), which are homologs of human lysine-specific demethylase 1

(LSD1), an amine oxidase that demethylates H3K4me2 and H3K4me1 in humans, act as *FLC* repressors (Jiang et al., 2007; Liu et al., 2007; Schmid et al., 2003). FLD reduces the H3K4me2 level in the central region of *FLC* chromatin to decrease its expression (Liu et al., 2007).

Mutations in genes that encode homologs of the components of yeast RNA polymerase II-associated factor 1 (Paf1) complex have resulted elevated *FLC* expression, even in the *FRI-*containing winter annual accessions. *EARLY FLOWERING 7* (*ELF7*, also known as *VERNALIZATION INDEPENDENCE 2*, *VIP2*), *ELF8* (*VIP6*), *VIP4*, and *VIP5* are components of the Paf1-complex, which has been shown to be required for a high level of *FLC* expression (He et al., 2004; Kim et al., 2005; Oh et al., 2004). In *Arabidopsis thaliana*, the *VIP4*, *VIP5*, *VIP6* (also called *ELF8*), and *ELF7* genes encode obvious homologs of *Leo1, Rtf1, Ctr9, and Paf1*, respectively, whereas *VIP3* encodes a protein closely related to hSki8 (He et al., 2004; Oh et al., 2004; Zhang et al., 2003; Zhang and van Nocker, 2002). Strong mutations in each of these genes has been shown to cause weak growth, defects in leaf and floral development, and acceleration of the natural phase transition from vegetative growth to flowering (He et al., 2004; Oh et al., 2004; Zhang et al., 2003; Zhang and van Nocker, 2002).

6.2 Sequence variation in *VERNALIZATION INDEPENDENCE5* (*VIP5*) between 4D1 and 11C1 plants

To assess the presence of mutations in *VIP5*, primers (Appendix A) were used in polymerase chain reactions (PCR) to amplify the coding region of *VIP5* alleles for 4D1 (4) and 11C1 (11) accessions used as the mapping parents. Products were sequenced with three sets of forward and reverse primers (Appendix A). The sequences of *VIP5* alleles from the parents are represented graphically in figure 23. Overall, 21 single nucleotide polymorphisms (SNPs) were found in the 4 allele, relative to Col-0 including 19 non-synonymous substitutions of

VIP5 for 4. There were four SNPs in the *VIP5* 11 allele relative to Col-0 including three non-synonymous substitutions (Fig.24).

Because the 4 and 11 alleles of *VIP5* had the potential to encode similar proteins, I also examined the expression level of *VIP5* in the two parents, to test whether different levels of expression might suggest differences in activity. RNA was extracted from young plants of both parents and used to make cDNA. The *VIP5* coding region was amplified for 20, 25 and 30 cycles of PCR and compared to amplification of the glutathione-dependent formaldehyde dehydrogenase gene, At5g43940, which should be expressed at similar level in all cells. There was no obvious difference in *VIP5* transcript levels for 4 and 11 alleles after 20 but a higher expression level was suggested for 11C1 for 25 and 30 cycles of PCR (Figure 25).
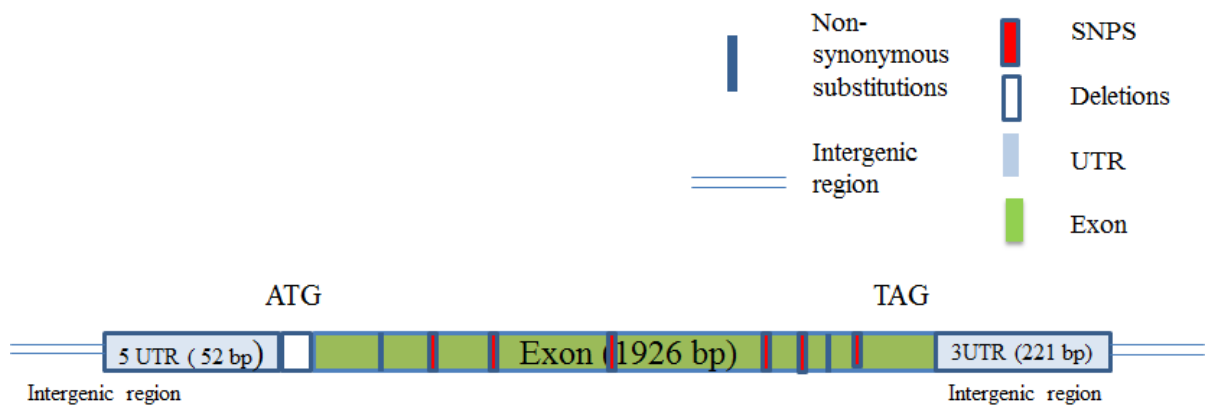
Figure 23: Graphical representation of VIP5 gene for 4D1 parent plant.

This figure shows different parts of genes in different colours.

```
VIP5-Col      AGATGCATCAAAAGGTGGTTCAGGTAGTCGAGATTTCTCATCAACGAAGAGGAAACCGTT       1

VIP5-11C1     AGATGCATCAAAAGGTGGTTCAGGTAGTCGAGATTTCTCGTCAAAGAGGAGGAAACCATT       1

VIP5-4D1      ------------------------------------------------------GCGGT       1

                                                                       a*bc*

VIP5-Col      AGCTTCCTCCAATTTGAGTAGTTCCAGCCAAAGTGACAGTGATAGTAGGTCTCAGAGTGA       2

VIP5-11C1     ACCTTCCTCCAATTTGAGTAGTTCCAGCCAAAGTGACAGTGATAGTAGGTCTCAGAGTGA       2

VIP5-4D1      GGCTTCCTCCAATTTGAGTAGTTCCAGCCAAAGTGACAGTGATAGTAGGTCTCAAAGTGA       2

              de***********************************************f*****

VIP5-Col      GTTTGAGGATGTTAAGGAAGTTACCATTAGACGGTCTAAGCTTGCCAAATGGCTAATGGA       4

VIP5-11C1     GTTTGAGGATGTTAAGGAAGTTACCATCAGACGGTCTAAGCTTGCCAAATGGCTAATGGA       4

VIP5-4D1      GTTTGAGGATGTTAAGGAAGTTACCATTAGACGGTCTAAGCTTGCCAAATGGCTAATGGA       4

              ***************************g********************************

VIP5-Col      GCCTTTCTTTGAAGAGCTTATAGTTGGGTGCTTTGTGAGGGTTGGGATCGGAAGGTCAAA       5

VIP5-11C1     GCCTTTCTTTGAAGAGCTTATAGTTGGGTGCTTTGTGAGGGTTGGGATCGGAAGGTCAAA       5

VIP5-4D1      GCCTTTCTTTGAAGAGCTTATAGTTGGGTGCTTTGTGAGAGTTGGGATCGGAAGGTCAAA       5

              ***************************************h********************

VIP5-Col      GAGTGGTCCAATTTACAGACTCTGCTGGGTGAAGAATGTTGATGCAACCGATCCTGACAA       6

VIP5-11C1     GAGTGGTCCAATTTACAGACTCTGCTGGGTGAAGAATGTTGATGCAACCGATCCTGACAA       6

VIP5-4D1      GAGTGGTCCAATTTACAGACTCTGCTGGGTGAAGAATGTAGATGCAACCGATCCTGACAA       6

              **************************************i*********************

VIP5-Col      GACCTACAAGCTAGAGAATAAAACTACACACAAGTACCTTAACGTCGTCTGGGGAAATGA       7

VIP5-11C1     GACCTACAAGCTAGAGAATAAAACTACACACAAGTACCTTAACGTCGTCTGGGGAAATGA       7

VIP5-4D1      GACTTACAAGCTAGAGAATAAGACTACACACAAGTATCTTAACGTCGTCTGGGGAAATGA       7

              ***j***************k***************l************************

VIP5-Col      AACCTCGGCGGCTCGATGGCAAATGGCTATGATCTCAGATGGTCATCCGCTGGAGGAAGA       8

VIP5-11C1     AACCTCGGCGGCTCGATGGCAAATGGCTATGATCTCAGATGGTCATCCGCTGGAGGAAGA       8

VIP5-4D1      AACCTCGGCGGCTCGATGGCAAATGGCTATGATCTCAGATGGTCATCCTCTGGAGGAAGA       8

              ************************************************m***********

VIP5-Col      CGAGAAAGATCGGCTTAGAAAAGAATTGGAAATTGCGCAGAGCAAAAACGATGAAGCAGG       12

VIP5-11C1     CGAGAAAGATCGGCTTAGAAAAGAATTGGAAATTGCGCAGAGCAAAAACGATGAAGCAGG       12

VIP5-4D1      CGAGAAAGATCGGCTTAGAAAAGAGTTGGAAATTGCGCAGAGCAAAAACGATGAAGCAGG       12

              ************************n***********************************
```

```
VIP5-Col      TGTAGAGAGGATCAAGTCGAAAATCAAACAGCTCGACGCTTCACGGAACAAGAAAGGGGT      13

VIP5-11C1     TGTAGAGAGGATCAAGTCGAAAATCAAACAGCTCGACGCTTCACGGAACAAGAAAGGGGT      13

VIP5-4D1      TGTAGAGAGGATCAAATCGAAAATCAAACAGCTCGACGCCTCACGGAACAAGAAAGGAGT      13

              ***************o*****************************p****************q**

VIP5-Col      AGATAAAAAAGCGCTTAAACTTGCTGAGATGAACAAGAAGAACAGAGCCGAGAATTTCAA      14

VIP5-11C1     AGATAAAAAAGCGCTTAAACTTGCTGAGATGAACAAGAAGAACAGAGCCGAGAATTTCAA      14

VIP5-4D1      CGATAAAAAAGCGCTCAAACTTGCCGAGATGAACAAGAAGAACAGAGCCGAGAATTTCAA      14

              ***************r********s***********************************

VIP5-Col      TCCGTTTTCAAGAAGATGGACCCGATCATCAAACTACTACAACGGGAAAAACAAGGGGAA      16

VIP5-11C1     TCCGTTTTCAAGAAGATGGACCCGATCATCAAACTACTACAACGGGAAAAACAAGGGGAA      16

VIP5-4D1      TCCGTTTTCAAGAAGATGGACCCGATCATCGAACTACTACAACGGGAAAAACAAGGGGAA      16

              ******************************t*****************************

VIP5-Col      AGATGGAGAAGAGAACGAGGCAGCGGTTGCAGCAGCGGTTGAGACCAATGGAGCAGATGC      17

VIP5-11C1     AGATGGAGAAGAGAACAAGGCAGCGGTTGCAGCAGCGGTTGAGACCAATGGAGCAGATGC      17

VIP5-4D1      AGACGGAGAAGAGAACGAGGCAGCGGTTGCAGCAGCGGTTGAGACCAATGGAGCAGATGC      17

              ***u***********v********************************************

VIP5-Col      AATAGACACAAGAGCTCCAATAGGTCAAGGAGCAGAACACAATCAGCTTCATAACTTTGA      19

VIP5-11C1     AATAGACACAAGAGCTCCAATAGGTCAAGGAGCAGAACACAATCAGCTTCATAACTTTGA      19

VIP5-4D1      AATAGACACAAGAGCTCCAATCGGTCGAGGAGCAGAACACAATCAGCTTCATAACTTTGA      19

              *********************w****x*********************************

VIP5-Col      GTTCATGGCGAGGAAGCAACTGACCGAAGCAACTGTGGGATGCAGAGTCGCAGAGAACGA      21

VIP5-11C1     GTTCATGGCGAGGAAGCAACTGACCGAAGCAACTGTGGGATGCAGAGTCGCAGAGAACGA      21

VIP5-4D1      GTTCATGGCGAGGAAGCAACTGACCGAAGCAACTGTGGGATGCAGAGTCGTAGAGAACGA      21

              **************************************************y*********

Col      CTGA      23

11C1     CTGA      23

4D1      CTGA      23

         ****
```

Figure 24: **Sequence alignment of 11 and 4 alleles of local accession for VIP5 gene compared to Col.0.**

Non-identical sequences are highlighted in grey, SNPs are shown in brown, start and stop codon in green. Only the parts of the gene which were sequenced are given here. Full alignment compared to Col.0 is shown appendix C.

160

Cycles 25 30 25 30 20 25 30 20 25 30

FDH-4D1 FDH-11C1 VIP5-11C1 VIP5-4D1
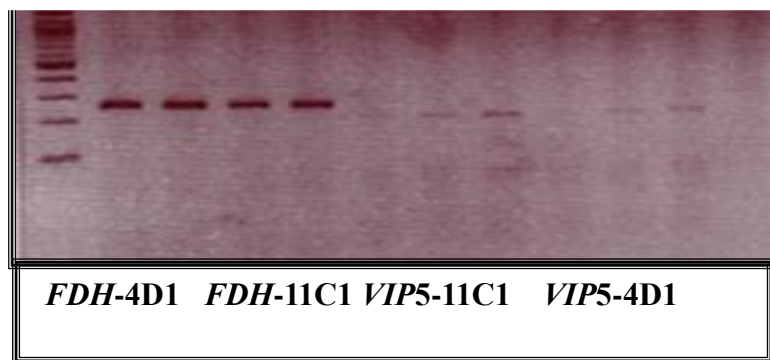
Figure 25: *VIP5* expression analysis for 20, 25 and 30 cycles relative to 25 and 30 cycles of glutathione-dependent formaldehyde dehydrogenase gene for parental lines, 11C1 and 4D1.

Difference in expression for *VIP5* between parental lines was suggested for 30 cycles which is not obvious for 20 and 25 cycles.

6.3 Sequence variation in *LSD1-LIKE 1* (*LDL1*) gene between 4D1 and 11C1 parents

To assess the presence of mutations in *LDL1*, *LDL1* primers (Appendix A) were used in polymerase chain reactions (PCR) to amplify the *LDL1* locus from 4D1 (4) and 11C1 (11). Products were sequenced with different sets of forward and reverse primers (Appendix A). *LDL1* is represented graphically in Figure 27. Overall, 28 single nucleotide polymorphisms (SNPs) distinguished the *LDL1* for 4 from the Col-0 reference sequence. These include three non-synonymous and twenty-five synonymous substitutions (Figure 26). Comparing the 11 allele for *LDL1* with Col-0 identified 16 differences, including an insertion of 882 bp in the promoter region of the 11 allele, and a deletion of 24 bp that included the transcriptional start site of the Col-0 allele. Neither the insertion nor the deletion was shared with the 4 allele.

This 882bp insertion in 11C1 is Stowaway-like MITEs. These have a TA target site duplications and terminal inverted repeats that look somewhat like this.

5' TCCTACTATATTATTTGGGAAGTA ……………TACCGCGGGTTAAAATCTAG 3'

5'   CTAGATTTTAACCCGCGGTA (inversion of 3'end)

This inverted repeat begins at base 6 and ends at the end. BLAST with our sequence gives a number of copies elsewhere in the Col-0 reference sequence and quite a few of the hits start at base 6. A list of loci with similar MITE sequence is given in the form of distance tree in Figure 28.

Transposon insertions in promoter regions often disrupt gene expression, which makes the 11 allele a candidate of loss of function mutation as also 24 bp deletion was found just before ATG in *LDL1* for 11C1 (Figure 25). This deletion might cause a mutation because it leads to deletion of part of the reported 5' UTR. In contrast, there is no evidence that the 4 allele might be non-functional.

162

The expression level was also examined for LDL1 with 20, 25 and 30 cycles of PCR for cDNA from both 11C1 and 4D1 parents compared to glutathione-dependent formaldehyde dehydrogenase gene (FDH). There was no detectable difference in expression level for 4 and 11 alleles of *LDL1* (Figure 28). Therefore the MITE insertion in the 11 allele does not appear to affect expression significantly.

```
11C1genomic     TTGACATAAAGATATAATCACTTATTTGTATAATGATTATGAAATCC ---1
4D1genomic       TTGACATAAAGATATAATCACTTATTTGTATAATGATTATGAAA--- 1
Col.genomic           ATAAAGATATAATCACTTATTTGTATAATGATTATGAAA--- 1

11C1genomic     TACTATATTATTTGGGAAGTACATATTAAATGTAACCTTAATTTTTGTAATTAATTACAT

11C1genomic     GACAATGCCATTAGAAAAATTTAATTAAAAACAAAATCCTTTAATGACGTAATTAAGGTT

11C1genomic     ACCAAAATCATTTAACGACAATATTTACTTCTTAATTATAGGGCTTATTAGATCTAAAGA

11C1genomic     CATGTCATCAAAGATTTCCTAAACTGAAGCAAAATATACCGAATATTCAAATATCTATCT

11C1genomic     GTTACCAAATTATAAACAAATAAGTTAATAAGTCAATAACTATTTGGAGGGACGGGTTTT

11C1genomic     TGAATCAACATTAATAAAAAAGTAAAATATAATTGATCCACCGTTTCAATACGGGTTAA

11C1genomic     ATCTTTAATTTATTATTTTTAAGACCACTGATATTAAACATATCAAATCATCCTAATTTA

11C1genomic     GAAAAGGTTATATAAAACCAAAAATGTTATGTGGTATGTATAATGTTACTATATATAAAA

11C1genomic     TTAAACTATAAAATATAAATGTATTAGAGAATAATACAATTTGTAAAACTTTTATATGTA

11C1genomic     ATAAATAATTCTCAAATTTTAAAAAATTACTACTTTAAAAACAAATCACGGGACGGGTAAA

11C1genomic     GAAATTAGAGAACAGATTTTATTTTGGAATTGAGTTATATGGTGGATGTATTTGAATCAA

11C1genomic     TATTTATAAAATTTTAAAATATTATTAATATGCTGTTTTAGTAAGGGTGAAAACTTCAGT

11C1genomic     TTTTTAACAATTGTCTCATGGATTCGTGGTATAGCGTTAGTTAATAACAATTATAAACTG

11C1genomic     TAAAATATAAATATTTTATAAAAATAAAATTTGCAAGTTTTAATATATATTATCTTTAAA

11C1genomic     AATAAATTATAACGCGGTATACCGCGGGTTAAAATCTAGTTTCATTTGTATTGTTTAAGC 2
4D1genomic                                              TTTCACTTGTATTGTTTAAGC 2
Colgenomic                                              TTTCACTTGTATTGTTTAAGC 2
                                                        ****  **************

11C1genomic     GAAAACACAGAATCAAAACCGATATTGTGCAAGCAATCTTTCAATGCAATAGTCCTAACA 3
4D1genomic      GAAAACAAAGGATCAAAACTGATATTGT----GCAATCTTTCAATGCAATAGTCCTAACA 3
Colgenomic      GAAAACAAAGGATCAAAACTGATATTGT----GCAATCTTTCAATGCAATAGTCCTAACA 3
                ******* ** ******** *******    *************************

11C1genomic     GTCCAGAAATCTAACGTTGCCGTATACCTTCCTTTTTTAGCTAAGAAAAGAAGGCTACAC 5
4D1genomic      GTCCAGAGATCAAAGGTTGCCGTATACCTTCCTTTTGTAGCTAAGAAAAGAAGGCTACAC 5
Colgenomic      GTGCAGAGATCAAACGTTGCCGTATACCTTCCTTTTTTAGCTAAGAAAAGAAGGCTACAC 5
                ** **** *** ** ********************* ****************

11C1genomic     TTCTT------------------------ACACAGAGCAAGAGCTATGTCAACAGAGAGT 6
4D1genomic      TTTGGTAGACTCGTCACTCACATCATCACACACAGAGCAAGAGCTATGTCAACAGAGAGT 6
Colgenomic      TTTGGTAGACTCGTCACTCACATCATCACACACAGAGCTAGAGCTATGTCAACAGAGACT 6
                **                           ******** *******************a*

11C1genomic     AAAGAAACCCGACCCGAAACTAAACCCGAAGACCCGGTTAACTCATACTACTGTGGATGTA 7
4D1genomic      AAAGAAACCCGACCCGAAACTAAACCCGAAGACCCGGCAACTCATACTACTGTGGATGTA 7
Colgenomic      AAAGAAACCCGACCCGAAACTAAACCCGAAGACCTGGGAACTCATACTACTGTGGATGTA 7
                ****************************** b**c*****************

11C1genomic     CCCGGTGAAGAACCTCTCGGAGAGCTTATCGCCGACGACGTGAACGAAGTCGTTTCTGAT 8
4D1genomic      CCCGGTGAAGAACCTCTCGGAGAGCTTATTGCCGACGACGTGAACGAAGTCGTTTCTGAT 8
Colgenomic      CCCGGTGAAGAACCTCTCGGAGAGCTTATCGCCGACGACGTGAACGAAGTCGTTTCTGAT 8
                ***************************** d*****************************

11C1genomic     GAAGACGGCCAAAACTCACTCGATGACCAATCGCCATTAACGGAGCTTCAGCCTCTTCCT 10
4D1genomic      GAAGACGGCCACAACTCACTCGATGACCAATCGCCATTAACGGAGCTTCAACCTCTTCCT 10
Colgenomic      GAAGACGGCCAAAACTCACTCGATGACCAATCGCCATTAACGGAGCTTCAGCCTCTTCCT 10
                ********** e**********************************f*********
```

```
11C1genomic   TCCGATCTGGTAACGGAGCAACAATCACAAAACCCTAATGCGGCGGAGCCTGGTCCTAGA 12
4D1genomic    TCCGATCTAGTAACGGAGCAACAATCACAAAACCCTAATGCGGCGGAGCCTGGTCCTAGA 12
Colgenomic    TCCGATCTGGTAACGGAGCAACAATCACAAAACCCTAATGCGGCGGAGCCTGGTCCTAGA 12
              ********g***************************************************

11C1genomic   GCAAGAAAAAGACGCCGTAGGAAACGTTTCTTCACTGAGATTAACGCAAACCCAGCTTTC 13
4D1genomic    GCAAGAAAACGACGCCGTAGGAAACGTTTCTTCACTGAGATTAACGCAAACCCAGCTTTC 13
Colgenomic    GCAAGAAAAAGACGCCGTAGGAAACGTTTCTTCACTGAGATTAACGCAAACCCAGCTTTC 13
              *********h*************************************************

11C1genomic   TCGGTAGGGTTTCCGGTTTATTCGCTCACGGAGGAAGAAATTGAAGCTAATGTGGTTTCG 15
4D1genomic    TCCGTAGGGTTTCCGGTTTATTCACTCACGGAGGAAGAAATCGAAGCTAATGTGGTTTCA 15
Colgenomic    TCGGTAGGGTTTCCGGTTTATTCGCTCACGGAGGAAGAAATTGAAGCTAATGTGGTTTCG 15
              **i*******************j****************k*****************l

11C1genomic   TGGCGATCTAATGTATCGAATTGGTTAACGCGAGATCATGCGCTTGAGTCTATACGTGCT 17
4D1genomic    TGGCGATCTAATGTATCGAATTGGTTAACGCGAGACCATGCGCTTGAGTCTATACGTGCT 17
Colgenomic    TGGCGATCTAATGTATCGAATTGGTTAACGCGAGATCATGCGCTTGAGTCTATACGTGCT 17
              ***********************************m***********************

11C1genomic   GAACACAAAACCTTAGTTGATACTGCTTACAATTTCCTTCTTGAGCATGGTTATATTAAC 18
4D1genomic    GAACACAAAACCTTAGTTGATACTGCTTACAATTTCCTCCTTGAGCATGGTTATATTAAC 18
Colgenomic    GAACACAAAACCTTAGTTGATACTGCTTACAATTTCCTTCTTGAGCATGGTTATATTAAC 18
              **************************************n*********************

11C1genomic   CCAAATGTTGTCGTTGTAGGGGCGGGTTTAGCTGGATTGGTTGCTGCTAGACAGTTGTTG 20
4D1genomic    CCTAATGTTGTCGTTGTAGGGGCGGGTTTAGCTGGTTTGGTTGCTGCTAGACAGTTGTTG 20
Colgenomic    CCAAATGTTGTCTGTTGTAGGGGCGGGTTTAGCTGGTTTGGTTGCTGCTAGACAGTTGTTG 20
              **o********p********************q***********************

11C1genomic   ACACGGAAGATGAAAGGTGGTGATGGTGTTGAGGCAATGGCTGATGTTGGTGGAAGCGTT 22
4D1genomic    ACTCGGAAGATGAAAGGTGGTGATGGTGTTGAGGCAATGGCTGATGTTGGTGGAAGTGTT 22
Colgenomic    ACACGGAAGATGAAAGGTGGTGATGGTGTTGAGGCAATGGCTGATGTTGGTGGAAGCGTT 22
              **r********************************************************s***

11C1genomic   CTCACCGGAATTAATGGGAATCCGCTTGGGGTTTTAGCGAGGCAACTTGGTTTGCCTCTT 23
4D1genomic    CTCACCGGAATCAATGGTAATCCGCTTGGGGTTTTAGCGAGGCAACTTGGTTTGCCTCTT 23
Colgenomic    CTCACCGGAATTAATGGGAATCCGCTTGGGGTTTTGGCGAGGCAACTTGGTTTGCCTCTT 23
              ***********t*****u****************v***********************

11C1genomic   CATAAGGTTAGAGATATTTGTCCTTTGTATCTTCCCAATGGAGAGCTTGCTGATGCTAGT 24
4D1genomic    CACAAGGTTAGAGATATTTGTCCTTTGTATCTTCCCAGTGGAGAGCTTGCTGACGCTGGT 24
Colgenomic    CATAAGGTTAGAGATATTTGTCCTTTGTATCTTCCCAATGGAGAGCTTGCTGATGCTAGT 24
              **w****************************************x***************y***z**

11C1genomic   ATATTTGTACATGCGTTAGCGGAAAATCTTCCGATTTTTTACGGGAGTACAGTTGAGAGC 30
4D1genomic    ATATTTGTACATGCTTTAGCGGAAAATCTTCCAATTTTTTACGGGAGTACAGTTGAGAGC 30
Colgenomic    ATATTTGTACATGCTTTAGCGGAAAATCTTCCAATTTTTTACGGGAGTACAGTTGAGAGC 30
              **************A*****************B*************************

11C1genomic   ATCAGATATGGAAGTAACGGGGTTCTGGTTTACACAGGTAACAAAGAGTTCCACTGCGAT 31
4D1genomic    ATCAGATATGGAAGCAACGGGGTTCTGGTTTACACAGGTAACAAAGAGTTCCACTGCGAT 31
Colgenomic    ATCAGATATGGAAGCAACGGGGTTCTGGTTTACACAGGTAACAAAGAGTTCCACTGCGAT 31
              **************C*****************************************

11C1genomic   ATGGCTCTTTGCACGGTTCCATTAGGTGTTCTGAAGAAAGGTTCAATTGAGTTTTATCCC 32
4D1genomic    ATGGCTCTTTGCACGGTTCCATTAGGTGTTCTGAAGAAAGGTTCGATTGAGTTTTATCCC 32
Colgenomic    ATGGCTCTTTGCACGGTTCCATTAGGTGTTCTGAAGAAAGGTTCGATTGAGTTTTATCCC 32
              *******************************************D**************

11C1genomic   ACCGAAGATCCGTCCACTAGAGGAGAATTCTTCTTGTTCTACAGCTATTCTTCTGTTTCC 35
4D1genomic    ACCGAAGATCCGTCCACTAGAGGAGAATTCTTCTTGTTCTACAGCTATTCTTCTGTTTCC 35
Colgenomic    ACCGAAGATCCGTCCACCAGAGGAGAATTCTTCTTGTTCTACAGCTATTCTTCTGTTTCC 35
              *****************E***************************************
```

165

```
11C1genomic    GCCGAGAGTGTCGGTGATGGAAGAGTGTTCTTTGCAGGTGAAGCTACTAACAGACAATAT 40
4D1genomic     GCCGAGAGTGTCGGTGATGGAAGAGTGTTCTTTGCAGGTGAAGCTACTAACAGACAATAT 40
Colgenomic     GCCGAGAGTGTTGGTGATGGAAGAGTGTTCTTTGCAGGTGAAGCTACTAACAGACAATAT 40
               ***********F************************************************

11C1genomic    GTTGCTAGAAGAAGAGCGTCATCATCGGCTTTAAATCCTAACCAGACCTGCATCGACAAA 42
4D1genomic     GTTGCTAGAAGAAGAGCGTCATCATCGGCTTTAAATCCTAACCAGACCTGCATCGACAAA 42
Colgenomic     GTTGCTAGAAGAAGAGCGTCATCATCGGCTTTAAATCCTAACCAGATCTGCATCGACAAA 42
               **********************************************G*************

11C1genomic    GAAGAAGAGGTAGAGGAAGAAGAAGACCGCTGTTTGGATCAGTTATTCGAGACACCCGAT 43
4D1genomic     GAAGAAGAGGTAGACGAAGAAGAAGACCGCTGTTTGGATCAGTTATTCGAGACACCCGAT 43
Colgenomic     GAAGAAGAGGTAGACGAAGAAGAAGACCGCTGTTTGGATCAGTTATTCGAGACACCCGAT 43
               *************H**********************************************

11C1genomic    TTAGTGACAAGGAAGCAAGCTATTGAGCTTGGTGAAATGGATGGAGATGAGTTGAGAAAT 46
4D1genomic     TTAGTGGCAAGGAAGCAAGCTATTGAGCTTGGTGAAATGGATGGAGATGAGTTGAGAAAT 46
Colgenomic     TTAGTGACAAGGAAGCAAGCTATTGAGCTTGGTGAAATGGATGGAGATGAGTTGAGAAAT 46
               ******I****************************************************

11C1genomic    GGGGAATCGATGATCTCTTCACTCAAAGCTGCAAGACTGAATCGACAGATCTTTGATTAG 48
4D1genomic     GGGGAATCGATGATCTCTTCACTCAAAGCTGCAAGACTGAATCGACAGATCTTTGATTAG 48
Colgenomic     GGGGAATCGATGATCTCTTCACTCAAAGCTGCAAGACTGAATCGACAGATCTTTGATTAG 48
               ************************************************************
```

**Figure 26: Sequence alignment of 11 and 4 alleles of *LDL1* gene compared to Col.0.**

Non-identical sequences are highlighted in grey, SNPs and deletions are shown in brown, start and stop codons in green.
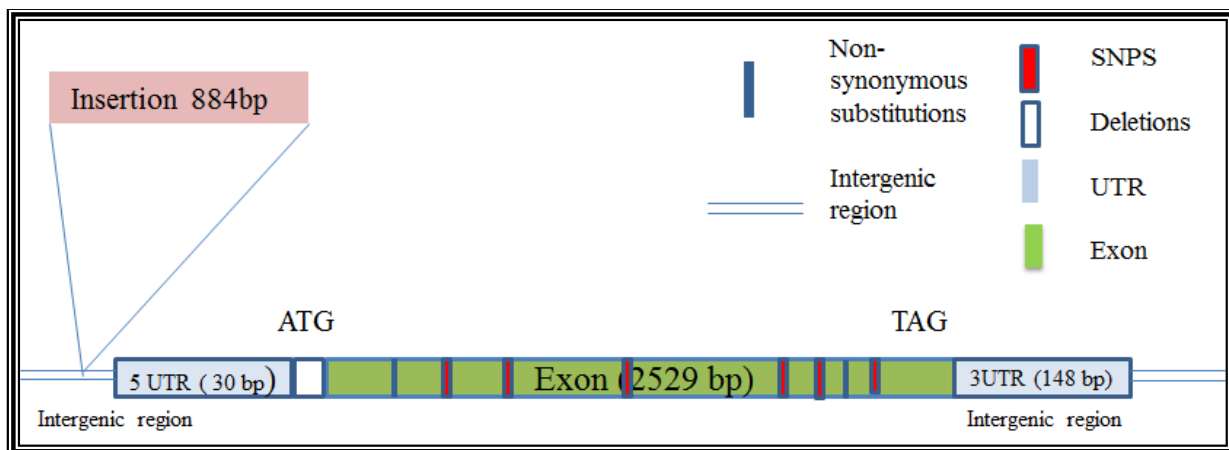
Figure 27: Graphical representation of the *LDL1* gene for the 11C1 parent.

This figure shows different parts of genes in different colours.
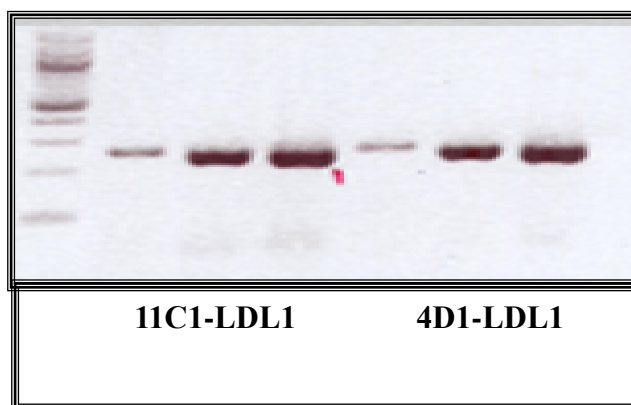
Figure 28: *LDL1* expression analysis for 20, 25 and 30 cycles of PCR.

No difference in expression of *LDL1* was found between the parents.
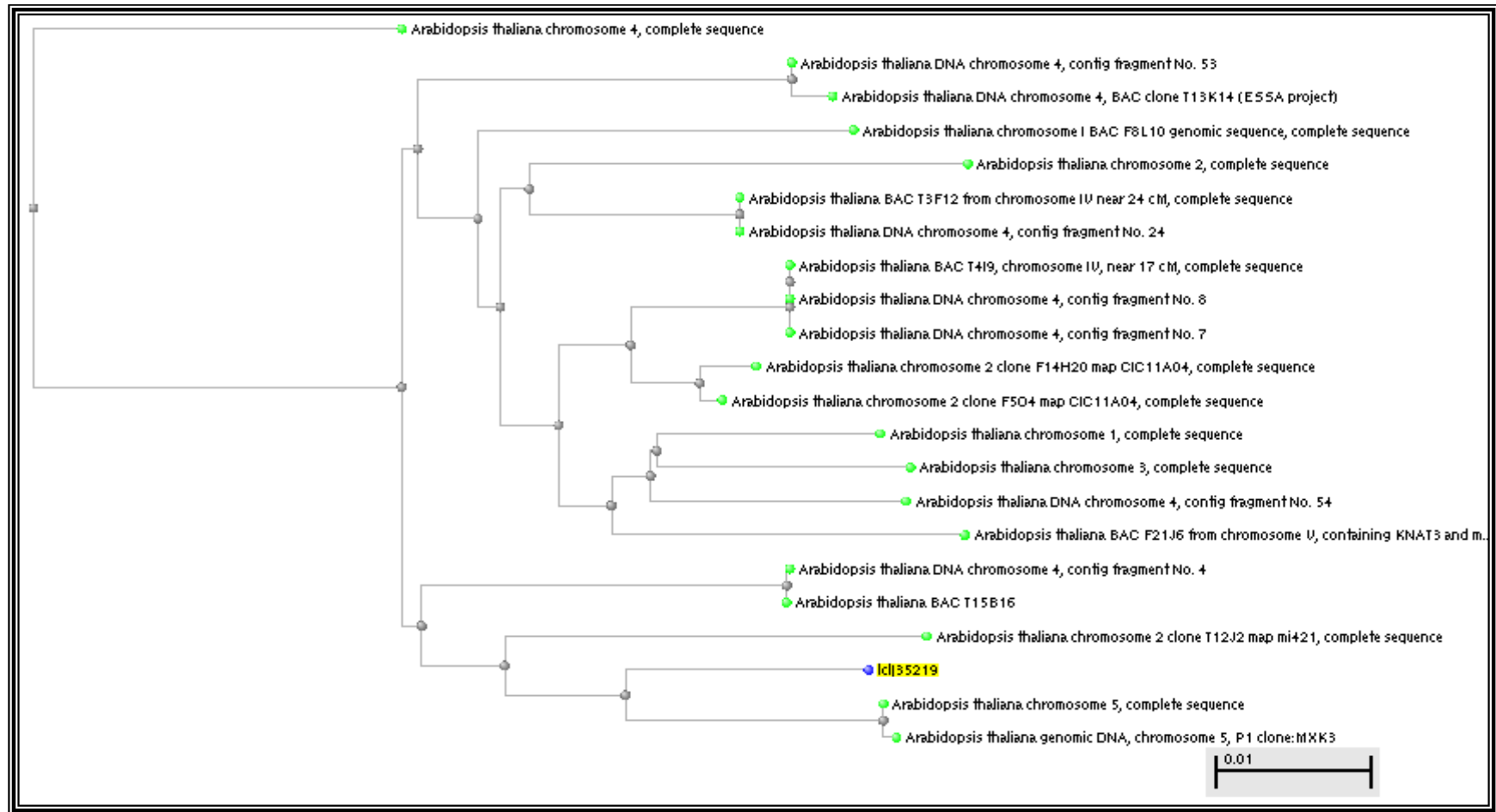
Figure 29: MITEs are found in the *Arabidopsis* genome at different positions.

Distance tree shows the relationship of the MITE in the *LDL1* allele of 11C1 to others found elsewhere in the Col-0 genome sequence.

6.4 The 11 allele of *LDL1* has a different transcription start site

No obvious difference had been detected in the level of *LDL1* expression in 4D1 and 11C1. However, this did not rule out the possibility that the MITE transposon or other 5' polymorphisms in 11C1 affected the sequence of the *LDL1*. For example, it is possible that transcription initiates in the MITE, leading to production of a transcript that encodes a longer, non-functional protein or that is incorrectly spliced. To test this, I amplified different regions of the *LDL1* transcript from 11C1. All had the same length as wild-type, suggesting that transcripts were not wrongly spliced. I therefore examined the transcription start sites using 5'RACE. A shorter RACE product was amplified from 11C1 compared to 4D1 (Figure 31). PCR products were gel purified and cloned in pJET1.2 and the largest inserts for each genotype were sequenced. This showed that transcription of the 4D1 allele began at the same position as Col-0, 31 bp upstream of the ATG (Figure 26). However, the 11C1 transcript extended only as far as the start of the deletion of the 5'UTR, 16 bp upstream of the ATG. It is possible that this affects translation of the LDL1 protein.

6.5 Association of *LDL1* polymorphisms with flowering time variation

The MITE insertion in the *LDL1* allele of 11C1 provided a convenient polymorphism for genotyping the F5 generation to compare the allelic effects of 11 and 4 on flowering time. *LDL1* was significantly associated (p=0.05) to flowering time in F5 plants in growth room conditions with the 11 allele causing later flowering (Table 37). However, this gene did not show an association to flowering time of plants grown in natural conditions.

6.6 The 11 allele of *LDL1* is found in other members of the same local population

To test for the presence of the MITE insertion in other local accessions, primers flanking the transposon insertion in the 11C1 allele were used (primers FN4 and RN4 in Appendix A). Genotyping results were analysed on agarose gel as described in 4.2.1 and scored as in Figure 15a. DNA from ninety-six local accessions were screened (Figure 30) and the insertion was only found in other accessions from the same site as 11C1 (site 11). Searching the 1001 Genomes database also showed that the 11 *LDL1* allele sequence has not been found in any other world-wide accession. These results showed that this polymorphism is not widely spread and therefore that it is likely to be a relatively recent insertion.

Gene	Trait	p-value	11/11	11/4	4/4
LDL1	GR2	0.05	34	0	30
LDL1	OS4	0.54	14	0	15

Table 37: Testing associations and allelic effects between Chromosome 1 genotypes and flowering time under natural and controlled conditions.

LDL1 showed significant association to flowering time of F5 plants in growth room but not in natural conditions.
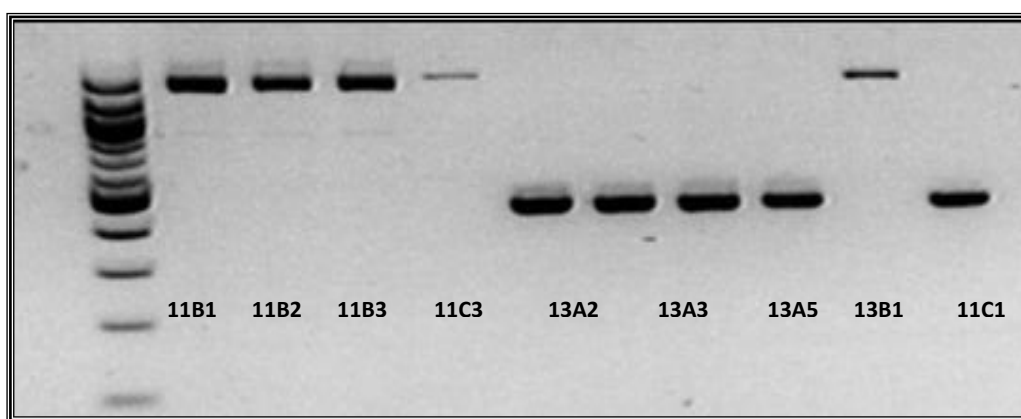
Figure 30: Genotyping of local parental accessions of F3 and F5 with LDL1 primers.

LDL1 primers detected the insertion in 11C1 and in other plants from the same local site.
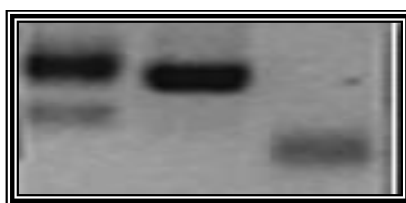
Figure 31: Analysis with 5' RACE PCR alleles and amplification of the inserts from pJET1.2 clones for 4 and 11 bands.

In first round 5' RACE PCR, no product was visible but difference was observed for 4 and 11 band in a second round of nested amplification (a). The 4 parent (first lane) produced a larger product than the 11 parent (second lane). The third lane is a no-cDNA control.

6.7 LDL1 amino acid sequence variation

Gene specific primers, GSP1 and GSP2 were designed for 5' RACE PCR and nested PCR, respectively, with NUP (Table 17). RACE PCR products for both 4 and 11 parental alleles were sequenced. This work is done with the help of Paulina Letmankova and Murray Wham, undergraduate project students working in the lab.

For 11C1 LDL1 protein, I found four amino acids substitutions compared to Col-0 LDL1 (Figure 32). These are threonine (T; polar and hydrophobic) is replaced by serine (S; polar), leucine (L; hydrophobic and aliphatic) by proline (P; unique/rare), glycine (G; tiny) by valine (V; small and aliphatic) and aspartic acid (D; negatively charged) is replaced by glutamic acid (E; negatively charged). Substitutions by proline and valine are important as they differ from leucine and glycine and these might change the properties of LDL1 protein.

For the 4D1 LDL1 protein, seven amino acids substitutions were also found compared to Col-0. These are threonine (T) is substituted by serine (S), leucine (L) by proline (P), glycine (G) by alanine (tiny and aliphatic), glutamine (Q; polar) by histidine (H; aromatic), asparagine (N; small) by serine (S; small), serine (S) by glycine (G) and threonine (T) by alanine (A) (Figure 29). However, none of these substituted amino acids are well conserved within *Arabidopsis* or between other *LDL1*-like genes.

Searching the *1001 Genomes* database of Arabidopsis ecotype sequences revealed that the 4D1 *LDL1* allele sequence was also present in two other accessions – An-1 from Antwerp in Belgium and Bg 2 from near Windsor, UK. The 11C1 *LDL1* allele did not appear to be present in any of the sequenced accessions. However, further work is needed to determine whether these substitutions might affect protein function.

```
                 a            bc
11    MSTESKETRPETKPEDPVTHTTVDVPGEEPLGELIADDVNEVVSDASATETDFSLSPSQS 1
4     MSTESKETRPETKPEDPATHTTVDVPGEEPLGELIADDVNEVVSDASATETDFSLSPSQS 1
Col0  MSTETKETRPETKPEDLGTHTTVDVPGEEPLGELIADDVNEVVSDASATETDFSLSPSQS 1
      ****:***********  *****************************************
               d
11    EQNIEEDGQNSLDDQSPLTELQPLPLPPPLPVEARISESLGEEESSDLVTEQQSQNPNAA 2
4     EQNIEEDGHNSLDDQSPLTELQPLPLPPPLPVEARISESLGEEESSDLVTEQQSQNPNAA 2
Col0  EQNIEEDGQNSLDDQSPLTELQPLPLPPPLPVEARISESLGEEESSDLVTEQQSQNPNAA 2
      ********:***************************************************

11    EPGPRARKRRRRKRFFTEINANPAFSRNRRTSVGKEVDSEALIAMSVGFPVYSLTEEEIE 3
4     EPGPRARKRRRRKRFFTEINANPAFSRNRRTSVGKEVDSEALIAMSVGFPVYSLTEEEIE 3
Col0  EPGPRARKRRRRKRFFTEINANPAFSRNRRTSVGKEVDSEALIAMSVGFPVYSLTEEEIE 3
      ***********************************************************

11    ANVVSIIGGKDQANYIVVRNHIIALWRSNVSNWLTRDHALESIRAEHKTLVDTAYNFLLE 4
4     ANVVSIIGGKDQANYIVVRNHIIALWRSNVSNWLTRDHALESIRAEHKTLVDTAYNFLLE 4
Col0  ANVVSIIGGKDQANYIVVRNHIIALWRSNVSNWLTRDHALESIRAEHKTLVDTAYNFLLE 4
      ***********************************************************

11    HGYINFGLAPVIKEAKLRSFDGVEPPNVVVVGAGLAGLVAARQLLSMGFRVLVLEGRDRP 5
4     HGYINFGLAPVIKEAKLRSFDGVEPPNVVVVGAGLAGLVAARQLLSMGFRVLVLEGRDRP 5
Col0  HGYINFGLAPVIKEAKLRSFDGVEPPNVVVVGAGLAGLVAARQLLSMGFRVLVLEGRDRP 5
      ***********************************************************
                                                           e
11    GGRVKTRKMKGGDGVEAMADVGGSVLTGINGNPLGVLARQLGLPLHKVRDICPLYLPNGE 6
4     GGRVKTRKMKGGDGVEAMADVGGSVLTGINGNPLGVLARQLGLPLHKVRDICPLYLPSGE 6
Col0  GGRVKTRKMKGGDGVEAMADVGGSVLTGINGNPLGVLARQLGLPLHKVRDICPLYLPNGE 6
      ********************************************************.**
          f
11    LADASVDSKIEASFNKLLDRVCKLRQSMIEENKSVDVPLGEALETFRLVYGVAEDQQERM 7
4     LADAGVDSKIEASFNKLLDRVCKLRQSMIEENKSVDVPLGEALETFRLVYGVAEDQQERM 7
Col0  LADASVDSKIEASFNKLLDRVCKLRQSMIEENKSVDVPLGEALETFRLVYGVAEDQQERM 7
      ****.******************************************************
11    LLDWHLANLEYANATLLGNLSMAYWDQDDPYEMGGDHCFIPGGNEIFVHALAENLPIFYG 8
4     LLDWHLANLEYANATLLGNLSMAYWDQDDPYEMGGDHCFIPGGNEIFVHALAENLPIFYG 8
Col0  LLDWHLANLEYANATLLGNLSMAYWDQDDPYEMGGDHCFIPGGNEIFVHALAENLPIFYG 8
      ***********************************************************

11    STVESIRYGSNGVLVYTGNKEFHCDMALCTVPLGVLKKGSIEFYPELPHKKKEAIQRLGF 9
4     STVESIRYGSNGVLVYTGNKEFHCDMALCTVPLGVLKKGSIEFYPELPHKKKEAIQRLGF 9
Col0  STVESIRYGSNGVLVYTGNKEFHCDMALCTVPLGVLKKGSIEFYPELPHKKKEAIQRLGF 9
      ***********************************************************
11    GLLNKVAMLFPCNFWGEEIDTFGRLTEDPSTRGEFFLFYSYSSVSGGPLLVALVAGDAAE 10
4     GLLNKVAMLFPCNFWGEEIDTFGRLTEDPSTRGEFFLFYSYSSVSGGPLLVALVAGDAAE 10
Col0  GLLNKVAMLFPCNFWGEEIDTFGRLTEDPSTRGEFFLFYSYSSVSGGPLLVALVAGDAAE 10
      ***********************************************************

11    RFETLSPTDSVKRVLQILRGIYHPKGIVVPDPVQALCSRWGQDKFSYGSYSYVAVGSSGD 11
4     RFETLSPTDSVKRVLQILRGIYHPKGIVVPDPVQALCSRWGQDKFSYGSYSYVAVGSSGD 11
Col0  RFETLSPTDSVKRVLQILRGIYHPKGIVVPDPVQALCSRWGQDKFSYGSYSYVAVGSSGD 11
      ***********************************************************

11    DYDILAESVGDGRVFFAGEATNRQYPATMHGAFLSGMREAANILRVARRRASSSALNPNQ 12
4     DYDILAESVGDGRVFFAGEATNRQYPATMHGAFLSGMREAANILRVARRRASSSALNPNQ 12
Col0  DYDILAESVGDGRVFFAGEATNRQYPATMHGAFLSGMREAANILRVARRRASSSALNPNQ 12
      ***********************************************************
```

```
                      g
11         TCIDKEEEVEEEEDRCLDQLFETPDLTFGNFSVLFTPNSDEPESMSLLRVRIQMEKPESG 13
4          TCIDKEEEVDEEEDRCLDQLFETPDLTFGNFSVLFTPNSDEPESMSLLRVRIQMEKPESG 13
Col0       ICIDKEEEVDEEEDRCLDQLFETPDLTFGNFSVLFTPNSDEPESMSLLRVRIQMEKPESG 13
            ********:****************************************************

                      h
11         LWLYGLVTRKQAIELGEMDGDELRNEYLREKLGLVPVERKSLSQEGESMISSLKAARLNR 14
4          LWLYGLVARKQAIELGEMDGDELRNEYLREKLGLVPVERKSLSQEGESMISSLKAARLNR 14
Col0       LWLYGLVTRKQAIELGEMDGDELRNEYLREKLGLVPVERKSLSQEGESMISSLKAARLNR 14
            *******:****************************************************

11         QIFD 15
4          QIFD 15
Col0       QIFD 15
            ****
```

Figure 32: Alignment of LDL1 amino acid sequences of 4D1 and 11C1 compared to Col-0.

The protein sequences of the sequenced LDL1 alleles, identical amino acids are in black, the presence of a substitution relative to the other sequences is identified in grey. The letters above the amino acid sequence refer to the nucleotide substitutions in Figure 24.

6.8 *Arabidopsis* transformation experiments

If an *ldl1* loss-of-function mutation contributes to the later flowering of 11C1, then transformation of 11C1 with a wild-type copy of the *LDL1* gene should cause earlier flowering. A T-DNA construct carrying the *LDL1* gene of 4D1 was therefore made, introduced into *Agrobacterium* and used to dip flowers of 11C1. Time did not permit selection of transformed plants or testing of their flowering times.

6.9 Discussion

Variation in the sequences of *LDL1* and *VIP5* between the 4D1 and 11C1 mapping parents were detected. Synonymous and non-synonymous single nucleotide substitutions were observed in *VIP5* for both 11C1 and 4D1 relative to Col-0. Difference in expression levels were also suggested for *VIP5* with 4D1 showed lower expression than 11C1. However, this finding needs to be further investigated with quantitative RT-PCR. *VERNALIZATION INDEPENDENCE 2* (*VIP2*), *ELF8* (*VIP6*), *VIP4*, and *VIP5* are members of the Paf1-complex, which has been shown to be required for a high level of *FLC* expression (He et al., 2004; Kim et al., 2005; Oh et al., 2004). The *Arabidopsis* PAF1-like complex components are required for the H3K4me3 enrichment in *FLC* chromatin and *FLC* expression, and thus for floral repression (He et al., 2004). Therefore if the *VIP5* allele from 11C1 is responsible for suppression of the *fri* mutant phenotype and late flowering, it would have to be more active than the 4D1 allele. 11C1 is a *fri* mutant but is late flowering and expresses *FLC* at a detectable level (McCulloch and Hudson, 2011). In the absence of a functional *FRI* allele, the autonomous-pathway genes, such as *FCA*, *FPA*, *FVE*, *LUMINIDEPENDENS*, and *FLOWERING LOCUS D* (*FLD*) can still repress *FLC* expression to accelerate flowering (Ausin et al., 2004; Lee et al., 1994a; Macknight et al., 1997; Schmid et al., 2003; Schomburg

et al., 2001b). *Arabidopsis* relatives of the human histone demethylase *LSD1* (*LDL1*, *LDL2*, and *FLD*) reduce the levels of H3K4 methylation in *FWA* and *FLC* chromatin and act to repress the expression of these two genes (Jiang et al., 2007). *FLD*, *LDL1*, and *LDL2* act in partial redundancy to repress *FLC* expression (Krichevsky et al., 2007). Importantly, an *ldl1* mutation can delay the early flowering of *fri* mutants (Jiang et al., 2007). Therefore an *ldl1* loss-of-function mutation in 11C1 is consistent with suppression of its *fri* mutant phenotype. Polymorphisms were found in *LDL1* between 4D1 and 11C1. Although a number of non-synonymous substitutions were found in the coding region, most of these were either conservative or were found in other world-wide accessions, suggesting that they did not affect LDL1 activity. The 11C1 allele was also found to carry a MITE insertion in its promoter. MITES tend to be found close to expressed genes. These often affect expression of the neighbouring genes, usually reducing expression (Fattash et al., 2013; Santiago et al., 2002). However, similar levels of mRNA were detected in 11C1 and 4D1 plants. A potential loss-of-function mutation involving deletion of the normal transcript initiation site of *LDL1* was also found in the 11C1 allele, causing transcription to start 16 bp upstream of the assumed ATG. The CAP-binding complex binds to 5' end of RNA. If it is too close to the ATG start codon, it could potentially prevent access of ribosomes to ATG to initiate translation, either resulting in reduced protein levels or initiation at a more downstream ATG and production of a truncated protein. This could be tested further by studies of protein expression, e.g., with Western blots. Co-segregation of in the F5 population supported the idea that *LDL1* might be the QTL and for a loss-of -function mutation (*ldl1*) in 11C1.

MITE transposons are considered to be not very active in *Arabidopsis* (Fattash et al., 2013), however, the insertion in *LDL1* in 11C1 is not found in other accessions (except plants from the same population), suggesting that it was active relatively recently. From evolutionary point of view- independent *fri* mutations have given rise to rapid cycling accessions from

winter annuals multiple times. The suppressor mutation in 11C1 might or not reverse this evolutionary trend.

A further possibility is that a loss-of-function *ldl1* mutation might also affect growth rate. Since, a growth rate QTL is also linked to *CIW1*; it is possible that a loss of LDL1 activity in 11C1 contributes to variation in both flowering time and growth rate. This is supported by the finding that an *ldl1* mutation affects root growth rate in Col-0 genetic background (Singh et al., 2012).

7.0 Critical appraisal

The simplest way to test the involvement of *LDL1* further would be to transform the 11C1 parent with a copy of the *LDL1* gene from 4D1. If 11C1 is an *ldl1* mutant, the transgene should complement it and advance flowering and reduce growth rate. If a shorter $5^{/}$ UTR (Figure 31) affects expression of LDL1 [11] allele, this can be tested by Western blotting with antibodies that are specific to LDL1 (11C1 should produce a shorter or less abundant protein than 4D1). However, antibodies are not available for LDL1, therefore an alternative approach would be to transform *ldl1* mutants with an *LDL1* construct in which the protein is epitope tagged (e.g., with Myc or GFP).

8.0 Concluding remarks

Extensive genetically determined variation in growth rate and flowering time in local accessions of *Arabidopsis thaliana* in Edinburgh (Scotland) had been observed on a similar scale to worldwide (Gazzani et al., 2003; Johanson et al., 2000).

Variation in plants growth rate was observed both in control and natural conditions and also in seasons. Variation in rosette areas of plants as growth rate has not been studied in *Arabidopsis* accessions compared to hypocotyl growth (Borevitz et al., 2002; Botto and Smith, 2002; Maloof et al., 2001). However, growth rate in different *Arabidopsis* accessions was studied and compared with other parameters like seed weight etc (Li et al., 1998). Significant correlation was found between seed weight and rosette area of plants (El-Lithy et al., 2004) High variation was observed in spring and this is consistent with studies that spring phenology has strong genetic component (Pellis et al., 2004; Tsarouhas et al., 2003; Weih, 2009) and it enables the plants to use favourable conditions of spring and this might affect productivity.

Significant correlations between traits studied in variable conditions suggest that these traits are predominantly controlled by same genetic factors. Two QTLs each for growth rate and flowering time were detected in analysis. However, these QTLs were not found in all subsequent experiments. Previous studies have shown that *FRI* and *FLC* are needed for very late flowering (Gazzani et al., 2003; Johanson et al., 2000; Michaels et al., 2003) and the 11C1 parent in our experiment is a *fri* mutant (McCulloch and Hudson, 2011) but none of these loci were detected in our study.

This analysis suggested that at least part of the late flowering behaviour of 11C1 parent could be attributed to the variation in the region of either *VERNALIZATION INDEPENDENCE 5*

178

(*VIP5*) or *LSD1-LIKE1* (*LDL1*). *vip5* mutants are known to flower earlier than *flc* null mutants, suggesting that other flowering-time genes are targeted by VIP5. Strong *vip* mutants also exhibit developmental pleiotropy, suggesting that the *VIP* genes also target mechanisms unrelated to flowering (Zhang et al., 2003), however 4D1 did not show any of the other characteristics of *vip5* mutants. Loss of *VIP5* function leads to down-regulation of not only *FLC*, but also other members of the *FLC/MAF* MADS-box gene family, all of which have the capacity to act as floral repressors (Ratcliffe et al., 2001; Scortecci et al., 2001). The other potential candidate gene in this locus is *LDL1* which is known to repress *FLC* and promotes flowering in *Arabidopsis* (Krichevsky et al., 2007). *LDL1/SWP1* gene also known to regulate lateral root initiation and elongation and mutation in *LDL1* increases both the density and length of lateral roots in *Arabidopsis* (Singh et al., 2012) . Though, 11C1 carries a loss-of-function *fri* mutation (McCulloch and Hudson, 2011) no QTL for flowering time was detected in the vicinity of *FRI* (at the top of chromosome 4). One explanation is that 4D1 is also a *fri* mutation, although no obvious loss of function mutation had been detected in the sequence of the *FRI* coding sequence from 4D1. Alternatively, the effects of the *fri* mutation from 11C1 might have been masked in the mapping population by segregating suppressors. Polymorphism was found between 11C1 and 4D1 alleles of *LDL1*. Besides SNPs in both genes, a 24 bp deletion at the start of the *LDL1* gene and 882 bp MITE insertion upstream of the gene were found for 11 allele in the late flowering parent. Alleles for early (4D1) and late flowering (11C1) parents expressed at same level with semi-quantitative RT-PCR but a shortened 5'-UTR was observed for 11C1 with 5' RACE PCR.

In long days conditions, *ldl1* mutants are known to flower later than the wild-type Columbia (Col-0) (Jiang et al., 2007). Transgenic lines have been produced that carry the 4D1 LDL1 allele in an *ldl1* Col-0 background (Jiang et al., 2007), to test whether the 4D1 allele is functional, and in the 11C1 background, to test whether an *ldl1* loss-of-function mutation

179

contributes to its late flowering. This work is on-going and will hopefully reveal whether *LDL1* is relevant to the differences in flowering behaviour seen between 11C1 and 4D1.

Bibliography

Alonso-Blanco, C., Aarts, M. G. M., Bentsink, L., Keurentjes, J. J. B., Reymond, M., Vreugdenhil, D. and Koornneef, M. (2009). What Has Natural Variation Taught Us about Plant Development, Physiology, and Adaptation? *The Plant Cell Online* **21**, 1877-1896.

Alonso-Blanco, C., Blankestijn-de Vries, H., Hanhart, C. J. and Koornneef, M. (1999). Natural allelic variation at seed size loci in relation to other life history traits of Arabidopsis thaliana. *Proc Natl Acad Sci U S A* **96**, 4710-7.

Alonso-Blanco, C., El-Assal, S. E., Coupland, G. and Koornneef, M. (1998). Analysis of natural allelic variation at flowering time loci in the Landsberg erecta and Cape Verde Islands ecotypes of Arabidopsis thaliana. *Genetics* **149**, 749-64.

Alonso-Blanco, C. and Koornneef, M. (2000). Naturally occurring variation in Arabidopsis: an underexploited resource for plant genetics. *Trends Plant Sci* **5**, 22-9.

Alvarez-Venegas, R. and Avramova, Z. (2005). Methylation patterns of histone H3 Lys 4, Lys 9 and Lys 27 in transcriptionally active and inactive Arabidopsis genes and in atx1 mutants. *Nucleic Acids Res* **33**, 5199-5207.

Amasino, R. M. and Michaels, S. D. (2010). The timing of flowering. *Plant Physiol* **154**, 516-20.

Andersson, C. R., Helliwell, C. A., Bagnall, D. J., Hughes, T. P., Finnegan, E. J., Peacock, W. J. and Dennis, E. S. (2008). The FLX gene of Arabidopsis is required for FRI-dependent activation of FLC expression. *Plant Cell Physiol* **49**, 191-200.

Andres, M. E., Burger, C., Peral-Rubio, M. J., Battaglioli, E., Anderson, M. E., Grimes, J., Dallman, J., Ballas, N. and Mandel, G. (1999). CoREST: a functional corepressor required for regulation of neural-specific gene expression. *Proc Natl Acad Sci U S A* **96**, 9873-8.

Arimura, G.-i., Kost, C. and Boland, W. (2005). Herbivore-induced, indirect plant defences. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **1734**, 91-111.

Armengaud, P., Sulpice, R., Miller, A. J., Stitt, M., Amtmann, A. and Gibon, Y. (2009). Multilevel analysis of primary metabolism provides new insights into the role of potassium nutrition for glycolysis and nitrogen assimilation in Arabidopsis roots. *Plant Physiol* **150**, 772-85.

Arvidsson, S., Perez-Rodriguez, P. and Mueller-Roeber, B. (2011). A growth phenotyping pipeline for Arabidopsis thaliana integrating image analysis and rosette area modeling for robust quantification of genotype effects. *New Phytol* **191**, 895-907.

Assuero, S. G., Mollier, A. and Pellerin, S. (2004). The decrease in growth of phosphorus-deficient maize leaves is related to a lower cell production. *Plant Cell Environ* **27**, 887-895.

Ausin, I., Alonso-Blanco, C., Jarillo, J. A., Ruiz-Garcia, L. and Martinez-Zapater, J. M. (2004). Regulation of flowering time by FVE, a retinoblastoma-associated protein. *Nat Genet* **36**, 162-6.

Balasubramanian, S. and Weigel, D. (2006). Temperature Induced Flowering in Arabidopsis thaliana. *Plant Signal Behav* **1**, 227-8.

Ballas, N., Battaglioli, E., Atouf, F., Andres, M. E., Chenoweth, J., Anderson, M. E., Burger, C., Moniwa, M., Davie, J. R., Bowers, W. J. et al. (2001). Regulation of neuronal traits by a novel transcriptional complex. *Neuron* **31**, 353-65.

Ballas, N. and Mandel, G. (2005). The many faces of REST oversee epigenetic programming of neuronal genes. *Curr Opin Neurobiol* **15**, 500-6.

Bandaranayake, C., Koumproglou, R., Wang, X., Wilkes, T. and Kearsey, M. (2004). QTL analysis of morphological and developmental traits in the Ler Ã— Cvi population of Arabidopsis thaliana. *Euphytica* **137**, 361-371.

Bastow, R., Mylne, J. S., Lister, C., Lippman, Z., Martienssen, R. A. and Dean, C. (2004). Vernalization requires epigenetic silencing of FLC by histone methylation. *Nature* **427**, 164-167.

Bäurle, I. and Dean, C. (2006). The Timing of Developmental Transitions in Plants. *Cell* **125**, 655-664.

Beemster, G. T., De Veylder, L., Vercruysse, S., West, G., Rombaut, D., Van Hummelen, P., Galichet, A., Gruissem, W., Inze, D. and Vuylsteke, M. (2005). Genome-wide analysis of gene expression profiles associated with cell cycle transitions in growing organs of Arabidopsis. *Plant Physiol* **138**, 734-43.

Bentsink, L., Alonso-Blanco, C., Vreugdenhil, D., Tesnier, K., Groot, S. P. and Koornneef, M. (2000). Genetic analysis of seed-soluble oligosaccharides in relation to seed storability of Arabidopsis. *Plant Physiol* **124**, 1595-604.

Bentsink, L., Yuan, K., Koornneef, M. and Vreugdenhil, D. (2003). The genetics of phytate and phosphate accumulation in seeds and leaves of Arabidopsis thaliana, using natural variation. *Theor Appl Genet* **106**, 1234-43.

Berger, S., Sinha, A. K. and Roitsch, T. (2007). Plant physiology meets phytopathology: plant primary metabolism and plant-pathogen interactions. *J Exp Bot* **58**, 4019-26.

Bertin, N., Causse, M., Brunel, B., Tricon, D. and Genard, M. (2009). Identification of growth processes involved in QTLs for tomato fruit size and composition. *J Exp Bot* **60**, 237-48.

Blazquez, M. A., Ahn, J. H. and Weigel, D. (2003). A thermosensory pathway controlling flowering time in Arabidopsis thaliana. *Nat Genet* **33**, 168-171.

Boege, K. and Marquis, R. J. (2005). Facing herbivory as you grow up: the ontogeny of resistance in plants. *Trends Ecol Evol* **20**, 441-8.

Borevitz, J. O., Maloof, J. N., Lutes, J., Dabi, T., Redfern, J. L., Trainer, G. T., Werner, J. D., Asami, T., Berry, C. C., Weigel, D. et al. (2002). Quantitative Trait Loci Controlling Light and Hormone Response in Two Accessions of Arabidopsis thaliana. *Genetics* **160**, 683-696.

Boss, P. K., Bastow, R. M., Mylne, J. S. and Dean, C. (2004). Multiple pathways in the decision to flower: enabling, promoting, and resetting. *Plant Cell* **16 Suppl**, S18-31.

Botto, J. F. and Smith, H. (2002). Differential genetic variation in adaptive strategies to a common environmental signal in Arabidopsis accessions: phytochrome-mediated shade avoidance. *Plant Cell Environ* **25**, 53-63.

Broekgaarden, C., Poelman, E. H., Steenhuis, G., Voorrips, R. E., Dicke, M. and Vosman, B. E. N. (2008). Responses of Brassica oleracea cultivars to infestation by the aphid Brevicoryne brassicae: an ecological and molecular approach. *Plant Cell Environ* **31**, 1592-1605.

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

Buescher, E., Achberger, T., Amusan, I., Giannini, A., Ochsenfeld, C., Rus, A., Lahner, B., Hoekenga, O., Yakubova, E., Harper, J. et al. (2010a). Natural Genetic Variation in Selected Populations of *Arabidopsis thaliana* Is Associated with Ionomic Differences. *PLoS One* **5**, e11081.

Buescher, E., Achberger, T., Amusan, I., Giannini, A., Ochsenfeld, C., Rus, A., Lahner, B., Hoekenga, O., Yakubova, E., Harper, J. F. et al. (2010b). Natural genetic variation in selected populations of Arabidopsis thaliana is associated with ionomic differences. *PLoS One* **5**, e11081.

Burn, J. E., Smyth, D. R., Peacock, W. J. and Dennis, E. S. (1993). Genes conferring late flowering inArabidopsis thaliana. *Genetica* **90**, 147-155.

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.

Campillo, C., García, M. I., Daza, C. and Prieto, M. H. (2010). Study of a Non-destructive Method for Estimating the Leaf Area Index in Vegetable Crops Using Digital Images. *HortScience* **45**, 1459-1463.

Cao, Y., Dai, Y., Cui, S. and Ma, L. (2008). Histone H2B monoubiquitination in the chromatin of FLOWERING LOCUS C regulates flowering time in Arabidopsis. *Plant Cell* **20**, 2586-602.

Castellarin, S. D., Pfeiffer, A., Sivilotti, P., Degan, M., Peterlunger, E. and G, D. I. G. (2007). Transcriptional regulation of anthocyanin biosynthesis in ripening fruits of grapevine under seasonal water deficit. *Plant Cell Environ* **30**, 1381-99.

Causse, M., Saliba-Colombani, V., Lecomte, L., Duffe, P., Rousselle, P. and Buret, M. (2002). QTL analysis of fruit quality in fresh market tomato: a few chromosome regions control the variation of sensory and instrumental traits. *J Exp Bot* **53**, 2089-98.

Chapin, F. S., Schulze, E. and Mooney, H. A. (1990). The Ecology and Economics of Storage in Plants. *Annual Review of Ecology and Systematics* **21**, 423-447.

Chinnusamy, V., Schumaker, K. and Zhu, J. K. (2004). Molecular genetic perspectives on cross-talk and specificity in abiotic stress signalling in plants. *J Exp Bot* **55**, 225-36.

Choi, K., Park, C., Lee, J., Oh, M., Noh, B. and Lee, I. (2007). Arabidopsis homologs of components of the SWR1 complex regulate flowering and plant development. *Development* **134**, 1931-41.

Chomczynski, P. and Mackey, K. (1995). Short technical reports. Modification of the TRI reagent procedure for isolation of RNA from polysaccharide- and proteoglycan-rich sources. *Biotechniques* **19**, 942-5.

Chong, J. A., Tapia-Ramirez, J., Kim, S., Toledo-Aral, J. J., Zheng, Y., Boutros, M. C., Altshuller, Y. M., Frohman, M. A., Kraner, S. D. and Mandel, G. (1995). REST: a mammalian silencer protein that restricts sodium channel gene expression to neurons. *Cell* **80**, 949-57.

Christie, P., Alfenito, M. 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.

Chung, C. T., Niemela, S. L. and Miller, R. H. (1989). One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution. *Proc Natl Acad Sci U S A* **86**, 2172-5.

Clarke, J. H. and Dean, C. (1994). Mapping FRI, a locus controlling flowering time and vernalization response in Arabidopsis thaliana. *Mol Gen Genet* **242**, 81-9.

Clarke, J. H., Mithen, R., Brown, J. K. and Dean, C. (1995). QTL analysis of flowering time in Arabidopsis thaliana. *Mol Gen Genet* **248**, 278-86.

Clauss, M. J. and Mitchell-Olds, T. (2003). Population genetics of tandem trypsin inhibitor genes in Arabidopsis species with contrasting ecology and life history. *Mol Ecol* **12**, 1287-99.

Clough, S. J. and Bent, A. F. (1998). Floral dip: a simplified method forAgrobacterium-mediated transformation ofArabidopsis thaliana. *The Plant Journal* **16**, 735-743.

Cole, B., Kay, S. A. and Chory, J. (2011). Automated analysis of hypocotyl growth dynamics during shade avoidance in Arabidopsis. *Plant J* **65**, 991-1000.

Coupland, G. (1995). Genetic and environmental control of flowering time in Arabidopsis. *Trends Genet* **11**, 393-7.

Crawley, M. J. (2009). Plant–Herbivore Dynamics. In *Plant Ecology*, pp. 401-474: Blackwell Publishing Ltd.

Darvasi, A. (1998). Experimental strategies for the genetic dissection of complex traits in animal models. *Nat Genet* **18**, 19-24.

Darvasi, A. and Soller, M. (1995). Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics* **141**, 1199-207.

Dicke, M., van Loon, J. J. A. and Soler, R. (2009). Chemical complexity of volatiles from plants induced by multiple attack. *Nat Chem Biol* **5**, 317-324.

Doerge, R. W. (2002). Mapping and analysis of quantitative trait loci in experimental populations. *Nat Rev Genet* **3**, 43-52.

Doyle, J. J. and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**, 11-15.

Dubcovsky, J., Lukaszewski, A. J., Echaide, M., Antonelli, E. F. and Porter, D. R. (1998). Molecular Characterization of Two Triticum speltoides Interstitial Translocations Carrying Leaf Rust and Greenbug Resistance Genes. *Crop Sci.* **38**, 1655-1660.

Dubnau, J. and Struhl, G. (1996). RNA recognition and translational regulation by a homeodomain protein. *Nature* **379**, 694-699.

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

El-Lithy, M. E., Clerkx, E. J., 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**, 444-58.

Fattash, I., Rooke, R., Wong, A., Hui, C., Luu, T., Bhardwaj, P. and Yang, G. (2013). Miniature inverted-repeat transposable elements: discovery, distribution, and activity1. *Genome* **56**, 475-486.

Ferreira, M. E., Satagopan, J., Yandell, B. S., Williams, P. H. and Osborn, T. C. (1995). Mapping loci controlling vernalization requirement and flowering time in Brassica napus. *Theoretical and Applied Genetics* **90**, 727-732.

Forneris, F., Binda, C., Vanoni, M. A., Mattevi, A. and Battaglioli, E. (2005). Histone demethylation catalysed by LSD1 is a flavin-dependent oxidative process. *FEBS Lett* **579**, 2203-7.

Frank, M. J., Preuss, D., Mack, A., Kuhlmann, T. C. and Crawford, N. M. (1998). The Arabidopsis transposable element Tag1 is widely distributed among Arabidopsis ecotypes. *Mol Gen Genet* **257**, 478-84.

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.

Gendall, A. R., Levy, Y. Y., Wilson, A. and Dean, C. (2001). The VERNALIZATION 2 gene mediates the epigenetic regulation of vernalization in Arabidopsis. *Cell* **107**, 525-35.

Godiard, L., Sauviac, L., Torii, K. U., Grenon, O., Mangin, B., Grimsley, N. H. and Marco, Y. (2003). ERECTA, an LRR receptor-like kinase protein controlling

development pleiotropically affects resistance to bacterial wilt. *The Plant Journal* **36**, 353-365.

Goldschmidt, V., Didierjean, J., Ehresmann, B., Ehresmann, C., Isel, C. and Marquet, R. (2006). Mg2+ dependency of HIV-1 reverse transcription, inhibition by nucleoside analogues and resistance. *Nucleic Acids Res* **34**, 42-52.

Granier, C., Aguirrezabal, L., Chenu, K., Cookson, S. J., Dauzat, M., Hamard, P., Thioux, J. J., Rolland, G., Bouchier-Combaud, S., Lebaudy, A. et al. (2006). PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in Arabidopsis thaliana permitted the identification of an accession with low sensitivity to soil water deficit. *New Phytol* **169**, 623-35.

Granier, C. and Tardieu, F. (2009). Multi-scale phenotyping of leaf expansion in response to environmental changes: the whole is more than the sum of parts. *Plant Cell Environ* **32**, 1175-84.

Greenup, A., Peacock, W. J., Dennis, E. S. and Trevaskis, B. (2009). The molecular biology of seasonal flowering-responses in Arabidopsis and the cereals. *Ann Bot* **103**, 1165-72.

Grennan, A. K. (2006). Variations on a Theme. Regulation of Flowering Time in Arabidopsis. *Plant Physiol* **140**, 399-400.

Grime, J. P. and Hunt, R. (1975). Relative Growth-Rate: Its Range and Adaptive Significance in a Local Flora. *Journal of Ecology* **63**, 393-422.

Grimes, J. A., Nielsen, S. J., Battaglioli, E., Miska, E. A., Speh, J. C., Berry, D. L., Atouf, F., Holdener, B. C., Mandel, G. and Kouzarides, T. (2000). The co-repressor mSin3A is a functional component of the REST-CoREST repressor complex. *J Biol Chem* **275**, 9461-7.

Gupta, P. K. and Rustgi, S. (2004). Molecular markers from the transcribed/expressed region of the genome in higher plants. *Funct Integr Genomics* **4**, 139-62.

Hakimi, M. A., Bochar, D. A., Chenoweth, J., Lane, W. S., Mandel, G. and Shiekhattar, R. (2002). A core-BRAF35 complex containing histone deacetylase mediates repression of neuronal-specific genes. *Proc Natl Acad Sci U S A* **99**, 7420-5.

Hall, S. E., Kettler, G. and Preuss, D. (2003). Centromere satellites from Arabidopsis populations: maintenance of conserved and variable domains. *Genome Res* **13**, 195-205.

Hanada, K., Zhang, X., Borevitz, J. O., Li, W. H. and Shiu, S. H. (2007). A large number of novel coding small open reading frames in the intergenic regions of the Arabidopsis thaliana genome are transcribed and/or under purifying selection. *Genome Res* **17**, 632-40.

Hasterok, R., Dulawa, J., Jenkins, G., Leggett, M. and Langdon, T. (2006). Multi-substrate chromosome preparations for high throughput comparative FISH. *BMC Biotechnology* **6**, 20.

He, Y. and Amasino, R. M. (2005a). Role of chromatin modification in flowering-time control. *Trends Plant Sci* **10**, 30-5.

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

He, Y., Doyle, M. R. and Amasino, R. M. (2004). PAF1-complex-mediated histone methylation of FLOWERING LOCUS C chromatin is required for the vernalization-responsive, winter-annual habit in Arabidopsis. *Genes Dev* **18**, 2774-84.

He, Y., Michaels, S. D. and Amasino, R. M. (2003). Regulation of flowering time by histone acetylation in Arabidopsis. *Science* **302**, 1751-4.

Henderson, I. R., Shindo, C. and Dean, C. (2003). The need for winter in the switch to flowering. *Annu Rev Genet* **37**, 371-92.

Hoffmann, W. A. P., Hendrik (2002). Avoiding Bias in Calculations of Relative Growth Rate. *Ann Bot* **90**, 37-42.

Hornyik, C., Duc, C., Rataj, K., Terzi, L. C. and Simpson, G. G. (2010). Alternative polyadenylation of antisense RNAs and flowering time control. *Biochem Soc Trans* **38**, 1077-81.

Huang, X., Paulo, M. J., Boer, M., Effgen, S., Keizer, P., Koornneef, M. and van Eeuwijk, F. A. (2011). Analysis of natural allelic variation in Arabidopsis using a multiparent recombinant inbred line population. *Proc Natl Acad Sci U S A* **108**, 4488-93.

Humphrey, G. W., Wang, Y., Russanova, V. R., Hirai, T., Qin, J., Nakatani, Y. and Howard, B. H. (2001). Stable histone deacetylase complexes distinguished by the presence of SANT domain proteins CoREST/kiaa0071 and Mta-L1. *J Biol Chem* **276**, 6817-24.

Hund, A., Trachsel, S. and Stamp, P. (2009). Growth of axile and lateral roots of maize: I development of a phenotying platform. *Plant and Soil* **325**, 335-349.

Jackson, M. B. (1980). Aeration in the nutrient film technique of glasshouse crop production and the importance of oxygen, ethylene and carbon dioxide. *Acta Hort.* **98**, 61-78.

Jansen, M., Gilmer, F., Biskup, B., Nagel, K. A., Rascher, U., Fischbach, A., Briem, S., Dreissen, G., Tittmann, S., Braun, S. et al. (2009). Simultaneous phenotyping of leaf growth and chlorophyll fluorescence via GROWSCREEN FLUORO allows detection of stress tolerance in Arabidopsis thaliana and other rosette plants. *Functional Plant Biology* **36**, 902-914.

Jansen, R. C. (1996a). A General Monte Carlo Method for Mapping Multiple Quantitative Trait Loci. *Genetics* **142**, 305-311.

Jansen, R. C. (1996b). A general Monte Carlo method for mapping multiple quantitative trait loci. *Genetics* **142**, 305-11.

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. *Theoretical and Applied Genetics* **91**, 33-37.

Jean Finnegan, E., Kovac, K. A., Jaligot, E., Sheldon, C. C., James Peacock, W. and Dennis, E. S. (2005). The downregulation of FLOWERING LOCUS C (FLC) expression in plants with low levels of DNA methylation and by vernalization occurs by distinct mechanisms. *The Plant Journal* **44**, 420-432.

Jensen, C. S., Salchert, K. and Nielsen, K. K. (2001). A TERMINAL FLOWER1-like gene from perennial ryegrass involved in floral transition and axillary meristem identity. *Plant Physiol* **125**, 1517-28.

Jiang, D., Yang, W., He, Y. and Amasino, R. M. (2007). Arabidopsis relatives of the human lysine-specific Demethylase1 repress the expression of FWA and FLOWERING LOCUS C and thus promote the floral transition. *Plant Cell* **19**, 2975-87.

Johanson, U., West, J., Lister, C., 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.

Karsai, I., Szucs, P., Meszaros, K., Filichkina, T., Hayes, P. M., Skinner, J. S., Lang, L. and Bedo, Z. (2005). The Vrn-H2 locus is a major determinant of flowering time in a facultative x winter growth habit barley (Hordeum vulgare L.) mapping population. *Theor Appl Genet* **110**, 1458-66.

Kearsey, M. J. and Farquhar, A. G. (1998). QTL analysis in plants; where are we now? *Heredity (Edinb)* **80 (Pt 2)**, 137-42.

Keddy, P. A. (2007). Plants and Vegetation. *Cambridge University Press*.

Kempel, A., Schadler, M., Chrobock, T., Fischer, M. and van Kleunen, M. (2011). Tradeoffs associated with constitutive and induced plant resistance against herbivory. *Proc Natl Acad Sci U S A* **108**, 5685-9.

Kikuchi, R., Kawahigashi, H., Ando, T., Tonooka, T. and Handa, H. (2009). Molecular and Functional Characterization of PEBP Genes in Barley Reveal the Diversification of Their Roles in Flowering. *Plant Physiol* **149**, 1341-1353.

Kim, D. H., Doyle, M. R., Sung, S. and Amasino, R. M. (2009). Vernalization: winter and the timing of flowering in plants. *Annu Rev Cell Dev Biol* **25**, 277-99.

Kim, S. Y., He, Y., Jacob, Y., Noh, Y. S., Michaels, S. and Amasino, R. (2005). Establishment of the vernalization-responsive, winter-annual habit in Arabidopsis requires a putative histone H3 methyl transferase. *Plant Cell* **17**, 3301-10.

Ko, J. H. and Han, K. H. (2004). Arabidopsis whole-transcriptome profiling defines the features of coordinated regulations that occur during secondary growth. *Plant Mol Biol* **55**, 433-53.

Kobayashi, Y. and Weigel, D. (2007). Move on up, it's time for change--mobile signals controlling photoperiod-dependent flowering. *Genes Dev* **21**, 2371-84.

Koornneef, M., Alonso-Blanco, C., Peeters, A. J. and Soppe, W. (1998). Genetic Control of Flowering Time in Arabidopsis. *Annu Rev Plant Physiol Plant Mol Biol* **49**, 345-370.

Koornneef, M., Alonso-Blanco, C. and Vreugdenhil, D. (2004). Naturally occurring genetic variation in Arabidopsis thaliana. *Annu Rev Plant Biol* **55**, 141-72.

Koornneef, M., Blankestijn-de Vries, H., Hanhart, C., Soppe, W. and Peeters, T. (1994). The phenotype of some late-flowering mutants is enhanced by a locus on chromosome 5 that is not effective in the Landsberg erecta wild-type. *The Plant Journal* **6**, 911-919.

Korves, T. M., Schmid, K. J., Caicedo, A. L., Mays, C., Stinchcombe, J. R., Purugganan, M. D. and Schmitt, J. (2007). Fitness effects associated with the major flowering time gene FRIGIDA in Arabidopsis thaliana in the field. *Am Nat* **169**, E141-57.

Krichevsky, A., Gutgarts, H., Kozlovsky, S. V., Tzfira, T., Sutton, A., Sternglanz, R., Mandel, G. and Citovsky, V. (2007). C2H2 zinc finger-SET histone methyltransferase is a plant-specific chromatin modifier. *Dev Biol* **303**, 259-69.

Kusnierczyk, A., Winge, P., Midelfart, H., Armbruster, W. S., Rossiter, J. T. and Bones, A. M. (2007). Transcriptional responses of Arabidopsis thaliana ecotypes with different glucosinolate profiles after attack by polyphagous Myzus persicae and oligophagous Brevicoryne brassicae. *J Exp Bot* **58**, 2537-52.

Laurie, D. A., Pratchett, N., Snape, J. W. and Bezant, J. H. (1995). RFLP mapping of five major genes and eight quantitative trait loci controlling flowering time in a winter x spring barley (Hordeum vulgare L.) cross. *Genome* **38**, 575-85.

Lee, I., Aukerman, M. J., Gore, S. L., Lohman, K. N., Michaels, S. D., Weaver, L. M., John, M. C., Feldmann, K. A. and Amasino, R. M. (1994a). Isolation of LUMINIDEPENDENS: a gene involved in the control of flowering time in Arabidopsis. *Plant Cell* **6**, 75-83.

Lee, I., Michaels, S. D., Masshardt, A. S. and Amasino, R. M. (1994b). The late-flowering phenotype of FRIGIDA and mutations in LUMINIDEPENDENS is suppressed in the Landsberg erecta strain of Arabidopsis. *The Plant Journal* **6**, 903-909.

Lee, M. G., Wynder, C., Bochar, D. A., Hakimi, M.-A., Cooch, N. and Shiekhattar, R. (2006). Functional Interplay between Histone Demethylase and Deacetylase Enzymes. *Molecular and Cellular Biology* **26**, 6395-6402.

Lee, M. G., Wynder, C., Cooch, N. and Shiekhattar, R. (2005). An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. *Nature* **437**, 432-5.

Leister, D., Varotto, C., Pesaresi, P., Niwergall, A. and Salamini, F. (1999). Large-scale evaluation of plant growth in Arabidopsis thaliana by non-invasive image analysis. *Plant Physiology and Biochemistry* **37**, 671-678.

Levy, Y. Y. and Dean, C. (1998). The transition to flowering. *Plant Cell* **10**, 1973-90.

Levy, Y. Y., Mesnage, S., Mylne, J. S., Gendall, A. R. and Dean, C. (2002). Multiple roles of Arabidopsis VRN1 in vernalization and flowering time control. *Science* **297**, 243-6.

Li, B., Suzuki, J.-I. and Hara, T. (1998). Latitudinal variation in plant size and relative growth rate in Arabidopsis thaliana. *Oecologia* **115**, 293-301.

Li, Y., Huang, Y., Bergelson, J., Nordborg, M. and Borevitz, J. O. (2010). Association mapping of local climate-sensitive quantitative trait loci in Arabidopsis thaliana. *Proc Natl Acad Sci U S A* **107**, 21199-204.

Lim, M.-H., Kim, J., Kim, Y.-S., Chung, K.-S., Seo, Y.-H., Lee, I., Kim, J., Hong, C. B., Kim, H.-J. and Park, C.-M. (2004). A New Arabidopsis Gene, FLK, Encodes an RNA Binding Protein with K Homology Motifs and Regulates Flowering Time via FLOWERING LOCUS C. *The Plant Cell Online* **16**, 731-740.

Lim, P. N. and Hudson, A. (2013). How does natural variotion play a role in adaptation. *PhD thesis submitted toThe University of Edinburgh*.

Lin, S.-I., Wang, J.-G., Poon, S.-Y., Su, C.-l., Wang, S.-S. and Chiou, T.-J. (2005). Differential Regulation of FLOWERING LOCUS C Expression by Vernalization in Cabbage and Arabidopsis. *Plant Physiol* **137**, 1037-1048.

Liu, C., Lu, F., Cui, X. and Cao, X. (2010). Histone methylation in higher plants. *Annu Rev Plant Biol* **61**, 395-420.

Liu, F., Quesada, V., Crevillen, P., Baurle, I., Swiezewski, S. and Dean, C. (2007). The Arabidopsis RNA-binding protein FCA requires a lysine-specific demethylase 1 homolog to downregulate FLC. *Mol Cell* **28**, 398-407.

Llorente, F., Alonso-Blanco, C., Sánchez-Rodriguez, C., Jorda, L. and Molina, A. (2005). ERECTA receptor-like kinase and heterotrimeric G protein from Arabidopsis are required for resistance to the necrotrophic fungus Plectosphaerella cucumerina. *The Plant Journal* **43**, 165-180.

Lorenc-Kukula, K., Jafra, S., Oszmianski, J. and Szopa, J. (2005). Ectopic expression of anthocyanin 5-o-glucosyltransferase in potato tuber causes increased resistance to bacteria. *J Agric Food Chem* **53**, 272-81.

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

Loudet, O., Chaillou, S., Krapp, A. and Daniel-Vedele, F. (2003a). Quantitative trait loci analysis of water and anion contents in interaction with nitrogen availability in Arabidopsis thaliana. *Genetics* **163**, 711-22.

Loudet, O., Chaillou, S., Merigout, P., Talbotec, J. and Daniel-Vedele, F. (2003b). Quantitative trait loci analysis of nitrogen use efficiency in Arabidopsis. *Plant Physiol* **131**, 345-58.

Loudet, O., Gaudon, V., Trubuil, A. and Daniel-Vedele, F. (2005). Quantitative trait loci controlling root growth and architecture in Arabidopsis thaliana confirmed by heterogeneous inbred family. *Theor Appl Genet* **110**, 742-53.

Lynch, J., Läuchli, A. and Epstein, E. (1991). Vegetative Growth of the Common Bean in Response to Phosphorus Nutrition. *Crop Sci.* **31**, 380-387.

Macknight, R., Bancroft, I., Page, T., Lister, C., Schmidt, R., Love, K., Westphal, L., Murphy, G., Sherson, S., Cobbett, C. et al. (1997). FCA, a gene controlling flowering time in Arabidopsis, encodes a protein containing RNA-binding domains. *Cell* **89**, 737-45.

Maloof, J. N., Borevitz, J. O., Dabi, T., Lutes, J., Nehring, R. B., Redfern, J. L., Trainer, G. T., Wilson, J. M., Asami, T., Berry, C. C. et al. (2001). Natural variation in light sensitivity of Arabidopsis. *Nat Genet* **29**, 441-446.

March-Diaz, R., Garcia-Dominguez, M., Florencio, F. J. and Reyes, J. C. (2007). SEF, a new protein required for flowering repression in Arabidopsis, interacts with PIE1 and ARP6. *Plant Physiol* **143**, 893-901.

Martin, C. and Zhang, Y. (2005). The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol* **6**, 838-849.

Masle, J., Gilmore, S. R. and Farquhar, G. D. (2005). The ERECTA gene regulates plant transpiration efficiency in Arabidopsis. *Nature* **436**, 866-870.

Massonnet, C., Vile, D., Fabre, J., Hannah, M. A., Caldana, C., Lisec, J., Beemster, G. T., Meyer, R. C., Messerli, G., Gronlund, J. T. et al. (2010). Probing the reproducibility of leaf growth and molecular phenotypes: a comparison of three Arabidopsis accessions cultivated in ten laboratories. *Plant Physiol* **152**, 2142-57.

Maue, R. A., Kraner, S. D., Goodman, R. H. and Mandel, G. (1990). Neuron-specific expression of the rat brain type II sodium channel gene is directed by upstream regulatory elements. *Neuron* **4**, 223-31.

McCulloch, H. L. and Hudson, A. (2011). The Genetic Basis for Natural Variation in Flowering Time in Local Populations of Arabidopsis thaliana. *PhD thesis submitted to The university of Edinburgh*.

McKay, J. K., Richards, J. H. and Mitchell-Olds, T. (2003). Genetics of drought adaptation in Arabidopsis thaliana: I. Pleiotropy contributes to genetic correlations among ecological traits. *Mol Ecol* **12**, 1137-51.

Mejía, N., Gebauer, M., Muñoz, L., Hewstone, N., Muñoz, C. and Hinrichsen, P. (2007). Identification of QTLs for Seedlessness, Berry Size, and Ripening Date in a Seedless x Seedless Table Grape Progeny. *American Journal of Enology and Viticulture* **58**, 499-507.

Melzer, S., Lens, F., Gennen, J., Vanneste, S., Rohde, A. and Beeckman, T. (2008). Flowering-time genes modulate meristem determinacy and growth form in Arabidopsis thaliana. *Nat Genet* **40**, 1489-92.

Mendez-Vigo, B., de Andres, M. T., Ramiro, M., Martinez-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**, 1611-23.

Menz, M. A., Klein, R. R., Mullet, J. E., Obert, J. A., Unruh, N. C. and Klein, P. E. (2002). A high-density genetic map of Sorghum bicolor (L.) Moench based on 2926 AFLP, RFLP and SSR markers. *Plant Mol Biol* **48**, 483-99.

Metzger, E., Wissmann, M., Yin, N., Muller, J. M., Schneider, R., Peters, A. H., Gunther, T., Buettner, R. and Schule, R. (2005). LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* **437**, 436-9.

Mewis, I., Appel, H. M., Hom, A., Raina, R. and Schultz, J. C. (2005). Major Signaling Pathways Modulate Arabidopsis Glucosinolate Accumulation and Response to Both Phloem-Feeding and Chewing Insects. *Plant Physiol* **138**, 1149-1162.

Michaels, S. D. and Amasino, R. M. (1999a). FLOWERING LOCUS C Encodes a Novel MADS Domain Protein That Acts as a Repressor of Flowering. *The Plant Cell Online* **11**, 949-956.

Michaels, S. D. and Amasino, R. M. (1999b). FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* **11**, 949-56.

Michaels, S. 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. *Proceedings of the National Academy of Sciences* **100**, 10102-10107.

Milla, R., Escudero, A. and Iriondo, J. M. (2009). Inherited variability in multiple traits determines fitness in populations of an annual legume from contrasting latitudinal origins. *Ann Bot* **103**, 1279-89.

Miller, N. D., Durham Brooks, T. L., Assadi, A. H. and Spalding, E. P. (2010). Detection of a gravitropism phenotype in glutamate receptor-like 3.3 mutants of Arabidopsis thaliana using machine vision and computation. *Genetics* **186**, 585-93.

Miller, N. D., Parks, B. M. and Spalding, E. P. (2007). Computer-vision analysis of seedling responses to light and gravity. *Plant J* **52**, 374-81.

Mitchell-Olds, T. and Pedersen, D. (1998). The molecular basis of quantitative genetic variation in central and secondary metabolism in Arabidopsis. *Genetics* **149**, 739-47.

Montes, J. M., Melchinger, A. E. and Reif, J. C. (2007). Novel throughput phenotyping platforms in plant genetic studies. *Trends Plant Sci* **12**, 433-6.

Motooka S, H. T., Mima Y, Konishi K. . (1991). Measurement of in-vitro plant growth by image processing. *J Jpn Soc Hortic Sci* **60**.

Mouchel, C. F., Briggs, G. C. and Hardtke, C. S. (2004). Natural genetic variation in Arabidopsis identifies BREVIS RADIX, a novel regulator of cell proliferation and elongation in the root. *Genes Dev* **18**, 700-14.

Mouradov, A., Cremer, F. and Coupland, G. (2002). Control of flowering time: interacting pathways as a basis for diversity. *Plant Cell* **14 Suppl**, S111-30.

Mousseau, T. A. and Fox, C. W. (1998). The adaptive significance of maternal effects. *Trends in Ecology & Evolution* **13**, 403-407.

Murray, S. C., Rooney, W. L., Mitchell, S. E., Sharma, A., Klein, P. E., Mullet, J. E. and Kresovich, S. (2008). Genetic Improvement of Sorghum as a Biofuel Feedstock: II. QTL for Stem and Leaf Structural Carbohydrates All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Permission for printing and for reprinting the material contained herein has been obtained by the publisher. *Crop Sci.* **48**, 2180-2193.

Nagel, M., Luu, O., Bisson, N., Macanovic, B., Moss, T. and Winklbauer, R. (2009). Role of p21-activated kinase in cell polarity and directional mesendoderm migration in the Xenopus gastrula. *Dev Dyn* **238**, 1709-26.

Ng, H. H., Robert, F., Young, R. A. and Struhl, K. (2003). Targeted Recruitment of Set1 Histone Methylase by Elongating Pol II Provides a Localized Mark and Memory of Recent Transcriptional Activity. *Molecular cell* **11**, 709-719.

Nilsson, H.-E. (1995). Remote Sensing and Image Analysis in Plant Pathology. *Annual Review of Phytopathology* **33**, 489-528.

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

Ögren, E. (1999). Fall frost resistance in willows used for biomass production. II. Predictive relationships with sugar concentration and dry matter content. *Tree Physiology* **19**, 755-760.

Oh, S., Zhang, H., Ludwig, P. and van Nocker, S. (2004). A mechanism related to the yeast transcriptional regulator Paf1c is required for expression of the Arabidopsis FLC/MAF MADS box gene family. *Plant Cell* **16**, 2940-53.

Olsen, P., Lenk, I., Jensen, C. S., Petersen, K., Andersen, C. H., Didion, T. and Nielsen, K. K. (2006). Analysis of two heterologous flowering genes in Brachypodium distachyon demonstrates its potential as a grass model plant. *Plant Science* **170**, 1020-1025.

Osborn, T. C., Kole, C., Parkin, I. A., Sharpe, A. G., Kuiper, M., Lydiate, D. J. and Trick, M. (1997). Comparison of flowering time genes in Brassica rapa, B. napus and Arabidopsis thaliana. *Genetics* **146**, 1123-9.

Patel, D., Basu, M., Hayes, S., Majláth, I., Hetherington, F. M., Tschaplinski, T. J. and Franklin, K. A. (2013). Temperature-dependent shade avoidance involves the receptor-like kinase ERECTA. *The Plant Journal* **73**, 980-992.

Paul, M. J. and Foyer, C. H. (2001). Sink regulation of photosynthesis. *J Exp Bot* **52**, 1383-400.

Pelletier, M. K., Murrell, J. R. and Shirley, B. W. (1997). Characterization of flavonol synthase and leucoanthocyanidin dioxygenase genes in Arabidopsis. Further evidence for differential regulation of "early" and "late" genes. *Plant Physiol* **113**, 1437-45.

Pellis, A., Laureysens, I. and Ceulemans, R. (2004). Genetic Variation of the Bud and Leaf Phenology of Seventeen Poplar Clones in a Short Rotation Coppice Culture. *Plant Biology* **6**, 38-46.

Pereyra-Irujo, G. A., Velazquez, L., Lechner, L. and Aguirrezabal, L. A. (2008). Genetic variability for leaf growth rate and duration under water deficit in sunflower: analysis of responses at cell, organ, and plant level. *J Exp Bot* **59**, 2221-32.

Pico, F. X., Mendez-Vigo, B., Martinez-Zapater, J. M. and Alonso-Blanco, C. (2008). Natural genetic variation of Arabidopsis thaliana is geographically structured in the Iberian peninsula. *Genetics* **180**, 1009-21.

Pien, S., Fleury, D., Mylne, J. S., Crevillen, P., Inze, D., Avramova, Z., Dean, C. and Grossniklaus, U. (2008). ARABIDOPSIS TRITHORAX1 dynamically regulates FLOWERING LOCUS C activation via histone 3 lysine 4 trimethylation. *Plant Cell* **20**, 580-8.

Platt, A., Horton, M., Huang, Y. S., Li, Y., Anastasio, A. E., Mulyati, N. W., Agren, J., Bossdorf, O., Byers, D., Donohue, K. et al. (2010). The scale of population structure in Arabidopsis thaliana. *PLoS Genet* **6**, e1000843.

Poorter, H. and Remkes, C. (1990). LEAF-AREA RATIO AND NET ASSIMILATION RATE OF 24 WILD-SPECIES DIFFERING IN RELATIVE GROWTH-RATE. *Oecologia* **83**, 553-559.

Poorter, H., van Rijn, C. P., Vanhala, T. K., Verhoeven, K. J., de Jong, Y. E., Stam, P. and Lambers, H. (2005). A genetic analysis of relative growth rate and underlying components in Hordeum spontaneum. *Oecologia* **142**, 360-77.

Poorter, H. L. a. H. (1992). Inherent Variation in Growth Rate Between Higher Plants: A Search for Physiological Causes and Ecological Consequences. *ADV ECOL RES* **23**, 187-261

Quesada, V., Macknight, R., Dean, C. and Simpson, G. G. (2003). Autoregulation of FCA pre-mRNA processing controls Arabidopsis flowering time. *EMBO J* **22**, 3142-52.

191

Rajendran, K., Tester, M. and Roy, S. J. (2009). Quantifying the three main components of salinity tolerance in cereals. *Plant Cell Environ* **32**, 237-49.

Ratcliffe, O. J., Kumimoto, R. W., Wong, B. J. and Riechmann, J. L. (2003). Analysis of the Arabidopsis MADS AFFECTING FLOWERING gene family: MAF2 prevents vernalization by short periods of cold. *Plant Cell* **15**, 1159-69.

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**, 122-32.

Rauh, L., Basten, C. and Buckler, S. t. (2002). Quantitative trait loci analysis of growth response to varying nitrogen sources in Arabidopsis thaliana. *Theor Appl Genet* **104**, 743-750.

Reymond, M., Svistoonoff, S., Loudet, O., Nussaume, L. and Desnos, T. (2006). Identification of QTL controlling root growth response to phosphate starvation in Arabidopsis thaliana. *Plant Cell Environ* **29**, 115-25.

Reymond, P., Bodenhausen, N., Van Poecke, R. M. P., Krishnamurthy, V., Dicke, M. and Farmer, E. E. (2004). A Conserved Transcript Pattern in Response to a Specialist and a Generalist Herbivore. *The Plant Cell Online* **16**, 3132-3147.

Ritter, K., Jordan, D., Chapman, S., Godwin, I., Mace, E. and Lynne McIntyre, C. (2008). Identification of QTL for sugar-related traits in a sweet × grain sorghum (Sorghum bicolor L. Moench) recombinant inbred population. *Molecular Breeding* **22**, 367-384.

Rivera-Pomar, R., Niessing, D., Schmidt-Ott, U., Gehring, W. J. and Jackle, H. (1996). RNA binding and translational suppression by bicoid. *Nature* **379**, 746-749.

Robertson A (1967). The nature of quantitative genetic variation. *Brink RA (ed) Heritage from Mendel. University of Wisconsin Press, Madison, WI*, 265–280.

Roopra, A., Qazi, R., Schoenike, B., Daley, T. J. and Morrison, J. F. (2004). Localized domains of G9a-mediated histone methylation are required for silencing of neuronal genes. *Mol Cell* **14**, 727-38.

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**, 263-76.

Ruiz D, L. P., Audergon JM, Dondini L, Tartarn S, Adam M, Gennar F, Cervellati C, De Fraanceschi P, Sansavn S, Bureau S, Gouble B, Reich M, Renard CMGC, Bass D, Testoln R. International Symposium on Apricot Breeding and culture. (2010). Identification of QTLs for fruit quality traits in apricot. *ISHS Acta Horticulturae* **862**.

Salome, P. A., Bomblies, K., Laitinen, R. A., Yant, L., Mott, R. and Weigel, D. (2011). Genetic architecture of flowering-time variation in Arabidopsis thaliana. *Genetics* **188**, 421-33.

Santiago, N., Herráiz, C., Goñi, J. R., Messeguer, X. and Casacuberta, J. M. (2002). Genome-wide Analysis of the Emigrant Family of MITEs of Arabidopsis thaliana. *Molecular Biology and Evolution* **19**, 2285-2293.

Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C. T., Schreiber, S. L., Mellor, J. and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. *Nature* **419**, 407-411.

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

Schlappi, M. R. (2006). FRIGIDA LIKE 2 is a functional allele in Landsberg erecta and compensates for a nonsense allele of FRIGIDA LIKE 1. *Plant Physiol* **142**, 1728-38.

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**, 1250-7.

Schneider, R., Bannister, A. J., Myers, F. A., Thorne, A. W., Crane-Robinson, C. and Kouzarides, T. (2004). Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nat Cell Biol* **6**, 73-77.

Schomburg, F. M., Patton, D. A., Meinke, D. W. and Amasino, R. M. (2001a). FPA, a Gene Involved in Floral Induction in Arabidopsis, Encodes a Protein Containing RNA-Recognition Motifs. *The Plant Cell Online* **13**, 1427-1436.

Schomburg, F. M., Patton, D. A., Meinke, D. W. and Amasino, R. M. (2001b). FPA, a gene involved in floral induction in Arabidopsis, encodes a protein containing RNA-recognition motifs. *Plant Cell* **13**, 1427-36.

Schonrock, N., Bouveret, R., Leroy, O., Borghi, L., Kohler, C., Gruissem, W. and Hennig, L. (2006). Polycomb-group proteins repress the floral activator AGL19 in the FLC-independent vernalization pathway. *Genes Dev* **20**, 1667-78.

Schranz, M. E., Quijada, P., Sung, S. B., Lukens, L., Amasino, R. and Osborn, T. C. (2002). Characterization and effects of the replicated flowering time gene FLC in Brassica rapa. *Genetics* **162**, 1457-68.

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

Shannon, S. and Meeks-Wagner, D. R. (1991). A Mutation in the Arabidopsis TFL1 Gene Affects Inflorescence Meristem Development. *Plant Cell* **3**, 877-892.

Sharbel, T. F., Haubold, B. and Mitchell-Olds, T. (2000). Genetic isolation by distance in Arabidopsis thaliana: biogeography and postglacial colonization of Europe. *Mol Ecol* **9**, 2109-18.

Sheldon, C. C., Burn, J. E., Perez, P. P., Metzger, J., Edwards, J. A., Peacock, W. J. and Dennis, E. S. (1999a). The FLF MADS box gene: a repressor of flowering in Arabidopsis regulated by vernalization and methylation. *Plant Cell* **11**, 445-58.

Sheldon, C. C., Burn, J. E., Perez, P. P., Metzger, J., Edwards, J. A., Peacock, W. J. and Dennis, E. S. (1999b). The FLF MADS Box Gene: A Repressor of Flowering in Arabidopsis Regulated by Vernalization and Methylation. *The Plant Cell Online* **11**, 445-458.

Sheldon, C. C., Conn, A. B., Dennis, E. S. and Peacock, W. J. (2002). Different Regulatory Regions Are Required for the Vernalization-Induced Repression of FLOWERING LOCUS C and for the Epigenetic Maintenance of Repression. *The Plant Cell Online* **14**, 2527-2537.

Sheldon, C. C., Hills, M. J., Lister, C., Dean, C., Dennis, E. S. and Peacock, W. J. (2008). Resetting of FLOWERING LOCUS C expression after epigenetic repression by vernalization. *Proceedings of the National Academy of Sciences* **105**, 2214-2219.

Sheldon, C. C., Rouse, D. T., Finnegan, E. J., Peacock, W. J. and Dennis, E. S. (2000). The molecular basis of vernalization: The central role of FLOWERING LOCUS C (FLC). *Proceedings of the National Academy of Sciences* **97**, 3753-3758.

Shi, Y.-J., Matson, C., Lan, F., Iwase, S., Baba, T. and Shi, Y. (2005). Regulation of LSD1 Histone Demethylase Activity by Its Associated Factors. *Molecular cell* **19**, 857-864.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A. and Casero, R. A. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941-53.

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**, 3079-83.

Sibout, R., Plantegenet, S. and Hardtke, C. S. (2008). Flowering as a condition for xylem expansion in Arabidopsis hypocotyl and root. *Curr Biol* **18**, 458-63.

Simpson, G. G. (2004). The autonomous pathway: epigenetic and post-transcriptional gene regulation in the control of Arabidopsis flowering time. *Curr Opin Plant Biol* **7**, 570-4.

Simpson, G. G. and Dean, C. (2002). Arabidopsis, the Rosetta stone of flowering time? *Science* **296**, 285-9.

Singh, S., Singh, A., Roy, S. and Sarkar, A. K. (2012). SWP1 negatively regulates lateral root initiation and elongation in *Arabidopsis*. *Plant Signal Behav* **7**, 1522-1525.

Sironi, M., Menozzi, G., Comi, G. P., Cagliani, R., Bresolin, N. and Pozzoli, U. (2005). Analysis of intronic conserved elements indicates that functional complexity might represent a major source of negative selection on non-coding sequences. *Hum Mol Genet* **14**, 2533-46.

Snoeren, T. L., Mumm, R., Poelman, E., Yang, Y., Pichersky, E. and Dicke, M. (2010). The Herbivore-Induced Plant Volatile Methyl Salicylate Negatively Affects Attraction of the Parasitoid Diadegma semiclausum. *Journal of Chemical Ecology* **36**, 479-489.

Sokal, R. R. and Rohlf, F. J. (1995). Biometry: W. H. Freeman.

Steppuhn, A., Schuman, M. C. and Baldwin, I. T. (2008). Silencing jasmonate signalling and jasmonate-mediated defences reveals different survival strategies between two Nicotiana attenuata accessions. *Mol Ecol* **17**, 3717-3732.

Strange, A., Li, P., Lister, C., Anderson, J., Warthmann, N., Shindo, C., Irwin, J., Nordborg, M. and Dean, C. (2011). Major-effect alleles at relatively few loci underlie distinct vernalization and flowering variation in Arabidopsis accessions. *PLoS One* **6**, e19949.

Stratton, D. A. (1998). Reaction norm functions and QTL-environment interactions for flowering time in Arabidopsis thaliana. *Heredity (Edinb)* **81 (Pt 2)**, 144-55.

Strauss, S. Y. and Agrawal, A. A. (1999). The ecology and evolution of plant tolerance to herbivory. *Trends Ecol Evol* **14**, 179-185.

Suk, J., Kim, S. and Ryoo, I. (2011). Non-contact plant growth measurement method and system based on ubiquitous sensor network technologies. *Sensors (Basel)* **11**, 4312-34.

Sung, S. and Amasino, R. M. (2005). Remembering winter: toward a molecular understanding of vernalization. *Annu Rev Plant Biol* **56**, 491-508.

Sung, S., Schmitz, R. J. and Amasino, R. M. (2006). A PHD finger protein involved in both the vernalization and photoperiod pathways in Arabidopsis. *Genes Dev* **20**, 3244-8.

Tachibana, M., Sugimoto, K., Fukushima, T. and Shinkai, Y. (2001). Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. *J Biol Chem* **276**, 25309-17.

Tadege, M., Sheldon, C. C., Helliwell, C. A., Stoutjesdijk, P., Dennis, E. S. and Peacock, W. J. (2001). Control of flowering time by FLC orthologues in Brassica napus. *Plant J* **28**, 545-53.

Theologis, A., Ecker, J. R., Palm, C. J., Federspiel, N. A., Kaul, S., White, O., Alonso, J., Altafi, H., Araujo, R., Bowman, C. L. et al. (2000). Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana. *Nature* **408**, 816-20.

Tian, D., Araki, H., Stahl, E., Bergelson, J. and Kreitman, M. (2002). Signature of balancing selection in Arabidopsis. *Proc Natl Acad Sci U S A* **99**, 11525-30.

Tisne, S., Reymond, M., Vile, D., Fabre, J., Dauzat, M., Koornneef, M. and Granier, C. (2008). Combined genetic and modeling approaches reveal that epidermal cell area and number in leaves are controlled by leaf and plant developmental processes in Arabidopsis. *Plant Physiol* **148**, 1117-27.

Tochio, N., Umehara, T., Koshiba, S., Inoue, M., Yabuki, T., Aoki, M., Seki, E., Watanabe, S., Tomo, Y., Hanada, M. et al. (2006). Solution structure of the SWIRM domain of human histone demethylase LSD1. *Structure* **14**, 457-68.

Todesco, M., Balasubramanian, S., Hu, T. T., Traw, M. B., Horton, M., Epple, P., Kuhns, C., Sureshkumar, S., Schwartz, C., Lanz, C. et al. (2010). Natural allelic variation underlying a major fitness trade-off in Arabidopsis thaliana. *Nature* **465**, 632-6.

Tsarouhas, V., Gullberg, U. and Lagercrantz, U. (2003). Mapping of quantitative trait loci controlling timing of bud flush in Salix. *Hereditas* **138**, 172-178.

Tucker, C. and Avila-Sakar, G. (2010). Ontogenetic changes in tolerance to herbivory in Arabidopsis. *Oecologia* **164**, 1005-15.

Turck, F., Fornara, F. and Coupland, G. (2008). Regulation and identity of florigen: FLOWERING LOCUS T moves center stage. *Annu Rev Plant Biol* **59**, 573-94.

Ullrich, H., Lattig, K., Brennicke, A. and Knoop, V. (1997). Mitochondrial DNA variations and nuclear RFLPs reflect different genetic similarities among 23 Arabidopsis thaliana ecotypes. *Plant Mol Biol* **33**, 37-45.

Ungerer, M. C., Halldorsdottir, S. S., Modliszewski, J. L., Mackay, T. F. and Purugganan, M. D. (2002). Quantitative trait loci for inflorescence development in Arabidopsis thaliana. *Genetics* **160**, 1133-51.

Val, E. and Dirzo, R. (2003). Does ontogeny cause changes in the defensive strategies of the myrmecophyte Cecropia peltata? *Plant Ecology* **169**, 35-41.

Valverde, P. L., Fornoni, J. and Núñez-Farfán, J. (2003). Evolutionary ecology of Datura stramonium: equal plant fitness benefits of growth and resistance against herbivory. *Journal of Evolutionary Biology* **16**, 127-137.

Vander Zwan, C., Brodie, S. A. and Campanella, J. J. (2000). The Intraspecific Phylogenetics of Arabidopsis thaliana in Worldwide Populations. *Systematic Botany* **25**, 47-59.

Vlad, D., Rappaport, F., Simon, M. and Loudet, O. (2010). Gene transposition causing natural variation for growth in Arabidopsis thaliana. *PLoS Genet* **6**, e1000945.

Walhout, A. J., Temple, G. F., Brasch, M. A., Hartley, J. L., Lorson, M. A., van den Heuvel, S. and Vidal, M. (2000). GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. *Methods Enzymol* **328**, 575-92.

Walter, A., Scharr, H., Gilmer, F., Zierer, R., Nagel, K. A., Ernst, M., Wiese, A., Virnich, O., Christ, M. M., Uhlig, B. et al. (2007). Dynamics of seedling growth acclimation towards altered light conditions can be quantified via GROWSCREEN: a setup and procedure designed for rapid optical phenotyping of different plant species. *New Phytol* **174**, 447-55.

Walter, A., Studer, B. and Kolliker, R. (2012). Advanced phenotyping offers opportunities for improved breeding of forage and turf species. *Ann Bot* **110**, 1271-9.

Weih, M. (2009). Genetic and environmental variation in spring and autumn phenology of biomass willows (Salix spp.): effects on shoot growth and nitrogen economy. *Tree Physiology*.

Weinig, C., Stinchcombe, J. R. and Schmitt, J. (2003). Evolutionary genetics of resistance and tolerance to natural herbivory in Arabidopsis thaliana. *Evolution* **57**, 1270-80.

Weinig, C., Ungerer, M. C., Dorn, L. A., Kane, N. C., Toyonaga, Y., Halldorsdottir, S. S., Mackay, T. F., Purugganan, M. D. and Schmitt, J. (2002). Novel loci control variation in reproductive timing in Arabidopsis thaliana in natural environments. *Genetics* **162**, 1875-84.

Werner, J. D., Borevitz, J. O., Uhlenhaut, N. H., Ecker, J. R., Chory, J. and Weigel, D. (2005a). 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. (2005b). Quantitative trait locus mapping and DNA array hybridization identify an FLM deletion as a cause for natural flowering-time variation. *Proc Natl Acad Sci U S A* **102**, 2460-5.

Wood, A. J. and Roper, J. (2000). A Simple & Nondestructive Technique for Measuring Plant Growth & Development. *The American Biology Teacher* **62**, 215-217.

Zeng, Z. B. (1994). Precision mapping of quantitative trait loci. *Genetics* **136**, 1457-68.

Zhang, H., Ransom, C., Ludwig, P. and van Nocker, S. (2003). Genetic analysis of early flowering mutants in Arabidopsis defines a class of pleiotropic developmental regulator required for expression of the flowering-time switch flowering locus C. *Genetics* **164**, 347-58.

Zhang, H. and van Nocker, S. (2002). The VERNALIZATION INDEPENDENCE 4 gene encodes a novel regulator of FLOWERING LOCUS C. *Plant J* **31**, 663-73.

Zhang, X., Hause, R. J., Jr. and Borevitz, J. O. (2012). Natural Genetic Variation for Growth and Development Revealed by High-Throughput Phenotyping in Arabidopsis thaliana. *G3 (Bethesda)* **2**, 29-34.

Zhao, Z., Yu, Y., Meyer, D., Wu, C. and Shen, W. H. (2005). Prevention of early flowering by expression of FLOWERING LOCUS C requires methylation of histone H3 K36. *Nat Cell Biol* **7**, 1256-60.

Appendix A

Primers used for sequencing *VIP5*

Name	Sequence	Tm (^{0}C)
VIP5-F1	GTGACAGTGAAAACCATTTAAATTC	50
VIP5-R1	TGACCTGTCATCACCTGTCTAGCA	57
VIP5-F2	GATTTCTCATCAACGAAGAGGAAAC	54
VIP5-R2	TTTTACCTCAGATGCGTTCTTGAAA	52
VIP5-F3	GTAGAGAGGATCAAGTCGAAAATCA	54
VIP5-R3	AGAGTAGCAAAACTGCAAAAGACTTG	54

Primers used for sequencing *LDL1*

Name	Sequence	Tm (^{0}C)
LDL-F1	ACTCACATCATCACACACAGAGCTA	55
LDL1-R1	CGCCCACCCGGTCTATCTCTACCTT	60
LDL-F2	GTCGTTTGATGGCGTAGAGCCGCCA	60
LDL1-R2	AGCTACAAGTGCTACAAGTAATGGA	54
LDL-F3	TCCGTGTAACTTCTGGGGCGAAGAG	60
LDL1-R3	AATAAGAAAAGTAAACTTCACTTCA	47
FN-1	TCATCGACTGTCTAAGCTTACAAACT	63
RN-1	TTATTAAAGAGGCGAAATTGAGGTCG	63
FN-2	TACCTTCAAGAACCAAAACCCTAAAC	63
RN-2	TCACTGAGATTAACGCAAACCCAGCT	66
FN-3	TAGCTTCAATTTCTTCCTCCGTGAGC	66
FN-4	CTTCTTCAATATTTTGCTCCGATTGG	63

RN-4	TGACATAAAGATATAATCACTTATTT	55
NF4-GP-580	TTACTACTTTAAAAACAAATCACGGG	60
NF4-GP-980	GGTTTGTTATTTGTTAGGACTATTGC	62
11C1-F-3610	ATCGAC AGATCT TTGATTAGTAGT AA	60
11-AH-F	TGTCGTTAAATGATTTTGGTAACCT	59

Primers for expression analysis of LDL1

Name	Sequence	Tm (^{0}C)
F	GAGGAGGAGGAAGAGGAAGAGGCTGA	62
R	ACTACTGTGGATGTACCCGGTGAAGA	62

pJET primers

Name	Sequence	Tm (^{0}C)
pJET1.2-F	5'-CGACTCACTATAGGGAGAGCGGC-3'	62
pJET1.2-R	5'-AAGAACATCGATTTTCCATGGCAG-3,	62

Appendix B

```
Col.0       ATGGGTGATTTAGAGAACTTGCTTTTGGAAGCTGCTGGGAGAACAAATTCAGCAGGGAGG
VIP5-11C1   ------------------------------------------------------------
VIP5-4D1    ------------------------------------------------------------

Col.0       AGTCGTCATCCTCCATCATCGAGGAGACGTGAGGGTTCTTACTCTGATGGTAGTAGCGAT
VIP5-11C1   ------------------------------------------------------------
VIP5-4D1    ------------------------------------------------------------

Col.0       TCAAGGGATGATTCTGATGAAGATCGTGGCTATGCTAGTAGAAAACCCTCTGGGTCTCAA
VIP5-11C1   ------------------------------------------------------------
VIP5-4D1    ------------------------------------------------------------

Col.0       GTTCCTTTGAAGAAGAGATTGGAGGCAGAGAGAGAAGATCGAGCTGCTCGAGTTGAAGGT
VIP5-11C1   ------------------------------------------------------------
VIP5-4D1    ------------------------------------------------------------

Col.0       GGTTATGGTGATGGACCATCTGATCGTGAAGGTGACAGCAGCGAGGAGTCTGATTTTGGA
VIP5-11C1   ------------------------------------------------------------
VIP5-4D1    ------------------------------------------------------------

Col.0       GATGACCTTTACAAGAATGAGGAAGACAGGCAGAAGCTTGCTGGAATGACTGAGTTTCAG
VIP5-11C1   ------------------------------------------------------------
VIP5-4D1    ------------------------------------------------------------

Col.0       AGAGAGATGATTCTCTCTGAACGTGCTGATAAGAAAGGTGATAAGAACTTCACTGAGAAA
VIP5-11C1   ------------------------------------------------------------
VIP5-4D1    ------------------------------------------------------------

Col.0       CTTAGGTCCAAGAGAGAAAGTGAGAAAACCCCTGTTTCTAAAAAGGAGACTCAGCCTCTT
VIP5-11C1   ------------------------------------------------------------
VIP5-4D1    ------------------------------------------------------------

Col.0       CCGGCCTCTCGTGGTGTGCGTTCATCTGCTAGATCTGCAGACAGAGCCGCTGCTAAGGAT
VIP5-11C1   ------------------------------------------------------------
VIP5-4D1    ------------------------------------------------------------

Col.0       GATGCCCTGAATGAATTGAGGGCGAAGCGTATGAAGCAGCAGGACCCAGCAGCTCTCAGG
VIP5-11C1   ------------------------------------------------------------
VIP5-4D1    ------------------------------------------------------------

Col.0       AAACTGAGAGATGCATCAAAAGGTGGTTCAGGTAGTCGAGATTTCTCATCAACGAAGAGG   1
VIP5-11C1   ----------ATGCATCAAAAGGTGGTTCAGGTAGTCGAGATTTCTCGTCAAAGAGGAGG   1
VIP5-4D1    ------------------------------------------------------------   1

Col.0       AAACCGTTAGCTTCCTCCAATTTGAGTAGTTCCAGCCAAAGTGACAGTGATAGTAGGTCT   2
VIP5-11C1   AAACCATTACCTTCCTC-AATTTGAGTAGTTCCAGCCAAAGTGACAGTGATAGTAGGTCT   2
VIP5-4D1    ---GCGGTGGCTTCCTCCAATTTGAGTAGTTCCAGCCAAAGTGACAGTGATAGTAGGTCT   2
               *    *  *******  ******************************************

Col.0       CAGAGTGATGATGAAGGCTCGAATGGAGGAATGCTAGACAGTGATGATGACAGGTCAGAT   3
VIP5-11C1   CAGAGTGATGATGAAGGCTCGAATGGAGGAATGCTAGACAGTGATGATGACAGGTCAGAT   3
VIP5-4D1    CAAAGTGATGATGAAGGCTCGAATGGAGGAATGCTAGACAGTGATGATGACAGGTCAGAT   3
```

199

```
                  **  *****************************************************************

Col.0             GTGCCTACGTTTGAGGATGTTAAGGAAGTTACCATTAGACGGTCTAAGCTTGCCAAATGG       4
VIP5-11C1         GTGCCTACGTTTGAGGATGTTAAGGAAGTTACCATCAGACGGTCTAAGCTTGCCAAATGG       4
VIP5-4D1          GTGCCTACGTTTGAGGATGTTAAGGAAGTTACCATTAGACGGTCTAAGCTTGCCAAATGG       4
                  *********************************** **********************


Col.0             CTAATGGAGCCTTTCTTTGAAGAGCTTATAGTTGGGTGCTTTGTGAGGGTTGGGATCGGA       5
VIP5-11C1         CTAATGGAGCCTTTCTTTGAAGAGCTTATAGTTGGGTGCTTTGTGAGGGTTGGGATCGGA       5
VIP5-4D1          CTAATGGAGCCTTTCTTTGAAGAGCTTATAGTTGGGTGCTTTGTGAGAGTTGGGATCGGA       5
                  *********************************************** ************


Col.0             AGGTCAAAGAGTGGTCCAATTTACAGACTCTGCTGGGTGAAGAATGTTGATGCAACCGAT       6
VIP5-11C1         AGGTCAAAGAGTGGTCCAATTTACAGACTCTGCTGGGTGAAGAATGTTGATGCAACCGAT       6
VIP5-4D1          AGGTCAAAGAGTGGTCCAATTTACAGACTCTGCTGGGTGAAGAATGTAGATGCAACCGAT       6
                  *********************************************** ************


Col.0             CCTGACAAGACCTACAAGCTAGAGAATAAAACTACACACAAGTACCTTAACGTCGTCTGG       7
VIP5-11C1         CCTGACAAGACCTACAAGCTAGAGAATAAAACTACACACAAGTACCTTAACGTCGTCTGG       7
VIP5-4D1          CCTGACAAGACTTACAAGCTAGAGAATAAGACTACACACAAGTATCTTAACGTCGTCTGG       7
                  *********** ***************** ************** ***************


Col.0             GGAAATGAAACCTCGGCGGCTCGATGGCAAATGGCTATGATCTCAGATGGTCATCCGCTG       8
VIP5-11C1         GGAAATGAAACCTCGGCGGCTCGATGGCAAATGGCTATGATCTCAGATGGTCATCCGCTG       8
VIP5-4D1          GGAAATGAAACCTCGGCGGCTCGATGGCAAATGGCTATGATCTCAGATGGTCATCCTCTG       8
                  ******************************************************** ***


Col.0             GAGGAAGAGTATAGGCAATGGATCAGAGAGGTTGAGCGAACAAATGGTCGCATGCCCACA       9
VIP5-11C1         GAGGAAGAGTATAGGCAATGGATCAGAGAGGTTGAGCGAACAAATGGTCGCATGCCCACA       9
VIP5-4D1          GAGGAAGAGTATAGGCAATGGATCAGAGAGGTTGAGCGAACAAATGGTCGCATGCCCACA       9
                  ************************************************************


Col.0             AAGCAAGATATATCGGAGAAGAAAGAAGCGATACAAAGAACAAACAGTTTTGTTTACTCT      10
VIP5-11C1         AAGCAAGATATATCGGAGAAGAAAGAAGCGATACAAAGAACAAACAGTTTTGTTTACTCT      10
VIP5-4D1          AAGCAAGATATATCGGAGAAGAAAGAAGCGATACAAAGAACAAACAGTTTTGTTTACTCT      10
                  ************************************************************


Col.0             GCGGAAACTGTTAAACAGATGCTGCAGGAGAAAAAATCTGCGTCAGTCAGGCCAATGAAT      11
VIP5-11C1         GCGGAAACTGTTAAACAGATGCTGCAGGAGAAAAAATCTGCGTCAGTCAGGCCAATGAAT      11
VIP5-4D1          GCGGAAACTGTTAAACAGATGCTGCAGGAGAAAAAATCTGCGTCAGTCAGGCCAATGAAT      11
                  ************************************************************


Col.0             GTTGCGGCCGAGAAAGATCGGCTTAGAAAAGAATTGGAAATTGCGCAGAGCAAAAACGAT      12
VIP5-11C1         GTTGCGGCCGAGAAAGATCGGCTTAGAAAAGAATTGGAAATTGCGCAGAGCAAAAACGAT      12
VIP5-4D1          GTTGCGGCCGAGAAAGATCGGCTTAGAAAAGAGTTGGAAATTGCGCAGAGCAAAAACGAT      12
                  ******************************** ***************************


Col.0             GAAGCAGGTGTAGAGAGGATCAAGTCGAAAATCAAACAGCTCGACGCTTCACGGAACAAG      13
VIP5-11C1         GAAGCAGGTGTAGAGAGGATCAAGTCGAAAATCAAACAGCTCGACGCTTCACGGAACAAG      13
VIP5-4D1          GAAGCAGGTGTAGAGAGGATCAAATCGAAAATCAAACAGCTCGACGCCTCACGGAACAAG      13
                  ********************** ******************** ****** ***********


Col.0             AAAGGGGTAGATAAAAAAGCGCTTAAACTTGCTGAGATGAACAAGAAGAACAGAGCCGAG      14
VIP5-11C1         AAAGGGGTAGATAAAAAAGCGCTTAAACTTGCTGAGATGAACAAGAAGAACAGAGCCGAG      14
VIP5-4D1          AAAGGAGTCGATAAAAAAGCGCTCAAACTTGCCGAGATGAACAAGAAGAACAGAGCCGAG      14
                  ***** ** ************* ******** ****** *********************


Col.0             AATTTCAAGAACGCATCTGAGGTAAAATCAATAACTGCTAGTCTCAAAGCCGGTGAAGCA      15
VIP5-11C1         AATTTCAAGAACGCATCTGAGGTAAAATCAATAACTGCTAGTCTCAAAGCCGGTGAAGCA      15
VIP5-4D1          AATTTCAAGAACGCATCTGAGGTARAATCAATAACTGCTAGTCTCAAAGCCGGTGAAGCA      15
                  *********************** *********************************


Col.0             GGGTATGATCCGTTTTCAAGAAGATGGACCCGATCATCAAACTACTACAACGGGAAAAAC      16
VIP5-11C1         GGGTATGATCCGTTTTCAAGAAGATGGACCCGATCATCAAACTACTACAACGGGAAAAAC      16
VIP5-4D1          GGGTATGATCCGTTTTCAAGAAGATGGACCCGATCATCGAACTACTACAACGGGAAAAAC      16
                  ************************************** ********************


                                                                        200
```

```
Col.0        AAGGGGAAAGATGGAGAAGAGAACGAGGCAGCGGTTGCAGCAGCGGTTGAGACCAATGGA        17
VIP5-11C1    AAGGGGAAAGATGGAGAAGAGAACAAGGCAGCGGTTGCAGCAGCGGTTGAGACCAATGGA        17
VIP5-4D1     AAGGGGAAAGACGGAGAAGAGAACGAGGCAGCGGTTGCAGCAGCGGTTGAGACCAATGGA        17
             *********** ************ ***********************************

Col.0        GCAGATGCAGGAGCAGGTGTTGAAGCGACAGAAGCAGCTTTAGAAGCAGCTGCAGAGGCA        18
VIP5-11C1    GCAGATGCAGGAGCAGGTGTTGAAGCGACAGAAGCAGCTTTAGAAGCAGCTGCAGAGGCA        18
VIP5-4D1     GCAGATGCAGGAGCAGGTGTTGAAGCGACAGAAGCAGCTTTAGAAGCAGCTGCAGAGGCA        18
             ************************************************************

Col.0        GGAAAGCTAATAGACACAAGAGCTCCAATAGGTCAAGGAGCAGAACACAATCAGCTTCAT        19
VIP5-11C1    GGAAAGCTAATAGACACAAGAGCTCCAATAGGTCAAGGAGCAGAACACAATCAGCTTCAT        19
VIP5-4D1     GGAAAGCTAATAGACACAAGAGCTCCAATCGGTCGAGGAGCAGAACACAATCAGCTTCAT        19
             ***************************** **** **********************

Col.0        AACTTTGAATTGTCGTTATCGCTAACGGCTTTACAGAAGTACGGAGGACCTCAAGGAGTA        20
VIP5-11C1    AACTTTGAATTGTCGTTATCGCTAACGGCTTTACAGAAGTACGGAGGACCTCAAGGAGTA        20
VIP5-4D1     AACTTTGAATTGTCGTTATCGCTAACGGCTTTACAGAAGTACGGAGGACCTCAAGGAGTA        20
             ************************************************************

Col.0        CAGAAAGCGTTCATGGCGAGGAAGCAACTGACCGAAGCAACTGTGGGATGCAGAGTCGCA        21
VIP5-11C1    CAGAAAGCGTTCATGGCGAGGAAGCAACTGACCGAAGCAACTGTGGGATGCAGAGTCGCA        21
VIP5-4D1     CAGAAAGCGTTCATGGCGAGGAAGCAACTGACCGAAGCAACTGTGGGATGCAGAGTCGTA        21
             ********************************************************** *

Col.0        GAGAACGATGGCAAGAGACATGGCCTTACGTTAACTGTTAGTGATTACAAGAGAAGGAGA        22
VIP5-11C1    GAGAACGATGGCAAGAGACATGGCCTTACGTTAACTGTTAGTGATTACAAGAGAAGGAGA        22
VIP5-4D1     GAGAACGATGGCAAGAGACATGGCCTTACGTTAACTGTTAGTGATTACAAGAGAAGGAGA        22
             ************************************************************

Col.0        GGTCTTCTCTGA        23
VIP5-11C1    GGTCTTCTCTGA        23
VIP5-4D1     GGTCTTCTCTGA        23
             ************
```

Appendix C

```
11C1genomic                          TTGACATAAAGATATAATCACTTATTTGTATAATGATTATGAAATCC 1
4D1genomic                           TTGACATAAAGATATAATCACTTATTTGTATAATGATTATGAAA--- 1
Colgenomic                              ATAAAGATATAATCACTTATTTGTATAATGATTATGAAA--- 1


11C1genomic   TACTATATTATTTGGGAAGTACATATTAAATGTAACCTTAATTTTTGTAATTAATTACAT

11C1genomic   GACAATGCCATTAGAAAAATTTAATTAAAAACAAAATCCTTTAATGACGTAATTAAGGTT

11C1genomic   ACCAAAATCATTTAACGACAATATTTACTTCTTAATTATAGGGCTTATTAGATCTAAAGA

11C1genomic   CATGTCATCAAAGATTTCCTAAACTGAAGCAAAATATACCGAATATTCAAATATCTATCT

11C1genomic   GTTACCAAATTATAAACAAATAAGTTAATAAGTCAATAACTATTTGGAGGGACGGGTTTT

11C1genomic   TGAATCAACATTAATAAAAAAGTAAAATATAATTGATCCACCGTTTCAATACGGGTTAA

11C1genomic   ATCTTTAATTTATTATTTTTAAGACCACTGATATTAAACATATCAAATCATCCTAATTTA

11C1genomic   GAAAAGGTTATATAAAACCAAAAATGTTATGTGGTATGTATAATGTTACTATATATAAAA

11C1genomic   TTAAACTATAAAATATAAATGTATTAGAGAATAATACAATTTGTAAAACTTTTATATGTA

11C1genomic   ATAAATAATTCTCAAATTTTAAAAATTACTACTTTAAAAACAAATCACGGGACGGGTAAA

11C1genomic   GAAATTAGAGAACAGATTTTATTTTGGAATTGAGTTATATGGTGGATGTATTTGAATCAA

11C1genomic   TATTTATAAAATTTTAAAATATTATTAATATGCTGTTTTAGTAAGGGTGAAAACTTCAGT

11C1genomic   TTTTTAACAATTGTCTCATGGATTCGTGGTATAGCGTTAGTTAATAACAATTATAAACTG

11C1genomic   TAAAATATAAATATTTTATAAAAATAAAATTTGCAAGTTTTAATATATATTATCTTTAAA

11C1genomic   AATAAATTATAACGCGGTATACCGCGGGTTAAAATCTAGTTTCATTTGTATTGTTTAAGC 2
4D1genomic                                              TTTCACTTGTATTGTTTAAGC 2
Colgenomic                                              TTTCACTTGTATTGTTTAAGC 2
                                                       ****  *************** 


11C1genomic   GAAAACACAGAATCAAAACCGATATTGTGCAAGCAATCTTTCAATGCAATAGTCCTAACA 3
4D1genomic    GAAAACAAAGGATCAAAACTGATATTGT----GCAATCTTTCAATGCAATAGTCCTAACA 3
Colgenomic    GAAAACAAAGGATCAAAACTGATATTGT----GCAATCTTTCAATGCAATAGTCCTAACA 3
              ******* ** ******** *******       ************************** 


11C1genomic   AATAACAAACCGAAGGGGTACTTTTCTAGATGGGTCTGGGCCAAGGCCCATACATTTCGG 4
4D1genomic    AATAACAAACCGAAGGGGTACTTTTCTAGATGGGTCTGGGCCAAGGCCCATACATTTCGG 4
Colgenomic    AATAACAAACCGAAGGGGTACTTTTCTAGATGGGTCTGGGCCAAGGCCCATACATTTCGG 4
              ************************************************************ 


11C1genomic   GTCCAGAAATCTAACGTTGCCGTATACCTTCCTTTTTTAGCTAAGAAAAGAAGGCTACAC 5
4D1genomic    GTCCAGAGATCAAAGGTTGCCGTATACCTTCCTTTTGTAGCTAAGAAAAGAAGGCTACAC 5
Colgenomic    GTGCAGAGATCAAACGTTGCCGTATACCTTCCTTTTTTAGCTAAGAAAAGAAGGCTACAC 5
              ** **** *** ** ******************** ********************** 


11C1genomic   TTCTT------------------------ACACAGAGCAAGAGCTATGTCAACAGAGAGT 6
4D1genomic    TTTGGTAGACTCGTCACTCACATCATCACACACAGAGCAAGAGCTATCTCAACAGAGAGT 6
Colgenomic    TTTGGTAGACTCGTCACTCACATCATCACACACAGAGCTAGAGCTATGTCAACAGAGACT 6
              **                          ******** ******************a* 
```

```
11C1genomic   AAAGAAACCCGACCCGAAACTAAACCCGAAGACCCGGTAACTCATACTACTGTGGATGTA 7
4D1genomic    AAAGAAACCCGACCCGAAACTAAACCCGAAGACCCGGCAACTCATACTACTGTGGATGTA 7
Colgenomic    AAAGAAACCCGACCCGAAACTAAACCCGAAGACCTGGGAACTCATACTACTGTGGATGTA 7
              *******************************b**c*********************


11C1genomic   CCCGGTGAAGAACCTCTCGGAGAGCTTATCGCCGACGACGTGAACGAAGTCGTTTCTGAT 8
4D1genomic    CCCGGTGAAGAACCTCTCGGAGAGCTTATTGCCGACGACGTGAACGAAGTCGTTTCTGAT 8
Colgenomic    CCCGGTGAAGAACCTCTCGGAGAGCTTATCGCCGACGACGTGAACGAAGTCGTTTCTGAT 8
              *****************************d******************************


11C1genomic   GCGTCGGCGACGGAGACAGACTTCTCACTCTCACCGAGCCAATCGGAGCAAAATATTGAA 9
4D1genomic    GCGTCGGCGACGGAGACAGACTTCTCACTCTCACCGAGCCAATCGGAGCAAAATATTGAA 9
Colgenomic    GCGTCGGCGACGGAGACAGACTTCTCACTCTCACCGAGCCAATCGGAGCAAAATATTGAA 9
              ***********************************************************


11C1genomic   GAAGACGGCCAAAACTCACTCGATGACCAATCGCCATTAACGGAGCTTCAGCCTCTTCCT 10
4D1genomic    GAAGACGGCCACAACTCACTCGATGACCAATCGCCATTAACGGAGCTTCAACCTCTTCCT 10
Colgenomic    GAAGACGGCCAAAACTCACTCGATGACCAATCGCCATTAACGGAGCTTCAGCCTCTTCCT 10
              **********e*************************************f*********


11C1genomic   CTTCCTCCTCCTCTTCCAGTCGAAGCACGAATCTCAGAATCGCTCGGTGAAGAAGAATCT 11
4D1genomic    CTTCCTCCTCCTCTTCCAGTCGAAGCACGAATCTCAGAATCGCTCGGTGAAGAAGAATCT 11
Colgenomic    CTTCCTCCTCCTCTTCCAGTCGAAGCACGAATCTCAGAATCGCTCGGTGAAGAAGAATCT 11
              ***********************************************************


11C1genomic   TCCGATCTGGTAACGGAGCAACAATCACAAAACCCTAATGCGGCGGAGCCTGGTCCTAGA 12
4D1genomic    TCCGATCTAGTAACGGAGCAACAATCACAAAACCCTAATGCGGCGGAGCCTGGTCCTAGA 12
Colgenomic    TCCGATCTGGTAACGGAGCAACAATCACAAAACCCTAATGCGGCGGAGCCTGGTCCTAGA 12
              ********g**************************************************


11C1genomic   GCAAGAAAAAGACGCCGTAGGAAACGTTTCTTCACTGAGATTAACGCAAACCCAGCTTTC 13
4D1genomic    GCAAGAAAACGACGCCGTAGGAAACGTTTCTTCACTGAGATTAACGCAAACCCAGCTTTC 13
Colgenomic    GCAAGAAAAAGACGCCGTAGGAAACGTTTCTTCACTGAGATTAACGCAAACCCAGCTTTC 13
              *********h*************************************************


11C1genomic   TCAAGAAACCGCCGCACTAGCGTCGGCAAAGAGGTGGATTCAGAAGCGCTAATCGCAATG 14
4D1genomic    TCAAGAAACCGCCGCACTAGCGTCGGCAAAGAGGTGGATTCAGAAGCGCTAATCGCAATG 14
Colgenomic    TCAAGAAACCGCCGCACTAGCGTCGGCAAAGAGGTGGATTCAGAAGCGCTAATCGCAATG 14
              ***********************************************************


11C1genomic   TCGGTAGGGTTTCCGGTTTATTCGCTCACGGAGGAAGAAATTGAAGCTAATGTGGTTTCG 15
4D1genomic    TCCGTAGGGTTTCCGGTTTATTCACTCACGGAGGAAGAAATCGAAGCTAATGTGGTTTCA 15
Colgenomic    TCGGTAGGGTTTCCGGTTTATTCGCTCACGGAGGAAGAAATTGAAGCTAATGTGGTTTCG 15
              **i*****************j****************k****************l


11C1genomic   ATCATCGGAGGTAAAGATCAAGCTAATTACATTGTTGTAAGGAATCACATTATTGCTCTG 16
4D1genomic    ATCATCGGAGGTAAAGATCAAGCTAATTACATTGTTGTAAGGAATCACATTATTGCTCTG 16
Colgenomic    ATCATCGGAGGTAAAGATCAAGCTAATTACATTGTTGTAAGGAATCACATTATTGCTCTG 16
              ***********************************************************


11C1genomic   TGGCGATCTAATGTATCGAATTGGTTAACGCGAGATCATGCGCTTGAGTCTATACGTGCT 17
4D1genomic    TGGCGATCTAATGTATCGAATTGGTTAACGCGAGACCATGCGCTTGAGTCTATACGTGCT 17
Colgenomic    TGGCGATCTAATGTATCGAATTGGTTAACGCGAGATCATGCGCTTGAGTCTATACGTGCT 17
              ***********************************m***********************


11C1genomic   GAACACAAAACCTTAGTTGATACTGCTTACAATTTCCTTCTTGAGCATGGTTATATTAAC 18
4D1genomic    GAACACAAAACCTTAGTTGATACTGCTTACAATTTCCTCCTTGAGCATGGTTATATTAAC 18
Colgenomic    GAACACAAAACCTTAGTTGATACTGCTTACAATTTCCTTCTTGAGCATGGTTATATTAAC 18
              *************************************n*********************


11C1genomic   TTCGGGCTTGCTCCGGTTATTAAAGAGGCGAAATTGAGGTCGTTTGATGGCGTAGAGCCG 19
4D1genomic    TTCGGGCTTGCTCCGGTTATTAAAGAGGCGAAATTGAGGTCGTTTGATGGCGTAGAGCCG 19
Colgenomic    TTCGGGCTTGCTCCGGTTATTAAAGAGGCGAAATTGAGGTCGTTTGATGGCGTAGAGCCG 19
              ***********************************************************
```

203

```
11C1genomic  CCAAATGTTGTCGTTGTAGGGGCGGGTTTAGCTGGATTGGTTGCTGCTAGACAGTTGTTG 20
4D1genomic   CCTAATGTTGTCGTTGTAGGGGCGGGTTTAGCTGGTTTGGTTGCTGCTAGACAGTTGTTG 20
Colgenomic   CCAAATGTTGTTGTTGTAGGGGCGGGTTTAGCTGGTTTGGTTGCTGCTAGACAGTTGTTG 20
             **o*********p*******************q*********************
11C1genomic  TCAATGGGGTTTAGGGTTTTGGTTCTTGAAGGTAGAGATAGACCGGGTGGGCGGGTTAAG 21
4D1genomic   TCAATGGGGTTTAGGGTTTTGGTTCTTGAAGGTAGAGATAGACCGGGTGGGCGGGTTAAG 21
Colgenomic   TCAATGGGGTTTAGGGTTTTGGTTCTTGAAGGTAGAGATAGACCGGGTGGGCGGGTTAAG 21
             ************************************************************

11C1genomic  ACACGGAAGATGAAAGGTGGTGATGGTGTTGAGGCAATGGCTGATGTTGGTGGAAGCGTT 22
4D1genomic   ACTCGGAAGATGAAAGGTGGTGATGGTGTTGAGGCAATGGCTGATGTTGGTGGAAGTGTT 22
Colgenomic   ACACGGAAGATGAAAGGTGGTGATGGTGTTGAGGCAATGGCTGATGTTGGTGGAAGCGTT 22
             **r***************************************************s***
11C1genomic  CTCACCGGAATTAATGGGAATCCGCTTGGGGTTTTAGCGAGGCAACTTGGTTTGCCTCTT 23
4D1genomic   CTCACCGGAATCAATGGTAATCCGCTTGGGGTTTTAGCGAGGCAACTTGGTTTGCCTCTT 23
Colgenomic   CTCACCGGAATTAATGGGAATCCGCTTGGGGTTTTGGCGAGGCAACTTGGTTTGCCTCTT 23
             ***********t*****u*****************v*****************
11C1genomic  CATAAGGTTAGAGATATTTGTCCTTTGTATCTTCCCAATGGAGAGCTTGCTGATGCTAGT 24
4D1genomic   CACAAGGTTAGAGATATTTGTCCTTTGTATCTTCCCAGTGGAGAGCTTGCTGACGCTGGT 24
Colgenomic   CATAAGGTTAGAGATATTTGTCCTTTGTATCTTCCCAATGGAGAGCTTGCTGATGCTAGT 24
             **w*******************************x*************y***z**
11C1genomic  GTTGATTCTAAGATTGAGGCATCGTTTAATAAGTTGTTGGATAGAGTTTGTAAGCTTAGA 25
4D1genomic   GTTGATTCTAAGATTGAGGCATCGTTTAATAAGTTGTTGGATAGAGTTTGTAAGCTTAGA 25
Colgenomic   GTTGATTCTAAGATTGAGGCATCGTTTAATAAGTTGTTGGATAGAGTTTGTAAGCTTAGA 25
             ************************************************************

11C1genomic  CAGTCGATGATAGAGGAGAATAAATCAGTTGATGTGCCTTTGGGAGAAGCGCTTGAAACA 26
4D1genomic   CAGTCGATGATAGAGGAGAATAAATCAGTTGATGTGCCTTTGGGAGAAGCGCTTGAAACA 26
Colgenomic   CAGTCGATGATAGAGGAGAATAAATCAGTTGATGTGCCTTTGGGAGAAGCGCTTGAAACA 26
             ************************************************************

11C1genomic  TTTCGATTGGTTTATGGGGTTGCTGAGGATCAGCAAGAGAGAATGCTCTTAGATTGGCAT 27
4D1genomic   TTTCGATTGGTTTATGGGGTTGCTGAGGATCAGCAAGAGAGAATGCTCTTAGATTGGCAT 27
Colgenomic   TTTCGATTGGTTTATGGGGTTGCTGAGGATCAGCAAGAGAGAATGCTCTTAGATTGGCAT 27
             ************************************************************

11C1genomic  TTAGCAAACTTGGAATATGCAAATGCTACATTGTTGGGGAATCTGTCAATGGCGTATTGG 28
4D1genomic   TTAGCAAACTTGGAATATGCAAATGCTACATTGTTGGGGAATCTGTCAATGGCGTATTGG 28
Colgenomic   TTAGCAAACTTGGAATATGCAAATGCTACATTGTTGGGGAATCTGTCAATGGCGTATTGG 28
             ************************************************************

11C1genomic  GATCAAGATGATCCGTATGAGATGGGTGGTGATCATTGTTTTATCCCAGGTGGGAACGAA 29
4D1genomic   GATCAAGATGATCCGTATGAGATGGGTGGTGATCATTGTTTTATCCCAGGTGGGAACGAA 29
Colgenomic   GATCAAGATGATCCGTATGAGATGGGTGGTGATCATTGTTTTATCCCAGGTGGGAACGAA 29
             ************************************************************

11C1genomic  ATATTTGTACATGCGTTAGCGGAAAATCTTCCGATTTTTTACGGGAGTACAGTTGAGAGC 30
4D1genomic   ATATTTGTACATGCTTTAGCGGAAAATCTTCCAATTTTTTACGGGAGTACAGTTGAGAGC 30
Colgenomic   ATATTTGTACATGCTTTAGCGGAAAATCTTCCAATTTTTTACGGGAGTACAGTTGAGAGC 30
             *************A******************B***********************
11C1genomic  ATCAGATATGGAAGTAACGGGGTTCTGGTTTACACAGGTAACAAAGAGTTCCACTGCGAT 31
4D1genomic   ATCAGATATGGAAGCAACGGGGTTCTGGTTTACACAGGTAACAAAGAGTTCCACTGCGAT 31
Colgenomic   ATCAGATATGGAAGCAACGGGGTTCTGGTTTACACAGGTAACAAAGAGTTCCACTGCGAT 31
             *************C**********************************************

11C1genomic  ATGGCTCTTTGCACGGTTCCATTAGGTGTTCTGAAGAAAGGTTCAATTGAGTTTTATCCC 32
4D1genomic   ATGGCTCTTTGCACGGTTCCATTAGGTGTTCTGAAGAAAGGTTCGATTGAGTTTTATCCC 32
Colgenomic   ATGGCTCTTTGCACGGTTCCATTAGGTGTTCTGAAGAAAGGTTCGATTGAGTTTTATCCC 32
             ********************************************D**************
```

```
11C1genomic    GAACTTCCTCATAAGAAGAAAGAAGCGATTCAGAGACTTGGATTCGGATTGTTGAACAAA 33
4D1genomic     GAACTTCCTCATAAGAAGAAAGAAGCGATTCAGAGACTTGGATTCGGATTGTTGAACAAA 33
Colgenomic     GAACTTCCTCATAAGAAGAAAGAAGCGATTCAGAGACTTGGATTCGGATTGTTGAACAAA 33
               ************************************************************

11C1genomic    GTGGCGATGTTGTTTCCGTGTAACTTCTGGGGCGAAGAGATTGATACTTTTGGGCGATTA 34
4D1genomic     GTGGCGATGTTGTTTCCGTGTAACTTCTGGGGCGAAGAGATTGATACTTTTGGGCGATTA 34
Colgenomic     GTGGCGATGTTGTTTCCGTGTAACTTCTGGGGCGAAGAGATTGATACTTTTGGGCGATTA 34
               ************************************************************

11C1genomic    ACCGAAGATCCGTCCACTAGAGGAGAATTCTTCTTGTTCTACAGCTATTCTTCTGTTTCC 35
4D1genomic     ACCGAAGATCCGTCCACTAGAGGAGAATTCTTCTTGTTCTACAGCTATTCTTCTGTTTCC 35
Colgenomic     ACCGAAGATCCGTCCACCAGAGGAGAATTCTTCTTGTTCTACAGCTATTCTTCTGTTTCC 35
               *****************E******************************************

11C1genomic    GGTGGTCCATTACTTGTAGCACTTGTAGCTGGAGACGCTGCGGAAAGATTCGAGACATTG 36
4D1genomic     GGTGGTCCATTACTTGTAGCACTTGTAGCTGGAGACGCTGCGGAAAGATTCGAGACATTG 36
Colgenomic     GGTGGTCCATTACTTGTAGCACTTGTAGCTGGAGACGCTGCGGAAAGATTCGAGACATTG 36
               ************************************************************

11C1genomic    TCGCCTACTGATTCCGTTAAACGGGTCTTGCAGATACTACGCGGAATATATCACCCAAAA 37
4D1genomic     TCGCCTACTGATTCCGTTAAACGGGTCTTGCAGATACTACGCGGAATATATCACCCAAAA 37
Colgenomic     TCGCCTACTGATTCCGTTAAACGGGTCTTGCAGATACTACGCGGAATATATCACCCAAAA 37
               ************************************************************

11C1genomic    GGAATTGTTGTTCCTGATCCGGTTCAAGCCCTCTGTTCCAGATGGGGACAAGACAAGTTT 38
4D1genomic     GGAATTGTTGTTCCTGATCCGGTTCAAGCCCTCTGTTCCAGATGGGGACAAGACAAGTTT 38
Colgenomic     GGAATTGTTGTTCCTGATCCGGTTCAAGCCCTCTGTTCCAGATGGGGACAAGACAAGTTT 38
               ************************************************************

11C1genomic    TCATACGGTTCTTACTCATATGTTGCGGTCGGATCATCAGGAGATGATTACGATATTTTA 39
4D1genomic     TCATACGGTTCTTACTCATATGTTGCGGTCGGATCATCAGGAGATGATTACGATATTTTA 39
Colgenomic     TCATACGGTTCTTACTCATATGTTGCGGTCGGATCATCAGGAGATGATTACGATATTTTA 39
               ************************************************************

11C1genomic    GCCGAGAGTGTCGGTGATGGAAGAGTGTTCTTTGCAGGTGAAGCTACTAACAGACAATAT 40
4D1genomic     GCCGAGAGTGTCGGTGATGGAAGAGTGTTCTTTGCAGGTGAAGCTACTAACAGACAATAT 40
Colgenomic     GCCGAGAGTGTTGGTGATGGAAGAGTGTTCTTTGCAGGTGAAGCTACTAACAGACAATAT 40
               ***********F************************************************

11C1genomic    CCAGCTACAATGCACGGAGCCTTCTTAAGTGGAATGAGAGAAGCAGCAAACATACTTAGA 41
4D1genomic     CCAGCTACAATGCACGGAGCCTTCTTAAGTGGAATGAGAGAAGCAGCAAACATACTTAGA 41
Colgenomic     CCAGCTACAATGCACGGAGCCTTCTTAAGTGGAATGAGAGAAGCAGCAAACATACTTAGA 41
               ************************************************************

11C1genomic    GTTGCTAGAAGAAGAGCGTCATCATCGGCTTTAAATCCTAACCAGACCTGCATCGACAAA 42
4D1genomic     GTTGCTAGAAGAAGAGCGTCATCATCGGCTTTAAATCCTAACCAGACCTGCATCGACAAA 42
Colgenomic     GTTGCTAGAAGAAGAGCGTCATCATCGGCTTTAAATCCTAACCAGATCTGCATCGACAAA 42
               **********************************************G*************

11C1genomic    GAAGAAGAGGTAGAGGAAGAAGAAGACCGCTGTTTGGATCAGTTATTCGAGACACCCGAT 43
4D1genomic     GAAGAAGAGGTAGACGAAGAAGAAGACCGCTGTTTGGATCAGTTATTCGAGACACCCGAT 43
Colgenomic     GAAGAAGAGGTAGACGAAGAAGAAGACCGCTGTTTGGATCAGTTATTCGAGACACCCGAT 43
               **************H*********************************************

11C1genomic    TTAACATTCGGGAATTTCTCAGTATTGTTTACTCCAAATTCGGATGAACCTGAATCCATG 44
4D1genomic     TTAACATTCGGGAATTTCTCAGTATTGTTTACTCCAAATTCGGATGAACCTGAATCCATG 44
Colgenomic     TTAACATTCGGGAATTTCTCAGTATTGTTTACTCCAAATTCGGATGAACCTGAATCCATG 44
               ************************************************************

11C1genomic    TCATTGTTGAGGGTTAGGATCCAAATGGAGAAACCCGAATCAGGTCTTTGGCTTTACGGT 45
4D1genomic     TCATTGTTGAGGGTTAGGATCCAAATGGAGAAACCCGAATCAGGTCTTTGGCTTTACGGT 45
Colgenomic     TCATTGTTGAGGGTTAGGATCCAAATGGAGAAACCCGAATCAGGTCTTTGGCTTTACGGT 45
               ************************************************************
```

```
11C1genomic    TTAGTGACAAGGAAGCAAGCTATTGAGCTTGGTGAAATGGATGGAGATGAGTTGAGAAAT 46
4D1genomic     TTAGTGGCAAGGAAGCAAGCTATTGAGCTTGGTGAAATGGATGGAGATGAGTTGAGAAAT 46
Colgenomic     TTAGTGACAAGGAAGCAAGCTATTGAGCTTGGTGAAATGGATGGAGATGAGTTGAGAAAT 46
               ******I*****************************************************

11C1genomic    GAATACTTGCGTGAGAAGTTAGGACTTGTTCCTGTCGAAAGGAAGAGTCTTTCTCAAGAA 47
4D1genomic     GAATACTTGCGTGAGAAGTTAGGACTTGTTCCTGTCGAAAGGAAGAGTCTTTCTCAAGAA 47
Colgenomic     GAATACTTGCGTGAGAAGTTAGGACTTGTTCCTGTCGAAAGGAAGAGTCTTTCTCAAGAA 47
               ************************************************************


11C1genomic    GGGGAATCGATGATCTCTTCACTCAAAGCTGCAAGACTGAATCGACAGATCTTTGATTAG 48
4D1genomic     GGGGAATCGATGATCTCTTCACTCAAAGCTGCAAGACTGAATCGACAGATCTTTGATTAG 48
Colgenomic     GGGGAATCGATGATCTCTTCACTCAAAGCTGCAAGACTGAATCGACAGATCTTTGATTAG 48
               **********************************************************

11C1genomic    TAGTAAAATAGATTTAGATTAAAATTGTATTCAGCTGCGTCTACGTTGTAAATTGTTGTT 49
4D1genomic     TAGTAAAATAGATTTAGATTAAAATTGTATTCAGCTGCGTCTACGTTGTAAATTGTTGTT 49
Colgenomic     TAGTAAAATAGATTTAGATTAAAATTGTATTCAGCTGCGTCTACGTTGTAAATTGTTGTT 49
               ************************************************************

11C1genomic    CTAATCTTTTTTGTACTTAACGCAAAACATTTTGGAAAACAAAGCTTTGTTACTTATTTG 50
4D1genomic     CTAATCTTTTTTGTACTTAACGCAAAACATTTTGGAAAACAAAGCTTTGTTACTTATTTG 50
Colgenomic     CTAATCTTTTTTGTACTTAACGCAAAACATTTTGGAAAACAAAGCTTTGTTACTTATTTG 50
               ************************************************************

11C1genomic    ACGTGAAG---------------------------------------------------- 51
4D1genomic     ACGTGAAG---------------------------------------------------- 51
Colgenomic     ACGTGAAG---------------------------------------------------- 51
               ********
```

206

Appendix D (1)

(Flowering time data of F5)

Family	GR2 No. of leaves to flower	OS4 No. of leaves to flower	Family	GR2 No. of leaves to flower	OS4 No. of leaves to flower	Family	GR2 No. of leaves to flower	OS4 No. of leaves to flower
31	31±5.7	14±0.7	362	36±20	12±1.3	612	28±14	13±0.2
33	37±8.5	13±0.8	364	37±7.3	13±9.2	613	33±8.7	13±0.3
51	33±24	14±3.3	366	32±58	13±1.8	614	35±6.7	12±1.7
52	13±5.3	15±1.7	392	39±4.3	15±2.2	615	28±22	13±0.8
54	17±11	15±0.7	393	34±28	14±2.3	632	22±2.8	15±1.8
56	39±1.7	15±1	394	37±5.6	17±1.8	634	33±1.8	14±0.3
61	35±4.3	14±3.2	395	34±24	17±4	635	25±18	14±1.5
64	34±9.8	13±5	396	40±30	14±6.8	651	18±19	14±0.9
65	36±10	14±5.5	421	30±39	14±2.8	652	26±35	15±1.5
66	29±10	14±3.2	422	33±2.2	15±4.7	656	30±58	14±1.5
72	30±22	15±1.2	441	31±34	14±0.5	671	31±9.2	15±3.7
76	20±13	17±1	443	34±15	15±4.3	676	37±11	15±6.2

86	37±16	14±0.8	**444**	31±1.7	12±0.2	**683**	25±90	14±2.7
91	35±10	14±1.7	**462**	30±18	15±2.2	**685**	35±15	14±1.3
92	42±10	17±1.7	**463**	31±20	15±1.3	**686**	30±14	16±1.3
93	34±15	15±1.7	**464**	33±10	14±0.8	**691**	42±6.5	14±1.8
94	32±19	13±0.3	**466**	29±3.7	15±2	**694**	38±24	15±4.5
95	36±11	16±1.3	**483**	30±2.2	14±6.2	**695**	39±6.8	16±9.7
101	35±8.3	13±2.2	**484**	25±43	14±0.8	**696**	37±11	15±8.2
104	33±3.7	14±4.8	**484b**	38±11	14±7	**706**	33±41	16±1.3
121	35±0.7	15±2.3	**485**	32±24	16±4.2	**524a**	35±20	16±4.3
124	34±8.5	15±4.7	**504**	27±43	15±2.8	**524b**	35±63	14±3.2
132	37±25	14±1.3	**511**	39±9.1	14±2.7			
134	34±11	16±3.7	**514**	27±20	14±3.7			
153	43±10	14±6.7	**516**	35±17	18±8			
155	43±14	15±2.8	**521**	31±11.	13±1.2			
156	14±2.7	14±1.7	**522**	36±9.3	15±0.2			
194	40±17	13±3.2	**523**	38±20	13±5.3			
241	25±19	14±0.7	**526**	35±34	14±2			
242	27±11.2	15±0.8	**581**	35±18.7	14±1.2			

243	29±16.7	14±1.2	**594**	39±42	14±6.3			
261	33±15	14±4.5	**601**	35±13.8	15±0.5			
341	29±34	14±1.6	**605**	33±26	17±1.2			
342	19±3.3	15±0.7	**611**	35±15	13±1.3			

* ± variance

Appendix D (2)

(Flowering time data of F3)

Family	Growth room –No. of leaves to flower	Family	Growth room –No. of leaves to flower
3	29±3.8	45	28±9.5
5	27±5.7	46	32±2.7
6	16±2.1	47	29±5
7	23±3.5	48	31±10.2
8	23±6.3	49	31±4.5
9	26±9.1	51	25±3.5
10	23±10.2	52	30±2.5
12	25±21.5	58	29±23.5
13	29±5.9	60	29±8.9
15	24±6.9	63	24±5.4
19	12±8.2	67	32±0.9
24	29±5.3	68	29±7.2
26	33±6.8	69	39±3.6
34	32±12.4	70	28±7.8
36	24±13	59a	30±7.8
39	32±1.8	65a	28±0.5
40	29±7.9		
41	26±0.6		
44	30±4.5		

* ± variance

Appendix D (3)

F3 flowering data (field)

Family	Natural conditions– No. of leaves to flower	Family	Natural conditions – No. of leaves to flower
9	26±9.1	47	29±5
10	23±10.2	60	29±8.9
12	25±21.5	63	24±5.4
13	29±5.9	67	32±0.9
15	24±6.9	68	29±7.2
19	12±8.2	69	39±3.6
24	29±5.3	70	28±7.8
26	33±6.8	59a	30±7.8
34	32±12.4	65a	28±0.5
36	24±13		
39	32±1.8		
40	29±7.9		
41	26±0.6		
44	30±4.5		
45	28±9.5		
46	32±2.7		

* ± variance

Appendix E (1)

Growth rate variation data of F5

Family	OS3 (S) (Area in mm^2)	OS4 (A) (Area in mm^2)	GR1(S) (Area in mm^2)	GR2(A) (Area in mm^2)
31	548±88	233±17	3558±281	365±38
33	666±63	218±24	4360±265	247±38
51	619±36	148±16	4221±445	247±37
52	520±82	228±11	3499±146	350±68
54	752±82	168±10	2801±272	308±53
56	660±46	198±18	4083±377	295±27
61	--	218±24	1701±730	308±58
64	552±88	202±20	2759±571	303±21
65	874±174	207±24	3172±963	376±31
66	452±74	249±40	2064±412	140±31
72	985±246	307±23	4465±245	322±26
76	1196±44	255±21	3737±247	341±15
86	828±115	225±17	4390±571	301±35
91	349±67	241±17	3633±319	413±65
92	550±78	194. ±7	3666±269	502±42
93	480 ±52	160±9	3926±435	282±39
94	474±84	229±24	2964±266	384±37

95	714±72	161±9	4542±346	268±25
101	587±76	185±51	4855±478	130±28
104	555±73	132±10	3300±226	160±26
121	1036±135	273±26	3540 ±571	331±105
124	982±150	218±14	4058±517	409±39
132	367±31	176±12	3608±115	347±57
134	1023±78	243±15	4241±338	328±44
153	491±75	224±32	2609±347	447±59
155	251±129	229±18	1922±477	514±36
156	680±196	240±30	1589±512	472±39
194	294±68	200±22	1388±227	396±56
241	629±114	229±18	3744±382	375±46
242	1153±158	259±40	4735±651	474±16
243	819±98	242±20	3615±357	271±43
261	687±59	207±23	2935±1103	110±37
341	508±46	208±19	4912±426	449±83
342	403±62	157±7	3686±329	386±45
362	1160±100	114±17	4077±565	217±67
364	1144±175	147±24	3681±256	327±22
366	811±97	153±5	2744±136	331±54
392	874±119	252±20	3542±375	358±26

393	538±207	258±41	1836±849	219±84
394	375±58	279±15	2701±468	275±31
395	954±116	293±28	4143±309	482±58
396	823±168	200±31	2885±441	424±34
421	1049±75	355±47	4614±290	431±42
422	803±73	181±27	3898±438	549±98
441	618±76	208±29	3420±374	436±29
443	1022±164	306±37	4212±444	401±29
444	638±63	140±35	3151±139	226±62
462	404±55	218±15	2299±164	315±66
463	784±85	222±10	3366±208	315±71
464	849±183	193±31	3024±361	363±47
466	699±82	219±15	3864±385	355±66
483	706±79	129±12	4025±761	326±41
484	881±101	189±15	3817±302	196±53
484b	471±98	186±6	4170±335	238±13
485	577±83	180±16	3726±190	261±45
504	831±74	341±42	4431±503	296±54
511	639±147	205±12	3424±309	384±40
514	922±89	230±27	4352±500	476±24
516	1048±115	249±24	4112±298	572±74

521	1097±135	236±8	3544±366	327±57
522	1279±123	312±19	7320±751	539±84
523	1216±175	211±13	5987±334	551±100
526	376±43	232±21	4097±393	456±70
581	853±131	267±19	3940±386	344±63
594	849±123	210±11	2998±368	287±52
601	932±116	217±31	4014±623	389±70
605	826±130	222±17	3384±249	263±53
611	840±76	211±4	3645±298	301±34
612	716±205	150±20	2484±516	185±18
613	687±118	145±14	3172±310	161±21
614	674±74	138±16	3246±419	180±23
615	675±147	149±14	2711±790	296±63
632	579±96	213±12	3851±268	418±48
634	898±123	230±18	4572±512	237±63
635	767±122	212±15	4394±237	475±69
651	1151±73	238±18	3130±300	322±48
652	1372±58	185±34	3692±278	259±77
655	1691±97	203±38	4939±735	352±74
671	1917±90	305±24	5714±463	525±97
676	1712±72	296±17	5408±788	389±37

683	888±84	234±27	3985±141	431±77
685	531±72	193±24	3831±378	390±40
686	746±89	233±12	4189±381	491±60
691	710±56	254±24	3171±189	483±19
694	619±77	247±22	3518±432	398±16
695	566±51	223±22	3261±315	392±35
696	662±70	225±16	3525±401	474±47
706	537±44	204±17	2690±394	313±40
524a	939±238	284±37	4852±647	600±35
524b	1244±86	287±22	4622±225	390±75

* ± std. error

Appendix E (2)

Growth rate variation data of F3

Family	OS1 (A) (Area in mm^2)	OS2 (S) (Area in mm^2)	GH1(A) (Area in mm^2)	GH2(S) (Area in mm^2)
2	1102±145	895±120	1053±132	1459±181
3	805±147	524±78	1277±280	1346±284
5	784±132	658±165	1073±66	1222±189
6	965±276	-	1516±236	387±95
7	711±46	-	-	315±75
8	1088±98	581±61	738±5	1119±122
9	835±138	1517±203	793±85	458±92
10	721±245	1193±272	-	1067±138
11	1126±87	1555±304	-	465±87
12	1164±104	1525±231	2060±682	914±77
13	854±186	1398±159	1117±129	715±89
14	678±162	1165±58	-	558±82
15	948±111	1506±79	992±275	674±156
16	668±143	978±163	1507±209	477±116
18	871±120	1507±204	1202±240	384±94
19	869±69	1224±117	1235±117	887±63

20	462±120	925±104	900±204	668±45
21	807±87	1055±193	373±82	570±104
22	1141±100	831±45	1083±364	536±79
23	914±77	1445±249	1431±244	539±125
24	874±52	1204±258	1268±273	673±58
25	829±132	1372±255	810±241	645±43
26	821±62	1253±311	650±95	642±85
27	1067±100	1433±271	1151±252	943±73
29	1228±142	1911±234	1221±240	512±77
30	819±271	1976±166	858±153	1125±134
31	663±154	1539±356	1107±148	860±79
32	746±41	978±246	1107±86	737±104
33	789±51	1311±255	921±85	802±86
34	864±163	1277±349	1334±213	616±178
35	1052±221	1791±237	1634±101	1052±71
36	842±157	1630±225	1168±75	1001±128
37	1315±196	1550±111	1253±320	1055±127
39	948±68	1694±248	1086±49	1256±112
40	896±42	2208±229	-	1328±85

41	793±41	1344±134	-	667±102
42	633±182	1184±153	1821±154	765±79
43	675±142	1218±93	1225±129	662±61
44	777±100	1097±116	-	595±82
45	861±226	1453±159	1251±335	804±81
46	713±73	1252±154	1122±148	454±24
47	947±77	979±221	1287±224	718±73
48	623±138	1160±85	877±59	445±80
49	789±138	1082±102	1271±246	198±42
50	1116±241	1533±228	--	199±47
51	910±127	1459±271	1788±318	857±128
52	794±22	1631±273	-	870±150
53	785±259	-	1146±220	-
54	-	1192±186	-	734±92
55	686±78	957±76	1555±102	225±77
56	902±293	1081±183	1342±443	395±95
57	1310±199	1289±273	1288±309	394±175
58	730±77	1550±164	1001±190	1149±108
59a	770±82	1482±294	1575±347	1147±112

59b	907±69	1669±63	741±122	768±180
60	763±167	1467±162	1011±330	944±155
61	874±120	1236±219	1122±483	826±102
62	901±184	1249±131	1438±323	817±131
63	1154±208	1127±101	396±71	402±98
64	1235±219	1194±83	1391±189	330±63
65a	844±239	1108±189	906±195	268±103
65b	577±171	706±83	1103±187	365±63
66	1019±108	1021±133	848±494	427±93
67	951±167	1300±179	-	351±88
68	968±113	922±326	1296±258	157±58
69	981±231	1211±270	-	365±128
70	935±131	1732±322	1355±241	149±43
71	1010±125	1041±327	1497±253	341±100
72	1069±110	1343±181	1359±244	510±68
73	975±104	1146±198	1056±328	458±130
75	825±61	1920±257	1741±255	212±96
76	850±131	865±264	294±172	302±91
78a	939±232	1191±122	1512±137	330±180

78b	930±88	1644±222	1102±186	538±50
79	996±159	1858±328	1776±164	733±141
81	978±121	1931±551	1123±497	714±135
82	483±87	1940±231	1066±154	604±155
83	887±210	1740±530	1399±289	553±201
84	975±69	1854±135	1805±112	341±104
85	856±76	1041±78	1091±68	641±73
4D1	700±84	1516±94	1575±260	1022±93
11C1	798±222	1665±73	1636±121	1643±88
11B1	351±	1328±145	1286±178	895±126

* ± std. error