The genetic and molecular basis of natural variation for plant growth and related traits in *Arabidopsis thaliana*

Bjorn Pieper

The genetic and molecular basis of natural variation for plant growth and related traits in *Arabidopsis thaliana*

Bjorn Pieper

Index

Chapter 1.

General introduction

Bjorn Pieper

Plant derived products play an important role in human society. They are important resources for industry, are an important part of our staple diet and that of animal stocks, and are becoming increasingly more important as a fuel source. Because of the ever-increasing size of the human population, the demand for plant-derived products can only be expected to rise. Millions of people suffer from hunger each year. Therefore, biotechnological innovations are both necessary and expected to improve the ability of plants to turn light energy into useful products in the coming 50 years (Borlaug, 2007). Biotechnology should be seen in a broad context as the application of fundamental knowledge acquired from research into plant growth and production. Increasing crop yield and biomass, with minimal use of other resources such as water, fertilizers and chemicals for crop protection is a major goal within this scope. A definition of plant growth is its increase in size or mass over time and is therefore directly involved in yield and biomass production where biomass production is concerned. That it includes the factor time brings it under the control of all relevant physiological processes, acting from the cellular level up to the level of the entire organism during growth. However, developmental aspects such as the transition from vegetative to generative growth also play a role in the control of vegetative growth (Cookson *et al.*, 2007; Tisné *et al.*, 2008). Plants are sessile organisms and need to be genetically adapted to their environment in order to ensure optimal growth and reproductive success. Plant growth is a complex trait that manifests itself from the combination of several component traits that each can have their own distinct genetic architecture that interact with environmental factors. There is a long-standing interest in understanding the genetic basis of complex traits, including yield, in crop plants. For most plant breeders yield is an important target for genetic improvement (Cooper *et al.*, 2009). This has also been the motivation for a considerable amount of (eco)physiological research on crop yield performance (Gonzalez *et al.*, 2008), particularly related to water-use efficiency (Tardieu, 2002; Condon *et al.*, 2004; Chenu *et al.*, 2008; Shao *et al.*, 2009). The current notion of a rapidly changing global climate has put further pressure on producing crop cultivars that will be able to cope with this without loosing productivity (Semenov and Halford, 2009). In contrast to the genetic and physiological basis, much less is known about the molecular basis of yield and plant growth and has only started to be elucidated. A detailed knowledge about the function of genes underlying variation for growth can provide a powerful tool to accelerate crop improvement.

The model plant *Arabidopsis thaliana*

Nowadays the study of plant genomics in crop species benefits from earlier studies on model species (Flavell, 2009). *Arabidopsis thaliana* (hereafter Arabidopsis), a small weed from the *Brassicaceae* family, became a major model species in plant research for its small size, rapid life-cycle, the small size of its fully sequenced genome and the efficiency of transforming it with the floral dip method (Somerville and Koornneef, 2002). This has led to a wealth of genetic and genomic resources currently available for this species, which includes artificially-induced mutants for almost every gene. Plant biology has benefited greatly from induced mutants and mutant screens continue as a powerful approach to the identification of gene effects (Page and Grossniklaus, 2002). Arabidopsis has a broad range of natural distribution throughout the Northern hemisphere from the north of Scandinavia down to Tanzania in Africa, at 2° latitude. Naturally occurring genetic variation caused by adaptation to the local environment provides an additional resource that can be used for the discovery of gene function. Furthermore, it has made Arabidopsis also attractive as model species for molecular ecology and evolution studies (Tonsor *et al.*, 2005; Mitchell-Olds and Schmitt, 2006; Shindo *et al.*, 2007). During vegetative growth, the Arabidopsis plant produces a flat rosette which size is solely dependent on the size and number of the individual leaves. The rosette leaf is the subject of many studies that contribute to the current understanding in the molecular basis of plant growth. Arabidopsis is currently the best characterized species with regard to growth-regulating molecular mechanisms (Gonzalez *et al.*, 2008).

Current status of the molecular basis of plant growth in Arabidopsis

Cell division and elongation rate are two fundamental processes that need to show variation to acquire differences in plant growth rate or final organ size. The cell cycle is regulated by a specific family of genes. Not long after the completion of the Arabidopsis genome sequence (AGI, 2000), a list of 61 core cell-cycle genes was published (Vandepoele *et al.*, 2002). The potential to influence plant growth by altering the expression of these genes has been demonstrated repeatedly in literature (reviewed in Beemster *et al.*, 2005). Although there seems to exist a relation between cell proliferation and cell expansion, this relation is poorly understood (Tsukaya, 2008). Cell expansion during growth depends on extension of the cell wall and is regulated by the expansin protein family (Li *et al.*, 2003). However, genetic control of plant growth is not limited to either the cell-cycle regulation machinery or mechanisms directly related to cell expansion. These processes themselves are rather regulated by both endogenous and environmental signalling pathways that converge on

them. A large number of genes that are involved in the control of organ growth have been identified, primarily by analysis of induced mutants in Arabidopsis, and their relevance for crop improvement has been evaluated (reviewed in Gonzalez *et al.*, 2008; and Krizek, 2009). Important conclusions in the above-cited recent reviews are that an integrated model of organ size is difficult to develop, but insight gained from studying growth regulation in Arabidopsis can be used for increasing crop productivity.

Natural variation and the analysis of its genetic and molecular basis

Natural variation as a resource

Induced mutagenesis has proven to be a powerful tool for the study of gene function, as illustrated by the number of genes discovered to be involved in control of plant growth. There are, however, certain limitations to this approach. Current mutant collections have typically been derived from mutagenesis of a limited number of genotypes, which excludes the possibility to discover genes that were intrinsically non-functional already. Furthermore, even within species, the gene content between accessions can vary substantially. These factors have been shown to add up to approximately 9.4% of all protein coding genes in the genome being naturally knocked-out in Arabidopsis accessions (Clark *et al.*, 2007). Secondly, many interesting traits have a complex genetic basis, which might include gene-by-gene and gene-by-environment interactions. Although components of such interactions can surely be identified by mutagenesis, the true genetic context remains elusive. The genetic basis of the tremendous diversity within species in nature provides a valuable asset that is less prone to the limitations of mutagenesis approaches. This natural variation has arisen from spontaneously occurring mutations that have been selected for within the context of the genetic background and the environment. Thus, naturally occurring genetic variation is a valuable resource for both crop improvement and fundamental research, in particular for traits related to adaptation (Benfey and Mitchell-Olds, 2008).

QTL analysis

Variation for many of the traits that are considered of interest is of quantitative nature. A characteristic of such a trait is continuous variation in a segregating population. The genetic basis of quantitative traits is characterised by multiple Mendelian loci, mostly of small and intermediate effect, commonly called quantitative trait loci (QTL). It was not until the 1980s that technological advances in the field of molecular biology made genotyping with

molecular markers readily available. This availability of large numbers of anchor points on the chromosomes facilitated that the genetic location of QTL and their contribution to trait variation could be readily estimated (Doerge, 2002). The analysis of the genetic basis of quantitative traits in populations derived from experimental crosses is known as QTL analysis, or QTL mapping. This procedure involves the evaluation of statistical associations between phenotypic variation and specific alleles at, and in between, marker loci. Statistically significant associations do thereby identify the QTL. In principle any population of offspring in which the alleles of the ancestor have segregated are suited for QTL analysis. The simplest population is an F2 but several more advanced types of QTL mapping populations are available that have certain advantages over the F2. In particular populations of inbred lines are advantageous. One example is the recombinant inbred line (RIL) population, which is produced through selfing and single seed descent of an F2 population, originally derived from a cross between two parental lines. Heterozygosity reduces by 50% after each generation of selfing with the result that after about 9 generations the individual lines are practically homozygous. This is a great advantage as it allows the exact same genetic material to be analyzed multiple times. Furthermore, the increased number of meiotic recombination events in each individual increases the resolution and power of QTL mapping without the need for an increased population size. Specialized software to perform QTL analysis has been available for about two decades now. The now classical methods of QTL analysis fit the effects of a single putative QTL (Lander and Botstein, 1989; Jansen and Stam, 1994). Through advances in, for instance, mixed model (Cooper *et al.*, 2009), and Bayesian methodology (Yi and Shriner, 2007) it is nowadays possible to assess genetic effects with far more complex models. The purpose has remained the same though; the identification and characterisation of loci at which allelic variation contributes to trait variation. However, the concept trait has broadened recently with advances in QTL mapping of high throughput "omics" data (Jansen *et al.*, 2009).

The motivation to develop methods of genetic analysis has been their application in the improvement of crop species (Alonso-Blanco *et al.*, 2009). In this field, a direct application of identified QTL is marker-assisted selection, which can increase the efficiency of plant breeding strategies (Hospital, 2009). Another purpose for QTL analysis can be found in the field of evolutionary biology. Examples of this found in literature include relating genetic variation to evolutionary changes in leaf shape and size in *Antirrhinum majus* (Langlade *et al.*, 2005), studying the regressive evolution of eyes in cave fish (Protas *et al.*, 2007), and the genetic architecture of plant-pollinator interactions (Noland *et al.*, 2008). The above exemplifies cases where information attained by QTL analysis is directly applied for practical purposes or to draw conclusions about biological systems. However,

there is an extended interest in elucidating the molecular basis of QTL to discover gene function and to study natural variation in the context of ecology and adaptation (Alonso-Blanco *et al.*, 2005; Tonsor *et al.*, 2005; Mitchell-Olds *et al.*, 2007; Benfey and Mitchell-Olds, 2008). The molecular basis can be defined as the gene(s), and more specifically the functional nucleotide polymorphism(s), causal to the QTL effects.

Elucidating the molecular basis of QTL

The resolution of QTL mapping is rather low so the physical genomic region covered by a QTL is large and could contain as many as 2000 genes (Price, 2006). Resolving a QTL to the functional polymorphisms is therefore not straightforward and often experimentally cumbersome. It requires that the multigenic nature of quantitative traits is first reduced to a situation where the QTL under scrutiny is the only source of genetic variation. One common strategy makes use of near isogenic lines (NILs), which carry an introgression of one genotype in the genetic background of the alternative genotype, in the region of the QTL. This NIL can then be analyzed for the presence of the expected QTL effects, so-called QTL validation. The NIL can subsequently be used to narrow down the region of the QTL by so-called fine-mapping. This involves associating the QTL phenotype to allelic values at marker loci in meiotic recombinants selected from progeny in which the initial introgression segregates. When the region containing the QTL is reduced sufficiently, candidate genes can be selected. However, further proof to positively identify a gene as causal to the QTL effects is needed when the QTL region contains multiple genes because DNA polymorphisms are rather ubiquitous between natural accessions. Several strategies have been described (Alonso-Blanco *et al.*, 2005), one of which is complementation of the QTL phenotype by transformation with the alternative allele of the genomic region containing the candidate gene. The elucidation of the molecular basis of QTL benefits greatly from the availability of genomics tools such as the genomic sequence for candidate gene selection, phenotypes of loss of function mutants and established protocols for molecular biology that e.g. allow efficient transformation. Thus, it comes of no surprise that in Arabidopsis most genes have been positively identified to be causal to QTL effects (Alonso-Blanco *et al.*, 2009). However, the number of sequenced plant genomes is increasing and genomic resources have been developed for other species, including rice, barley and tomato, which will aid in the identification of genes and functional polymorphisms underlying QTL for traits of interest (Sakamoto and Matsuoka, 2008; Sang, 2009).

Quantification and genetic analysis of natural variation for plant growth

The complexity of plant growth makes it hard to define it as a trait. It is therefore better described in terms of multiple growth traits, which are specifically related to growth rate and final plant size (Maloof, 2002). This potentially spans a broad range of traits but in this introduction some that are directly quantified on the whole plant or specific organs are described in some detail. The growth rate of a plant is a variable; derived from modelling the change in mass or size over time. One method for determining plant growth rate is by estimation of the relative growth rate (RGR), classically determined as the net increase in biomass per unit biomass per unit time, which reduces to the unit (e.g.) "per day (d^{-1})". In an eco-physiological context, plant species have been known to display large inherent differences for this trait when grown under optimal conditions (Grime and Hunt, 1975; Poorter and Remkes, 1990; Hunt and Cornelissen, 1997). On the contrary, within species the variation is much more constrained, as was demonstrated for Arabidopsis (Li *et al.*, 1998). For QTL analysis, the determination of RGR based on biomass is unpractical because a large population of plants needs to be grown simultaneously in order to acquire good estimates of biomass at several points in time. Nevertheless, this method has been successfully applied and QTL analysis for this trait has been performed in Arabidopsis (Meyer *et al.*, 2007) and barley (Elberse *et al.*, 2004; Poorter *et al.*, 2005). Static measurements of biomass have also been subjected to QTL analysis in Arabidopsis (Meyer *et al.*, 2007; Lisec *et al.*, 2008), including in the context of heterosis (Kusterer *et al.*, 2007a; Kusterer *et al.*, 2007b; Melchinger *et al.*, 2007). Digital image analysis has provided a more efficient method to quantify plant growth. Leister *et al*. (1999) have shown that the projected rosette area (PRA) of Arabidopsis, as measured from images, correlates well to actual biomass. This method has subsequently also been adopted to efficiently determine RGR, reducing the number of plants that need to be grown because the same plant can be measured repeatedly. QTL for RGR, based on PRA, have been detected in Arabidopsis, as well as QTL for PRA itself (El-Lithy *et al.*, 2004; Melchinger *et al.*, 2007). The size and shape, or morphometry, of individual rosette leaves of Arabidopsis has also been subjected to QTL analysis in the light of leaf architecture and development (Perez-Perez *et al.*, 2002; Juenger *et al.*, 2005a), but these traits can also be considered as growth related.

QTL analysis studies rarely include only a single trait and often QTL for different traits map to the same region. This genetic correlation can be either due to the QTL for different traits being discreet but closely genetically linked, or it can mean that a single QTL has pleiotropic effects on various traits. This co-location has also been shown to occur for flowering time QTL and plant growth traits, quantified on the rosette during vegetative

growth (El-Lithy *et al.*, 2004; Tisné *et al.*, 2008). The relation between timing of flowering and plant growth was described long ago, and was verified more recently by studying plant growth in Arabidopsis while modulating flowering time non-genetically (Cookson *et al.*, 2007 and references therein). The study of Tisné *et al.,* (2008) is the first to demonstrate that genetic effects on rosette and leaf level are directly correlated to variation at the cellular level. Using a different approach, a clear relation between the developmental switch to flowering and secondary growth has also been shown to exist in Arabidopsis (Sibout *et al.*, 2008). Thus, in Arabidopsis, flowering time should also be considered a growth-related trait.

The proper timing of flowering, hence reproductive growth, is important to assure maximum fitness but in the same line one could postulate that the preceding vegetative growth phase must also be successful for this purpose. In part, common genetic determinism of both growth phases does therefore not seem unreasonable. Over 80 genes have been described to be involved in the regulation of flowering time in Arabidopsis (Koornneef *et al.*, 1998; Simpson and Dean, 2002; Boss *et al.*, 2004; Turck *et al.*, 2008). Only nine of these genes are known be involved in natural variation for this trait (Alonso-Blanco *et al.*, 2009). Although, this does not exclude that for other flowering time genes, natural variation can exist. Co-location of flowering time and vegetative growth QTL in regions containing flowering time genes known to show natural variation has been observed on the top and bottom of chromosome 5 in two separate RIL populations (El-Lithy *et al.*, 2004; Tisné *et al.*, 2008). Two genes located on the top of chromosome 5 are *FLC* (Michaels and Amasino, 1999), and *HUA2* (Doyle *et al.*, 2005; Wang *et al.*, 2007) have been suggested as important candidates. Located on the bottom end of chromosome 5 are *MAF2-MAF5*, a tandem cluster of four MADS family genes that are close homologues of *FLC*, which were recently shown to harbour extensive genomic rearrangements (Caicedo *et al.*, 2009). Direct proof of natural variation at the nucleotide level in these genes being linked to phenotypical variation has not been demonstrated though. The discovery of co-locations between flowering time and growth traits might be the first explicit clues, a genuine common genetic determinism for both traits has not been proven.

The molecular basis of natural variation for plant growth

Only a single, small-effect, plant growth (biomass) QTL has the causal gene been positively identified. This QTL was discovered circumstantially while fine-mapping another QTL for an unrelated trait, but could be fine-mapped to a single gene that encoded a putative serine/threonine protein kinase (Kroymann and Mitchell-Olds, 2005). The authors found that this gene interacted epistatically with a second locus within a 210 kb interval of its location,

but also with other QTL in the genetic background. They pointed out that if this finding was representative, it predicted a highly polygenic, epistatic genetic architecture of complex traits, such as plant growth.

Scope of this thesis

The aim of the works presented in this thesis is the elucidation of the genetic and molecular basis of natural variation for plant growth in *Arabidopsis thaliana*. QTL analysis of plant growth forms the basis by the identification of loci involved in variation for this trait. QTL for plant growth that had been detected within a broader scope of the genetic analysis of plant performance (El-Lithy *et al.*, 2004) were to serve as starting material for the present study. Validation of these QTL failed to reproduce the predicted QTL effects on plant growth. Where they were validated, they appeared genetically correlated to strong effects on flowering time. Pleiotropic effects of already known flowering time pathways on plant growth was to be avoided but this can not be distinguished from the possibility of closely linked QTL affecting both traits separately. One such QTL was chosen to be fine-mapped and cloned, in part because natural genetic variation for flowering time of this locus had also not yet been described. Furthermore, if the observed effects on flowering time and plant growth were pleiotropic, this would establish that flowering time pathways do also directly control plant growth. A QTL mapping experiment was performed that specifically focussed on the analysis of plant growth. Two Arabidopsis RIL populations were analyzed simultaneously to capture a larger portion of the potential genetic architecture and multi-trait QTL mapping methodology was applied to account for the complexity of plant growth. The detected QTL were to provide a solid basis for the elucidation of the molecular basis of plant growth. The development of a set of NILs and their analysis for validation of the QTL is a further step towards satisfying this aim.

Chapter 2

Quantitative trait loci analysis of plant growth in two *Arabidopsis thaliana* RIL populations by using multi-trait methodology based on mixed-models

Bjorn Pieper[1], Marcos Malosetti[2], Maarten Koornneef[1,3], Fred A. van Eeuwijk[2], Matthieu Reymond[1]

[1]Department of Plant Breeding and Genetics, Max Planck Institute for Plant Breeding Research, Cologne, Germany. [2]Biometris, Wageningen UR, Wageningen, The Netherlands. [3]Laboratory of Genetics, Wageningen UR, Wageningen, The Netherlands.

Abstract

The *Arabidopsis thaliana* recombinant inbred line (RIL) populations derived from crosses of Kashmir-2 and Shakdara to Landsberg *erecta* respectively have been analyzed for 28 plant growth related traits. Projected rosette area (PRA), Feret diameter of the rosette, and area, length and width of the largest leaf of each plant were quantified using non-destructive digital image analysis. A logistic model was fitted to PRA to model the entire rosette expansion phase in order to determine maximum growth rate and duration of increasing rosette expansion rate. Additionally, the relative growth rate (RGR) was estimated from the initial exponential expansion phase only. Furthermore, flowering time and leaf were quantified as well as chlorophyll content index and plant height. Mixed-model based multi-trait QTL analysis, performed on a selection of 15 traits simultaneously was used to include information of the extensive correlation structure of the traits. A total of 18 QTL were detected over both RIL populations. Plant growth and flowering time were found to be strongly genetically correlated. Nevertheless, seven QTL specific to plant growth were also detected. In addition, 20 significant 2-way interactions were found between the 18 detected QTL. Particularly large amounts of variation for traits directly related to plant growth such as RGR, PRA and the dimensions of the largest leaf were explained by these interactions. The detected QTL for plant growth are to be subjected to fine-mapping in order to clone the underlying genes and to describe their molecular basis.

Introduction

Plant growth is a complex trait that is regulated by both genetic and environmental factors. Both these aspects are continuously exploited for improved crop production to meet our ever increasing demand for plant derived products. Knowledge of the physiological, genetic and molecular basis of plant growth can greatly facilitate crop improvement and production, as well as our understanding of biological systems. The basis of differences in plant growth both between and within species has been described at the level of physiology and morphology from the whole plant down to the cell level. The analysis of growth at the molecular level has been limited mainly because of the complexity of the trait. Nevertheless, progress has been made over the last decade, mainly, but certainly not exclusively, through reverse genetic approaches using mutants and gene over expression of candidate genes in *Arabidopsis thaliana* (see Krizek, 2009 and references therein).

Fundamental to variation in size between plants is variation in either cell number, cell size, or both. Plant growth is characterized by the rate at which a plant increases its size over time and is thus a function of the rate of cell division and cell expansion. The complexity in plant growth arises from the plethora of signalling and regulatory pathways and their interactions that are all integrated to control these two processes. Plant hormones are important regulators of plant growth and development that mediate both the endogenous and environmental signals (Gray, 2004). Molecular mechanisms involved in hormonal control have been elucidated (Wolters and Jürgens, 2009). However, the discovery of genes such as *BIG BROTHER* (Disch *et al.*, 2006) and KLUH (Anastasiou *et al.*, 2007) indicate that there are yet undiscovered signalling compounds involved in plant growth. Reverse genetic approaches using the Arabidopsis leaf as a model have led to the discovery of these, and many more genes involved in its growth. Additionally, there is evidence that growth is also regulated at the whole plant level, closely related to the developmental state of the plant (Cookson *et al.*, 2007; Tisné *et al.*, 2008). These signals seem to originate from the apical meristem as its removal leads to responses on the cellular level in developing organs (Tsukaya, 2002; Cookson *et al.*, 2007).

The study of induced mutants in Arabidopsis has proven a powerful tool for the discovery of genes involved in plant growth. A drawback to this method is that naturally non-functional genes in the mutated genetic material are hard to discover. Furthermore, there are indications that plant growth has a complex genetic architecture that includes epistatic interactions (Kroymann and Mitchell-Olds, 2005). Although single components of epistatic interactions can be identified through mutagenesis, the true genetic context remains elusive. Forward genetics using naturally occurring variation provides the possibility to overcome

such limitations. Natural variation for plant growth has a quantitative genetic basis and reflects a genetic adaptation to the environment. Such variation has been demonstrated to exist at the shoot level, and the root level (Beemster *et al.*, 2002) in Arabidopsis. This resource of variation could provide a valuable complement to the information gained through reverse methods using Arabidopsis.

QTL mapping is a key tool for studying the genetic architecture of complex traits in plants (Holland, 2007). Several reports on QTL analysis of growth-related traits of the shoot in Arabidopsis have been published. QTL for rosette area, relative growth rate and dry weight of the rosette have been detected (El-Lithy *et al.*, 2004) as well as QTL for leaf and floral organ size (Juenger *et al.*, 2005b), and for morphometric parameters of the leaf (Perez-Perez *et al.*, 2002). QTL for growth quantified as biomass were detected in a study of heterosis (Kusterer *et al.*, 2007). Furthermore, QTL analysis of biomass and metabolite levels aimed at studying the relationship between these traits has also been performed (Meyer *et al.*, 2007; Lisec *et al.*, 2008). A QTL analysis of cell number and cell area of the 6th rosette leaf, and total rosette leaf area and leaf number has also been performed (Tisné *et al.*, 2008). This study underlined the dependency of leaf variables on both leaf and whole plant control mechanisms. A strong genetic correlation between QTL effects on the cellular level and the size of the rosette was observed and, additionally, the majority of variance was explained by epistatic interactions. The ultimate goal of QTL mapping in this context is the identification of the genes harbouring the DNA sequence polymorphism(s) causal to the observed QTL effects. This is generally accomplished through map-based cloning by fine-mapping, followed by genomic complementation. The number of genes that has been cloned using this strategy is rising steadily, however, in the majority of cases the cloned QTL were of major effect, accounting for most of the explained variance (Alonso-Blanco *et al.*, 2009). Although this suggests only such QTL are suited for map-based cloning, this is not necessarily true. The choice to clone such QTL is obvious because they pose the most important contribution to the observed trait variation. However, also QTL of smaller effect, can be successfully fine-mapped (Röder *et al.*, 2008), even down to a single gene, providing direct prove of its causality of the QTL effects (Kroymann and Mitchell-Olds, 2005). Moreover, the latter reference is the only report that describes the identification of a gene underlying a QTL for plant growth in Arabidopsis.

The aim of the study presented herein was to elucidate the genetic basis of plant growth in Arabidopsis using state of the art multiple trait methods and multiple population analysis. The long term goal is the elucidation of the underlying molecular basis of the identified allelic variation. Two RIL populations have been analyzed to capture a larger part of the potential genetic architecture. Several growth-related traits were quantified that were chosen

such that in concert they covered growth at the whole plant level. Non-destructive digital image analysis was adopted to quantify the projected rosette area (PRA) and Ferret diameter of the Arabidopsis rosette repeatedly in time, from the seedling until a maximum was reached. This method for quantification of growth was first demonstrated to be adequate for Arabidopsis by Leister *et al.* (1999), and has since been successfully applied (El-Lithy *et al.*, 2004; Granier *et al.*, 2006). Measurements of PRA were known to be strongly genetically correlated to actual plant biomass (El-Lithy *et al.*, 2004). Two different models were fitted to the PRA data in order to quantify the growth rate of the plants and the duration of rosette expansion. Genetic control of leaf growth is multi-factorial with respect to the rank of the leaf (Perez-Perez *et al.*, 2002). Early measurements of PRA do capture differences of low ranked leaves but not of higher ranked ones because PRA gets increasingly confounded by leaf overlap. Therefore, the largest leaf of each plant was harvested and its dimensions were quantified. Finally, the dependency of plant growth variables on the developmental state of the plant were taken into account by also quantifying flowering time and related traits such as leaf numbers. Genetic correlations could be expected and were often found as co-locations of QTL for different traits in the studies referenced above. By modelling these correlations while fitting QTL effects, they provide increased power to detect QTL and the data is analyzed more realistically as opposed to a collection of single trait analyses that assume no genetic correlations (Malosetti *et al.*, 2008). By applying multi-trait mixed-model methodology for this purpose, 18 QTL could be detected considering the two Arabidopsis RIL population that were analyzed. The results show that QTL for flowering time affect plant growth rate in a predictable manor as a strong trend for negative correlation between QTL effects for both traits was found at these loci. However, a considerable amount of genetic variance was explained by plant growth-specific QTL, which showed that this trait also has a distinct genetic basis. These detected QTL will provide the basis for future studies aimed at the elucidation of the underlying molecular basis.

Materials and Methods

Genetic material and environmental conditions

Two recombinant inbred line (RIL) populations, derived from the cross between Landsberg *erecta* (L*er*) x Kashmir-2 (Kas-2) (El-Lithy *et al.*, 2006) and L*er* x Shakdara (Sha) (Clerkx *et al.*, 2004) respectively were kindly provided by the laboratory of genetics of Wageningen University (Wageningen, The Netherlands). The experiment was performed using F10 and F11 generations of the L*er* x Kas-2 & L*er* x Sha respectively. Of the L*er* x Kas-2 population,

a selection of 144 RILs was made, out of the 164 lines in total, excluding lines with poor genotyping information. Seeds were stratified at 4°C on water saturated filter paper for 4 days prior to sowing. The seeds were sown in square pots measuring 7cm that contained a sand peat mixture enriched with slow releasing nutrients. Three plants per RIL and parental accession were analyzed in the experiment. Pots were distributed randomly over 2 growth chambers (Percival AR95L/3). The day-length was set for a 12 hour light period and a 12 hour dark period. During the first and last 15 minutes of the light period only incandescent light bulbs were switched on to simulate dawn and dusk while for the rest of the light period fluorescent lamps were also switched on. The total illumination at plant level with all lamps switched on was 170-200 $\mu mol \cdot m^{-2} \cdot s^{-1}$ depending on the position in the chamber. The pots were randomly re-distributed over both growth chambers every 2-3 days to minimize positional effects. The temperature was set at 22°C during the day and 18°C during the night and the relative air humidity at a constant 70%. Constant environmental conditions could be guaranteed by continuous monitoring with data loggers (HOBO[®] U12-012).

Digital imaging and growth modelling

Starting from 8 days after sowing (DAS) the trays were photographed every 2-4 days from above with a ccd-camera (Sony DSC-F828). When only very late flowering plants were left to be photographed, time intervals were relaxed somewhat to 5-8 days. The images measured 8 megapixels and never captured more than 35 pots. They were stored in low-compression JPEG format. Imaging was continued until at least 1-1.5 weeks after the plants had flowered. The projected rosette area (PRA) and Feret diameter (FD; maximum diameter) of each plant was quantified with a dedicated image analysis package (Image-Pro analyzer 6.0, Media Cybernetics). PRA an FD were measured at up to 12 time-points, depending on the flowering time of the plants, starting from the seedling at 8 DAS through 14, 18, 22, 24, 27, 30, 34, 37, 44 and 55 DAS. The final PRA (PRAmax) and FD (FDmax) of each plant were measured from an image taken about 1-1.5 weeks after flowering time.

A logistic curve was fitted to the development of PRA in time using the FITCURVE directive of Genstat version 10 (Payne *et al.*, 2007). PRA measuring was stopped after a maximum size had been reached but to allow enough information for the curve to be fitted, this final measurement was duplicated at an additional time-point. This point was placed at a distance from the date of the actual measurement, equal to the average time between all preceding PRA measurements. The following model was fitted to the PRA data and the time of measurement in DAS:

$$PRA = A + \frac{C}{1+e^{-B*(time-M)}} + \epsilon$$

Where A is the lower asymptote, this estimate was not used for further analysis. The parameter *C* provided an estimate of the upper asymptote, the final PRA reached by each plant. The parameter *M* provided an estimate of the inflection point of the PRA growth curve and the parameter *B* an estimate of the maximum achieved expansion rate. The R^2 of the fitted logistic model ranged from 0.95 to 1.00 indicating a good fit.

Additionally, linear regression of the natural logarithm of the PRA was performed on time for each plant. The regression coefficient provided an estimate of the relative growth rate (RGR) (d^{-1}). For each individual plant only PRA measurements that preceded its estimate of the inflection point from the logistic regression were included to ensure that PRA was still expanding exponentially. Furthermore, PRA at 8 DAS was excluded from the estimation of the RGR because it consistently did not fit well to the exponential model. At this date all plants were still seedlings and consisted mainly of the cotyledons. Before the first 2 leaves started to dominate the PRA, rosette expansion rate deviated from an exponential model.

Additional phenotypes quantified

Bolting time (BT) was recorded as the time, in DAS, at which the bolt started to rise above the rosette. Flowering time (FT) was scored as the DAS at which the first flower opened. After flowering time the rosette leaf number (RLN) was counted, as well as the cauline leaf number (CLN), the sum of both yielded the total leaf number (TLN). At flowering time the CCI (Opti-Sciences CCM200) was measured on three of the largest rosette leaves from each plant. The average of the three measurements was taken to represent the CCI of each plant. After the plants were photographed for determination of the final PRA, the largest leaf of each plant was harvested. The leaves were stuck to double-sided tape and were scanned at 300 DPI using a colour scanner. The resulting images were analyzed by digital image analysis (Image-Pro analyzer 6.0, Media Cybernetics) in order to determine area (LLA), maximum length (LLL) and maximum width (LLW) of the largest leaves of each plant. Finally, during harvesting of the largest leaves, the height of each plant was measured (plH).

Genotyping and genetic map construction

Plant material was harvested from the each RIL and DNA was isolated using a BioSprint Workstation and the appropriate reagent/disposable kit (Qiagen). Both RIL populations had

been genotyped already and genetic maps were available (El-Lithy *et al.* 2004; 2006). However, 3 SSLP markers (T2N18, T3K9 and F17A22) were added to the bottom end of linkage group 2 of the genetic map of the L*er* x Sha RIL population as marker density was relatively low in this region. Both RIL populations were genotyped with a CAPS marker designed specifically for the L*er hua2-5* allele (Doyle *et al.*, 2005) which was included on the genetic maps. Additionally, several SSLP markers were added to both populations' genetic maps to facilitate map integration. In particular SSLP markers already present on the L*er* x Sha map were used to genotype the L*er* x Kas-2 population for this purpose, including the newly added T2N18 and T3K9. In addition, two markers (med24d and msat5-17) were added in both populations' genetic maps. A single integrated genetic map was produced from the genotypic data of both populations with JoinMap 4.0 (Stam, 1993). Map integration was facilitated by 46 markers common to both populations and 6 markers pairs that were considered as a single markers due to very linked physical positions (<100 Kb apart). The integrated genetic map consisted of a total of 119 markers.

Statistical analysis and QTL mapping

All statistical analyses were performed using Genstat version 10 (Payne *et al.*, 2007). Diagnostic biplots were produced using the BIPLOT procedure after using the MULTMISSING procedure to estimate the missing values in the dataset. All other analyses were performed using the actual data without estimating missing values.

A set of 15 correlated traits was selected from the total of 28 quantified phenotypic traits and growth model parameters for multi-trait QTL mapping. Several traits in the total dataset were so highly correlated to each other that they provided redundant information, in which case only one of the correlated traits was retained. None of the FD measurements was included in the selected set due to high correlation with the respective PRA measurements. Only 3 PRA measurements were selected to prevent over weighting this trait in the correlation structure. Additionally, all PRA measurements were already represented in the model parameters. All 3 estimated parameters of the logistic model were selected to be able to account for residual correlation between them. The set of selected traits consisted of: PRA at 8, 14 and 27 DAS, RLN, CLN, FT, LLA, LLL, LLW, the parameters B, M and C, RGR, CCI and plH.

Mixed model methodology was applied for statistical analysis of the phenotypical data and multi-trait QTL mapping. The methodology presented here is derived from recent advances in multi-trait multi-environment QTL mapping using mixed models (Malosetti *et al.*, 2008).

The following phenotypic model served as the basis for extension to a model including QTL. The convention to underline random variables is adopted.

$$\underline{Y}_{ijt} = \mu_t + \underline{u}_{tj} + \underline{e}_{tij} \tag{1}$$

Where $\underline{Y}_{ijt}$ is the observed phenotype of individual i ($i=1…I$) with genotype j ($j=1…J$), for trait t ($t=1…T$). The trait-dependent intercept term μ_t represents the population mean for trait t ($t=1…T$). The random genetic effect on trait t of the individuals with genotype j is modelled by the term $\underline{u}_{tj}$. These effects were assumed multi-Normally distributed. The last term $\underline{e}_{tij}$ models the non-genetic residuals, effectively collecting the deviations from the predicted values of the model for each trait per individual. These are also assumed normally distributed.

For multi-trait simple interval mapping (MTSIM) this model was extended to include the fixed effect of a single putative QTL, leading to the following model:

$$\underline{Y}_{ijt} = \mu_t + x_i\alpha_t + \underline{u}_{tj}^* + \underline{e}_{tij} \tag{2}$$

Where x_i is the additive fixed genetic predictor for the locus under evaluation in individual i ($i=1…I$) and α_t is the additive effect of this locus on trait t ($t=1…T$). From equation 2 onwards the random genetic effects per trait $\underline{u}_{tj}^*$ are caused by QTL effects that are not linked to the genetic predictors in the fixed terms. This is denoted by the asterisk (*).

The model was extended further to perform multi-trait composite interval mapping (MTCIM):

$$\underline{Y}_{ijt} = \mu_t + \sum_{s=1}^{S} x_{is}\alpha_{ts} + x_{iq}\alpha_{tq} + \underline{u}_{tj}^* + \underline{e}_{tij} \tag{3}$$

Here, significant QTL effects detected at loci $1…S$ were fitted to correct for genetic background effects while testing for a QTL at locus q.

The presence of non additive two-way interactions between the detected QTL was assessed with the following, final, extension of the model:

$$\underline{Y}_{ijt} = \mu_t + \sum_{v,w=1}^{V,W} x_{iv}x_{iw}\alpha_{tvw} + \sum_{s=1}^{S} x_{is}\alpha_{ts} + \underline{u}_{tj}^* + \underline{e}_{tij} \tag{4}$$

The additive effects of all detected QTL are modelled by the QTL indexed by 1…*S*. The significant two-way interactions are the product of genetic predictors v (v=1…V and V ϵ S) and w (w=1…W and W ϵ S) in individual i, where $v \neq w$, and non-additive effects α_{tvw} on trait t (t=1…T).

The models described above were fitted to the phenotypic data using the REML directive of Genstat. An unstructured covariance model for both random terms in the models was used as it gave the best results. To avoid numerical problems, traits were rescaled to have a similar range of magnitude. Equation 1 was fitted to estimate the genetic variance component σ^2_g and the non-genetic variance component σ^2_e. The heritability for each trait was calculated as $H^2 = \sigma^2_g/(\sigma^2_g + \sigma^2_e/n)$ with n the average number of observations per genotype.

Genetic predictors were calculated from the molecular marker data and the integrated genetic map at intervals no larger than 2 cM. At marker loci, the homozygous L*er* allele genotype took the value -1 and the homozygous genotype for the alternative allele the value 1. Firstly, MTSIM genome scans were performed for both RIL populations by fitting equation (2) at each genetic predictor. At each interval the null hypothesis of no QTL at the respective genetic predictor was tested by a Wald test. The $-\log_{10}$ of the χ^2 probability for the Wald test statistic with 15 degrees of freedom was calculated. The Wald test statistics was also computed for each separate trait by dividing their squared regression coefficients by the diagonal of the estimated variance-covariance matrix of the genetic predictor under evaluation (Buse, 1982). The corresponding $-\log_{10}(p)$ values were calculated (χ^2 distribution with 1 degree of freedom). The significance threshold for QTL mapping, was calculated according to Li and Yi (2005) with a genome-wide type I error rate α=0.05. The results of QTL mapping were visualized by plotting $-\log_{10}(p)$ values against the genetic position. Secondly, MTCIM genome scans were performed. Genetic predictors with the highest $-\log_{10}(p)$ values at QTL detected by MTSIM were chosen as co-factors in equation 3. When the genetic predictor under evaluation was located less than 8cM from a co-factor, then this co-factor was temporarily removed. MTCIM scans were repeated with newly discovered QTL added to the co-factor set after every subsequent genome scan until no more new QTL were detected. A backwards elimination was performed on all detected QTL to end up with the final additive model, containing only significant QTL. The Wald test statistic of each QTL was evaluated for a number of degrees of freedom equal its number of significant allelic effects. The σ^2_g explained for each trait by the final additive QTL model (Eq.3) was calculated as the percentage difference from σ^2_g estimated with the final model and with equation (1). Each detected QTL was dropped from the full model to determine its

contribution to the total σ^2_g explained per trait conditional on the other QTL in the model. This was calculated as the difference in σ^2_g per trait as compared to the full model, in percent of σ^2_g estimated by equation (1). Significance of allelic effects on individual traits were assessed by their standard errors using $\alpha=0.05$ (effect is significant when ≥ 1.96*standard error). Due to the range of the values used as genetic predictors, the allelic effects represent a single allele substitution of a L*er* allele with an allele of the other accession. The allelic effects are expressed as the percentage of the respective population mean. Finally, non additive effects of two-way epistatic interactions between the detected QTL were determined by fitting equation 4. A backwards elimination of non-significant interactions was performed starting from a model including all possible pair wise combinations. Eventually, weakly significant interactions with marginal allelic effect on few traits, accounting for little genetic variance explained were also discarded. The σ^2_g explained by the model including interactions was calculated like for the additive model. Likewise, the contribution of each detected interaction was determined as was done for the additive QTL. It should be noted that the sign of the significant allelic effects for the interactions has lost its property to predict phenotypic effects of allele substitutions.

Results

Phenotypic correlations between the quantified traits

A set of 28 traits has been used to quantify growth and development of all the plants. The biplots in figure 1 show the correlation between traits. Clustering in the correlation structure reflected the parts of the plant, time of measurement and the relation to developmental aspects of the traits that were quantified. These could therefore be sub-divided in the following three distinct groups: i) the rosette during vegetative growth, ii) the final state of the rosette, and iii) flowering time (fig. 1). Correlations were similar between both RIL populations. However, in the LK population the correlations between traits from the final state of the rosette and those related to flowering time were stronger than in the LS population. RLN and TLN clustered tightly together and were strongly positively correlated to FT. A third tight cluster was formed by the parameter C and PRAmax, which were strongly positively correlated to both FDmax and to RLN. The parameter B was very strongly negatively correlated to the flowering time traits in both mapping populations. On the other hand, the correlation between RGR and those traits was much weaker, even though both are estimates for plant growth rate. RGR showed a strong positive correlation to PRA and FD during vegetative growth. PRA and FD from 8 till 30 DAS formed a tight cluster in

the LK population. However, in the LS population a negative correlation was found between PRA (and FD) at 8 DAS, and FT that respectively decreased in strength for PRA and FD measured at later time-points. Those later measurements on the rosette in the LS population, and all measurements on the rosette in the LK population (PRAmax and FDmax excepted in both populations), were only weakly negatively correlated to FT. The group representing measurements on the final state of the rosette was moderately positively correlated to traits from both the groups i and ii. The largest differences between both populations were observed for CLN and plH. In the LS population CLN was strongly correlated to RLN while in the LK population CLN was positioned very close to the origin of the biplots indicating a lack of correlations with the other traits. In the LS population plH was most strongly, and positively, correlated to FDmax and less strongly to the flowering time traits. In the LK population, on the other hand, plH was weakly correlated to FDmax and a moderately strong negative correlation with the flowering time traits was found.
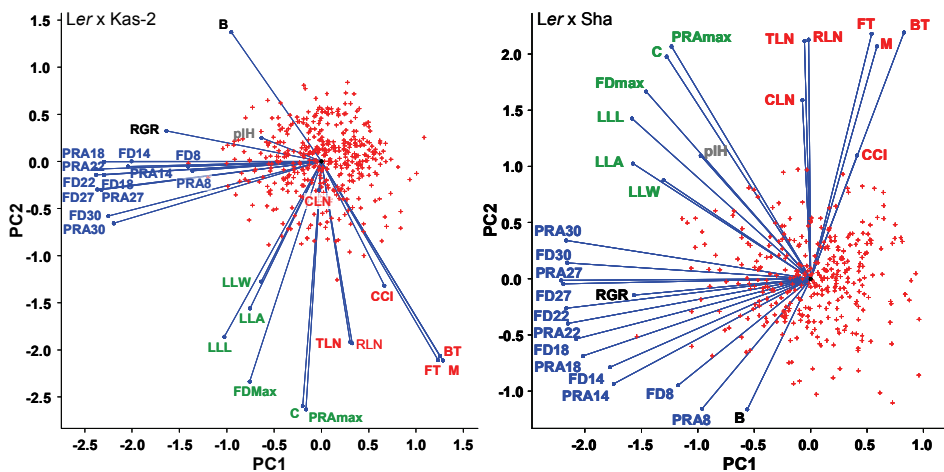
Figure 1. Diagnostic biplots of the 28 traits that were quantified in the L*er* x Kas-2 (left) and L*er* x Sha (right) populations. Three distinct clusters of traits are indicated by the colors: blue – the rosette during vegetative growth, green – final state of the rosette, red – flowering time related traits. The growth rate traits RGR and parameter B are indicated in black and plant height in grey. Notice that both biplots have a similar structure but that they are mirrored along a horizontal axis.

The selection of 15 traits for multi-trait QTL mapping

A reduced set of 15 traits that represented the correlation structure but had roughly equal weight in the three groups of traits was selected to perform multi-trait QTL mapping, which are summarized in table 1. All traits showed transgressive segregation in both populations. For each of the 15 selected traits and for both RIL populations the genetic variance component was significantly greater than zero. The genetic variance components for RLN, FT, LLA, LLW, parameters B and M, CCI and plH were much larger in the LK population than in the LS population, in which it was much larger for CLN. Heritabilities (H^2) were calculated to determine the genetic contribution to the observed phenotypic variance. Overall, the H^2 of the 15 traits were high (>0.50) in both mapping populations with flowering time related traits showing the highest values (0.96/0.97 for RLN, and 0.91/0.93 for FT, LK/LS respectively). The parameters M and C were also among the traits with the highest heritabilities whereas H^2 for the plant growth-rate traits parameter B and RGR where amongst the lowest. Of both, RGR showed the lowest H^2 (0.61) in the LK population while the growth rate parameter B showed the lowest H^2 in the LS population (0.55).

QTL detection by multi-trait mapping

QTL analysis for the 15 selected traits simultaneously using a multi-trait composite interval mapping method resulted in the detection of 18 significant main effect QTL over both RIL populations (figure 2). Seven of these QTL were common to both populations (QTL5,6,7,10,15,17,18), four were unique to the LS population (QTL2,4,9,13) and seven were unique to the LK population (QTL1,3,8,11,12,14,16). The total genetic variance explained by the detected QTL for each trait is summarized in table 1. The highest explained genetic variances observed were 80.6% for FT in the LK population and 83.1% for the parameter B in the LS population. The lowest percentage of genetic variance explained was 31.4% for PRA at 8 DAS in the LS population and 34% for PRA at 27 DAS in the LK population. For most QTL the significance of the multi-trait test was considerably higher than for any of the individual traits (figure 1). In particular for QTL7 in both populations and for QTL16 in the LK population the significance of the multi-trait test was greatly increased over the significance of any individual trait. These were also the most significant QTL detected. QTL1 and QTL11 were not significant in the multi-trait test but the single trait tests indicated suggestive QTL (adjustment of the threshold for multiple testing would have been required for proper interpretation of the single trait tests). These QTL were maintained in the final QTL model because they did have significant and relatively strong effects on

Table 1. Summary of QTL analysis. The ranges, means, and variance components presented in this table are derived from the data with adjusted range. The range of the traits is given as the minimum and maximum trait value observed; μ - trait means of the respective RIL populations estimated by fitting the phenotypic mixed model, and arithmetic trait means of the three founder accessions; SE- standard error of the mean. σ^2_g – genetic variance component; σ^2_e – residual variance component. H^2 –heritability. σ^2_g explained – percent reduction in σ^2_g after fitting the additive QTL model (Main effect) and additive + two-way interaction QTL model (Main eff. + interaction).

| trait | pop | Ler | | Accession | | RIL population | | | | | | | σ^2_g explained | |
		μ	SE	μ	SE	min	max	μ	SE	σ^2_g	σ^2_e	H^2	Main effect	Main eff. + interaction
PRA at 8 DAS	LK	36.5	4.2	52.9	9.2	7.8	87.6	49.4	0.9	68.6	104.9	0.66	52.8	69.7
	LS	59.0	6.8	45.1	5.4	23.1	126.2	58.6	1.0	67.6	127.9	0.61	31.4	50.0
PRA at 14 DAS	LK	41.7	2.9	90.1	21.2	13.2	93.2	49.4	1.0	102.7	121.4	0.72	44.7	56.8
	LS	56.5	3.9	52.8	0.3	27.4	126.3	58.6	1.1	97.8	119.3	0.71	46.4	76.6
PRA at 27 DAS	LK	81.7	5.9	117.0	8.1	4.5	170.0	49.6	1.7	306.8	363.3	0.71	34.0	57.9
	LS	64.0	4.6	73.4	8.3	12.4	140.2	59.0	2.2	418.7	324.9	0.79	60.7	75.5
RLN	LK	30.9	1.0	82.1	23.6	19.5	192.6	47.5	2.2	678.1	73.1	0.96	60.8	76.6
	LS	46.8	1.6	59.1	2.1	29.5	132.9	57.9	1.8	371.8	33.9	0.97	61.1	74.2
CLN	LK	47.7	3.7	52.3	6.2	18.4	101.4	48.6	1.0	124.7	54.9	0.85	53.9	65.5
	LS	56.4	4.4	43.7	0.0	10.9	120.2	58.3	1.5	227.4	66.5	0.90	49.0	64.6
FT	LK	44.7	1.5	56.4	6.2	33.6	88.7	49.3	0.8	83.3	18.2	0.91	80.6	90.8
	LS	57.4	1.9	58.1	1.5	43.2	79.4	58.6	0.6	41.1	7.3	0.93	61.7	75.2
LLA	LK	75.5	3.3	56.1	12.5	9.9	103.1	48.9	1.4	212.5	125.7	0.78	36.8	64.9
	LS	69.3	3.0	78.7	16.1	11.3	97.8	58.6	1.3	164.9	68.4	0.83	54.2	61.7
LLL	LK	54.7	1.8	58.1	9.3	21.7	79.0	49.1	0.9	93.1	45.3	0.81	48.6	67.1
	LS	61.6	2.0	69.8	6.9	22.1	89.1	58.6	1.0	102.3	24.7	0.89	69.4	78.5
LLW	LK	61.8	1.0	51.2	4.2	23.1	98.0	49.1	0.8	69.4	40.0	0.78	41.8	68.6
	LS	63.9	1.0	67.2	7.4	29.5	80.7	58.6	0.7	46.7	16.8	0.85	44.2	51.1
B	LK	58.0	3.4	40.5	7.9	26.8	79.2	49.3	0.7	55.8	44.9	0.72	60.3	72.2
	LS	56.9	3.3	58.3	6.7	38.5	93.8	58.1	0.7	30.5	51.1	0.55	83.1	94.2
M	LK	47.2	0.3	48.5	0.7	28.0	71.0	49.5	0.6	53.9	12.5	0.90	74.7	84.7
	LS	59.4	0.4	59.5	0.4	39.5	77.2	59.1	0.5	27.4	8.7	0.87	60.0	81.9
C	LK	62.0	4.6	84.9	0.6	6.0	144.1	48.3	2.1	582.8	173.6	0.87	61.5	76.0
	LS	62.0	4.6	82.9	19.8	5.0	136.8	59.6	2.3	549.0	115.3	0.91	69.2	77.2
RGR	LK	62.8	2.1	54.2	5.5	23.2	95.7	49.4	0.6	30.8	58.0	0.61	67.5	91.1
	LS	59.8	2.0	65.6	2.8	33.0	81.4	58.6	0.7	34.6	44.0	0.70	48.1	69.2
CCI	LK	34.6	1.7	55.6	18.3	17.9	210.6	49.1	2.1	505.2	249.3	0.80	67.8	80.5
	LS	53.3	2.7	42.7	9.8	24.9	108.8	58.8	1.2	126.7	79.9	0.75	61.1	74.4
plH	LK	67.9	13.0	51.0	18.0	3.0	139.6	49.9	2.6	723.5	215.9	0.84	75.6	75.3
	LS	59.8	11.4	100.4	5.3	10.6	126.9	58.0	2.5	647.4	138.6	0.90	64.9	68.2

plant growth-specific traits nevertheless. As more cofactors were included for subsequent composite interval mapping genome scans, some significant multi-trait QTL were detected eventually that were not included in the final QTL models. These QTL were found to have negligible contribution to the total explained genetic variance.

All of the 18 detected QTL had effects on multiple traits except QTL6, which only had a significant effect on CLN in the LK population. Table 2 summarises the significant allelic effects and the genetic variance explained by each of the detected main effect QTL for each trait in both mapping populations. Although for most traits QTL with large effects were detected, this was not the case for both growth-rate parameters B and RGR. The largest allelic effects found were 5.2% for RGR and 6.7% for parameter B although the genetic variance explained per QTL for these traits did show extreme values.
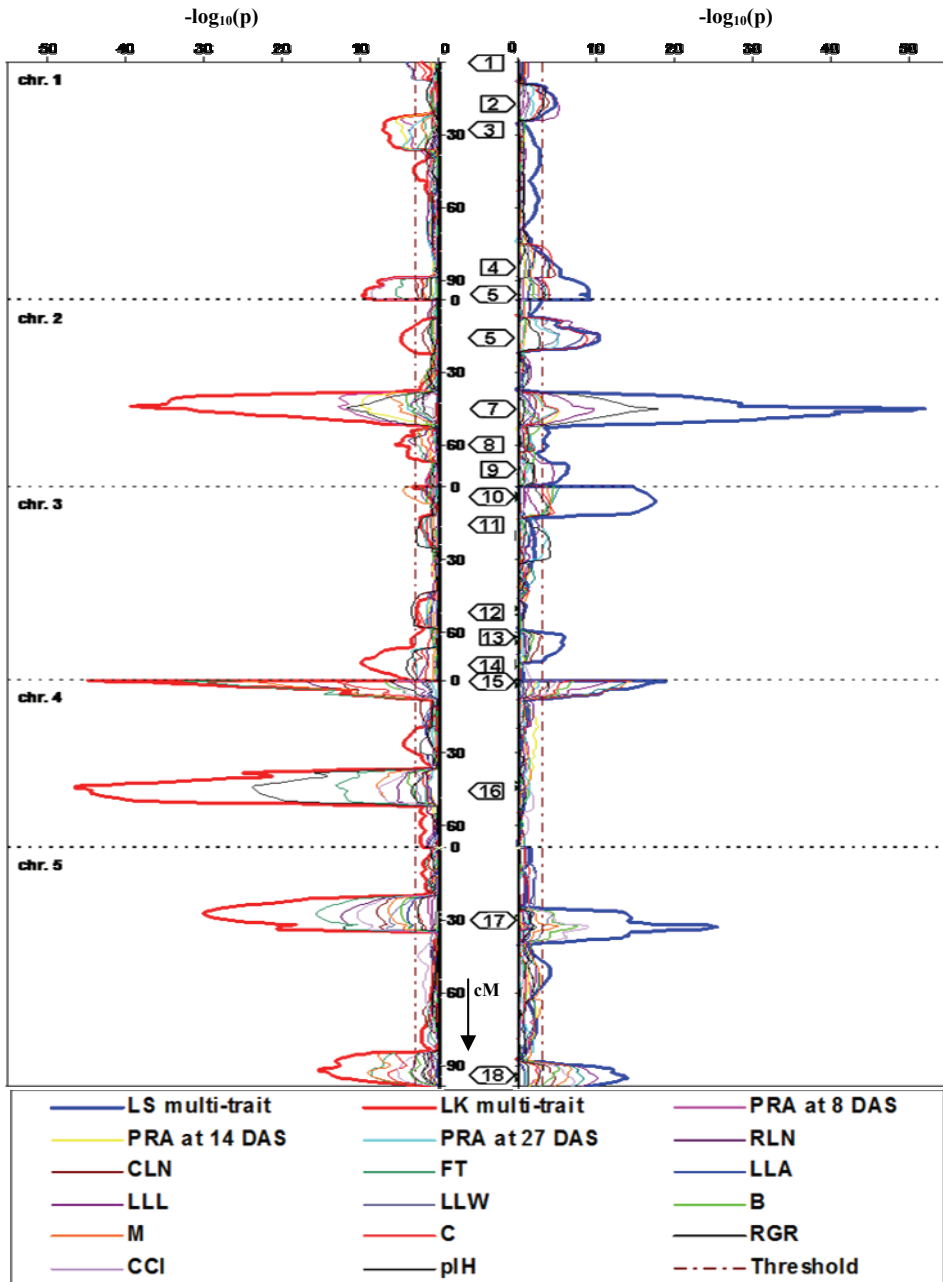
Figure 2. Results of the multi-trait composite interval mapping genome scan. −log₁₀ transformed p-values are plotted on the horizontal axis against the genetic position in centi-Morgan on the vertical axis. The individual linkage groups representing the 5 chromosomes of Arabidopsis are indicated. To the left are plotted the results from the L*er* x Kas-2 population and to the right the L*er* x Sha population. The multi-trait test statistic is represented by the bold trace and the test statistics of the individual traits by the other traces. The positions of the 18 detected QTL are indicated along the linkage groups with flags that point at the graph of the population in which it has been detected.

Table 2. The significant allelic effects of the detected QTL (figure 2) expressed as the percentage of the respective population means (table 1), and the genetic variance explained by these effects (σ^2_g expl.). The chromosome number and the respective map positions of the genetic predictors used in the final QTL models for each population are indicated in the column headers.

Trait	pop	QTL1 (chr1, 0.0)		QTL2 (chr1, 16.4)		QTL3 (chr1, 28.3)		QTL4 (chr1, 82.0)		QTL5 (chr1, 96.2/95.6)		QTL6 (chr2, 45.2)		QTL7 (chr2, 58.3)		QTL8 (chr2, 69.7)		QTL9 (chr2, 4.6)		QTL10 (chr3)		QTL11 (chr3, 14.5)		QTL12 (chr3, 50.5)		QTL13 (chr3, 65.8)		QTL14 (chr3, 73.4)		QTL15 (chr4, 44.1)		QTL16 (chr4, 27.0/33.0)		QTL17 (chr5, 91.6/96.0)		QTL18 (chr5)	
		effect	expl	effect	expl	effect	expl	effect	expl	effect	expl	effect	expl	effect	expl	effect	expl	effect	expl	effect	expl	effect	expl	effect	expl	effect	expl	effect	expl	effect	expl	effect	expl	effect	expl	effect	expl
PRA at 8 DAS	LK					-5.9	8.3					-13.3	36.7	3.4	0.2					-4.1	5.7							3.6	5.1	3.6	1.4			-5.5	10.4		
PRA at 14 DAS	LK					-8.3	10.7					-13.5	23.2	4.2	-0.5					-3.9	3.0							3.9	-0.4			-5.6	7.3		
PRA at 27 DAS	LK	7.5	3.6			-12.6	10.3			-14.6	9.6	11.6	10.6					11.2	4.3			11.4	7.7	8.2	3.8								
	LS	-9.6	4.0	8.6	3.8	-12.7	11.6	-13.6	10.6	-7.0	2.1			-9.2	3.1			7.8	4.3	-30.6	28.8	-14.9	8.5	-25.6	3.3	6.6	-2.3	15.0	10.9		
RLN	LS	-10.4	5.3	9.5	5.1			-7.8	14.4	-11.4	5.0	-10.6	6.2			7.6	4.2	-15.6	20.7	-5.7	2.2	-9.9	25.1	9.7	19.2						
	LK	8.4	3.8													4.7	6.3							-10.5	15.9	-1.9	2.6	4.9	14.8		
CLN	LS	7.9	8.2	-9.3	8.0	7.6	6.2	5.9	3.5	-6.4	4.0			6.0	5.4	-11.7	21.4	-10.5	15.9	9.9	11.3				
	LK	4.9	2.9	4.4	2.6	3.1	1.7			-8.4	10.2	-11.7	21.4	10.6	10.2				
FT	LS	2.5	1.6	-3.4	5.5	2.8	4.3	-1.8	1.6	4.2	11.6	-6.3	8.1	-11.3	32.4	-5.8	27.1	-6.5	11.0	-7.7	14.1	5.5	7.2		
LLA	LS	10.1	9.8	-5.6	3.3	4.2	1.9	-5.4	3.7	5.8	1.7	-3.0	2.5	-7.7	4.6	-1.9	2.6		
LLL	LK	6.0	8.8	6.3	6.8	3.8	3.8	-6.6	6.4	2.7	1.7	4.5	2.7	-11.7	4.6	-7.7	4.6	2.6	4.9		
	LS	5.1	7.3	-5.0	4.6	-7.5	16.6	-8.3	16.4	1.7			-3.1	2.3	-4.7	6.2	7.3	14.8		
LLW	LK	5.6	9.3	2.3	2.3	-4.8	14.2	-5.0	6.4	4.1	3.3	-2.1	1.9	3.4	7.7		
B	LK			-2.8	5.6			-4.6	19.3	2.3	3.7	6.7	16.8	4.4	5.6	10.6	6.2		
	LS			2.2			3.9	3.3	2.8	4.9	26.2	5.1	29.1	-4.3	16.1		
M	LK			2.2	1.1	-2.1	2.9	3.9	3.2	10.3	-8.9	30.7	-4.5	7.5	-4.3	9.3	3.7	12.9		
	LS			2.2	3.9	3.9	-2.3	1.0	-5.0	31.0	-2.8	9.3		
C	LK	9.5	2.4	8.5	3.8	-13.0	6.3	11.2	5.4	-15.3	13.3	-13.0	7.6	10.7	5.6	8.9	0.7	-11.2	2.0	-26.8	23.8	-6.8	0.7	-10.1	2.3	17.6	8.8		
	LS			2.1	2.8	8.2	1.3	8.3	1.0	3.3	2.8	7.2	-124	9.4	8.4	3.0		
RGR	LK	3.6	10.9					-3.2	8.1	-2.6	3.1	3.2	10.3	4.0	6.6	-4.3	8.9	5.2	18.6	3.6	8.9	3.3	6.6	5.1	-2.2		
CCI	LK	18.5	16.4	6.1	6.9	3.8	2.0			4.3	11.8	3.3	9.8	2.4				
	LS							6.0			-8.4	10.2	-15.5	12.0	-17.4	15.5	-22.3	21.3	-9.9	25.1		
pH	LK	-8.0	2.1	-9.0	42.3							-7.6	14.4	-9.9	25.1	10.7	2.1		
	LS	-7.7	1.6					-29.5	20.4	-30.5	34.7	12.1	5.9	6.4	1.7	37.1	42.3	-8.3	3.3	10.8	4.4		

Genetic variance explained by QTL that co-locate between both populations

The similarity in genetic architecture between the both populations was investigated by comparing the amount of genetic variance explained by QTL that co-located (co-location between population in this context) between both RIL populations to the total variance explained (figure 3A). Although more population specific QTL were found in the LK population, co-locating QTL accounted for the larger part (>50%) of the genetic variance explained for most traits in both populations. In addition, whenever the co-locating QTL had allelic effects on the same traits in both RIL populations, these were of the same sign. Only for PRA at 27 DAS, RGR and plH in the LK population did specific QTL explain more genetic variance than did co-locating QTL. For the traits PRA at 8 DAS, PRA at 14 DAS and CCI, all genetic variance explained in the LS population was accounted for by QTL that co-located between both populations. Typical for the LS population was that a relatively large part of the genetic variance for CLN was explained by population specific QTL. Interestingly, in both populations, the RGR was among those traits for which a relatively large fraction of the genetic variance explained was attributed to specific QTL.

The effects of flowering time QTL on other traits

QTL with significant effects on FT always also had effects on other traits. In the LK population eight of the detected QTL (3, 5, 7, 8, 10, 15, 16, 17 and 18) and in the LS population seven QTL (4, 5, 9, 10, 15, 17, 18) had significant effects on FT. Most of the explained genetic variance for FT was accounted for by QTL that were detected between both populations (figure 3A). QTL 15, 17 and 16 explained most of the genetic variance for FT in the LK population while in the LS population QTL 15, 18 and 10 explained most of the genetic variation for this trait. These QTL also had a high contribution to the explained genetic variance for RLN, the parameters B and M, and CCI. In the LK population, a significant allelic effect of 3.1% for FT was found for QTL7, at the *ERECTA* locus but this only explained 1.7% of the genetic variance. To prevent confusion due to the large effect of *ERECTA* on other traits, the genetic variance explained for each trait by this locus was considered separately (figure 3B). For the following comparisons QTL effects of *ERECTA* will be considered not include to FT in the LK population. In the LS population, however, most of the genetic variance for these traits was explained by other QTL, which also had a strong effect on FT. For PRA at 27 DAS about one third of the genetic variance was explained by QTL that also had an effect on FT in both RIL populations. As expected from phenotypic correlations mentioned above, in both populations QTL with an effect on FT also

explained most of the genetic variance for RLN, CLN, CCI, parameter M, and parameter B. For the LK population the same was true for the parameter C and plH but not for the LS population. The allelic effects on RLN, CLN, CCI and the parameter C were always positively correlated to the effect on FT for these QTL (table 2). Also for RGR most of the variance explained was due to QTL involved in variation of FT, however, considerably less than for parameter B. The allelic effects on parameter B or RGR of FT QTL were generally negatively correlated to the effect on FT. In each population there was a single exception for RGR, notably QTL5 in the LK population and QTL9 in the LS population. The allelic effects on RLN and CLN of QTL9 in the LS population were relatively strong in relation to its effect on FT, indicating possible effetcs on plastichron. These effects were accompanied by effects on PRA ay 27 DAS and RGR accordingly. At three more detected QTL the allelic effects deviated from the generally strong genetic correlation between FT and RLN. A weak allelic effect on FT and parameter M was found for QTL3 but an effect on RLN was not detected. This QTL did have strong effects on the three PRA measurements. QTL10 presented a similar situation in both mapping populations with effects on FT but not on RLN and CLN. Finally, the before mentioned allelic effect on FT detected at the *ERECTA* locus in the LK population was not accompanied by an effect on RLN.

QTL4 and 5 presented a special situation in the LS population. A single QTL was detected in this region in both RIL populations with simple interval mapping. However, while selecting co-factors in the LS population, using QTL4 as co-factor led to the detection of QTL5 and *vice versa*, while the QTL that was chosen as co-factor was no longer significant. Both QTL4 and 5 were included in the final QTL model for the LS population. These QTL had allelic effects on largely the same traits, of similar magnitude, but of opposing sign. They had relatively large effects on RLN, PRA at 27 DAS and the parameter C but also on FT, CLN, LLA, LLL, and the parameter M. However, only QTL5 also had effects on PRA at 14 DAS, parameter B, and CCI. This was the only QTL for which genetically correlated effects on both PRA and either one of the growth rate traits, were of opposing sign. QTL8, detected in the LK population only, very closely linked to QTL7 was the only QTL with effects specific to FT. QTL effects on PRA at 8 and 14 DAS were also detected for this QTL but they did not explain any genetic variance. These QTL effects could bee caused by the close linkage to QTL7, which was the QTL with the largest effects on PRA at 8 and 14 DAS in the LK population.
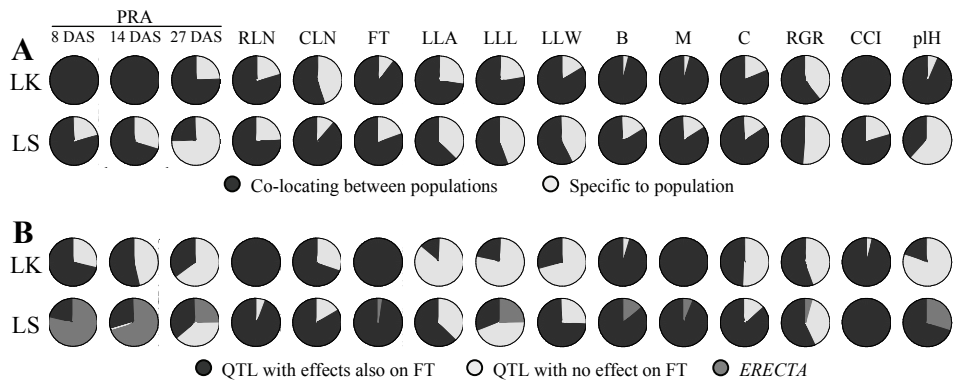
Figure 3. Comparison for each trait of the genetic variance explained by: A) QTL that co-located between both populations versus population specific QTL, and B) QTL with effects on flowering time and plant growth versus plant growth-specific QTL. The sections represent the respective percentage of the total genetic variance explained by the additive QTL models. In figure 3B the genetic variance explained by QTL7, linked to the *ERECTA* locus is indicated separately (refer to text). The sections indicate the percentages of the total genetic variances explained for each trait by all main effect QTL.

Plant growth specific QTL

QTL with allelic effects on flowering time related traits often also had genetically correlated effects on any of the other traits. QTL were also detected, however, that had specific effects on rosette-related traits and growth rate, without having an effect on FT. In particular the PRA measurements and LLL were the traits for which most explained genetic variance was due to QTL that did not have an effect on FT in the LK population. In the LS population most of the genetic variance explained for LLL, LLA and LLW was accounted for by QTL with no effect on FT, followed by the parameter C. QTL1, specific to the LK population only had relatively strong effects on LLA, LLL, LLW, and the parameter C; thus specific to the final state of the rosette. QTL2 in the LS population had similar allelic effects on these traits and, in addition, significant effects on PRA at 27 DAS, CLN and RGR were found. The effect on RGR was average, being 3.6% but it accounted for 10.9% of the genetic variance explained.

The allelic effect of QTL6 on CLN in the LK population was 4.4%, explaining a mere 2.6% of the genetic variance. In the LS population this QTL had moderate to strong effects on PRA at 27 DAS, the measurements on the largest leaves, parameter C, and RGR. In fact, it accounted for the highest explained genetic variance for LLA, LLL, LLW and the parameter C by any single QTL. QTL7 had a very strong effect on plH in both populations. A single allele substitution by L*er* encompassed an effect of about -30.0%. However, in the

LK population this accounted for 20.4% of the genetic variance while in the LS population it accounted for 34.7%. Only in the LK population did this QTL have an effect on both RGR and the parameter B. In the LS population no significant effect on these traits was found. Furthermore, this QTL had strong effects on the PRA in both populations. While the effect was constant at about -13% from the 8 DAS till 27 DAS in the LK population, it increased from -4.5% at 8 DAS to reach -13.6% at 27 DAS in the LS population. The effects on PRA at 8, 14 and DAS accounted for 36.7%, 23.2% of explained genetic variance in the LK population while in the LS population it was more moderate with 6.2% and 14.4% respectively.

QTL11 was detected on chromosome 3 in the LK population, closely linked to QTL10. When both loci were used as a cofactor during MQM mapping the significance of QTL11 dropped below the threshold. However, when fitting the final QTL model including both QTL on the top of chromosome 3 it did have significant allelic effects. It was found to have positive effects on PRA at 27 DAS, LLL, parameter C and RGR. For RGR this QTL explained 6.6% of the genetic variance and for the other traits less than that. Three additional QTL were detected on chromosome 3, QTL12 and 14 in the LK population at and QTL13 in the LS population. All three QTL had significant effects on one of the growth rate traits. At QTL13 in the LS population weak positive and effects on the parameter B and LLW were found respectively that also did not explain much variance. This QTL had slightly stronger positive effects on CLN and plH. For CLN it accounted for 5.5% of genetic variance and for the other traits less. QTL12 in the LK population had significant negative effects on PRA, parameter C and RGR only. The magnitude of the effects on PRA increased in time to be -11.2% for parameter C (final PRA). Although the effects were of considerable magnitude, the genetic variance explained for these traits by this QTL was marginal. Nevertheless, this QTL had an effect of -4.3% on RGR that accounted for 8.9% genetic variance explained. QTL14 had significant positive effects on PRA at 27 DAS and parameter C, being the strongest on the former with 11.4%. The relatively large effect on RGR of 4.3% accounted for 11.8% of the genetic variance. Additionally, this QTL had moderately strong positive effects on RLN and CLN.

Non-additive 2-way interactions between the detected QTL

In addition to the final QTL model with all detected QTL in main effect a second model including two-way interactions between the detected QTL was tested. Backwards selection from a model with all QTL as main effects plus all possible two-way combinations resulted in 20 significant interactions for each RIL population. Hence, a final QTL model including

these two-way interactions explained considerably more genetic variance explained for all traits, except for plH (table 1). For many traits numerous significant effects were detected but the magnitude, and in particular, the genetic variance explained for these was often low. Only particular aspects are described in detail below, a complete summary of the detected allelic effects and genetic variances explained is presented in Table 3. No significant two-way interaction was found for QTL11 but all other QTL were component of at least one interaction when considering the results from both populations. In the LK population the highest number of interactions was found for QTL18, which interacted with 7 other QTL. In the LS population QTL6 and QTL15 showed the highest number of interactions as both interacted with 6 other QTL. Only the interactions QTL6xQTL18, QTL10xQTL18, QTL15xQTL17 and QTL17xQTL18 were common between both populations. Overall the interactions had effects on groups of traits that were before found to be highly correlated (figure 1).

As was observed for the allelic effects of the additive model, interactions with significant effects on FT often also had effects on traits from other groups. However, FT was one of the traits with the smallest increase in genetic variance explained for the model including interactions over the additive model. This is reflected in the low genetic variance explained for FT by each relevant interaction. Only one interaction had effects specifically on FT and related traits in each population. The interaction QTL7xQTL16 in the LK population had an effect of 12.3% on RLN but also of lower magnitude on CLN and FT. In the LS population the interaction QTL9xQTL15 had an effect on the same traits with the strongest being 8.8% on CLN.

In both populations RGR was among those traits with the largest increase in genetic variance explained attributed to interactions, being 24% and 21% for the LK and LS populations respectively. As was observed for the main effect QTL also the magnitude of the additive effects on RGR of the interactions was limited. The largest effect observed was 3.2%. Specific to the LK population PRA at 27 DAS also showed a large increase in genetic variance explained but also for the LLA, LLL and LLL a large increase was found. In the LS population the largest increase in genetic variance explained due to including interactions was found for the traits PRA at 14 DAS and parameter M, apart from RGR. It was for those traits that interactions explaining substantial genetic variance were found.

In the LS population 7 interactions had significant effects on RGR. Three of those had effects specifically on growth at rosette or leaf levels and explained relatively large amounts of genetic variance. Interestingly, QTL2 was a component in all three, interacting with QTL17, QTL6 and QTL4. Furthermore, these three interactions had effects on the measurements on the largest leaf and, in accordance with effects on RGR, PRA. An

interaction between the two closely linked QTL4 and QTL5 also had a significant effect on RGR. The interaction QTL2xQTL10 had an effect of on RGR and also on PRA at 14 DAS, that accounted for a relatively high amount of 8.1% genetic variance explained. Together with the 10.6% explained genetic variance for PRA at 14 DAS of interaction QTL4xQTL9, most of the increased variance explained for PRA at 14 DAS by interactions was accounted for.

The interaction QTL1 x QTL10 was found to explain 6.2% genetic variance for RGR in the LK population which was the highest value observed for an interaction in this population. The only other trait on which this interaction had a significant effect was the parameter B. Again, for this trait it also posed the largest contribution to the increased genetic variance explained of all interactions. The interaction QTL3 x QTL12 came in second place for explaining 4.5% genetic variance for RGR. This also accounted for the second highest genetic variance explained for PRA at 27 DAS, LLA and LLL observed for any interaction in this population, accompanied by relatively strong effects. The interaction with the largest genetic variance explained for PRA at 27 DAS was QTL16xQTL18 with 7.7%, and an effect of 11.7%. It also accounted explained the highest amount of variance for LLL and LLW of all the interactions. QTL3xQTL18 had effects on the growth rate, and largest leaf traits, the allelic effect on LLA of -9.1% and an explained variance of 7.1% was the highest for any single interaction for this trait.

Table 4 <on the following pages>. Significant effects and genetic variances explained (σ^2_g) by those effects of the significant two-way interactions between the detected QTL. The first components of the interactions are presented in the left-most column and the second components are presented in the top row. The population in which significant effects were found is mentioned below the second components of the interactions. Only the traits for which significant effects were found are shown, in the second column.

The following table is rotated 90° on the page. Column headers are grouped by QTL (QTL1–QTL18); each QTL column reports a pair of values: **LS effect / σ²g** (or **LK effect / σ²g**). Row blocks are labelled QTL1 through QTL6 with their associated traits.

	QTL4 LS effect	σ²g	QTL5 LS effect	σ²g	QTL6 LK effect	σ²g	QTL6 LS effect	σ²g	QTL9 LS effect	σ²g	QTL9 LS effect	σ²g	QTL10 LK effect	σ²g	QTL10 LS effect	σ²g	QTL10 LS effect	σ²g	QTL12 LK effect	σ²g	QTL13 LS effect	σ²g	QTL14 LK effect	σ²g	QTL15 LK effect	σ²g	QTL15 LS effect	σ²g	QTL16 LK effect	σ²g	QTL17 LS effect	σ²g	QTL18 LK effect	σ²g	QTL18 LS effect	σ²g
QTL1 PRA at 8 DAS																							3.0	2.6												
FT																							-1.7	1.4												
B									4.1	6.2													-1.6	1.5												
M																																				
RGR									3.2	6.7																										
QTL2 PRA at 8 DAS							-3.8	5.1					-5.4	8.1																						
PRA at 14 DAS	-8.8	3.7											-8.0	3.2																						
PRA at 27 DAS													1.7	1.2																						
FT																															5.3	5.1				
LLL																															3.3	3.2				
LLA																															2.5	3.8				
C																															4.9	1.0				
RGR	-2.6	4.2					2.8	5.5					-2.5	3.9																	3.0	7.7				
QTL3 PRA at 8 DAS					-4.3	4.1											-4.7	3.3																		
PRA at 14 DAS																	-11.2	6.6																		
PRA at 27 DAS					-8.4	3.9											1.8	1.4																		
FT																	-8.2	5.5													-9.1	7.1				
LLL																	-4.9	4.4													-4.7	4.3				
LLW																	-3.4	3.0													-4.1	4.6				
B					-2.9	3.5																									-3.6	3.8				
C																	-8.7	1.9																		
RGR					-2.3	2.8							-2.9	4.5																	-2.6	3.4				
QTL4 PRA at 8 DAS									5.8	11.3															-7.0	2.1									
PRA at 14 DAS									6.5	10.6															-6.6	2.6									
PRA at 27 DAS																																				
RLN													-5.9	1.4																						
CLN													-7.0	2.6																						
FT							-2.5	3.1																	-1.9	1.6									
LLA													5.7	3.2																						
LLL			-3.5	5.8									3.4	1.8											-3.7	3.1									
LLA			1.8	1.9									2.9	2.9																						
B							-4.2	6.5			-2.4	5.1													2.3	2.3										
M							12.0	3.5																	-2.0	3.1										
C			-2.6	2.7																					-7.7	2.5										
RGR			5.0	3.2									-4.1	2.2																						
pH													8.9	1.5											-6.6	7.6										
QTL5 PRA at 14 DAS							11.8	2.1																			4.2	4.0								
PRA at 27 DAS							-4.8	4.0																					3.9	4.7						
LLL							-11.6	4.8																												
LLW							-3.8	2.2																												
C							-10.5	3.1																					-7.5	0.1						
CCI																																				
pH							-9.5	1.6																												
QTL6 PRA at 8 DAS									3.6	2.9					-6.2	2.4											4.2	4.2								
PRA at 14 DAS									-8.4	4.6	-6.2	1.1					-8.0	1.3	-5.9	2.4					4.3	2.9								
PRA at 27 DAS									-5.3	2.0	-3.8	1.6							-7.9	7.0											7.3	0.9		
RLN									-2.5	3.4	-1.9	1.7							-1.7	1.6											5.2	2.4		
CLN																							-1.7	1.3							2.5	1.9		
FT									-2.5	1.2																					6.2	4.3		
LLA																					-3.4	2.6									3.3	1.7		
LLL													3.0	2.6																	3.7	5.7		
LLA																			2.7	6.1														
B									-2.3	4.5													-2.3	2.2										
M									-6.2	1.4	-8.5	2.4					-5.8	1.5																
C											-2.1	2.6																						
RGR																							-8.1	-0.1							-4.6	3.0		
CCI																																				
pH																																				

The table below reports QTL effects. Columns QTL4, QTL5, QTL6, QTL6, QTL9, QTL10, QTL10, QTL12, QTL13 and QTL14 contain no entries (shaded/empty in the original). Only the data‑bearing columns are shown; for each QTL the sub‑columns are the allelic "effect" and the genetic variance σ^2_g.

Trait	QTL15 LK effect	σ^2_g	QTL15 LS effect	σ^2_g	QTL16 LK effect	σ^2_g	QTL17 LK effect	σ^2_g	QTL17 LS effect	σ^2_g	QTL18 LK effect	σ^2_g	QTL18 LS effect	σ^2_g
QTL7														
PRA at 14 DAS											4.9	4.2		
M											-2.9	3.5		
C											-7.6	1.2		
RGR											2.3	2.8		
QTL8														
RLN					12.3	4.1								
CLN					4.1	2.1								
FT					2.7	1.7								
QTL9														
RLN			8.2	4.4										
CLN			8.8	7.1										
FT			2.1	2.3										
QTL10														
PRA at 8 DAS													4.7	5.7
PRA at 27 DAS											-7.8	3.3		
RLN													-6.1	1.7
CLN			4.7	1.1							-3.7	1.3		
FT			1.7	1.1										
LLA														
LLL			-2.6	1.1							-5.2	3.0		
M			1.4	1.3							-3.1	2.4		
C														
RGR			-3.0	5.3							-7.3	1.3	-6.7	6.3
CCI											-2.6	3.7		
QTL12														
PRA at 27 DAS	6.6	2.6												
LLA	4.9	3.0												
LLL	3.0	2.4												
LLW	3.3	4.1												
C	7.5	1.4												
QTL13														
PRA at 14 DAS			3.3	3.4										
LLA			2.4	3.3										
RGR														
QTL14														
RGR														
QTL15														
PRA at 8 DAS														
PRA at 27 DAS							4.1	3.7			2.0	2.4		
RLN					9.7	2.2	15.0	5.6	-4.8	1.5			4.2	4.6
CLN							4.1	1.6					-6.6	5.6
FT					2.1	1.7	4.2	4.7	-2.1	2.7			-10.0	2.7
LLA							-5.3	3.1	-3.9	2.3			-6.6	2.1
LLL														
LLW														
M							-3.0	3.2	-2.3	3.1				
C									-1.4	2.0				
RGR					2.4	3.6	10.6	2.5	-8.2	3.7			-8.0	2.3
CCI					10.3	2.2							-2.9	4.4
pH					-15.2	5.3								
QTL16														
PRA at 8 DAS											3.5	2.5		
PRA at 14 DAS											4.6	3.3		
PRA at 27 DAS											11.7	7.7		
RLN							-7.5	2.6						
FT							12.2	3.1						
LLA							3.5	3.0			7.0	4.6		
LLL							-6.9	3.9			4.6	4.5		
LLW							-4.6	4.0			4.6	5.9		
CCI							-3.5	3.2			-10.3	2.4		
							12.3	4.0						
QTL17														
PRA at 8 DAS											-4.7	4.7		
PRA at 27 DAS														
CLN											4.4	1.1	6.5	1.8
LLA											6.5	3.3		
LLL													2.4	1.1
LLW											3.5	2.8		
M											2.5	2.4		
C											8.2	1.5		

Discussion

Segregation of large effect QTL is rare per se (Salvi and Tuberosa, 2005) and up to now, no such QTL involved in plant growth have been described in Arabidopsis. Only a single, small effect, QTL for plant growth has been successfully identified in this species (Kroymann and Mitchell-Olds, 2005). Failure to detect large effect QTL for growth indicates that such alleles might not be common in nature. Furthermore, epistatic QTL effects could be more important for fitness-related traits, such as plant growth, than additive effects (Malmberg *et al.*, 2005). A multi-trait QTL mapping method was used in the present study in order to gain power and to overcome the limitations of small effect QTL. Several multi-trait QTL mapping approaches have been described based on a variety of methods such as maximum-likelihood (Jiang and Zeng, 1995), principal component analysis (Mähler *et al.*, 2002), or Bayesian inference (Banerjee *et al.*, 2008), amongst others. Many methods were not particularly suited for the aim of cloning the QTL because of reasons like data reduction, limitation in the number of traits that can be analyzed together efficiently, or the requirement for advanced programming skills. Recent advances in mixed-model methodology made this an efficient alternative for multi-trait QTL mapping (for instance: Malosetti *et al.*, 2008). This method was particularly suitable because it did not suffer from any of the above mentioned restrictions and tests for non-additive effects between QTL could easily be implemented, while explicitly modelling the genetic covariance between traits. Our particular dataset required fitting an unstructured covariance model between traits. This is the most demanding model in terms of number of parameters and made model fitting somewhat computationally intensive. However, no major convergence problems were observed.

The implementation of two-way interactions on the full data-set using the multi-trait methodology is controversial at the moment. By doing so, the assumption was made that epistatic interactions affect multiple traits simultaneously while this does not necessarily need to be the case, in which case such a model would not be suited. However, certain genetic correlations could be expected due to the nature of the traits in the dataset. For instance, any genetic effect on growth rate, whether additive or epistatic, should in principle lead to an effect on PRA. The detected interactions did indeed generally affect sets of traits in such a sensible manner. Nevertheless, future work on these statistical interactions will have to include a more careful analysis with less complicated models. An approach that extends multi-trait QTL models with interactions step-wise, starting from a single-trait, single-QTL situation is to be preferred. In addition, the detected plant growth-specific QTL

have been selected for QTL validation using near isogenic lines (NILs) (chapter 3). Some of the detected interactions can also be tested for validity in the process. To have information about potential epistatic interactions is useful for QTL validation and fine-mapping because it might allow to designs NILs with a genetic background that contrast the effects of a certain QTL. It also provides information how QTL affect phenotypical variation in concert. It is therefore worthwhile to further explore the detected interaction by both statistical and biological means. If the identified statistical interactions indeed represent epistatic interactions between the QTL, they indicate a complex genetic architecture, in particular for the traits related to vegetative growth. This would be representative of reported results from the precise dissection of a plant growth QTL (Kroymann and Mitchell-Olds, 2005).

The detection of plant growth-specific QTL was a major aim of the experiment described, as opposed to QTL for plant development. In this respect, the choice to analyze two RIL populations has been justified by the doubling of the number of interesting QTL for follow-up studies. This went on the cost of the number of replicates analyzed per RIL but the loss of power was made up by the multi-trait methodology. Large effect QTL for plant growth related traits could be detected although not in the order suggesting monogenic control. Detected QTL effects in the order of 20%, considering homozygous allele substitutions, are in principle suitable for efficient map-based cloning. Such effects were found for traits at the rosette level. However, the magnitude of QTL effects on the growth rate traits RGR and parameter B was constrained to a maximum of around 10%. The explanation for this is the relatively large number of detected QTL that are underlying the small amount of variation that was found for them. This is supported by the high amount of genetic variance explained by these small QTL effects. The small variation is in line with previously reported constrains on plant growth rate within this species (Li *et al.*, 1998). The larger effects on PRA and often also to other traits of QTL with effects on plant growth rate can be exploited for map-based cloning of the underlying growth-rate QTL. Validation of these QTL using near isogenic lines or heterogeneous inbred families will be the first step towards physiological characterisation of the effects and positional cloning of the underlying genes.

An estimate of the growth rate of the plants was considered a necessity in order to accurately quantify their growth. Dicotyledonous leaves expand their tissue area exponentially for the larger part of the duration of growth (Granier and Tardieu, 2009). A small variation in time of germination can therefore lead to relatively large variability of rosette size during vegetative growth in Arabidopsis. Modelling the change in plant size over time allows estimating its growth-rate independent of such variation. Digital image analysis was adopted to quantify PRA repeatedly from the seedling until the maximum PRA was reached. It offers the advantage over destructive measurements that a single plant can be

measured repeatedly over time. This reduced greatly the number of plants that need to be analyzed. The change in PRA over time was modelled with two models that both provided an estimate of plant growth rate (RGR and parameter B). The moderate correlation between RGR and the parameter B shows that, although both provide estimates for plant growth rate, they only partially described the same process. The explanation for this can be the dependency on flowering time of the shape of the growth curves. The logistic model was forced to fit the data by over-weighting the maximum PRA of the plants. The moment at which maximum PRA was reached depended on its term on the duration of vegetative growth, hence flowering time. This showed from the strong correlation between parameter M and flowering time and is likely to have increased residual correlations between the estimated parameters. Additionally, it indicates that a longer duration of vegetative growth leads to reduced growth rate. The constrained variability of growth rate makes this trait particularly prone to be biased by this. Indeed, a much stronger negative correlation was observed between the parameter B and flowering time than between RGR and flowering time. Such induced correlation with flowering time could be minimized for RGR to some extend (not-shown) by making sure to estimate it before rosette expansion reached the inflection point. The negative correlation between leaf expansion rate and flowering time was recently demonstrated in a physiological study using Arabidopsis (Cookson *et al.*, 2007). The authors pointed out that this phenomenon can pose problems for the analysis of growth rate in Arabidopsis. Both populations analyzed showed significant segregation of flowering time. Therefore, RGR provides a better estimate of plant growth rate than the parameter B as this model is less prone to induce residual correlation with flowering time. This might be achieved also for the parameter B by correcting on flowering time.

QTL analysis emphasized the very strong negative genetic correlation between growth rate and flowering time in both RIL populations. The almost perfect genetic correlation between the parameter B and flowering time further supports the bias of the former by the latter. However, even though RGR could be de-correlated from flowering time, still most of the genetic variance explained was genetically correlated to flowering time. Such a correlation has been found before by using single trait QTL analysis for a similar set of traits in the L*er* x Sha population (El-Lithy *et al.*, 2004). The same QTL were detected again in the present experiment and the negative correlation between effects on flowering time and plant growth rate were more consistent. Such a correlation between flowering time and plant growth-related traits on the cellular level has also been demonstrated recently (Tisné *et al.*, 2008). These genetic correlations can be caused by both pleiotropy and close linkage of independent QTL. However, the observed consistency strongly suggests the former to be mostly underlying the observed correlations. Based on

their results, Cookson *et al* (2007) suggested that a signal from the shoot apex in response to the induction of flowering could be responsible for the effects they observed on leaf growth. The existence of such a signal was proven by its stimulating effect on secondary growth in Arabidopsis (Sibout *et al.*, 2008). A similar, if not the same signal is therefore likely to be underlying the correlation between growth rate and flowering time that was found in the present study. This suggests that a mechanism related to the developmental switch to flowering is thus the main regulator of RGR in Arabidopsis. However, it might also merely invoke two discrete stages of differing growth rate on the Arabidopsis rosette, defined by the developmental state of the apical meristem. Exactly this kind of regulation was found by Sibout *et al.* (2008). The same plant growth-specific QTL can in principle underlie variation for both distinct stages. Under this hypothesis the observed genetic correlation between growth rate and flowering time reduces to an artefact of sampling. To investigate this, a clear distinction between both stages must be made which is not possible with the data of the current experiment. Due to increasing overlap of rosette leaves, a second stage of increased RGR can not at all be independently quantified from the PRA. However, minimizing the correlation between RGR and flowering time did allow this to some extent for the initial vegetative growth phase. The detection of QTL, specific to RGR, not genetically correlated to flowering time proves that a discrete genetic architecture for this trait does exist.

Two QTL were detected with a positive genetic correlation between flowering time and growth rate. These can be cases where the correlation is due to closely linked QTL, which would mean that the effects on growth rate might have had to overcome those caused by flowering time. An intriguing alternative would be that these are pleiotropic effects of the same QTL. Such a contrast to the general trend could provide the opportunity for closer investigation of the relation between plant growth rate and flowering time. The genetic correlations between significant allelic effects for QTL9 in the L*er* x Sha population offer an alternative explanation. The very weak effect on flowering time that was accompanied by a fairly large effect on rosette leaf number can be interpreted as an effect on plastochron. This, on its term could be responsible for the observed effects on RGR. El-Lithy *et al* (2004) reported the detection of a QTL for RGR at the *ERECTA* locus, where we did not observe such effects. It is quite possible that the same QTL effects were detected in both experiments but that these could be resolved to an individual QTL thanks to the additional markers added to the genetic map in this region. In the L*er* x Kas-2 population the *ERECTA* locus did have an effect on RGR so it could also be that this was not detected in the L*er* x Sha population due to the smaller size of this population.

The 15 traits that were analyzed by multi-trait QTL analysis could de grouped in three classes relating to growth and development: i) the rosette during vegetative growth, ii)

the final state of the rosette, and iii) flowering time. These three classes formed separated groups in the correlation structure, yet, were correlated to each other. The detected QTL reflected this structure nicely by having effects on a specific group, on two groups, or even on all three groups. QTL specific to plant growth (groups i & ii) showed different modes of action judging from the traits they affected. Firstly, effects limited to the final state of the rosette were found, such as for QTL1. The relatively large effects on the parameters of the largest leaf but not on growth rate and earlier PRA indicate that this QTL acts only late during vegetative growth. Secondly, effects of increasing strength on PRA over time accompanied by effects on growth rate but no effects on the largest leaf. Such effects could be explained by differences in plastochron in case of QTL14. However, the absence of such effects, as for QTL 12, suggests an intrinsic QTL effect on growth rate during early vegetative growth only. Finally, QTL with increasingly strong effects on PRA accompanied by effects on growth rate and the largest leaf were found. These QTL effects last during the entire phase of vegetative growth. Furthermore, a relatively large number of significant non-additive interactions were detected; in particular for rosette growth specific traits these explained considerable genetic variance. The properties of the detected QTL effects might therefore be partly defined by the genetic background.

This complex genetic architecture of plant growth could only be elucidated by studying it at the whole plant level. The complexity could be expected from the large number of genes that can potentially have an effect on growth (Beemster *et al.*, 2005; Gonzalez *et al.*, 2008; Wolters and Jürgens, 2009; and many more) and from experiences in the field of plant breeding (Holland, 2007; Cooper *et al.*, 2009). It makes hypothesising about candidate genes that might underlie the detected QTL effects not very meaningful. Two RIL populations were analyzed to get a broader view of the potential genetic architecture. The results show a considerable heterogeneity for plant growth because QTL specific to growth of the rosette tended to be also largely specific to the populations. The genetic determinism for flowering time, on the other hand, was mainly common between both populations. If this reflects the general trend, naturally occurring genetic variation could provide a vast resource of QTL for plant growth.

Five of the detected QTL that co-locate in the RIL populations are commonly detected QTL for flowering time (Koornneef *et al.*, 2004; El-Lithy *et al.*, 2006; Simon *et al.*, 2008). These QTL had effects mainly on the same sets of traits, and of the same sign, in both RIL populations. This is an indication that allelic variation for the same genes might be underlying these QTL in both populations. Candidate genes include *MAF1* for QTL5 (Werner *et al.*, 2005), *FRIGIDA* for QTL15 (Johanson *et al.*, 2000), and *MAF2-MAF5* for QTL18 (Ratcliffe *et al.*, 2003; Caicedo *et al.*, 2009; this thesis, chapter 4). The opposite

allelic effects of similar magnitude detected for QTL4, closely linked to QTL5 in the L*er* x Sha, could not be explained by problems with the genetic map. Furthermore, results from an independent experiment were reported that showed the same phenomenon (El-Lithy *et al.*, 2004). A candidate gene for this QTL could be FT (Corbesier *et al.*, 2007). The Landsberg *erecta* parent used for both RIL populations is known to carry the *hua2-5* point mutation in the gene *HUA2* (Doyle *et al.*, 2005). A CAPS marker designed for this SNP was added to the genetic map to take this into account. QTL17 mapped to this region in both RIL populations. In the L*er* x Sha population the QTL mapped distinctly to this marker and the relatively weak allelic effects observed on flowering time fit with the expectation from the report by Doyle *et al.* (2005). This makes the *HUA2* a good candidate for the observed QTL effects in this population. However, the much stronger allelic effects found in the L*er* x Kas-2 population and the distance of 6 cM from the *hua2-5* marker suggest a different genetic basis. Although *HUA2* can still contribute, a different gene underlying these QTL effects in this population must be considered. Several other candidate genes for flowering time located on the top of chromosome 5 are *FLC* (Michaels and Amasino, 1999)*, FY* (Simpson and Dean, 2002), *CO* (Putterill et al., 1995), *FRL1* (Michaels *et al.*, 2004), and At5g23460 (Keurentjes *et al.*, 2007) but none map directly in the region of the QTL.

QTL7, common to both populations at the *ERECTA* locus, can be expected to be caused by the L*er* allele of the *ERECTA* gene. The effects of the mutant allele carried by the L*er* accession on plant morphology are well described (van Zanten *et al.*, 2009) and fit with the allelic effects observed for this QTL. However, the observed effects on flowering time and the very strong effect on PRA at 8 and 14 DAS observed in the L*er* x Kas-2 population are not easily explained by the *ERECTA* gene. One or more QTL closely linked to *ERECTA* could be contributing these effects or Kas-2 carries a particular allele of *ERECTA*. This QTL accounted for the bulk of the genetic variance explained for PRA at 8 and 14 DAS whereas in the L*er* x Sha population QTL effects on these traits were observed mainly for the main flowering time QTL. This distinctly different genetic architecture between Sha and Kas-2 explains the lack of correlation between early PRA and flowering time in the L*er* x Kas-2 population, one of the few subtle differences that was found between the correlation structures of both RIL populations. This makes the Kas-2 allele at this QTL an interesting target for fine-mapping in order to elucidate the nature observed effects. Another trait showing a difference in its correlation to other traits between both populations that stood out was plant height. This can be attributed to the segregation of the Kas-2 allele at the *GA5* locus in the L*er* x Kas-2 population, which confers a semi-dwarf phenotype (El-Lithy *et al.*, 2006). Plant height was not accurately determined and was maintained in the selected set of traits specifically to account for the effects of *ERECTA* and *GA5*. However, it deserves to be

pointed out that the QTL at the *GA5* locus had allelic effects on almost every, emphasising its impact throughout the entire life-cycle of the plant.

Finally, QTL6 is a novel QTL that was detected in both mapping populations. Actually, only thanks to the multi-trait mapping method was it detected in the L*er* x Kas-2 population. Its single odd allelic effect on cauline leaf number could suggest a false discovery. However, the finding of this QTL being a component of non-additive two-way interactions with 5 other QTL supports its validity could explain its detection by effects on the genetic covariance structure. Both Sha and Kas-2 might harbour different alleles for this QTL because in the L*er* x Sha population this is one of the QTL with large additive effects plant growth. This QTL was nevertheless component of 6 interactions also in this population of which only one was with the same partner as in the L*er* x Kas-2 population. The genetic background could therefore also play a role determining whether allelic variation for this QTL leads to measurable additive effects. Either way, this QTL is a particularly intriguing subject for elucidation of the underlying molecular nature.

Acknowledgements

Chapter 3

Validation of QTL for plant growth and related traits in a selected set of near isogenic lines.

Bjorn Pieper[1], Samija Amar[1], Mohamed E. El-Lithy[2], Matthieu Reymond[1]

[1]Max Planck Institute for Plant Breeding Research, Cologne, Germany. [2]Menoufia University, Faculty of Science, Botany Department, Cairo, Egypt

Abstract

Near isogenic lines (NILs) with introgressions of Kas-2 or Sha alleles at QTL in a L*er* genetic background were selected to validate previously detected QTL for plant growth and related traits (El-Lithy *et al.*, 2004; this thesis, chapter 2). At all eight loci that were analyzed with these NILs, QTL effects were validated by comparing phenotypes of the NIL to those of L*er*. Generally, the NILs varied from L*er* for several traits including RGR, flowering time and final leaf size. However, in several cases effects were observed in the NILs that had either not been detected with QTL analysis or their magnitude deviated from that expected according to that analysis. This is suggested to be due to increased power to detect such small effects in NILs as compared to RILs or that this difference is attributable to variation between experiments. The results indicate that epistatic interactions between QTL that had been identified (chapter 2) could also account for modification of expected QTL effects in the NILs. An effect on plant morphology was observed in one NIL, which was sufficiently penetrant to allow efficient fine-mapping of this QTL down to 151 genes. Overall, the results demonstrate that the QTL analysis was highly accurate and that the detected QTL can be fine-mapped, in particular when exploiting epistasis, QTL by environment interactions and strong pleiotropic effects.

Introduction

Naturally occurring variation is a valuable resource for the identification of gene function. Most interesting traits and in particular complex traits such as plant growth have a quantitative genetic basis in which multiple loci of small or moderate effect contribute to the overall phenotype. Quantitative trait loci (QTL) analysis can identify those loci by which they become liable to elucidation of their underlying molecular basis that is characterized by the genes and, more specifically, the functional DNA polymorphisms causal to the detected QTL effects.

The resolution of QTL mapping is generally low and therefore a QTL can represent a genetic interval covering up to tens of centi-Morgan (cM). Depending on genome size, one cM can correspond from 120 kb on average for small genomes like Arabidopsis and rice to 5 Mb for barley (Alonso-Blanco *et al.*, 2005). These estimates are likely optimistic because they also vary within species depending on whether the QTL maps to heterochromatic or euchromatic genomic regions. To identify the genes that are underlying the QTL effects is therefore not straight forward. Initially, the size of the QTL needs to be reduced to a smaller genomic region by fine-mapping. This requires a population that segregates only for the region containing the QTL. Two common strategies for this make use of either near isogenic lines (NILs) or heterogeneous inbred families (HIFs). The latter depends on residual heterozygosity in the region of a QTL in individuals from a population of inbred lines, which is not always available. NILs are usually derived by marker assisted selection from a backcross population derived from a cross between a line from the mapping population and one of the founder genotypes. The fine-mapping procedure involves selecting individuals carrying recombination events within the segregating region of a studied QTL in large populations of progeny from NILs or HIFs. A further generation after selfing might be required in order to select homozygous recombinant lines. Quantifying these lines for the phenotypes of interest then allows associating the QTL effects with a smaller genomic region. Ideally, fine mapping allows reducing the region sufficiently to a single gene that would thereby be proven to underlie the QTL effects. However, due to a lack of polymorphic markers and/or a lack of recombinant plants this is rarely the case. In practise fine-mapping is often followed by a candidate gene approach when the region containing the QTL is reduced enough to allow choosing genes based their annotation, yet other strategies exist (Alonso-Blanco *et al.*, 2005). The availability of DNA sequence information, either of a reference species or first hand, is crucial in this process. Naturally occurring sequence polymorphisms are quite ubiquitous so further prove that positively identifies the candidate gene as causal to the QTL effects is required. Complementation by plant transformation

provides convincing prove that this is indeed the case. QTL analysis and subsequent cloning has played a large role for most of the nearly 100 genes for which natural allelic variation has been positively identified to contribute to trait variation (Alonso-Blanco *et al.*, in press).

To make fine-mapping of a QTL feasible it is required that its effects can be confirmed in NILs or HIFs. Fine-mapping is therefore preceded by validation of the QTL in this selected genetic material. To facilitate fine mapping of a QTL (i.e. to prevent the necessity of analyzing large numbers of replicates of an already large population of recombinants), the QTL effects need to be of sufficient magnitude and heritability. Indeed, most of positively identified genes underlying QTL had large allelic effects (Alonso-Blanco *et al.*, in press), which certainly made fine mapping easier. However, variation for a complex trait such as plant growth results from the integration of different mechanisms from the cellular to the whole plant level over time. The genetic determinism of variation of plant growth is therefore typically not exemplified by few large effect QTL but by many small effect QTL (this thesis, chapter 2). The risk of false-positive QTL detection is higher in the case of small effect QTL. In addition, epistatic interactions and non-genetic variation might start to play a relatively large role, making validation and fine-mapping more difficult. An example of this is the situation where 14 QTL detected in an Arabidopsis RIL population were not detected using a NIL population, 10 of these QTL had been shown to interact with other QTL (Keurentjes *et al.*, 2007). So far, only one small effect QTL for plant growth has been fine-mapped to a resolution at which the underlying gene could be identified in Arabidopsis. This QTL was not detected by QTL analysis but during fine-mapping of another locus, yet, again strong epistatic interactions with a closely linked locus and the genetic background were found (Kroymann and Mitchell-Olds, 2005).

Previously, multi-trait QTL mapping has been performed specifically aimed at elucidating the genetic basis of plant growth in Arabidopsis (this thesis, chapter 2). The present work reports on the selection and analysis of NILs for validation of the QTL that were identified. Whereas QTL mapping relied on mixed-model methodology, the phenotypes of the NILs were evaluated in a straight forward way, and validation of the QTL performed by comparing the traits means to those of the accession representing the genetic background. The reason for this was that a clear segregation of the individual phenotypes should be visible in order to potentially allow fine-mapping. QTL effects could be validated in all cases which confirmed the results of the multi-trait QTL mapping methodology that was applied. The QTL effects were, however, not always as the QTL analysis predicted. It has become clear that, in particular, for small phenotypical differences between the NILs and the control, variability between experiments plays a large role. Such variability can also be expected between replicates of the QTL mapping experiment which was, however, only

performed once and thus the prediction of small QTL effects was not precise. Indications for epistatic interactions have also been revealed in the NILs, including complete reversal of expected QTL effects, which underlined the complex genetic architecture of plant growth. The validated QTL can provide the starting material for fine-mapping and cloning of genes involved in the variation of plant growth. This is already advanced for a single QTL by exploiting penetrant morphological differences between the selected NIL and L*er* accession. The results presented here not only demonstrate that the growth QTL detected using a multi-trait approach were validated but also provides a set of promising lines for determining the genes and the molecular basis involved in the variation of plant growth.

Materials and Methods

Genetic material

Several NILs with Sha introgressions in the genetic background of L*er* were generated by marker assisted selection in the progeny of crosses from specific L*er* x Sha RILs (El-Lithy *et al.*, 2004) with L*er*. A NIL with an introgression in the top of chromosome 1 (1a.105-19), and two further NILs with introgressions in the top of chromosome 5 (5a.36-8 and 5a.92-27) could be selected. In a single case a line with heterozygous Sha introgressions in the top of chromosome 3 and bottom chromosome 5 provided by Dr Carlos Alonso-Blanco (CNB, CSIC Madrid, Spain) was used as starting material. From the progeny of this line two NILs, each with the single respective introgressions were selected of which the one on chromosome 3 (3a.39) is described here. Previously, three recombinant NILs with smaller but overlapping introgressions had been derived from this NIL by co-workers (3a.226-3, 3a.94-12 and 3a.203-10). Line 56 from the L*er* x Sha RIL population (El-Lithy *et al.*, 2004) and Line 55 from the L*er* x Kas-2 RIL population (El-Lithy *et al.*, 2006) were backcrossed to L*er*. Through selfing of selected progeny NILs were selected with Sha introgressions in the bottom of chromosome 1 (K142), in the top of chromosome 2 (K231-22) and with both these introgressions (K231-31). By the same procedure NILs were selected with Kas-2 introgressions in the bottom of chromosome 2 (AM205) and in the bottom of chromosome 3 (AM242). Two NILs with Kas-2 introgressions in chromosome 3 into the genetic background of L*er* (DOG6-8 and DOG6-9) were kindly provided by Dr Leónie Bentsink, Laboratory of Molecular Plant Physiology, Utrecht, The Netherlands. Detailed information about the genotypes of the NILs is provided in figure 1. A L*er* line with the wild type *HUA2* allele (Doyle *et al.*, 2005) was provided by Dr Rick Amasino, Univ. of Wisconson, USA.

Experimental conditions

Seeds were stratified at 4°C on water saturated filter paper for 3-4 days prior to sowing. The seeds were sown in 7x7 cm square pots measuring that contained a sand peat mixture enriched with slow releasing nutrients. Four experiments were performed in Percival AR95L/3 growth chambers that exactly reproduced the environmental conditions used for QTL mapping (experiments A-D). These conditions included a day-length of 12 hours of which during the first and last 15 minutes only incandescent light bulbs were switched on to simulate dawn and dusk while for the rest of the light period fluorescent lamps were also switched on. The temperature was set at 22°C during the day and 18°C during the night at a constant relative air humidity of 70%. During a fifth experiment the chamber was shared with colleagues that required a constant temperature of 20°C but other conditions were set as described above (experiment E). A sixth experiment was performed in a custom build growth chamber (Elbanton BV) that was set-up with the same environmental conditions as for experiments A-D (experiment F). Irradiation quantity and air flow velocity at plant level was lower than in the other growth chambers. A seventh experiment was performed in a Percival AR95L/3 growth chamber under 8 hour day-length with temperature and air humidity like mentioned before (experiment G). Constant environmental conditions during validation were guaranteed by continuous monitoring with data loggers (HOBO[®] U12-012).

Plant phenotyping

The plants were photographed every 2-4 days with a ccd camera (Sony DSCF828), starting from 10-14 days after sowing (DAS). Photographing was continued until bolting time. The images resolution was 8 mega pixels and they were stored in low-compression JPEG format. The projected rosette area (PRA) of each plant was quantified with a dedicated image analysis package (Image-Pro analyzer 6.0, Media Cybernetics). Linear regression of the natural logarithm of the PRA was performed on time for each plant. The regression coefficient provided an estimate of the relative growth rate (RGR, in d^{-1}). RGR was estimated from PRA measurements taken between 14 – 25 DAS. Flowering time (FT) was quantified in DAS at which the first flower opened. The number of rosette (RLN) and cauline (CLN) leaves were counted after flowering time. Chlorophyll content index (CCI; Opti-Sciences CCM200) or, in this case equivalent, SPAD (Konica Minolta SPAD-502) was measured at flowering time. The average of single measurements on three of the largest rosette leaves was determined for each plant and used for further analysis. At least one week after flowering time, the largest leaf of each plant was harvested. The leaves were stuck to

double-sided adhesive tape and were scanned at 300 DPI using a colour scanner. The resulting images were analyzed by digital image analysis (Image-Pro analyzer 6.0, Media Cybernetics) in order to determine area (LLA), length (LLL) and width (LLW) of the largest leaf of each plant. Depending on experiment, either the full set or only subsets of these phenotypes were quantified.

Data analysis

The effect of the introgression in a NIL on each trait was tested in pair-wise comparison to the accession L*er*. The response was modelled as: trait-dependent intercept + effect on trait + residual, with intercept and effect as fixed terms and the latter a factor with two levels corresponding to L*er* and NIL. The residual was modelled explicitly as random term with diagonal variance-covariance matrix and was assumed Normally distributed. This model was fitted, on all traits simultaneously for convenience only, using the REML directive of Genstat version 10 (Payne *et al.*, 2007). The regression coefficients provided an estimate of the effects on each trait. The Wald test statistic for each trait was calculated through division of the respective squared regression coefficients by the genetic variance. From these, the corresponding p-values were calculated (chi-square distribution with 1 degree of freedom).

Results

Selection and analysis of NILs

The selection of NILs and their use to validate the detected QTL are currently ongoing processes in Cologne. Figure 1 shows the graphical genotypes of NILs that have been analyzed for validation of the respective QTL they were selected for. A recurring problem with the analysis of NILs newly selected from segregating populations was high non-genetic variability, in particular for plant growth traits. Large segregating populations for the selection of NILs were typically grown densely in the greenhouse in long days which was likely to cause poor quality seeds. Because growth and related traits are highly sensitive to environmental variation, care has been taken to obtain seed batches of uniform quality for the selected lines. This was done by growing the mother plants of the lines in the tightly controlled conditions of the growth chambers used for QTL mapping and validation. Despite this, variation of quantified traits within some selected lines persisted over more than one generation. Only results from NILs with uniform phenotypes among analyzed individuals are presented in this chapter.
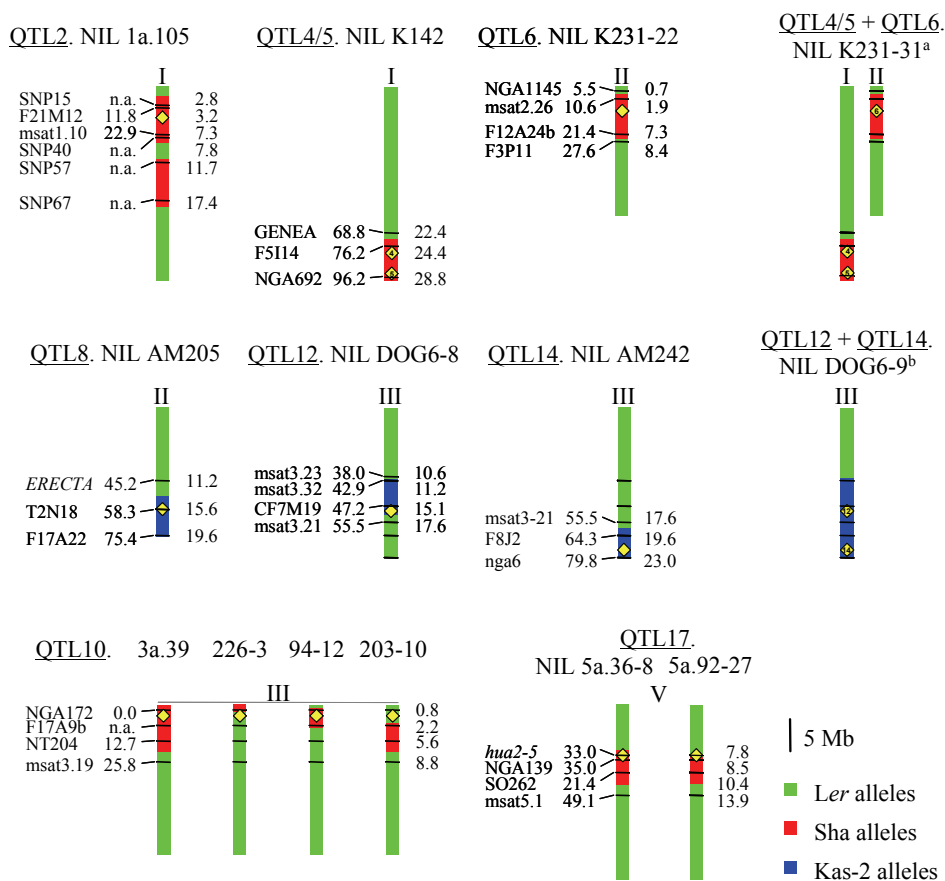
Figure 1. Graphical genotypes of the NILs selected for QTL validation. Only the chromosomes that carry an introgression of either Sha or Kas-2 alleles (in red and blue respectively) are presented and their numbers are indicated by the roman numbers. The genetic background of all NILs is from the accession L*er* (in green). Markers used for genotyping the selected lines are mentioned with their genetic position indicated on the left (in cM) and their physical position (in Mb) indicated on the right of the chromosome. The approximate positions of the QTL for which the NILs have been selected are indicated by the yellow boxes. The representations of the chromosomes themselves are scaled to their respective physical maps. [a] - Marker information is omitted because the genotype of these two NILs is the combination of the two NILs pictured to their left respectively. [b] – Marker information is omitted but the markers are the same as used for the two NILs displayed to the right. n.a. – not available, the genetic position of markers that are not on the genetic map is unknown.

Validation of QTL2

NIL 1a.105 carries two introgressions of Sha alleles on the top of chromosome 1, in the region where QTL2 has been detected, in the genetic background of L*er*. Initially, this NIL was assumed to have a single introgression of Sha alleles on the top of chromosome 1 but genome-wide SNP genotyping later revealed the second introgression below the region of QTL2 on chromosome 1 (figure 1). However, no QTL had been detected in the region of this second introgression though (El-Lithy *et al.*, 2004; this thesis, chapter 2). The NIL was analyzed in Experiment A that reproduced the environmental conditions of the QTL mapping experiment and in experiment G at a day length of 8 hours (table 1). No differences in early PRA were observed between the NIL and L*er*. The NIL did, however, have an increased RGR and in accordance PRA at 23 DAS and was significantly larger than L*er*. In experiment G only PRA at 42 DAS was quantified which is comparable to PRA at 23 DAS in experiment A, because of the difference in the photoperiod between the two experiments. However, no differences were found for PRA at 42 DAS. The NIL did have a reduced FT and CLN in both experiments. Furthermore, RLN was only reduced under 8 hours day length and SPAD, only measured under these conditions, was also reduced. These results validate the presence of a QTL in the region of the Sha introgression. However, the observed effects in the NIL do not resemble the effects detected with QTL analysis (chapter 2). The detected QTL effects predicted a reduced RGR and an accordingly reduced PRA in later phases of vegetative growth. In addition, they predicted that the NIL should not have a different FT than L*er*. On the other hand, the effect on CLN in the NIL was indeed predicted by QTL analysis.

Validation of QTL4/5

The two closely linked QTL4 and QTL5 on the bottom end of chromosome 1 (hereafter QTL4/5), detected in the L*er* x Sha population had effects of similar magnitude on approximately the same set of traits, yet, of opposite sign. NIL K142 had a single Sha introgression that covered the region of both QTL (figure 1). This NIL was analyzed in experiments C and E (table 2). In experiment C no differences for PRA at 14 and 23 DAS was found but in experiment E PRA at 11 DAS was smaller for the NIL compared to L*er*. Nevertheless, a significant positive effect on RGR was found in the NIL in experiment C. Slight differences in leaf shape were observed during experiment C when visually comparing the NIL to L*er* but no significant effects on the dimensions of the largest leaves

were found. However, in experiment E the NIL had a significantly increased length of the largest leaves (LLL). In both experiments but more strikingly in experiment E, higher values were found on RLN, CLN and FT the NIL. However, on closer inspection it was actually L*er* that had much lower values in experiment E, with its higher night temperature and possibly increased day-length, whereas the NIL behaved similar in both experiments. Finally, a significant positive effect on CCI/SPAD was found in the NIL only in experiment E. These results validated the presence of a QTL at the bottom end of chromosome 1 and the significant effects on FT, RLN and CLN found in experiment C fitted the expectations from QTL4 (chapter 2) both in sign and magnitude. The Sha introgression suppressed the response to environmental condition in the NIL which explains the much stronger effects in experiment E. A main effect on RGR was not predicted based on QTL analysis but the interaction between QTL4 and QTL5 did have a significant effect on this trait (chapter 2). Other effects predicted by QTL4/5, on PRA, LLA, LLL and CCI were not observed in the NIL in experiment C. The effects on LLL, PRA and CCI that were found in experiment E can not be compared to QTL analysis due to the effects on L*er* of different environmental conditions.

Table 1. Validation of QTL2 using NIL 1a.105-19. The experiments in which the plants were analyzed are indicated in the left column, followed by the traits that were quantified. The detected QTL effects (chapter 2) are presented in the third column and represent a substitution of the homozygous L*er* allele with a homozygous Sha allele. QTL effects that were significant are indicated by bold type. The mean trait value and the effect with respect to that value are given for the control accession L*er* and the NIL respectively. The significances of the observed effects are indicated by asterisks where $0.01 < p < 0.05$ for [*] and $p < 0.01$ for [**]. The numbers of replicates from which these estimates were derived are mentioned in the columns headed with 'n'.

exp	trait	QTL2 effect	L*er* mean	n	1a.105-19 effect	n
A	PRA8\10	0.01	0.14	22	0.03[*]	14
	PRA14\13	0.0	0.38	27	0.02	14
	PRA27\23	**-2.9**	9.2	22	2.7[**]	14
	RGR	**-0.016**	0.306	22	0.014[**]	14
	RLN	-0.8	15.0	21	-0.4	11
	CLN	**-0.8**	6.10	21	-0.9[**]	11
	FT	-0.1	36.8	21	-1.7[**]	11
G	PRA42	**nd**	35.8	9	-1.6	7
	RLN	**0.0**	27.4	9	-6.3[**]	10
	CLN	**-0.8**	11.2	9	-2.4[*]	11
	FT	-0.1	60.4	10	-3.2[*]	10
	CCI\SPAD	-0.5	30.1	6	-2.3[*]	10

Validation of QTL6

NIL K231-22, with a Sha introgression in the region of QTL6 on chromosome 2, was analyzed in the same 2 experiments as was the NIL K142, selected for QTL4/5 described above (table 2). Only a significant positive effect on RGR was found for NIL K231-22 in experiment C, and in experiment E significant negative effects on only PRA at 23 DAS and CCI\SPAD were found. Although a QTL for RGR can be validated by these results, QTL analysis predicted a stronger effect on this trait. However, based on QTL analysis the NIL was also predicted to have increased PRA at 23 DAS and increased dimensions of the largest leaf but this was not found. The effects found on PRA at 23 DAS and CCI found in experiment E can not be directly compared to QTL analysis because of the different environmental conditions. They do, however, validate the presence of a QTL for these traits in this region.

The combined effects of QTL4/5 and QTL6

NIL K231-31 was selected for having both the introgressions of NILs K142 and K231-22 that should contain QTL4/5 and QTL6. This NIL was also analyzed in experiments C and E, allowing direct comparison with K142 and K231-22 (table 2). In experiment C only, PRA at 23 DAS and RGR as well as LLL were increased in this NIL, compared to L*er*. In experiment E, the NIL also had increased LLA and LLW, in addition to LLL. A different leaf shape from both L*er* and K142 could be noticed for this NIL. In addition, the rosette leaves were often bend or twisted as if under tension. In both experiments NIL K231-31 flowered later, and had increased leaf numbers compared to L*er*. CCI was strongly reduced in the NIL in experiment C while no effect on this was found in experiment E. The PRA at 23 DAS and RGR of NIL K231-31 did not show obvious deviation expected for additive effects of QTL4/5 (NIL K142) and QTL6 (NIL 231-22). For the remaining traits, however, this was less obvious. In both experiments the FT and RLN of NIL 231-31 were substantially reduced when compared to K142 even though in NIL K231-22 no effects on these traits had been found. The same was true for the dimensions of the largest leaf. In experiment E, the much larger LLA and LLL of NIL K231-31 could not be achieved by addition of the effects on these traits found in NILs K142 and K231-22. A non-additive effect on LLL was detected by QTL analysis for the interaction between QTL5 and QTL6 (chapter 2) but not for FT, RLN, LLA and LLW. These results do indicate that a QTL is located on the top of chromosome 2 that can affect the traits for which strong QTL effects

had been detected. However, in this particular case they are discovered as a circumstantial interaction with the environment.

Table 2. Validation of QTL4, QTL5 and QTL6 using NILs K142, K231-22 and K231-31. The experiments in which the plants were analyzed are indicated in the left column, followed by the traits that were quantified. The detected QTL effects (chapter 2) are presented in the third column and represent a substitution of the homozygous L*er* allele with a homozygous Sha allele. QTL effects that were significant are indicated by bold type. The mean trait values and the effects with respect to those values are shown for the control accession L*er* and the NILs respectively. The significances of the observed effects are indicated by asterisks where $0.01 < p < 0.05$ for * and $p < 0.01$ for **. The numbers of replicates from which these estimates were derived are mentioned in the columns headed with 'n'.

| | | | | K142 | | | | | K231-22 | | K231-31 (QTL4/5+ QTL6) | |
| | | **QTL4** | **QTL5** | L*er* | | (QTL4/5) | | **QTL6** | (QTL6) | | | |
exp	trait	effect	effect	mean	n	effect	n	effect	effect	n	effect	n
C	PRA14\14	0.06	**-0.09**	1.1	22	-0.04	22	0.05	-0.02	22	0.08*	22
	PRA27\23	**3.8**	**-3.4**	9.5	21	0.29	22	**5.0**	0.41	22	1.02*	22
	RGR	0.005	-0.001	0.239	21	0.008*	22	**0.014**	0.009**	22	0.022**	22
	LLA	**0.9**	**-0.7**	12.3	11	-0.75	13	**1.8**	0.20	13	0.42	12
	LLL	**0.6**	**-0.5**	7.7	11	0.04	13	**0.9**	0.26	13	0.79**	12
	LLW	0.10	-0.03	2.5	11	-0.11	13	**0.21**	-0.03	13	-0.09	12
	RLN	**3.3**	**-3.0**	15.2	11	3.20**	13	-0.19	0.28	13	1.15*	12
	CLN	**1.0**	**-0.8**	5.6	11	1.83**	13	-0.09	0.36	13	1.28**	12
	FT	**2.3**	**-1.9**	35.3	11	2.14**	12	-0.05	0.73	12	1.48*	12
	CCI	-0.3	**-1.9**	16.1	11	-1.08	11	-0.74	-0.01	11	-3.35**	11
E	PRA14\11	0.06	**-0.09**	0.3	10	-0.08**	12	0.05	-0.04	13	-0.04	13
	PRA27\23	**3.8**	**-3.4**	11.3	10	-1.69	12	**5.0**	-1.44*	13	-1.23	13
	RGR	0.005	-0.001	0.291	10	0.008	12	**0.014**	0.000	13	0.002	13
	LLA	**0.9**	**-0.7**	10.7	10	0.13	10	**1.8**	-0.10	13	2.46**	8
	LLL	**0.6**	**-0.5**	6.3	10	0.56**	10	**0.9**	0.14	13	1.56**	8
	LLW	0.10	-0.03	2.5	10	0.04	10	**0.21**	-0.13	13	0.13*	8
	RLN	**3.3**	**-3.0**	10.7	10	8.63**	12	-0.19	-0.24	13	7.13**	12
	CLN	**1.0**	**-0.8**	3.4	10	3.35**	12	-0.09	0.52	13	3.30**	10
	FT	**2.3**	**-1.9**	29.8	10	10.53**	12	-0.05	0.05	13	7.20**	11
	CCI\SPAD	-0.3	**-1.9**	29.2	10	3.51**	12	-0.74	-1.41*	13	1.35	10

Validation of QTL8

QTL8 was detected in the L*er* x Kas-2 population, closely linked to QTL7, at the *ERECTA* locus. NIL AM205 has been selected with a Kas-2 introgression in the region of QTL8 but not including the *ERECTA* locus. So far this NIL has only been analyzed in experiment C

(table 3). The NIL had a smaller PRA at 10 and 13 DAS than L*er* but it also had an increased RGR and accordingly the PRA at 23 was no longer significantly different. Furthermore, the NIL had an increased RLN. The reduced PRA and increased RLN could be predicted by QTL analysis (chapter 2) but the effect size found for these traits in the NIL was larger and smaller respectively. An effect on RGR was not predicted by QTL analysis even though in that analysis an effect on later PRA was absent.

Table 3. Validation of QTL8 using NIL AM205. The experiment in which the plants were analyzed is indicated in the left column, followed by the traits that were quantified. The detected QTL effects (chapter 2) are presented in the third column and represent a substitution of the homozygous L*er* allele with a homozygous Kas-2 allele. QTL effects that were significant are indicated by bold type. The mean trait values and the effects with respect to those values are shown for the control accession L*er* and the NIL respectively. The significances of the observed effects are indicated by asterisks where $0.01 < p < 0.05$ for [*] and $p < 0.01$ for [**]. The numbers of replicates from which these estimates were derived are mentioned in the columns headed with 'n'.

		QTL8		L*er*	AM205
exp	V2	effect	SE	mean	effect
C	PRA8/10	**-0.01**	0.005	0.3	-0.09[**]
	PRA14/13	**-0.10**	0.049	1.1	-0.22[**]
	PRA27/23	-1.0	1.0	9.5	-0.9
	RGR	-0.001	0.004	0.239	0.016[**]
	RLN	**4.4**	1.5	15.18	2.2[*]
	CLN	-0.2	0.2	5.64	0.2
	FT	**2.2**	0.8	35.27	1.2
	LLA	0.06	0.36	12.3	-0.24
	LLL	0.14	0.20	7.73	-0.08
	LLW	-0.05	0.06	2.55	-0.06

Validation of QTL10

Four NILs with Sha introgressions in the genetic background of L*er* have been analyzed in order to validate and to narrow down QTL10 located on the top of chromosome 3. NIL 3a.39 had a large Sha introgression covering the QTL and the three recombinant NILs 3a.226-3, 3a.94-12 and 3a.203-10 had smaller but overlapping introgressions. These lines were phenotyped in experiments A and C but from the latter, good data could only be collected from NILs 3a.39 and 3a.226-3 (table 4), due to bad performance of the other two lines. NILs 3a.39, 3a.226-3 and 3a.203-10 had a reduced early PRA (10/13 DAS) when compared to L*er* and this effect was strongest in NIL 3a.203-10. The latter NIL was the only one that also had a significantly reduced PRA at 23 DAS. Nevertheless, all three NILs had an increased RGR when compare to that of L*er*. NILs 3a.39 and 3a.226-3 had a reduced FT, RLN and CCI, yet,

in NIL 203-10 RLN and FT were significantly increased compared to L*er*. A small reduction in CLN was also found as the only difference between NIL 3a.94-12 and L*er*. NILs 3a.39 and 3a.226-3 furthermore had decreased LLA and LLL. These traits were not quantified for the two other NILs. The differences observed between L*er* and both the NILs 3a.39 and 3a.226-10 match to a large extend what could be predicted from QTL analysis. Furthermore, the phenotypes of both NILs were highly similar to each other. The introgressions of NILs 3a.226-3 and 3a.203-10 did not overlap but the introgression of NIL 3a.39 included both of them. This suggested the presence of 2 QTL and that the one present in the introgressed region of NIL 3a.226-3 is epistatic to the one present in the introgressed region of NIL 3a.203-10. The strongly increased RGR, and the reduced LLA and CCI were not expected based on QTL analysis.

Table 4. Validation of QTL10 using NILs 3a.39, 3a.226-3, 3a.94-12 and 3a.203-10. The experiments in which the plants were analyzed are indicated in the left column, followed by the traits that were quantified. The detected QTL effects (chapter 2) are presented in the third column and represent a substitution of the homozygous L*er* allele with a homozygous Sha allele. QTL effects that were significant are indicated by bold type. The mean trait values and the effects with respect to those values are shown for the control accession L*er* and the NILs respectively. The significances of the observed effects are indicated by asterisks where $0.01 < p < 0.05$ for [*] and $p < 0.01$ for [**]. Twelve to fourteen plants were analyzed per NIL, per trait and 21-22 plants for L*er*.

exp	trait	QTL10 effect	L*er* mean	3a.39 effect	3a.226-3 effect	3a.94-12 effect	3a.203-1 effect
A	PRA8/10	**0.01**	0.14	-0.04[**]	-0.03[*]	-0.02	-0.07[**]
	PRA14/13	**0.08**	0.38	-0.11[**]	-0.06[*]	-0.02	-0.20[**]
	PRA27/23	0.5	9.2	-1.0	0.7	0.2	-2.9[**]
	RGR	0.001	0.306	0.025[**]	0.019[**]	0.001	0.025[**]
	RLN	-1.4	15.0	-1.3[*]	-2.1[**]	-0.5	2.9[**]
	CLN	0.1	6.1	-0.5[*]	-0.7[**]	-0.5[*]	0.4
	FT	**-2.8**	36.8	-3.1[**]	-3.0[**]	-0.3	2.1[**]
C	PRA8/10	**0.01**	0.326	-0.08[**]	-0.10[**]	-	-
	PRA14/13	**0.08**	1.086	-0.22[**]	-0.26[**]	-	-
	PRA27/23	0.5	9.5	0.0	-0.8	-	-
	RGR	0.001	0.239	0.028[**]	0.022[**]	-	-
	LLA	-0.31	12.29	-1.12[*]	-1.32[**]	-	-
	LLL	**-0.35**	7.73	-0.89[**]	-0.77[**]	-	-
	LLW	-0.06	2.55	-0.01	-0.07	-	-
	RLN	-1.4	15.2	-2.8[**]	-3.2[**]	-	-
	CLN	0.1	5.6	-0.6	-0.9[**]	-	-
	FT	**-2.8**	35.3	-3.2[**]	-3.3[**]	-	-
	CCI	-0.7	16.1	-1.7[**]	-1.8[**]	-	-

Validation of QTL12

NIL DOG6-8, with an introgression of Kas-2 alleles in the region of QTL12, was analyzed in the experiments B and C (table 5). This NIL had a reduced PRA at 23 and 24 DAS and although this was accompanied by a trend of reduced RGR these differences were not significant. The NIL also had a consistently lower LLA, LLL and LLW than L*er*. Typically, in both experiments the RLN of the NIL was decreased while its CLN was increased. A reduced FT and an increased CCI were only observed in the second experiment though. These differences between the NIL and L*er* clearly confirm the presence of a QTL in the region of introgressed Kas-2 alleles. However, the differences are contradictory to what could be predicted by QTL analysis. In particular, an increase in PRA and RGR was predicted and all other traits should not show any difference when comparing the NIL to L*er*.

Validation of QTL14

NIL AM242, with Kas-2 alleles at the bottom end of chromosome 3, was selected for validation of QTL14, which was specific to the L*er* x Kas-2 population. This NIL was analyzed in the same two experiments as NIL DOG6-8, described above (table 5). In both experiments the NIL had a reduced PRA at 23 or 24 DAS and in experiment B also PRA at 13 DAS was reduced. This reduction in PRA at 23 and 24 DAS was consistently accompanied by a reduced RGR. On the contrary, the NIL had consistently increased LLA and LLW, but its LLL was not different from L*er*. The RLN of the NIL was higher than L*er* in experiment B while CLN was not different but in experiment C it was quite the opposite were CLN was higher than L*er* and RLN was not different. The NIL did flower later than L*er* but its CCI was increased in both experiments. The phenotypic differences in PRA and RGR between the NIL and L*er* validate the QTL effects that were detected for these traits. However the different RLN and CLN observed between the NIL and L*er* did not match the prediction based on QTL analysis. QTL analysis predicted the NIL to be similar to L*er* for the remaining traits. The NIL furthermore showed particular morphological differences when compared to L*er*. These differences were not quantified but included the notion of generally larger organs, in particular thicker stems and larger flowers. In addition, reduced apical dominance was observed in both experiments that led to the total height of the plant being determined by axillary branches rather than the main inflorescence. The overall posture of the plant was thereby distinctly different from L*er* (figure 2).

NIL AM242 was selected from an F2 population derived from a cross between a RIL and L*er* in which furthermore the entire chromosome 2 and the bottom of chromosome 4 segregated for Kas-2 alleles. During selection under long-day length conditions in the greenhouse, differences in morphology as described above, including the effects on LLW that were validated in the two experiments described above, could be clearly identified (figure 2). These differences were highly penetrant and could be discriminated regardless of the allelic value at *ERECTA*, which also clearly affects plant morphology. This has been exploited to fine-map the QTL for this trait. The effect on morphology was scored as either present, absent or segregating in the progenies of 39 F2 plants that were selected for being recombinant in the region of QTL14. Genotyping of these plants with additional markers that were designed for this region allowed narrowing down the location of this QTL to a region between 23.04 Mb and 23.47 Mb, the end of the chromosome, containing 151 genes. These results were later confirmed by analyzing the progeny of two lines selected from the initial fine-mapping. The presence of the above mentioned distinct morphological differences from L*er* confirmed that this QTL is located at very end of chromosome 3 (figure 2).

The combination of QTL12 and QTL14

QTL12 and QTL14 are genetically linked and NIL DOG6-9 allowed the analysis of the combined effects of both QTL in the same genetic background (table 5). The PRA at 23 DAS of the NIL was lower than L*er* in both experiments but earlier PRA was only lower in the first experiment. These observations agreed with a significant but weak reduction in RGR only in the second experiment. The NIL further displayed a reduction in LLL in both experiments while the LLA of the NIL was only significantly lower than L*er* in the first experiment. A reduction in RLN and CLN was also found in both experiments but not for FT, suggesting this to be a plastochron effect. Finally, the CCI was increased in experiment C but not significantly in experiment B. In absence of any significant interaction between QTL12 and QTL14 (chapter 2), the only expectation for the phenotype of NIL DOG6-9 would be the additive effect of the differences found between L*er* and, NILs DOG6-8 and AM242. This seems to hold, roughly, for FT, leaf numbers and the dimensions of the largest leaf, in particular in experiment C. PRA and RGR of NIL DOG6-9, on the other hand, are highly similar to the individual values found for DOG6-8 or AM242, when significant, and thus do not appear to be additive in DOG6-9.

Table 5. Validation of QTL12 and QTL14 using NILs DOG 6-8, AM242 and DOG6-9. The experiments in which the plants were analyzed are indicated in the left column, followed by the traits that were quantified. The detected QTL effects (chapter 2) are presented in the third column and represent a substitution of the homozygous L*er* allele with a homozygous Kas-2 allele. QTL effects that were significant are indicated by bold type. The mean trait values and the effects with respect to those values are shown for the control accession L*er* and the NILs respectively. The significances of the observed effects are indicated by asterisks where $0.01 < p < 0.05$ for * and $p < 0.01$ for **. The numbers of replicates from which these estimates were derived are mentioned in the columns headed with 'n'.

exp	trait	QTL12 effect	QTL14 effect	L*er* mean	n	DOG6-8 (QTL12) effect	n	AM242 (QTL14) effect	n	DOG6-9 (QTL12+QTL14) effect	n
B	PRA14\13	**0.12**	-0.03	0.64	15	-0.11	15	-0.15**	14	-0.16**	14
	PRA27\24	**2.38**	**-2.94**	15.1	15	-4.07**	14	-3.62**	14	-3.40**	15
	RGR	**0.015**	**-0.015**	0.293	15	-0.007	15	-0.011**	14	-0.002	14
	LLA	0.59	-0.16	18.3	15	-5.16**	14	1.53*	15	-1.89**	15
	LLL	0.37	-0.01	9.5	15	-1.67**	14	-0.42	15	-0.86**	15
	LLW	0.12	-0.05	3.2	15	-0.36**	14	0.47**	15	0.03	15
	RLN	0.48	**-2.95**	15.9	15	-0.74*	16	1.82**	16	-0.80**	15
	CLN	0.21	**-0.63**	6.7	15	1.21**	16	-0.48	16	-1.04**	16
	FT	-0.77	0.75	37.4	15	-0.46	16	4.85**	16	0.79	16
	CCI	1.28	0.16	16.3	15	0.79	15	1.03	16	1.28	14
C	PRA8\10	0.01	0.01	0.3	21	-0.01	22	0.00	22	-0.04	22
	PRA14\14	**0.12**	-0.03	1.1	22	-0.10	22	0.08	22	-0.09	22
	PRA27\23	**2.38**	**-2.94**	9.5	21	-1.55**	22	-0.83	22	-1.51**	22
	RGR	**0.015**	**-0.015**	0.239	21	-0.007	22	-0.016**	22	-0.007*	22
	LLA	0.59	-0.16	12.3	11	-3.56**	12	1.84**	12	-0.70	13
	LLL	0.37	-0.01	7.7	11	-1.34**	12	-0.02	12	-0.47**	13
	LLW	0.12	-0.05	2.5	11	-0.32**	12	0.38**	12	0.05	13
	RLN	0.48	**-2.95**	15.2	11	-1.18**	13	-0.57	13	-2.26**	13
	CLN	0.21	**-0.63**	5.6	11	0.67**	13	-1.41**	13	-0.48*	13
	FT	-0.77	0.75	35.3	11	-1.43**	13	1.48**	12	0.42	13
	CCI	1.28	0.16	16.1	11	1.84**	12	3.28**	11	1.46*	13

Figure 2. Features of plant morphology observed in NIL AM242, associated with the Kas-2 introgression in the bottom of the chromosome 3 in the region of QTL14 and fine-mapping of this QTL. (A) Plants grown in long-days in the greenhouse, note how the leaf shape of AM242 (bottom) is more circular than the lancet shape of leaves from L*er* (top). (B) Plants from a QTL validation experiment in 12 hour day length, arrows indicate the reduced apical dominance in AM242 (middle and right) compared to L*er* (left). (C) Graphical genotypes of the 38 selected recombinant F2 and the score indicating how the morphological characteristics of QTL14 segregated in their progeny. Only the bottom end of chromosome 3 is drawn and each bar represents a single line. The lines 'a' and 'b', which were chosen after additional genotyping to verify the results by studying their progeny (F3), are drawn to the right. The markers used to genotype the plants are mentioned next to the chromosomes, which are colour coded based on their allelic values: green – L*er*, blue – Kas-2, yellow – heterozygous, grey – unknown.

Validation of QTL17

For validation of QTL17, the NILs 5a.36-8 and 5a.92-27 were available, both with Sha introgressions. These NILs were initially selected for the validation of this QTL after its detection in an earlier study (El-Lithy *et al.*, 2004). Although this QTL was again discovered in the present study (chapter 2), it was located slightly more towards the top of the

chromosome. The introgressions in NILs 5a.36-8 and 5a.92-27 (figure 1) did not fully cover the QTL detected on the middle of chromosome 5 as they were selected based on the results by El-Lithy *et al.* (2004). The NILs were analyzed in the two experiments A and G (table 6). The only difference in PRA observed was a reduced PRA at 42 DAS for NIL 5a.36-8 under short days in experiment G. NIL 5a.92-27 nevertheless had a higher RGR than L*er* in experiment A. The CLN of both NILs was higher than that of L*er* in both experiments but only NIL 5a.36-8 also had consistently increased RLN which did not differ for NIL 5a.92-27 in short days. Only NIL 5a.36-8 had an increased CCI compared to L*er*, in both experiments.

Table 6. Validation of QTL17 using NILs 5a.36-8 and 5a.92-27. The experiments in which the plants were analyzed are indicated in the left column, followed by the traits that were quantified. The detected QTL effects (chapter 2) are presented in the third column and represent a substitution of the homozygous L*er* allele with a homozygous Sha allele. QTL effects that were significant are indicated by bold type. The mean trait values and the effects with respect to those values are shown for the control accession L*er* and the NILs respectively. The significances of the observed effects are indicated by asterisks where $0.01 < p < 0.05$ for [*] and $p < 0.01$ for [**]. The numbers of replicates from which these estimates were derived are mentioned in the columns headed with 'n'.

		QTL17		L*er*		5a.36-8		5a.92-27	
exp	trait	effect	SE	mean	n	effect	n	effect	n
A	PRA8/10	-0.005	0.004	0.1	22	-0.004	12	-0.008	13
	PRA14/13	-0.04	0.03	0.4	22	-0.01	15	-0.04	14
	PRA27/23	-2.00	1.09	9.2	22	0.02	12	0.59	13
	RGR	**-0.011**	0.004	0.306	22	0.003	12	0.013[**]	13
	RLN	**1.8**	0.67	15.0	21	5.0[**]	12	4.9[**]	11
	CLN	0.2	0.22	6.1	21	1.3[**]	12	1.0[**]	11
	FT	**1.3**	0.49	36.8	21	2.8[**]	12	-0.5	11
G	PRA42	nd	-	35.8	9	-5.3[*]	8	1.6	8
	RLN	**1.8**	0.67	27.4	9	3.0[*]	11	1.3	11
	CLN	0.2	0.22	11.2	9	1.1[*]	11	1.3[**]	11
	FT	**1.3**	0.49	60.4	10	0.7	11	-0.4	11
	CCI	**3.0**	0.49	30.1	6	2.6[*]	8	0.8	8

Considering that the L*er* parent of both L*er* x Kas-2 and L*er* x Sha population carries a mutant allele of the gene *HUA2* (Doyle *et al.*, 2005), located in the region of QTL17, a L*er* strain with a wild-type *HUA2* allele was also analyzed (table 7).

Table 7. Validation of *HUA2* as candidate gene for QTL17 using a L*er* strain with a wild-type *HUA2* allele and the control L*er* with the mutant *hua2-5* allele. The experiments in which the plants were analyzed are indicated in the left column, followed by the traits that were quantified. The detected QTL effects (chapter 2) are presented in the third column and represent a substitution of the homozygous L*er* allele with a homozygous Sha allele. QTL effects that were significant are indicated by bold type. The mean trait values and the effects with respect to those values are shown for the control accession L*er* and the L*erHUA2* respectively. The significances of the observed effects are indicated by asterisks where $0.01 < p < 0.05$ for [*] and $p < 0.01$ for [**]. The numbers of replicates from which these estimates were derived are mentioned in the columns headed with 'n'.

| | | QTL17 | | | | L*er* | | L*erHUA2* | |
| | | L*er* x Sha | | L*er* x Kas-2 | | | | | |
exp	trait	effect	SE	effect	SE	mean	n	effect	n
D	RLN	**1.8**	0.67	**10.0**	1.43	14.9	10	3.9[**]	9
	CLN	0.2	0.22	**1.1**	0.19	5.7	10	1.6[**]	9
	FT	**1.3**	0.49	**5.7**	0.71	34.0	10	3.9[**]	9
	CCI	**3.0**	0.49	**8.8**	1.35	16.1	10	2.9[*]	9
F	RLN	**1.8**	0.67	**10.0**	1.43	8.3	18	2.8[**]	23
	CLN	0.2	0.22	**1.1**	0.19	3.6	18	1.1[**]	23
	FT	**1.3**	0.49	**5.7**	0.71	37.2	18	1.3[*]	23
	LLA	-0.5	0.28	**-1.3**	0.34	3.7	18	1.3[**]	23
	LLL	0.3	0.13	-0.3	0.19	4.2	18	1.1[**]	23
	LLW	**-0.1**	0.04	**-0.3**	0.06	1.6	18	0.13[*]	23

The NILs and the L*er HUA2* lines were analyzed in separate experiments but the environmental conditions were again set-up as they had been during QTL analysis. L*erHUA2* consistently had increased values over the control L*er* for all traits analyzed, which included RLN, CLN, FT, CCI, LLA, LLL and LLW. The sign of the effects on RLN, CLN and FT matched the prediction based on QTL analysis for QTL17 in both RIL populations. However, apart from CLN, only a prediction based on QTL analysis from the L*er* x Sha population would resemble the phenotype of L*erHUA2* while the predicted trait values based on the L*er* x Kas-2 population were much higher. Considering this, *HUA2* presents a good candidate gene for QTL17 in the L*er* x Sha population. Although the CLN of L*erHUA2* did more resemble a prediction based on the effects of QTL17 as detected in the L*er* x Kas-2 population, the large deviation from expectation for RLN and FT speak against *HUA2* being, solely, responsible in this population. Furthermore, the size of the largest leaves from L*erHUA2* was contradictory to predictions from QTL analysis, particularly in the L*er* x Kas-2 population. NIL 5a.36-8 carried the Sha allele of *HUA2*, without the L*er* mutation, and had similar values for RLN, CLN and FT as L*erHUA2*. Although this also supports *HUA2* as candidate gene for QTL17 in this population, leaf

numbers (but not FT) were equally increased in NIL 5a.92-27, which carries the L*er* allele of *HUA2*. The presence of an additional QTL (or more than one), apart from *HUA2* in the region of QTL17 in the L*er* x Sha population is therefore likely and this QTL would confer increased RGR.

Discussion

With few exceptions, all QTL for which the underlying genes and functional polymorphisms have been identified so far were of large effect. However, QTL of large effect are rare (Salvi and Tuberosa, 2005), and as far as we know, no major QTL involved in plant growth have been detected in Arabidopsis (El-Lithy *et al.*, 2004; this thesis, chapter 2). Nevertheless, it is possible to fine-map QTL with small effects on plant growth (Kroymann and Mitchell-Olds, 2005). The QTL mapping experiment presented in this thesis (chapter 2) was conducted specifically to detect QTL involved in the control of plant growth. The aim of the work presented herein was to validate some of those QTL by selecting NILs covering the QTL regions and to evaluate the feasibility of fine-mapping them. At all 8 loci for which NILs have been selected and analyzed so far, QTL effects could be attributed to the respective introgressions. In general, more, and weaker significant QTL effects were revealed by analysis of the NILs than were detected by QTL analysis (considering only the traits quantified per case). This result is in accordance with an the increased power to detect main effect QTL in NILs as opposed to RILs (Keurentjes *et al.*, 2007; Reif *et al.*, 2009). With respect to the analysis presented in this chapter, there are either additional small effect QTL present in the introgressed regions in the NILs or additional pleiotropic effects of the detected QTL have been discovered. In particular for the growth trait RGR, effects were found in NILs that were selected for QTL at which no such effects had been detected. All but one of the validated QTL with effects on RGR also had effects on flowering time, leaf number, and/or leaf size, which fits with plant growth being a complex trait resulting from such different mechanisms (Tisné *et al.*, 2008; Alonso-Blanco *et al.*, 2009). This despite the fact that the QTL regions targeted in the NILs were selected primarily for loci not directly involved in flowering time (i.e. not those QTL accounting for most variance of this trait). The observed were only weak and it could be that during QTL mapping their contribution to the total genetic variance for flowering time was negligible. An interesting result was that the negative correlation between flowering time and RGR (increased RGR when decreased FT) found for flowering time QTL (chapter 2) was not present in these NILs. These flowering time effects are likely to be secondary effects of the plant growth QTL.

Although the presence of QTL could be validated at the loci for which NILs have been developed, in no single case did the observed effects match exactly the prediction from QTL analysis for all traits. A known issue with the estimation of QTL effects is the so-called Beavis effect. Beavis (1998) demonstrated that the estimated genetic variances of correctly identified QTL were greatly overestimated when a population of 100 progeny were evaluated, and that it took 1000 progeny to get estimates fairly close to the actual magnitude. With the small population size of 100 progeny the effects were typically overestimated to be 10 higher than the actual magnitude. When generally comparing the observed effects in the NILs to the detected QTL effects, no systematic overestimation of these effects appear. In fact, observed effects in the NILs did regularly match or even exceed the prediction based on QTL analysis, hence the Beavis effect seems to play little role in this case. The variation found between the various experiments provides a reasonable explanation why predicted QTL effects are not always validated or why validated effects deviate from prediction by QTL effects. The variation in phenotypical values for the accession L*er* between the various experiments was considerable although the experimental conditions of QTL mapping were most often exactly reproduced. This indicates that although the same growth chamber, the same soil, the same temperature, etc. were used there is a source of growth variation that was not controlled. This source could consist of (e.g.) deterioration of lights sources in the chamber over time, variation between soil batches or variation in seed quality between experiments. The weakness in predictive power of the QTL analysis could be due to the fact that the experiment was only performed once and that many detected QTL effects were of small magnitude and thus prone to variation between experiments.

QTL analysis included a test for two-way interactions between the detected main effect QTL (this thesis, chapter 2). A high number of interactions, for some traits accounting for a relatively high amount of explained genetic variance, were detected. However, these results need to be interpreted with care because the implementation of the test assumed an interaction would affect multiple trait did not necessarily have to be the case. However, subsequent analysis with two-segment NILs allows validating the detected interactions. RGR was one of the traits for which the explained genetic variance increased most when including two-way interactions (this thesis, chapter 2). This apparent importance of epistasis is likely explained by the strong physiological constraint on this trait within species (Li *et al.*, 1998). In the NILs the magnitudes of effect on RGR were smaller than 10% (of the L*er* trait value), similar to the largest QTL effects that were detected. Nevertheless, if the validated QTL effects on RGR could be additively combined in a single NIL, this plant should grow so fast it would quickly lead to a rosette size that is not observed for Arabidopsis. A more elaborate study using two-segment NILs led to the conclusion that additive-by-additive epistasis does

indeed strongly influence the genetic architecture of growth-related traits (Reif *et al.*, 2009). In the present study the interaction that was detected between QTL4/5 and QTL6 in the L*er* x Sha provides a strong indication that indeed there is an epistatic interaction between both loci. This supports the validation of QTL6, in particular since the predicted strong main effects on several traits were not found in the NIL for this QTL. Nevertheless, an effect on RGR could be validated, as the only demonstration of a QTL affecting this trait without affecting flowering time. QTL6 was among those with most interactions detected during QTL analysis in both mapping populations. A large number of unlinked interacting loci would provide a genetic background in which allelic variation at QTL6 could appear as a main effect, when those interactions are similar to the one with QTL4/5. For future investigation of this hypothesis, the NIL selected for QTL6 has been crossed to several other NILs with introgressions at loci where QTL that interacted with QTL6 were detected. The possibility of an interaction between QTL4 and QTL5 being responsible for the effect on RGR in the NIL K142 also needs further investigation by analyzing recombinant NILs. This should first prove that indeed two linked QTL are located in this region because this NIL validated the presence of QTL4 only. The interactions that were detected for QTL2 (this thesis, chapter 2) can also provide an explanation for the radically different phenotype of the NIL 1a.105-19 that was selected for validation of this QTL. The interactions of this QTL with four other loci accounted for a large amount of genetic variance explained for RGR and for two of them a large effect on PRA at 27 DAS was also detected. The effects on flowering time and CLN in NIL 1a.105-19, on the other hand, could have been too weak to be detected as QTL effect. The detected interactions can not provide an explanation for the phenotype of DOG6-8, selected for validation of QTL12, because only a single significant interaction with mild effects was detected for this QTL. However, interactions with other loci than the detected main effect QTL can not be excluded. In particular an interaction with a closely linked locus could explain the results, such as was the case for another plant growth-rate QTL (Kroymann and Mitchell-Olds, 2005). This interaction might have been broken in the NILs that were analyzed. Taken together, these results indicate that epistatic interactions do indeed play a large role in trait variation for planet growth. However, because for most QTL two-way interactions with several other QTL were detected, in reality it is not unlikely that higher order interactions play a role in determining the phenotypical effect. In prospect of fine-mapping and cloning QTL, interactions can be exploited to provide a genetic background in which a QTL has a large effect, such as for the QTL6.

The experiment that suffered from influences due to different environmental conditions in other parts of the chamber revealed strong differences in flowering time between L*er* and the NIL K142 with Sha alleles in the region of QTL4/5. Although

circumstantial, the effects can be attributed to temperature and photoperiod variations. Induction of flowering in long-days occurs by up-regulation of *FT* and *TSF* through perception of day-length by *CONSTANS* (Turck *et al.*, 2008). Induction of flowering by changes in ambient temperature also affect expression of *FT* (Blazquez *et al.*, 2003). *FT* and *TSF* are repressed by *FLC* and *MAF1/FLM* (Sung *et al.*, 2006). Both *FT* and *MAF1/FLM* are located in the introgression of Sha alleles in NIL K142 and are thus both good candidate genes for the QTL that is/are validated in this region. The Sha allele of *MAF1/FLM* has been hypothesized to be weak based on the detection of transposon insertions upstream of the gene, and on the fact that the Sha allele of a QTL mapped to the *MAF1/FLM* region in the Bay-0 x Sha RIL population confers early flowering (Werner *et al.*, 2005). However, this hypothesis has never been tested and *MAF1/FLM* remains a good candidate gene for this QTL. The QTL that has been validated in this region can be an important repressor of flowering in the Shakdara accession. In particular, because this accession has a weak allele of *FLC* (Michaels *et al.*, 2003), and also lacks negative regulation of flowering by *MAF2* and *MAF3* (this thesis, chapter 4). Interestingly, apart from increased flowering time, RGR was also increased in the NIL. This was one example that opposes the general trend of negative correlation between both traits found with QTL mapping (this thesis, chapter 2). The tomato orthologue of the gene *FT* has recently been tied directly to plant growth in that species (Shalit *et al.*, 2009). Considering *FT* as candidate gene for the validated QTL effects would imply attenuated or even loss of function because the NIL did not respond to presumed longer day-length due to light leakage. It is intriguing to think that also the effects of plant growth are then directly related to this gene. Although no distinction between closely linked QTL and pleiotropy can be made, fine-mapping this QTL and candidate gene approach will offer better insight into the molecular basis of this QTL.

QTL14 was one of the most interesting QTL with respect to plant growth in the L*er* x Kas-2 because of its effect on RGR without having an effect on flowering time. Despite this, phenotyping NIL AM242 revealed that this QTL has also an effect on flowering time. The explanation for not detecting a FT QTL in the QTL analysis is that the effect on flowering time is rather weak. Furthermore, the results from NIL DOG6-9 showed that in combination with the effect of the linked QTL12, its effect is cancelled out. Considering the strong effects on plant growth and morphology, then, when pleiotropic, flowering time is likely not the primary trait affected also by this QTL. In fact, the validated effects resemble in particular detail those described for the gene *BIG BROTHER* (*BB*) located in the region to which this QTL was fine-mapped. This gene was identified as a negative regulator of organ size by characterisation of a mutant in Arabidopsis that displayed the formation of larger-than-normal floral organs (Disch *et al.*, 2006). Apart from larger flowers, the authors

reported thicker stems and significantly wider leaves in a *bb* knock-out mutant, which matches the results obtained with NIL AM242. However, they also reported unaltered leaf area due to slightly shorter leaves and effects on flowering time, leaf numbers and apical dominance were not mentioned. The possibility of natural variation for this gene underlying the QTL is particularly tantalizing since it is directly involved in plant growth, the primary trait under investigation. If true, Kas-2 would carry a non functional allele, at least partially. Both further fine-mapping of this QTL and a candidate gene approach are currently being performed and this will answer whether *BB* is the gene underlying the effects observed in the NIL AM242, keeping in mind that closely linked genes could be responsible for the different effects observed.

The results of QTL validation are promising with respect to elucidating the molecular basis of the detected QTL for plant growth. Pleiotropic effects on other traits can be used to potentially fine-map these QTL more efficiently. Eventually it must be tested whether indeed these effects and those on growth are pleiotropic or due to closely linked QTL. In addition, efficiency of fine-mapping can be greatly enhanced by exploiting objectively scalable phenotypes such as demonstrated for QTL10, although also here pleiotropy between those phenotypes and RGR still needs to be clearly established. Finally, the strong enhancement of phenotypes found circumstantially in one experiment show how QTL effects can be enhanced by contrasting environmental conditions. Efficient fine-mapping of a QTL can therefore be facilitated by increasing its effects when growing lines in specific range of environmental conditions. Currently, the fine-mapping of QTL4/5, QTL6 and QTL14 are ongoing making use of these observations and the developments of NILs for further QTL involved in plant growth is underway.

Chapter 4

Fine-mapping and complementation studies identify the *MAF2-MAF5* cluster as causal to natural variation for flowering time in *Arabidopsis thaliana*

Bjorn Pieper and Matthieu Reymond

Abstract

The cluster of *MADS AFFECTING FLOWERING* family members *MAF2-MAF5* have gathered attention as potentially contributing to natural variation for flowering time in *Arabidopsis thaliana*. QTL for flowering time, rosette and cauline leaf numbers (RLN, CLN), chlorophyll content index (CCI) and relative growth rate (RGR) have been previously been detected in the region of this cluster in the Landsberg *erecta* (L*er*) x Shahdara (Sha) RIL population. These QTL have been validated and fine-mapped using near isogenic lines (NILs). The resolution achieved with recombinant NILs proves that allelic variation in the *MAF2-MAF5* cluster on the bottom arm of chromosome 5 is causal to the detected QTL effects. Recombination events mapped within the *MAF* cluster, in the progeny of NILs, indicated that *MAF5* is likely not contributing to the observed effects. The Sha accession was found to carry an insertion homologous to the 3' part of *MAF3* in the 3' portion of *MAF2* that was identical to the published sequence for the alleles of the accessions Kas-1 and Chi-1 (Caicedo *et al.*, 2009). The uniqueness of the Sha allele sequenced in the present work resides in the absence of an intact *MAF3* gene. In accordance, *MAF2* or *MAF3* transcripts could not be detected in plants carrying Sha *MAF* cluster. A novel chimeric transcript was isolated coding for a full length chimeric MADS-box protein with M, I and truncated K domain of *MAF2*, and complete the K and C domains of *MAF3*. Sequence analysis and the weak effects of transformation with a genomic construct of *MAF4* from L*er* made this gene an unlikely candidate for the QTL effects. Transformation of the NIL with the L*er* allele of *MAF3* did not restore the L*er* phenotype completely. In contrast, transformation with the L*er* allele of *MAF2* increased the flowering time of the NIL beyond that of L*er*. We hypothesize that both *MAF2* and *MAF3* might be required for proper control of flowering time.

Introduction

The correct timing of the vegetative and reproductive growth phases is critical for flowering plants to optimize their reproductive fitness. In particular the genetic and molecular bases that control flowering time, the developmental switch between both phases, is well studied in the model species *Arabidopsis thaliana*. Studying mutants and natural variation in this species has led to the identification of over 70 genes that are involved in the regulation of flowering time (Koornneef *et al.*, 1998; Simpson and Dean, 2002; Boss *et al.*, 2004; Turck *et al.*, 2008). Genetic analysis has identified four pathways that promote the transition to flowering (Boss *et al.*, 2004). These pathways all converge on floral integrators, a small number of genes that determine meristem identity (Blazquez and Weigel, 2000; Samach *et al.*, 2000; Michaels *et al.*, 2005). Transcription of the floral integrator *FT* is promoted on perception in the leaf of increasing day length in spring and early summer by the photoperiod pathway. The *FT* protein is then transported to the apical meristem in the phloem where it promotes the transition to flower development. Under specific conditions Arabidopsis accessions might display winter-annual behaviour (Wilczek *et al.*, 2009), which requires exposure of the rosette to a prolonged period of cold temperature to relief repression of flowering by the vernalization pathway.

The MIKC type MADS-box transcription factor *FLC* acts quantitatively as a floral repressor and plays a central role in the regulation of flowering time. Its expression in the phloem of the leaf and in the apical meristem represses the expression of *FT* and the floral integrator *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1* (*SOC1*) respectively, by directly binding to *cis*-elements (Searle *et al.*, 2006). The induction of flowering by the autonomous pathway acts through the repression of *FLC* expression (Marquardt *et al.*, 2006). The winter-annual behaviour of Arabidopsis is established by an epistatic interaction between the *FRIGIDA* (*FRI*) gene and *FLC*, in which *FRI* activates the expression of *FLC* (Johanson *et al.* 2000; Michaels and Amasino 1999). The resulting high levels of *FLC* transcript can suppress flowering promoting signals from the photoperiod pathway (Searle *et al.*, 2006). In response to vernalization, a prolonged period of cold temperature, *FLC* transcript and protein levels are reduced (Michaels and Amasino 1999; Sheldon *et al.* 1999; Sheldon *et al.* 2000), and flowering is promoted. Reduction of *FLC* expression is achieved through epigenetic silencing of *FLC* chromatin mediated by VIN3, and which requires the action of VRN2 and VRN1 to reach a stabile inactive state (Gendall *et al.* 2001; Levy *et al.* 2002; Sung and Amasino 2004).

Most of the natural variation for flowering time in Arabidopsis has been suggested to be explained by *FRI* and *FLC*. Active alleles at both genes confer a vernalization

requirement and thereby a winter-annual flowering strategy. The requirement for vernalization ensures that the plant overwinters as a rosette and flowers in early spring. In regions with milder winters, a spring-annual, rapid cycling flowering strategy due to the loss of vernalization requirement would be beneficial. Several non-functional alleles of *FRI* have been identified that are suggested to have arisen independently (Johanson *et al.*, 2000; Le Corre *et al.*, 2002; Shindo *et al.*, 2005). Natural allelic variation at *FRI* has been shown to explain 23-70% of the phenotypic variation for flowering time (Lempe *et al.*, 2005; Shindo *et al.*, 2005; Werner *et al.*, 2005b). Null alleles such as commonly found for *FRI*, or other polymorphisms that clearly impair gene function are rare for *FLC*. However, two haplotype groups (*FLC^A* and *FLC^B*) have been identified based on a common polymorphism in the first intron, which affect flowering time in absence of a functional *FRI* allele (Caicedo *et al.*, 2004; Scarcelli *et al.*, 2007). A field experiment with 136 Arabidopsis accessions under spring and winter-annual conditions did show fitness effects of *FRI* functionality that correlated with *FLC* haplotypes (Korves *et al.*, 2007). However, the authors did not exclude the possibility that the genetic background of the respective haplotypes might also explain the results to some extend. More recently, artificial selection for flowering time under simulated spring and winter-annual conditions led to altered allele frequencies for *FRI* only under the former condition while no differences were found for *FLC* (Scarcelli and Kover, 2009).

The strong dependency of the *FRI FLC* genotypes on vernalization to flower has established *FLC* as the main component in this pathway. However, loss of *FLC* activity did not lead to complete abolishment of the vernalization response (Michaels and Amasino, 2001). This indicates other genes act in the vernalization pathway in the control of flowering time. *FLC* belongs to a clade of six related genes in Arabidopsis that further include the five *MADS AFFECTING FLOWERING* (*MAF1-MAF5*) genes. All five genes have been shown to be capable of acting as floral repressors and to be regulated by vernalization (Ratcliffe *et al.*, 2003). The deletion of the *MAF1/FLM* gene in the accession Niederzenz (Nd) has been identified as the polymorphism underlying a QTL for flowering time (Werner *et al.*, 2005b). A *maf2* null mutant flowers early compared to its wild-type Columbia (Col) and was shown to respond much quicker to vernalization than wild-type, which led to the conclusion that an active *MAF2* can prevent vernalization by short periods of cold (Ratcliffe *et al.*, 2003). The same study demonstrated that the expression of *MAF3* and *MAF4* responded similarly to *MAF2* under vernalization but the *MAF5* expression levels increased. It was later confirmed that for *MAF1-MAF4* in a Col *FRI* genetic background, under vernalization transcript levels decrease slower than for *FLC* (Sheldon *et al.*, 2009). The authors found that in contrast to *FLC*, *MAF1-MAF5* expression was reduced in vernalized seeds actually increased after

vernalization of 13 day old plants. *MAF2-MAF5* are located in a tandem gene cluster on the bottom arm of chromosome 5. A recent sequence analysis of the cluster in several accessions has led to the discovery of chimeric fusions between *MAF2* and portions of *MAF3* (Caicedo *et al.*, 2009). These events occurred at moderate frequency in a sample of 169 accessions and the authors could demonstrate an association between them and accelerated flowering. QTL analysis of several Arabidopsis RIL populations has led to the detection of QTL for flowering time in that region (Ungerer *et al.*, 2002; El-Lithy *et al.*, 2004; El-Lithy *et al.*, 2006; Simon *et al.*, 2008). Although the *MAF2-MAF5* gene cluster is a likely candidate to be underlying these QTL in particular given the knowledge now available, no direct proof that variation in *MAF2-MAF5* contributes to natural variation for flowering time has been available yet.

In the present study we validated and fine-mapped the QTL detected on the bottom arm of chromosome 5 in the L*er* x Sha RIL population (El-Lithy *et al.*, 2004; this thesis, chapter 2). The achieved resolution unambiguously proves that natural allelic variation in the *MAF2-MAF5* cluster is causal to the QTL effects for flowering time, leaf numbers, and chlorophyll content index (CCI). The QTL effect on relative growth rate (RGR) detected in this region has also been validated. RGR could be fine-mapped to a slightly larger region that contains three additional genes. None of these genes are better candidates than the *MAF* for this trait. Sequence analysis of the *MAF* cluster revealed that Sha carried a chimeric fusion allele for *MAF2* that is identical to insert type s2 found in the accessions Kas-1 and Chi-1 (Caicedo *et al.*, 2009). However, contrary to those accessions, the intact *MAF3* gene was not present in the Sha accession. The high homology of the chimeric fusion gene suggests that a single re-organization event might have led to both s2 and Sha type alleles. Transformation of a NIL with the L*er* allele *MAF4* excluded this gene to be a likely candidate for the QTL effects due to only a very weak effect on flowering time. The effect of transformation with *MAF3* was stronger but it still did not fully restore the L*er* phenotype. In contrast, transformation with the L*er* allele of *MAF2* increased flowering time beyond that of the L*er* accession. We therefore speculate that both *MAF2* and *MAF3* are required for proper regulation of flowering time.

Materials and Methods

Genetic material and environmental conditions

The homozygous NIL 5b.17 as well as three additional heterozygous NILs, all with Sha introgressions in the bottom arm of chromosome 5 were kindly provided by the Laboratory

of Genetics, Wageningen University, The Netherlands. Seeds were stratified at 4°C on water saturated filter paper for 4 days prior to sowing. The seeds were sown in square pots measuring 7cm that contained a sand peat mixture enriched with slow releasing nutrients. QTL validation an fine-mapping experiments under 12 hour day length were performed in Percival AR95L/3 growth chambers. During the first and last 15 minutes of the light period only incandescent light bulbs were switched on to simulate dawn and dusk while for the rest of the light period fluorescent lamps were also switched on. The temperature was set at 22°C during the day and 18°C during the night at a relative air humidity of 70%. A single fine-mapping experiment under 8 hour day length, but with all other conditions the same, was performed in a custom Elbanton growth chamber (Elbanton BV, The Netherlands). The pots were randomly re-distributed over both growth chambers every 2-3 days to minimize positional effects in the QTL validation experiments. Constant environmental conditions during validation were guaranteed by continuous monitoring with data loggers (HOBO® U12-012). Selection of recombinant NILs was performed in a greenhouse under long day conditions. With the aid of supplementary light the day length was kept at than 16 hours or more. The greenhouse was climate controlled for 20°C during the day and 18°C during the night and for 60% relative air humidity. Additionally, selection of NILs was performed in growth rooms under 8 hour day length conditions and a temperature of 22°C during the day and 18°C during the night at fixed relative air humidity of 70%.

Plant phenotyping

During validation experiments the plants were photographed every 2-4 days with a ccd camera (Sony DSCF828), starting from 10-14 days after sowing (DAS). Photographing was continued until bolting time. The images measured 8 mega pixels and were stored in low-compression JPEG format. The projected rosette area (PRA) of each plant was quantified with a dedicated image analysis package (Image-Pro analyzer 6.0, Media Cybernetics). Linear regression of the natural logarithm of the PRA was performed on time for each plant. The regression coefficient provided an estimate of the relative growth rate (RGR) (d^{-1}). RGR was estimated from PRA measurements taken between 10 – 25 DAS. Flowering time was quantified in DAS at which the first flower opened. During experiments aimed at the selection of recombinant NILs, all plants were always phenotyped for flowering time. The number of rosette and cauline leaves were counted after flowering had commenced. Chlorophyll content index (CCI; Opti-Sciences CCM200) or, in this case equivalent, SPAD (Konica Minolta SPAD-502) was measured at flowering time. These measurements were

taken once on the three largest leaves of each plant and the average of them was used for analyses.

QTL validation and fine-mapping strategies

From each of the three heterozygous NILs that were provided, 100 progeny were grown under long day conditions. Homozygous progeny of two out of the three NILs were selected and subjected to a validation experiment under 12 hour day length together with the NIL 5b.17 was analyzed. After a further generation of selfing, twelve homozygous recombinant NILs as well as 23 recombinant NILs that still remained heterozygous were also selected using marker assisted selection. Twelve progeny plants of each line in this selection were grown in a follow-up experiment performed under 12 hour day length. From this experiment, three plants that were still heterozygous but had different, yet overlapping introgressions, were selected. Of each, 300 progeny were grown under 8 hour day length conditions for the selection of novel recombinants. Recombinants selected from previous experiments that still remained heterozygous were included also in order to select homozygous NILs. A selection of 28 recombinant NILs from the various experiments, either homozygous or heterozygous, was analyzed for flowering time under 12 hour day length. NILs 5b.3-12, 5b.3-6 and 5b.1-5 were selected from this experiment and analyzed under 8 hour day length for flowering time, leaf numbers, SPAD and PRA. Finally, a sibling of NIL 5b.3-12 that remained heterozygous was selected and 1000 progeny were grown for the selection of novel recombinants. This led to the selection of 25 recombinants, among which was the informative NIL 5b.136. Of each recombinant, 25 progeny plants were analyzed for flowering time, leaf numbers and SPAD under 12 hour day length.

Molecular techniques and sequence analysis

Unless otherwise specified all DNA extraction, cloning and manipulation was performed using standard molecular biology techniques (Sambrook and Russell, 2001). Large scale DNA extraction was performed using a BioSprint 96 Workstation and the BioSprint 96 Plant kit according to the manufacturer's instructions (Qiagen). Unless otherwise specified all DNA sequencing was done by the MPIZ DNA core facility on Applied Biosystems (Weiterstadt, Germany) Abi Prism 377, 3100 and 3730 sequencers using BigDye-terminator v3.1 chemistry. Premixed reagents were from Applied Biosystems. DNA sequence alignment and analysis was done using the MegAlign, SeqMan Pro and SeqBuider modules of the Lasergene package (DNASTAR).

Genotyping and molecular marker design

Commonly available markers used are described on the TAIR website (www.Arabidopsis.org) or on the INRA website (http://www.inra.fr/internet/Produits/vast/msat.php). Markers not previously described are listed in the table below. Molecular markers were designed based on published sequencing data (Nordborg *et al.*, 2005; Clark *et al.*, 2007). The datasets were queried using the web-based MSQT/SBE tool (Warthmann *et al.*, 2007; http://msqt.weigelworld.org/cgi-bin/nordborg/msqt-sbe.cgi; http://polymorph.weigelworld.org/ cgi-bin/msqt-sbe.cgi). Indels were used for the design of SSLP markers and SNPs were used for the design of CAPS or dCAPS markers. For the latter the web-based tool dCAPS Finder was used (Neff *et al.*, 2002).

BAC selection and sequencing of the MAF cluster

A BAC library of the Sha accession was available in house and the BIBAC library of the L*er* accession was used (Chang *et al.*, 2003). BACs containing the *MAF2-MAF5* cluster were isolated by Southern blot hybridization. An 113bp PCR product was amplified from the L*er* allele of the gene At5g65090 with the primers 5'-TACTCTGCTTCAAGAACAACTAC-3' and 5'-CTGCTCAAGACCTTAAGATTC-3' and used as probe to scan the BAC libraries. The identified positive colonies were incubated overnight on LB agar. A single colony was chosen of both libraries after verifying the presence of the entire *MAF* cluster on the BAC by colony PCR. In addition to using the primers of the probe for this, a second PCR was performed with primer pair 5'-GTAAATGTCACGATGATGGC-3' and 5'- TAGGAAAG CGTCATGGATC-3', which amplified a 218bp fragment of gene At5g65040. Both colonies were incubated o/n at 37°C in 2 ml LB medium, which served as starter cultures for larger 2l cultures in LB medium that were incubated for 24 hours at 28°C. A maxi-prep was used to isolate the BACs from the cultures (Qiagen). The L*er* and Sha alleles of the entire *MAF* cluster and the 5'and 3' intergenic regions were sequenced by Qiagen Genomic Services using the respective selected BACs as template. The Colombia (Col) sequence was used as reference in sequence analysis (TAIR8; www.Arabidopsis.org). In addition, the L*er* and Sha sequences of the *MAF* cluster were compared to published sequences of this region from other accessions (Caicedo *et al.*, 2009; PopSet Genbank records EU980614-EU980630).

Molecular markers that were designed and used for fine-mapping of the QTL

marker	phys.pos	primer 1	primer 2	type	enzyme
MUL3	23115167	ACGGATTGTTCAAGAAAGGAG	CCAAAGTTCAAACCGAATAATC	dCAPS	*Hpy*I
K19M22	23815848	TGGAGATCATGCATACAACTTG	TCCTGATATTCAAGAAAGGCTG	CAPS	*Ssp*I
MMN10	24131190	TCAAACTGTTCTCTTGCAGGAC	ATCCGAAACCTTTAGGTCTCTG	SSLP	
MJH22	25301048	AGCTCTGAGCATCCATTAACC	TACAGTCGAGGAACATACCTGAG	CAPS	*Dde*I
MBM17	25616974	AACTGGCGGGAATCGGAGTG	GACAGCTTGCCACCATAGTCGG	CAPS	*Hpy*99I
MSJ1	25715052	TCTTCTTATCTCCTTTTCCTATC	CGCCTTTCTTTGATGACTGGAG	dCAPS	*Bsp*HI
MSJ1-1	25761900	TTGCATCTTAGGCGTATCGAG	CATCCCTCTCTCTATGCTGTCG	SSLP	
MVP7-1	25908320	TATCAGCTAAACCTCGTCC	CAACAACGTGAGTCAAAGC	CAPS	*Dde*I
MVP7-2	25916395	GAAGAGGAAATGAGGTTAGG	TCAAACTTTCACGTTATCAAC	CAPS	*Hpy*188I
MVP7-4	25917986	CTATCTTAAGGGTCAATTCAC	CTCATGTTCCTTCAAGTCAC	CAPS	*Alu*I
MVP7-3	25919601	CTAATTATCATCCATCGTCG	GCTTTCATTATATTGGTGTGTC	CAPS	*Bsm*AI
MVP7-5	25920201	CTATTACAACGACCACCC	CTAGTGAATGAATTCTCGG	CAPS	*Bsm*AI
MXK3-4	25925765	CTTGGTGTGCACTGTAAAG	GCTTGGTTAAGAATCTTGC	CAPS	*Ban*I
MXK3	25934850	TTCCCGGTTTCCAATCG	AGCGCGTCTTGCTTTCTG	CAPS	*Mse*I
MXK3-5	25996884	TGTCTTGTCTCACTATTTCCTC	TAACAAGGCCCATATAACTG	CAPS	*Hpy*188I
Mrec1	25999361	CACTACTTTTCGATATTTGTATCC	TTTAACACTCGTCGTTGTGAA	dCAPS	*Mbo*II
M2d[a]	26002550	GCTAGGAAGGTATTTTGTACTGC	CTTGGTTACAGGATCTGCAGAGAAAACCTAAACAGTATCTGCAG		
M3M1	26006870	CACTCTATAAATTCCTCCAACTG	AAGAGGGAGATAAAAGGTGTC	CAPS	*Bpm*I
M34del	26008900	ATTCATCATTACACTTGCGT	TGACTTATCATTTGTCCGG	SSLP	
F15O5-1	26008971	TACCGATCTTTGTTGTACC	CGTGACTTATCATTTGTCC	CAPS	*Hinf*I
M5M1	26015030	GATCTCCGACCAGTTTATACAG	AGGTGCATTGAACTAGTGAGAG	CAPS	*Bst*BI
F15O5-2	26020974	AAACAGAGGCACGTGAAG	TGTTAATACCTTCGAAATCG	CAPS	*Bst*Z17I
MQN23	26029439	ACACTTAATACTGGGCATGACAC	ATTGACAACGACGTGTCCAC	CAPS	*Bsa*HI
MQN23-1	26042826	AACTCGGATTCGAAATTC	CACAACAACGTACGCATAG	CAPS	*Mfe*I
MQN23-2	26062266	ATCGGATGATTAGACAACG	CTTCTGTTCTGAATCGACG	CAPS	*Acl*I
MQN23-3	26078291	CTCTGCGGCGGAATTAG	CGTAGCGATTGAGAGACTAAG	CAPS	*Bcg*I
MQN23-4	26113783	CTAACCGGAAGATCCATAG	GATCCCATCACTCATTCC	CAPS	*Hpy*188I
MNA5-1	26150926	GATAATCTATGAGTTCCACTTG	AGAACAGGGAAAACAACAC	CAPS	*Cvi*AII
K21L13	26235134	GAATCTGTTGGTTGCTGAC	CCGATGAAGATGGAACAG	CAPS	*Mse*I
K2A18	26469955	AACGATCCATCACAGACAC	TGGCTGCTGCTATAAGTG	CAPS	*Bsr*I
K9I9	26988887	CCCAAACAACGAGCCAG	TCAACATCTGGTGCCTCC	CAPS	*Sau*3AI

All primers are presented in 5' – 3' orientation. The physical position on chromosome 5 is given in base-pair according to the TAIR8 reference sequence (www.Arabidopsis.org). [a] This marker was designed to identify the presence or absence of the Sha *MAF2/MAF3* fusion and depends on PCR amplification with the 3 primers specified.

Genomic complementation

The genomic regions containing *MAF2*, *MAF3*, and *MAF4* respectively were amplified by PCR with *Pfu* polymerase (Stratagene *PfuULTRA* AD) by using the isolated BAC as template. The primer pairs used were, respectively: 5'-GTGATATATCATAAGCATTA GGAAC-3'/5'-AAATCTTATATCGTCTACGAAGG-3', 5'-ATTCTTTCAGTCCGGTTTA TC-3'/5'-TGACTTATCATTTGTCCGG-3' and 5'-CCTTAGTAGACAAATATCAGAGT

TC-3'/5'-AATACTATATCATCCTGTCTCCG-3'. These amplified fragments of 7177 bp, 7308-bp and 4810-bp for the 3 genes respectively that contained the ORFs, UTRs and the entire 5' intergenic regions in case of *MAF3* and *MAF4* while for *MAF2* 3214-bp of 5' intergenic region was amplified. The blunt ended PCR products were A-tailed by incubating them for 15 minutes at 72°C in the presence of 1 nmol·µl^{-1} dATP and *Taq* polymerase (Roche). After precipitation and re-suspension in millipore water the fragments were cloned into pGEM T-Easy and introduced into *E.coli* DH5α by heat-shock according to the manufacturer's protocol (Promega). Positive clones were picked and verified by colony PCR. A single clone for *MAF2* and 3 clones for both *MAF3* and *MAF4* were sequenced in order to exclude potential PCR amplification errors (Qiagen Genomic Services). The insert DNA of the selected clones was excised with *Aat*II / *Sac*I (*MAF2*), *Swa*I / *Fsp*I (*MAF3*) or *Eco*RI (*MAF4*) and ligated into appropriately linearised pCAMBIA2300 (www.cambia.org). The reaction mixture was precipitated and re-suspended in millipore water, which was used for transformation of *E.coli* DH5α by electroporation. Positive clones were picked and verified by colony-PCR. The plasmids were isolated from the *E.coli* and used for transformation of *A.tum* GV3101 by electroporation (Weigel and Glazebrook, 2006b). Positive clones were identified by colony PCR. Transformation of Arabidopsis was performed by vacuum infiltration (Weigel and Glazebrook, 2006a). Seeds from transformed plants were surface sterilized by soaking in 70% ethanol for 1 minute, followed by soaking in 99% ethanol for 1 minute. In a sterile down-flow hood, the seeds were transferred to sterile filter paper by pipetting to dry on the air. When dry, the seeds were distributed on 1% agar plates with full strength MS medium (Murashige and Skoog, 1962) containing 50 mg·µl^{-1} kanamycin. The plates were sealed with microporous tape and incubated under 12 hour day-length at 20°C in a growth chamber. Kanamycin resistant plants were selected, transferred to soil and moved to long-day conditions in a greenhouse. To select plants homozygous for the transgene, 32 progeny plants were grown in the greenhouse under long day conditions. The seeds harvested from these plants were subjected to kanamycin selection as described above. Seed batches that did not segregate and were resistant to kanamycin were used for subsequent analysis.

Gene transcript analyses

RNA was isolated from 8 day old seedlings (Qiagen RNeasy) of L*er*, Sha and NIL 5b.3-12 grown under 12 hour day length at 22°C during the day and 18°C during the night with constant 70% relative air humidity. cDNA was synthesized using the isolated RNA as template with the SuperScript III kit according to the manufacturer's instructions

(Invitrogen). Qualitative RT-PCR was done in 20 µl total volume using 1 µl of the first-strand cDNA synthesis reaction mix as template. Primers designed and used to amplify the ORFs of *MAF2-MAF5* were: 5'-ATGGGTAGAAAAAAAGTCGAG-3'/5'-TTACTTGAGC AGCGGAAGAG-3', 5'-ATGGGAAGAAGAAAAGTCGAG-3'/5'-TTACTTGAGCAGCG AAAGAGT-3', 5'-ATGGGAAGAAGAAAAGTAGAGATC-3'/5'-TTACTTGAGAAGCA GGAGAGTCT-3', 5'-ATGTGTCGGAAGAGTGAAGC-3'/5'-TTACTTGAGAAGCGGGA GAG-3' respectively. A PCR program of 40 cycles was used with an annealing temperature of 59°C. PCR products were separated by electrophoresis on 1.5% standard agarose gel (Bio-budget). PCR products were cloned into pGEM T-Easy and introduced into *E.coli* DH5α by heat-shock according to the manufacturer's instructions (Promega). Positive clones were verified by colony PCR and a selection of them was subsequently sequenced using the standard T7 and SP6 primers, as provided by Promega.

Results

Validation and fine-mapping of the QTL

Three independent NILs with Sha introgressions in the region of the QTL at the bottom end of chromosome 5 (figure 1; 5b.17, 5b.1-30, 5b.2-74) were phenotyped in order to validate some of the QTL effects that were detected in this region (table 1). The environmental conditions were identical to those used during the QTL mapping experiment (this thesis). Although the QTL effects on early PRA detected in this region were positive, in all three NILs a significant negative effect was found. Strong negative effects were found for FT, RLN, and CLN in the NILs when compared to L*er*. These effects were highly significant and in the same order of magnitude as the detected QTL effects. Highly significant positive effects on RGR were also found in all three NILs. These were stronger than the previously detected QTL effects (this thesis). These results validated the presence of QTL for FT, leaf numbers and RGR in the region under scrutiny. Furthermore, the size of the introgression of Sha alleles in NIL 5b.2-74 was smaller than in the other two NILs. The similarity in the observed effects on the phenotype of the NILs allowed reducing the size of the QTL to a region of 23.1 Mb at the end of chromosome 5.

In order to fine-map the QTL further, several large populations of NILs segregating for the introgression on chromosome 5 were grown. FT, RLN, CLN and CCI were quantified on all plants during these experiments although they were primarily aimed at the selection of recombinant NILs, based on marker phenotypes. Individual experiments have been performed at long day lengths of 16 hours and short day lengths 8 hours. Co-segregation of
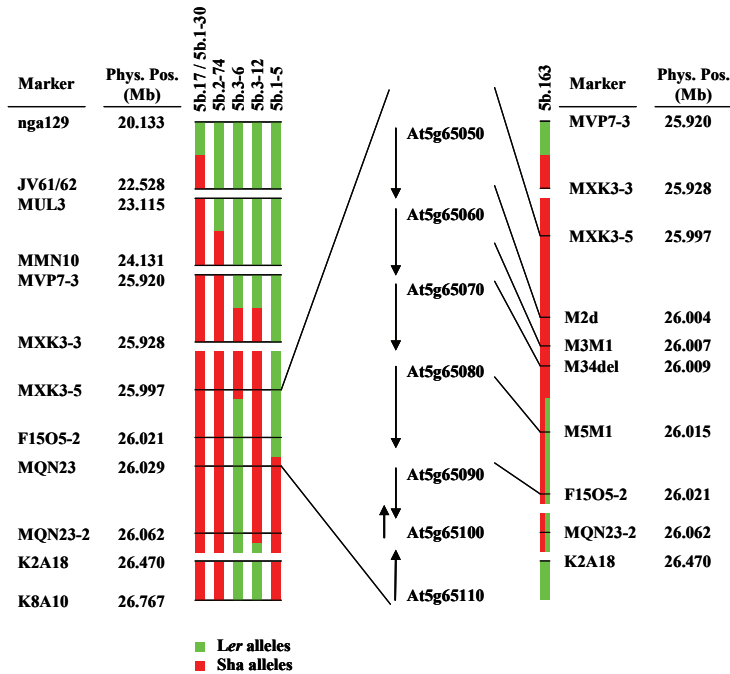
Figure 1. Graphical genotypes of the near isogenic lines (NILs) selected for validating (table 1) and fine-mapping (table 2) of QTL detected on the bottom of chromosome 5. Each NIL is represented by a vertical bar with L*er* alleles represented in green and Sha alleles in red. The genes located in the fine-mapped region are shown on next to the bars. Marker names and their physical position on the Col reference sequence (TAIR8) are indicated next to the bars. The physical position of the markers used to genotype NIL 5b.163 with respect to the *MAF* genes (At5g65050 – At5g65080) are indicated by black lines.

recombinant NILs, based on marker phenotypes. Individual experiments have been performed at long day lengths of 16 hours and short day lengths 8 hours. Co-segregation of the phenotypes with the allelic value of each marker was analysed in both experiments. Genotypic variation at the marker MQN23 (26.029 Mb) was consistently found to explain the observed variation of the phenotypes (table 2) the best. Under a 16h photoperiod a significant negative effect on FT and RLN was observed for the class of plants with homozygous Sha alleles at MQN23, but not for CLN. These effects were -0.4 days and -0.6 leaves respectively and thereby much weaker than the effects observed under 12 hour day length (table 1). Plants heterozygous for this marker did show a slightly weaker significant effect on FT but no significant effect on RLN or CLN was found. Under a day length of 8 hours strong negative effects were observed on FT, RLN, CLN and CCI for plants from both the homozygous Sha and heterozygous at MQN23. The observed effects on FT, RLN and CLN for plants with homozygous Sha alleles at MQN23 under these conditions were about

twice as large as those under a 12 h day length of (table 2). The effects on CCI validated also the presence of a QTL for this trait closely linked to MQN23. The phenotypical means of FT, RLN, CLN and CCI for plants heterozygous for this marker were approximately half the strength of the homozygous class. These results suggested that the validated QTL were all closely linked to the marker MQN23 and that their effects were semi-dominant and dependant to the photoperiod.

Table 1. Validation of QTL for FT, RLN, CLN, PRA at 14 DAS and RGR detected on the bottom of chromosome 5 using selected NILs. The expected effects from QTL analysis (this thesis) are given for the substitution of a homozygous L*er* allele with the homozygous Sha allele. The effects of the introgressed Sha alleles at this locus in the NILs (genotypes presented in figure 1) are with respect to the phenotypical mean of L*er*. SE indicates the standard error of the effects. The level of significance of the observed effects is indicated by asterisks: * $0.001 < p < 0.05$; ** $p \leq 0.001$.

| | QTL analysis | | QTL validation | | | | | | |
| | | | L*er* | 5b.17 | | 5b.1-30 | | 5b.2-74 | |
Trait	effect	SE	mean	effect	SE	effect	SE	effect	SE
PRA at 14 DAS	0.12	0.039	0.38	-0.14**	0.034	-0.12**	0.030	-0.13**	0.032
RLN	-5.0	0.76	15.0	-5.2**	0.49	-4.3**	0.45	-4.5**	0.49
CLN	-1.1	0.25	6.1	-1.5**	0.18	-1.0**	0.20	-0.7**	0.17
FT	-3.3	0.56	36.8	-3.7**	0.58	-3.3**	0.55	-3.0**	0.58
RGR	0.010	0.005	0.306	0.021*	0.006	0.016**	0.005	0.016**	0.005

Table 2. Co-segregation of QTL effects with marker MQN23 in NILs. Data from two segregating populations, grown at day lengths of 16 hours and 8 hours respectively is shown. The effects are estimated with respect to the mean of the homozygous L*er* allele genotypic class. SE and n indicate the standard error of the effects and the number of observations in each genotype x trait combination respectively. The significance of the estimated effects is indicated by asterisks: * $0.001 < \alpha < 0.05$; ** $\alpha \leq 0.001$.

| | | homozygous L*er* | | heterozygous | | | homozygous Sha | | |
day length	trait	mean	n	n	effect	SE	n	effect	SE
16 hours	FT	20.4	61	131	-0.3*	0.14	76	-0.4*	0.18
	RLN	5.1	44	84	-0.2	0.11	48	-0.6**	0.12
	CLN	1.8	44	84	-0.1	0.13	48	0.0	0.14
8 hours	FT	56.0	65	55	-4.7**	0.69	198	-10.4**	0.50
	RLN	25.1	72	55	-4.1**	0.65	203	-9.2**	0.39
	CLN	9.7	72	55	-1.9**	0.28	203	-3.5**	0.18
	CCI	12.3	72	55	-2.6**	0.31	198	-4.4**	0.18

For the selection of recombinant NILs, the focus was put on the region around MQN23. Three recombinant NILs (figure 1; 5b.3-6, 5b.3-12, 5b.1-5) carrying different but overlapping introgressions of Sha alleles on the bottom of chromosome 5 were selected. These NILs were analyzed under short day length conditions of 8 hours. The stronger effects on FT, RLN and CLN under these conditions were exploited to gain resolution. In addition, the effect on rosette growth was also increased under this condition. In table 3, the observed effects in the recombinant NILs are compared to QTL analysis (performed under 12 hour day length) and to L*er* and NIL 5b.17 that were included as controls. Strong negative effects on FT, RLN, CLN and SPAD were found for NIL 5b.17 when compared to L*er*. All these effects were highly significant and in the same order of magnitude as found from marker analysis during selection of recombinant NILs. Of the three selected recombinant NILs that were phenotyped in the same experiment, only 5b.3-12 showed significant effects. No significant effects were found for PRA at 42 DAS in any of the NILs, likely due to the large variability of this trait. However, a clear trend of increased PRA, correlated to decreased values for the other traits could be observed. Indeed, when the data was compared in two groups, based on presence (NILs 5b.17 and 5b.3-12) and absence (L*er*, and NILs 5b.3-6 and 5b.1-5) of significant QTL effects on the other traits, these two groups were significantly different for PRA at 42 DAS (6.77 cm^2, p=0.005). This effect could be explained by the validated QTL for RGR in this region. Hence, these results strongly suggested that the QTL for RGR was also located in the same fine-mapped region as the other traits mentioned above. Two follow-up experiments under the same conditions as used for QTL mapping were performed in which NIL 5b.3-12 was analyzed. In both cases effects for FT, RLN, CLN, CCI and RGR were observed in the NIL that were consistent with QTL mapping and with the results of the previous experiments (not shown).

The best candidate genes in the region to which the QTL had been fine-mapped were *MAF*2 – *MAF*5, four members of the *MAF* (MADS Affecting Flowering) family arranged in a tandem cluster. A final selection was performed on a population of 1000 segregating plants in order to obtain NILs with recombination events within the *MAF* cluster. A recombinant NIL (figure 1; NIL 5b.168) could be selected that was homozygous for Sha alleles from *MAF*2 until at least the intergenic region between *MAF*3 and *MAF4*, but was heterozygous from the 5' UTR of *MAF5* onwards towards the bottom end chromosome 5. When 25 progeny of this NIL were analyzed, no segregation for FT, RLN (figure 2), CLN or SPAD (not shown) was found and all plants showed a phenotype similar to NIL 5b.3-12, which was used as positive control. This result allowed excluding the *MAF5* ORF to be underlying the effect of the QTL and to conclude that the QTL for FT, RLN, and CLN was fine-mapped to a genomic region of 18.289-kbp containing *MAF*2 – *MAF*4.

Table 3. Phenotypes of three selected recombinant NILs (genotype presented in figure 1). The detected QTL effects are shown for reference but note that they were obtained under day length conditions of 12 hours while the NILs have been grown at 8 hours. Both L*er* and NIL 5b.17 were used as control. The phenotypical means were estimated for L*er*. The effects and their standard errors (SE) observed on the phenotypes in the NILs were estimated with respect to this mean. The significance of the effects is indicated by asterisk: * $0.001 < p < 0.05$; ** $p \leq 0.001$. [b] QTL for CCI were mapped in the RIL population but in this case SPAD was quantified in the NILs; both traits are equivalent.

Trait	QTL analysis[a]		L*er*	5b.17		5b.3-6		5b.3-12		5b.1-5	
	effect	SE	mean	effect	SE	effect	SE	effect	SE	effect	SE
RLN	-5.0	0.76	27.4	-9.4**	0.99	-0.1	1.20	-11.1**	0.92	-1.6	1.06
CLN	-1.1	0.25	11.2	-4.1**	0.49	0.2	0.52	-4.4**	0.51	0.2	0.47
FT	-3.3	0.56	60.4	-6.9**	1.39	2.0	1.20	-8.7**	0.73	0.9	1.21
CCI/SPAD[b]	-3.0	0.56	30.1	-6.5**	0.61	0.5	0.49	-6.6**	0.53	-1.1	0.75
PRA at 42 DAS	nd	nd	35.8	6.3	3.85	-4.4	3.28	4.9	3.22	-1.0	2.86
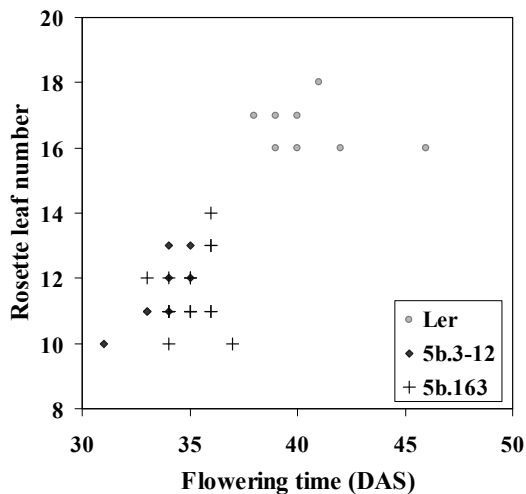
Figure 2. A scatterplot of RLN against FT quantified on 25 progeny of L*er*, NIL 5b.3-12 and NIL 5b.163 (see figure 1) grown under 12 hour day length.

Sequence analysis of the MAF2-5 genomic region in Ler and Sha

The entire genomic regions containing the four *MAF* genes including the 5' and 3' intergenic regions of *MAF2* and *MAF5* respectively were sequenced from BACs of both the L*er* and Sha accession. The average sequencing coverage per base pair was 3.4x and 4.0x for the L*er* and Sha BACs respectively. The L*er* sequence for this region was 25.840-Kb in length and

was found to be highly similar (99.4%) to the Col reference sequence. An alignment of both sequences revealed 96 SNPs and 24 small indels of which the largest were of 5-bp long. None of these polymorphisms occurred in protein coding regions. A single larger indel of 23-bp was located in a di-nucleotide repeat in the 5' UTR of *MAF2*. The sequence length of Sha for the *MAF2 – MAF5* region was considerably shorter than that of L*er* with 22.057-kbp.

Alignment of this sequence with L*er* revealed major structural differences (figure 3). From the 5' intergenic region of *MAF2* until base-pair 95 of intron 5 in this gene, the Sha sequence did show high similarity to L*er*. The remaining part of *MAF2* was found to be deleted (region I in figure 3) and instead 2240-bp of sequence highly homologous to *MAF3* followed (region III in figure 3). Preceding this part of *MAF3* were 10-bp of sequence that did not show homology to any *MAF*-related sequence from either L*er* or Col (region II in figure 3). The part of *MAF3* fused to *MAF2* spanned from the last 645-bp of intron 1 until and including 200-bp of the 3'UTR of this gene. However, following the 3'UTR of this *MAF3* sequence were 166-bp that showed 100% identity to the last 14-bp of the *MAF2* 3'UTR and 152-bp of the *MAF2 – MAF3* intergenic region. In this respect the *MAF3* sequence could be described as being inserted into *MAF2*, accompanied by the deletion of 873-bp of *MAF2* sequence. The only polymorphisms found between L*er* and Sha in original *MAF* protein coding sequence were 3 SNPs located in Exon 7 of the *MAF3* sequence inserted into *MAF2*. These three SNPs made this exon identical to exon 7 of *MAF2*, as found in L*er* (* a,b,c in figure 3). No intact *MAF3* gene was found in Sha (region IV in figure 3) and the sequence immediately following the chimeric *MAF2/MAF3* fusion was homologous to the intergenic sequence between *MAF3* and *MAF4*. The first 5' 199-bp of the *MAF3-MAF4* intergenic sequence, as found in L*er* was not present in Sha. Furthermore, in Sha an additional 25-bp deletion was found in the intergenic region between the chimeric *MAF2/MAF3* fusion and *MAF4*, 153-bp before the *MAF4* 5'UTR (region V in figure 3). The remaining sequence of Sha, which covered the open reading frames of *MAF4* and *MAF5* and their UTRs and intergenic sequence as well as the 3' intergenic region of *MAF5,* showed high homology to the complementary L*er* sequence.

Comparing the Sha sequence with published sequencing results (Caicedo *et al.*, 2009) revealed a 99.99% homology between the Sha allele of the chimeric *MAF2/MAF3* fusion and that of the s2 type allele found in the accessions Kas-1 and Chi-1. However, the s2 type allele of the *MAF2-MAF5* cluster does include an intact *MAF3* gene and therefore two alignments with the Sha allele were possible. Indeed, the entire Sha sequence starting from the *MAF3* portion of the *MAF2/MAF3* fusion gene until the 3' UTR of *MAF5* aligned without gaps to the s2 type allele, starting from the equivalent portion of the intact *MAF3*
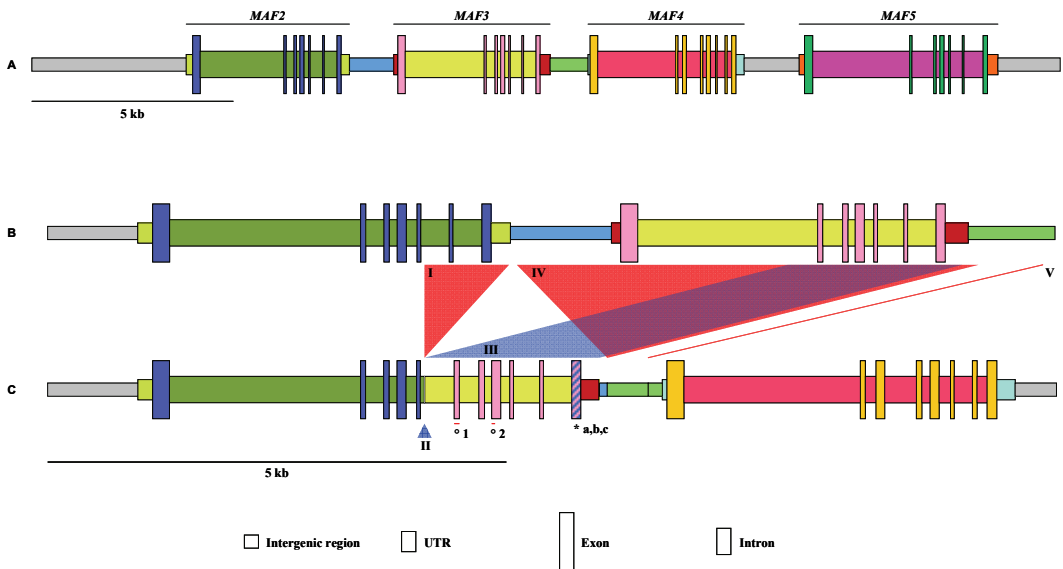
Figure 3. The results from BAC sequencing of the L*er* and Sha alleles of the *MAF* cluster on chromosome 5. The gene models represented are At5g65050.3 (*MAF2*), At5g65060.1 (*MAF3*), At5g65070.1 (*MAF4*) and At5g65080.1 (*MAF5*). A) Results of sequencing the L*er* allele; structurally invariant to the Col reference sequence (TAIR8). B) Magnification of the region homologous to *MAF2* and *MAF3* sequence in L*er*. C) Magnification of the region homologous to *MAF2*, *MAF3* and *MAF4* sequence in Sha. Deletions in the Sha sequence, when compared to L*er* are indicated in red and insertions are indicated in blue. I) 873 bp deletion of *MAF2* sequence. II) 10 bp insertion with no known homology. III) 2240 bp insertion of *MAF3* sequence in *MAF2*. IV) Deletion of the entire *MAF3* gene. V) 25 bp deletion in the 5' intergenic region of *MAF4*. Three SNPs were found in the last exon of *MAF3* in Sha, indicated in the figure by "* a,b,c", that caused this exon to be 100% homologous to the last exon of *MAF2* in L*er*. When transcripts of the Sha *MAF2/MAF3* chimeric fusion gene were sequenced, two differential splicing events of *MAF3* exons were observed. These events are indicated by: ° 1) exon skipping of *MAF3* exon 2, and ° 2) partial splicing of *MAF3* exon 4.

gene. This opened a single 4952-bp gap from the 3' flank of the 10-bp with no known homology until 645-before the 3' end of *MAF3* intron 1, both positions as described above. Alternatively, the Sha allele of the *MAF2/MAF3* fusion gene aligned perfectly with the s2 type *MAF2/MAF3* fusion gene. This opened a gap of the same size from the 3' flank of the *MAF2/MAF3* chimeric fusion until the intergenic sequence 5' of *MAF4*, both these positions also as described above. A single SNP was found in intron 5 of *MAF3* where in Sha a 'C' was called instead of a 'T' in Kas-1 and Chi-1. A second SNP constituted a G/T conversion in the 3'UTR of *MAF3* where for Chi-1 a 'T' was called and a 'G' for Sha and Kas-1. Both these SNPs occurred identically in either possible alignment of the Sha allele with the Kas-1 and Chi-1 alleles. Therefore the level of homology between the Sha, and the Kas-1 and Chi-1 alleles was identical for both alignments.

Sequence analysis of chimeric fusion transcripts

RNA was extracted from 8 day old seedlings of L*er*, Sha and the NIL 5b.3-12 in order to investigate the expression of the *MAF* genes. Qualitative RT-PCR was used to amplify cDNA of the four *MAF* genes as well as the chimeric gene fusion (figure 4). Both *MAF4* and *MAF5* transcripts were amplified in both L*er* and Sha, and the NILs. No *MAF2* or *MAF3* transcript was detected in either the NIL 5b.3-12 or Sha, regardless of the saturating PCR conditions. Instead, the primers designed to amplify the *MAF2* and *MAF3* cDNAs amplified larger products in the NIL and Sha as compared to L*er*. These products were indistinguishable in size from the amplicon produced by the combination of forward and reverse primers designed for *MAF2* and *MAF3* (figure 4).
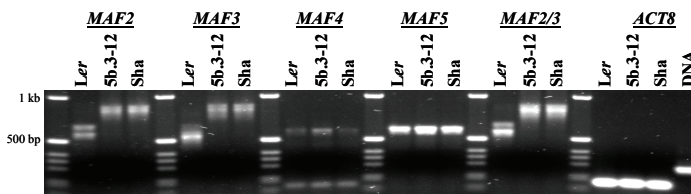
Figure 4. Qualitative RT-PCR of the *MAF2-MAF5* transcripts performed on L*er*, NIL 5b.3-12 and Sha. Primers were designed for each of the four genes separately and to amplify the fusion transcript, the forward and reverse primer of *MAF2* and *MAF3* were combined respectively. The *ACT8* gene was used for reference.

The amplified cDNA of the *MAF2/MAF3* fusion transcript was cloned and 19 individual clones were sequenced. Analysis of the sequence confirmed that the transcripts originated from a chimeric fusion of *MAF2* and *MAF*3. All introns, including the chimeric intron consisting partly of *MAF2* intron 5 and partly of *MAF3* intron 1 were properly spliced. However, exon skipping was observed for *MAF3* exon 2 (figure 3; ° 1). In 11 out of the 19 sequenced cDNAs this exon was spliced out. Furthermore, 38-bp of *MAF3* exon 4 was found to be spliced out, also in 11 out of 19 cDNAs (figure 3; ° 2). Both splicing events occurred independently because only 5 cDNAs were found where both had taken place. Conversely, none of these events had taken place in 3 cDNAs. Figure 5 shows the predicted proteins for these cDNA sequences. All contained the full amino-acid sequence predicted from the *MAF2* portion of the chimeric transcript. This included the MAF2 MADS-box domain, I domain, and a slightly truncated K-box domain. Translation of mRNA with exon 2 of *MAF3* not spliced out would lead to an early stop codon just after *MAF2* exon 5. On the predicted protein the slightly truncated K-box domain of MAF2 would be followed by 5 additional *MAF3* sequence derived amino acids. In case *MAF3* exon 2 and the partial *MAF3* exon 4

were spliced out, 37 *MAF3* derived amino acids would be translated before a stop codon would appear. This sequence showed homology to a splice variant of *MAF3* (Ratcliffe *et al.*, 2003). Both predicted proteins described above were also found by Caicedo *et al.* (2009). The *MAF* fusion transcript variant with only *MAF3* exon 2 spliced out, and its predicted protein, has not been reported before. This transcript can putatively be translated into a full length protein that includes entire K-box domain and C domain of MAF3.

```
        1         11        21        31        41        51        61        71        81        91
        |         |         |         |         |         |         |         |         |         |
A: MGRKKVEIKRIENKSSRQVTFSKRRNGLIEKARQLSILCESSIAVLVVSGSGKLYKSASGDNMSKIIDRYEIHHADELEALDLAEKTRNYL
B: MGRKKVEIKRIENKSSRQVTFSKRRNGLIEKARQLSILCESSIAVLVVSGSGKLYKSASGDNMSKIIDRYEIHHADELEALDLAEKTRNYL
C: MGRKKVEIKRIENKSSRQVTFSKRRNGLIEKARQLSILCESSIAVLVVSGSGKLYKSASGDNMSKIIDRYEIHHADELEALDLAEKTRNYL
D: MGRKKVEIKRIENKSSRQVTFSKRRNGLIEKARQLSILCESSIAVLVVSGSGKLYKSASGDNMSKIIDRYEIHHADELEALDLAEKTRNYL

          101       111       121       131       141       151       161       171       181
           |         |         |         |         |         |         |         |         |
A: PLKELLEIVQSKLEESNVDNASVDTLISLEEQLETALSVTRARKTELMMGEVKSLQKTHVKDH*
B: PLKELLEIVQSKLEESNVDNASVDTLISLEEQLETALSVTRARKTELMMGEVKSLQKTDLAEKIRNYLPHKELLEIVQRFSNIYGGTARDC
C: PLKELLEIVQSKLEESNVDNASVDTLISLEEQLETALSVTRARKTELMMGEVKSLQKTDLAEKIRNYLPHKELLEIVQSKLEESNVDNVSV
D: PLKELLEIVQSKLEESNVDNASVDTLISLEEQLETALSVTRARKTELMMGEVKSLQKTENLLREENQTLASQ...................

            191       201       211       221       231       241       251       261
             |         |         |         |         |         |         |         |
B: SVSN*
C: DSLISMEEQLETALSVIRAKKTELMMEDMKSLQEREKLLIEENQILASQVGKKTFLVIEGDRGMSWENGSGNKVRETLSLLK*
D: ................................................VGKKTFLVIEGDRGMSWENGSGNKVRETLPLLK*
```

Figure 5. Predicted proteins for the sequenced chimeric fusion transcripts with Col *MAF2* gene model At5g65050.3 (D) for reference. A) chimeric transcripts with *MAF3* exon 2 not spliced out; B) chimeric transcript with both *MAF3* exon 2 and the partial exon 4 spliced out; C) chimeric transcript with only exon 2 spliced out. From residue 149 onward the amino-acid sequence is derived from the *MAF3* part of the chimeric transcripts.

Genomic complementation with the Ler alleles of MAF2, MAF3, and MAF4

The genomic regions of the *MAF2, MAF3* and *MAF4* genes were individually cloned from the L*er* BAC. Verification sequencing of the cloned *MAF2* revealed repeat-number variations of 2 and 1 base-pair length in 3 separate poly-A/T tracks (nA/T > 10-bp) that were all located 5' of the ORF. Clones with 100% homology to the L*er* reference sequence were identified for *MAF3* and *MAF4*. Both NIL 5b.3-12 and L*er* were transformed with the three constructs. Flowering time, leaf numbers and CCI have been quantified on a preliminary selection of homozygous transgenic lines (table 4). Transformation of L*er* with its own *MAF2* allele had strong effects on flowering time and rosette leaf numbers. Of the 5 lines analyzed the weakest effects observed, were a 4.6 day delay in flowering with 3.4 rosette leaves more than untransformed L*er*. In this line the effects on cauline leaf number and CCI were not significant but in the others positive effects on all four traits were observed. A single line displayed particularly strong effects with flowering being 15.3 days delayed and

Table 4. FT, RLN, CLN and CCI of L*er* and NIL 5b.3-12 transformed with the L*er* allele of *MAF2*, *MAF3* or *MAF4*. The first three columns indicate the transgene, the genetic background and the individual line designators respectively. The remaining columns present the estimated means of the four traits and their standard errors (SEM). The significance of the observed effects, tested against the untransformed L*er* ($^{a/A}$) and NIL 5b.3-12 ($^{b/B}$), are indicated in superscript where $0.01 < {}^{a/b} < 0.05$ and $^{A/B} < 0.01$.

gene	back-ground	line	FT mean	FT SEM	RLN mean	RLN SEM	CLN mean	CLN SEM	CCI mean	CCI SEM
	L*er*		34.0^B	0.49	14.9^B	0.18	5.7^B	0.21	16.1^B	0.78
	5b.3-12		29.3^A	0.37	10.3^A	0.21	5.0^A	0.32	11.8^A	0.40
MAF2	L*er*	A12	39.6^{AB}	0.62	18.8^{AB}	0.47	7.5^{AB}	0.27	19.8^{AB}	1.13
	L*er*	A18	41.3^{AB}	0.53	21.3^{AB}	1.24	7.8^{AB}	0.70	21.1^{AB}	1.86
	L*er*	A37	49.3^{AB}	1.27	29.7^{AB}	0.50	9.8^{AB}	0.20	39.5^{AB}	0.64
	L*er*	A43	41.7^{AB}	0.56	22.3^{AB}	1.02	9.2^{AB}	0.31	24.3^{AB}	1.61
	L*er*	A48	38.6^{AB}	0.64	18.3^{AB}	0.67	6.3^B	0.40	18.9^B	1.45
	5b.3-12	A11	39.4^{AB}	0.58	18.1^{AB}	0.59	6.4^B	0.44	19.7^{AB}	0.99
	5b.3-12	A12	40.4^{AB}	0.51	23.0^{AB}	1.41	8.2^{AB}	0.49	21.9^{AB}	0.90
	5b.3-12	A13	40.1^{AB}	0.40	22.3^{AB}	1.04	8.0^{AB}	0.44	25.1^{AB}	1.46
	5b.3-12	A3	33.4^B	0.65	12.2^{AB}	0.25	5.2	0.25	13.8^{aB}	0.57
	5b.3-12	A6	39.8^{AB}	0.68	18.3^{AB}	0.37	6.7^{aB}	0.37	19.8^{AB}	0.99
MAF3	L*er*	A1	40.1^{AB}	0.46	18.7^{AB}	0.94	7.5^{AB}	0.48	20.8^{AB}	1.34
	L*er*	A2	34.6^B	0.43	14.5^B	0.56	5.5	0.34	15.3^B	0.79
	L*er*	A25	34.6^B	0.43	13.8^B	0.63	5.1	0.28	14.9^B	0.45
	L*er*	A9	37.6^{AB}	0.40	16.0^B	0.58	5.9^b	0.38	18.1^B	1.04
	5b.3-12	A3	32.2^{aB}	0.55	10.1^A	0.41	5.0	0.30	12.9^A	0.64
	5b.3-12	A7	31.6^{AB}	0.45	10.9^A	0.23	4.8^A	0.20	13.9^{aB}	0.67
	5b.3-12	A8	30.6^{Ab}	0.43	10.8^A	0.36	5.2^a	0.13	13.4^{Ab}	0.51
MAF4	L*er*	A1	36.3^B	1.25	15.7^B	0.50	6.0^B	0.37	16.6^B	0.94
	L*er*	A7	35.9^{AB}	0.48	15.1^B	0.48	5.4	0.34	16.0^B	0.89
	5b.3-12	A1	30.1^A	0.50	8.6^{AB}	0.31	3.7^{AB}	0.33	10.9^A	0.36
	5b.3-12	A4	30.4^A	0.47	8.5^{AB}	0.27	4.1^{AB}	0.23	10.8^A	0.67

the rosette leaf number increased by 14.8 leaves compared to untransformed L*er*. Flowering time, rosette leaf numbers and CCI were also increased in 4 out of 5 NILs transformed with *MAF2* and the values of these traits were similar to the ones observed in L*er* transformed with *MAF2*. In a single line the effects on flowering time and cauline leaf number were not different from untransformed L*er* while RLN and CCI did have lower values. Four lines homozygous for the *MAF3* transgene in the L*er* background were analyzed of which only two differed from the untransformed L*er*. One of them only showed an increased flowering time of 3.6 days. In the second line flowering time was increased by 6.1 days and also the values for the other traits were increased compared to untransformed L*er*. All three NILs

transformed with *MAF3* also had significantly increased flowering time, when compared to the untransformed NIL. However, this effect ranged from 1.3 – 2.9 days whereby it did not reach the value of untransformed L*er*. Furthermore, only a significant effect on CCI was found in two out of the three lines while none of them showed differences in the other traits compared to L*er*. Unfortunately, for both L*er* and the NIL only 2 homozygous transgenic lines could be selected for *MAF4* so far. Both lines with the genetic background of L*er* had increased flowering times but for one the effect was not significant. No significant effect with respect to untransformed L*er* was found for any of the other traits in these lines. Both NILs transformed with *MAF4* did not show effects on flowering time and CCI when compared to the untransformed NIL. The effects on rosette, and cauline leaf number were significant. However, both lines had reduced numbers of leaves compared to the untransformed NIL, which was opposite to the general trend observed so far.

Discussion

QTL for flowering time, RGR, chlorophyll fluorescence and several other traits were detected on the bottom of chromosome 5 in the Arabidopsis RIL population derived from the cross between Landsberg *erecta* and Shahdara (El-Lithy *et al.*, 2004; this thesis). Furthermore, QTL for flowering time have been identified in this region in several other RIL populations (Ungerer *et al.*, 2002; El-Lithy *et al.*, 2006; Simon *et al.*, 2008; this thesis). The QTL effects on flowering time were detected in both long day and short day conditions (El-Lithy *et al.*, 2004). These effects could be validated under both conditions. The observation of significant effects on PRA at 14 DAS that were found in a single experiment with NILs could not be repeated and validated. This could be due to suppression of this phenotype in the L*er* genetic background. Accurate measurements of RGR on the most discriminative NIL were not yet available so the validated QTL effects on this trait could not be definitely attributed to the *MAF* genes. However, none of the remaining three genes to which the QTL has been fine-mapped are better candidates. The developmental switch to flowering has shown to be linked to variation in vegetative growth (Cookson *et al.*, 2007; Tisné *et al.*, 2008). In addition, its effects on secondary growth has been demonstrated (Sibout *et al.*, 2008). It is therefore likely that the validated effects on RGR and the other traits are pleiotropic effects of the same QTL. The trait CCI has been shown to correlate well to actual chlorophyll content (Richardson *et al.*, 2002). QTL effects on chlorophyll fluorescence that were detected in the L*er* x Sha RIL population (El-Lithy *et al.*, 2004) can thus also be attributed to the *MAF* cluster.

The QTL was fine-mapped precisely to the cluster of four *MAF* genes located in this region. These genes belong to a clade of 6 MIKC-type MADS-box genes. The two remaining members, known for their contribution in natural variation to flowering time, are *FLC* (Michaels and Amasino, 1999) and *MAF1/FLM* (Werner *et al.*, 2005a). It has been shown that the remaining four *MAF* genes can all act as negative regulators of flowering time (Ratcliffe *et al.*, 2003), as do *FLC* and *MAF1/FLM*. A specific function for *MAF2* in preventing flowering in response to short periods of cold was defined by mutant and over expression analysis in this study. Furthermore, a statistical association between flowering time and extensive molecular variation found among different accessions at this locus has been demonstrated (Caicedo *et al.*, 2009). However, direct proof of their contribution to natural variation was not yet reported. The result of fine-mapping reported in this study proves that the gene cluster from *MAF2 –MAF4* harbours the gene(s) that underlie natural variation for flowering time, leaf numbers and CCI. The NIL with which the QTL was ultimately fine-mapped did not exclude the 5' intergenic sequence of *MAF5* from the QTL. The single SNP that was found in this region could therefore be responsible for the observed QTL effects through modification of *MAF5* expression. Similarly, the deletions in the 5' intergenic sequence of *MAF4* could affect the expression pattern of this gene. However, the *MAF2/MAF3* fusion allele and the deletion of *MAF3* are both the best candidate polymorphisms to be underlying the observed QTL effects.

The weak, or even absent, effects of genomic complementation with *MAF3* and *MAF4* in the NIL excludes these genes to be solely responsible for the observed QTL effects. At this point it can not be excluded that absence of effects in the two L*er* lines transformed with *MAF3* is because these are actually not transformed. Although ambiguous, the results of complementation with *MAF2* do show that this gene can have strong effects on flowering time. The particularly strong effects in a single L*er* line can be due to multiple copies and high expression of the transgenes. The weaker effects in a transgenic NIL, on the other hand, might be explained by a reduced expression level of the transgene. This could for instance be due to the position of insertion into the genome. The generally strong effects are therefore a good indication that loss of *MAF2* function in Sha is the major causal polymorphism underlying the detected QTL effects. This would be in accordance with the effects of a *maf2* null mutation in the Col genetic background. This mutant caused quantitatively similar effects as reported herein in the NILs with a Sha introgression (Ratcliffe *et al.*, 2003). Yet, the deletion of *MAF3* must also be considered because complementation with this gene did result in significantly increased flowering time. The QTL effects are therefore more likely to be caused by the combined loss of function of both genes. MADS-box proteins act in multimeric complexes that interact with the promoter

sequence of the target gene and they can also function in auto-regulatory feedback loops (de Folter and Angenent, 2006). It is therefore conceivable that both genes interact epistatically in their proper regulation of flowering time. In this case the effects of both transgenes expressed in the NIL could be synergistic and would lead to even more severe late flowering than caused by transformation with *MAF2* alone. This could be in agreement with the observation that the relative effect of transformation with *MAF2* was stronger in the NIL than in L*er*. The latter accession obviously does express a functional *MAF3*, which could thus be involved in tempering the effects of *MAF2*. Mis-regulation by a feedback loop or by other MADS complexes normally involving *MAF3* could also lead to an elevated expression of *MAF2* causing the overly late flowering in the complementation lines. The transgenic lines for the three genes have been crossed to acquire the genetic material for further investigation. It must be noted, however, that above all more transgenic lines need to be analyzed still before drawing final conclusions.

A recent analysis of the *MAF* cluster has shown that insertions of *MAF3* segments into the 3' portion of *MAF2* occur at moderate frequency among accessions (Caicedo *et al.*, 2009). Six different insert types, s1 to s6, were found by the authors after sequencing the entire (s2, s4) or partial (s1, s3, s5) *MAF* cluster from several accessions. It seemed likely that an intact *MAF3* would be present downstream of *MAF2* in all insert types, like found in s2 and s4 (Caicedo *et al.*, 2009). However, Sha carries a novel allele, hereby suggested to be designated s7, with no intact *MAF3* present that shows that this does not necessarily need to be the case. The very high homology between the Sha allele of the chimeric fusion and the s2 type allele as found in Kas-1 and Chi-1 does suggest that they share a common origin. The Sha allele could have arisen by deletion of the intact *MAF3* from an s2 type ancestral allele, however, this seems unlikely. It would require incredible precision, or coincidence, to occur without leaving a footprint such as an additional gap in the Sha sequence when aligned to the s2 alleles. An alternative explanation can be hypothesized when assuming that the s2 type and Sha alleles arose from through same event from a common ancestral allele. This common ancestral allele might still have an intact *MAF2* gene but the *MAF3* allele would already be similar to the allele as presently found in the s2 type alleles. Under this assumption the event that led to the re-organization, for instance mis-alignment during homologous recombination, could have resulted in a heterozygote consisting of the s2 and Sha type alleles. These would then have segregated to found the ancestors of the present day populations carrying either allele. Future studies on the effects of these reorganizations should also consider the *MAF3* deletion because the results from complementation show it could be important and the Sha accession might not be an isolated case.

A novel fusion transcript was isolated from the Sha allele. In contrast to all other *MAF2/MAF3* fusion transcripts described so far, translation of this transcript would lead to a full-length protein. Furthermore, it would be the only example where the amino-acid sequence is fully homologous to domains from the MADS-box transcription factor consensus. Depending on transcript stability and on successful translation, all chimeric transcripts (alternatively spliced or not) can lead to functional proteins. The M- and I-domains, are necessary and can be sufficient for homo-dimerisation and target binding. The K- and C-domains are invoked mainly in hetero-dimerisation and higher order interactions, as well as in additional functions of the established complex (see Kaufmann *et al.*, 2004 for detailed description). All chimeric transcripts that have been described contain the M and I domains from *MAF2* and can thus, potentially, bind their targets. Although their action might be impaired by the truncated K-domain and the lacking C-domain, they could compete for binding sites with functional complexes. This could explain why the L*er* allele was not dominant over the Sha allele but it could also be a dosage dependant effect. Translation of the novel transcript would result in a protein that does contain the intact K-domain and C-domain of *MAF3*. This protein could thus also be functional and it could engage in higher order interactions. No convincing phenotypical evidence that can be attributed to such functional complexes involving chimeric proteins can be pointed out in the data. The only indication that leaves the possibility, though, is the result obtained from transformation with *MAF4*. Whereas transformation with *MAF2* and *MAF3* led to comparable effects in both L*er* and the NIL, this was not the case for *MAF4*. The negative effects on only RLN and CLN in the NIL but not on FT might indicate an effect on plastochron, the time between initiation of leaf primordia by the apical meristem (Wang *et al.*, 2008). Such effects were not observed in L*er* and could thus be linked to increased expression levels of *MAF4* in a genetic background that expresses the fusion transcripts. However, alternative hypotheses such as differences in expression level between complementation lines or interaction with other genes in the Sha introgression are equally valid at the moment.

Chapter 5

Summarizing discussion

Bjorn Pieper

Plant growth is a complex trait that results from the integration of many relevant physiological processes that act during the time of growth but in which also developmental aspects play a role. It is an important trait in agriculture as it can directly determine yield as biomass. Improving plant production to keep up with the ever increasing demand for plant derived products is a major challenge. This has been, and is the motivation to gain knowledge about the genetic and molecular basis of complex traits such as plant growth. A large number of genes involved in the control of plant growth have been identified, mainly by studying induced mutants and transgenic lines over-expressing specific genes in the model plant Arabidopsis (reviewed in Gonzalez *et al.*, 2008; and Krizek, 2009). Induced mutagenesis has as a drawback that gene function in the context of the genetic background and the environment remains elusive, and that genes with natural loss of function alleles in the parental genotype can not be identified. Naturally occurring genetic variation reflects the selection for beneficial mutations in the context of the genetic background and the environment. It thus provides a valuable additional resource to study the genetic and molecular basis of plant growth. Arabidopsis populations occur naturally in very different environments and climates (Hoffmann, 2005), to which they are expected to be genetically adapted. In addition, its small size, its rapid life cycle, its small and fully sequenced genome, and the availability of genetics and genomics tools make Arabidopsis an attractive species for studying the genetic and molecular basis of natural variation for plant growth.

The aim of the work presented in this thesis was the elucidation of the genetic and molecular basis of natural variation for plant growth in Arabidopsis. A QTL mapping experiment was performed in which the two Arabidopsis RIL populations L*er* x Kas-2 (El-Lithy *et al.*, 2006) and L*er* x Sha (El-Lithy *et al.*, 2004) were analyzed simultaneously (chapter 2). To take into account the complexity of plant growth, the plants were phenotyped for a panel of 28 growth-related traits. To estimate the growth rate of each plant a logistic model was fitted to projected rosette area (PRA) data quantified at up to nine time-points during the entire expansion phase of the rosette. However, it was found that the estimate of growth rate (parameter B) by this model was strongly negatively correlated to flowering time. It was in fact much stronger than the negative correlation between flowering time and relative growth rate (RGR), estimated from PRA during the exponential expansion phase only. RGR provided a better estimate of growth rate because it was less influenced by variation of flowering time. The fitted logistic model furthermore did provide an accurate estimate of the final rosette area (parameter C) and the estimate of its inflection point (parameter M) indicated when the rosette had reached its maximum expansion rate. The modelling thereby

allowed the joint analysis of all PRA measurements dynamically as a limited number of comprehensible plant growth traits versus multiple static PRA measurements.

All traits that were quantified showed a high heritability and they formed a complex correlation structure. This indicated that they could partly share a common genetic basis. Fifteen traits were chosen for QTL mapping such that they did not provide redundant information due to too high correlations but would still represent the original correlation structure. A multi-trait mixed-model method was used for QTL mapping that directly modelled the genetic correlations, which provided a more powerful and more realistic analysis than a collection of single trait analyses (Malosetti *et al.*, 2008). A total of 18 QTL were detected in both mapping populations together. Two of the, QTL1 and QTL11, were not significant during the multi-trait genome scan but tests for the individual traits were, indicating that they could be putative false-positive QTL. Because these QTL were found to have significant and relatively strong effects exclusively on vegetative growth–related traits in the final QTL model, they were nevertheless included in the analysis. However future validation using selected NILs is required to prove that they are truly affecting growth-related traits in a significantly. Almost all of the identified QTL had effects on multiple traits, which were often both affecting development and vegetative growth. This further supports that vegetative plant growth and the developmental switch from vegetative to reproductive growth are in part under common genetic control. This was also concluded from a recent detailed QTL study (Tisné *et al.*, 2008), and had been indicated by co-location of QTL (El-Lithy *et al.*, 2004). From the present QTL analysis it could therefore be concluded that there exists a strong negative genetic correlation between plant growth rate and flowering time in Arabidopsis.

Seven of the detected QTL mapped to common positions in both RIL populations and were considered to be the same loci. It should be taken into account though that at the resolution of QTL mapping no distinction can be made between closely linked QTL and pleiotropic effects of a single QTL. Most of these QTL allowed candidate genes to be suggested based on their location and their effects on the various traits. One of the common QTL mapped to the *ERECTA* locus and is likely this gene because the mutant allele of L*er*, which has well described effects on plant morphology (van Zanten *et al.*, 2009), segregated in both populations. Additionally, strong positive effects on early PRA were found for the Kas-2 allele at this locus and also weakly positive and negative effects on flowering time and plant growth rate respectively. However, the strong negative effect of the L*er* allele on final rosette area found in the L*er* x Sha population was not found in the L*er* x Kas-2 population. This indicated that in this population there is possibly another QTL, closely linked to

ERECTA, which might be interesting for further research for its effects on vegetative growth. The common QTL5, 10, 15, 17 and 18 had effects on flowering time and accounted for most of the explained variance for this trait, but they also had effects on vegetative growth traits. They mapped to chromosomes 1, 3 and 5, at loci where QTL for this trait had been detected before in these, and other Arabidopsis RIL populations (Ungerer *et al.*, 2002; El-Lithy *et al.*, 2004; El-Lithy *et al.*, 2006; Simon *et al.*, 2008). QTL15 and QTL17 accounted for most of the variance explained for rosette leaf number and flowering time in the L*er* x Kas-2 population. Their strong additive effects explained the much larger range for these traits in this population. QTL10, QTL15 and QTL18 explained most variance for flowering time in the L*er* x Sha population but their additive effects were much weaker than in the L*er* x Kas-2 population, in accordance with the smaller range for this trait in this population.

QTL5 mapped in the region of the candidate gene *MAF1/FLM* (Werner *et al.*, 2005) and the Sha allele increased flowering time. The L*er* allele increased flowering time at QTL10. This QTL was also detected for flowering time under long day conditions and the L*er* allele again increased the trait value (El-Lithy *et al.*, 2006). No candidate gene had been proposed yet but considering the results of initial fine-mapping with NILs (chapter 3) it could be *CONSTANS-LIKE 2,* located at the very top of chromosome 3, which is similar to *CONSTANS*. Given the direction of the effects on flowering time it would mean that L*er* carries an non functional allele (at least partially) for this gene, when it plays a similar role as *CONSTANS* (Turck *et al.*, 2008). QTL15 mapped to the *FRI* locus, in which gene L*er* carries a deletion (Johanson *et al.*, 2000) and is therefore likely this gene itself. QTL17 mapped to the top of chromosome 5 but not in the region of *FLC*, which is in-line with published results that showed that all three accessions have weak alleles for this gene (Michaels and Amasino, 1999; Michaels *et al.*, 2003; Hagenblad *et al.*, 2004). The L*er* strain that was used to develop the RIL populations carries a mutant allele of *HUA2* and its effects on flowering time have been described (Doyle *et al.*, 2005). This gene has been suggested as candidate for the QTL detected in this region (El-Lithy *et al.*, 2006). However, QTL validation with two NILs that were respectively mutant and wild-type for this gene because of different Sha introgressions showed *HUA2* could explain their differing flowering time but not rosette leaf number. In conjunction with the analysis of two L*er* strains with a mutant and wild-type allele at *HUA2* respectively, it was found that these polymorphisms do cause effects like those detected QTL (chapter 2), making it a strong candidate. However, the NIL with the mutant *hua2* allele still differed from L*er* with *hua2* allele, which suggested that another QTL is located below this gene in the L*er* x Sha population. Furthermore, in the L*er* x Kas-2 population QTL17 had strong effects on FT, RLN, CLN, LLA, LLW, parameter C, RGR and CCI that could not be explained by the results of the two isogenic L*er* strains. In

addition, this QTL mapped to a locus slightly above the QTL in the L*er* x Sha population which suggested that this could after all be a population specific QTL. Finally, QTL18 co-located in both populations and the fine-mapping of this QTL by using NILs with Sha introgressions is discussed below.

The detection of QTL for plant growth related traits other than those that also had major effects on flowering time was of major interest for the study presented in this thesis. Although growth and flowering time were controlled by the same QTL in many cases, these accounted for only about 50% of the total explained genetic variance for RGR and even less for PRA at 27 DAS. In addition, most genetic variance for the dimensions of the largest leaf was also explained by other QTL than the major flowering time QTL in the L*er* x Sha population. These QTL plant growth were largely specific to the respective RIL populations, which showed that the three accessions have a largely distinct genetic basis for these traits. Some of these QTL did have mild effects on flowering time, which were weak in relation to the effects on the size of the vegetative rosette. From this it could be concluded that these QTL are more important for rosette growth than for flowering time. Some QTL had relatively strong effects on leaf numbers without having an effect on flowering time at all, or only had a very weak effect. These therefore control plastochron, which describes the number of leaves produced, hence biomass, per unit time. In the L*er* x Kas-2 population QTL1, QTL3, QTL11, QTL12 and QTL14 were of particular interest for having relatively strong effects on RGR and several other plant growth traits without affecting flowering time. In the L*er* x Sha population QTL2 and QTL6 were of particular interest for the same reasons. These QTL from both populations were chosen for validation with NILs, with the prospect of fine-mapping them for the elucidation of their molecular basis.

QTL-by-QTL interactions have been shown to play a large role in the only plant specific growth QTL for which the causal gene has been identified in Arabidopsis (Kroymann and Mitchell-Olds, 2005). A test revealed 20 significant two-way interactions between the detected main effect QTL in each RIL population (chapter 2). In both populations, the genetic variance explained for RGR and, in accordance, PRA increased substantially when including the interactions. Furthermore, in the L*er* x Kas-2 population also the explained genetic variance for the size of the largest leaf increased substantially by including epistatic interactions in the QTL model. This indicated that indeed epistatic interactions also played a large role in the variation for plant growth traits in the two RIL populations analyzed in the scope of this thesis. However, there is some controversy about how the test for interactions was implemented in the multi-trait model. By testing for multi-trait interactions the

assumption was made that an interaction that affects one trait should also affect other traits, which in principle need not necessarily be true. However, for certain combinations of traits this could indeed be expected, for instance, that an interaction with an effect on RGR would also affect PRA. Most of the significant interactions had effects on such groups of traits for which the correlation made sense from a biological point of view which gave the model credibility. Furthermore, with the NILs that are, and will become available these statistical interactions can be empirically tested. When the epistatic interactions that were predicted can be demonstrated in NILs, the detected statistical interactions gain trust and can provide useful information. This can then be used to design NILs with a genetic background optimized for contrasting the effects of a particular QTL. In addition, validation of epistatic interactions in NILs will provide insight in how to the interpret tests for interactions with multi-trait mixed models. Alternatively, the data presented in this thesis allows a more careful implementation of such test and its use in further research on this topic will be pursued in the future. Already, first indications that epistatic interactions indeed play a role where they have been detected have already been found during validation experiments with NILs (chapter 3).

By using the NILs that have been selected or provided, the presence of QTL has now been validated for nine loci (chapters 3 and 4). The observed phenotypic effects in the NIL did, however, not always match the detected QTL effects for each trait, for which several reasons have been suggested. Firstly, even though the growth chambers that were used allow tight control of the environment, there were factors such as soil quality, deterioration of light sources and seed quality that were likely to be responsible for small variations between experiments. QTL mapping has only been performed once and it is thus unknown whether, in particular, QTL effects of small magnitude would be consistently detected if the experiment would be repeated. Secondly, it has been demonstrated that NILs provide increased power to detect main effect QTL over RILs (Keurentjes *et al.*, 2007; Reif *et al.*, 2009), and thus additional small QTL effects could have been identified in the NILs. Finally, epistatic interactions between the detected main effect QTL, but also potentially involving undetected QTL could have played a role. QTL-by-QTL (or QTL x background interactions) could be responsible for small modifications of the QTL effects. More drastic deviations from the detected QTL effects in the NILs selected for QTL2 and QTL6 in the L*er* x Sha population could be directly tied to the results of testing for two-way interactions (chapters 2 and 3).

The observed effects in the NIL covering QTL2 were strong but of opposite sign compared to the detected QTL effects. However, previously published results of QTL

analysis in the L*er* x Sha RIL population had detected this QTL also with very similar effects as in the present study (El-Lithy *et al.*, 2004), which supports the validity of the QTL mapping results. Four two-way interactions were detected of which QTL2 was a component. These accounted for the larger part of the additional genetic variance explained for RGR, indicating their potential importance. A similar situation was found for QTL12 in the L*er* x Kas-2 population but only one single interaction, with QTL3, that could explain this phenomenon was detected statistically. Interactions with undetected QTL in the genetic background might also be involved in reversing the phenotypic effects of QTL in the NILs though.

With the NIL covering QTL6 from the L*er* x Sha population, only a weak effect on RGR was validated initially although strong effects on largest leaf size were expected. However, QTL6 could be more firmly validated by it modifying the effects of QTL 4 and QTL5 in a NIL with two introgressions in both these respective regions. Furthermore, a QTL-by-environment interaction was discovered circumstantially using these two NILs when elsewhere in the growth chamber different conditions were set-up. These results did provide an important clue about how the effect of QTL6 in the NILs could be accentuated. In interaction with the introgression in the region of QTL4 and 5, a strong positive effect on leaf size could indeed be observed, as expected from QTL analysis. The selection of recombinant NILs is currently in progress for QTL6 in a genetic background that contains a fixed Sha introgression in the region of QTL4 and 5. QTL6 is intended to be fine-mapped making use of the interaction with QTL4 and 5 as well as the increased effect size in a contrasting environment.

Obvious differences in plant morphology were observed in a NIL with Kas-2 alleles covering QTL14. QTL effects for RGR could also be validated in this NIL as well as increased area and width of the largest leaf. By directly associating the effects on plant morphology with Kas-2 alleles in a segregating population, the QTL underlying these effects could efficiently be fine-mapped to a region containing 151 genes at the very end of chromosome 3. A selection of 12 NILs with recombinant events in this region is currently being analyzed to confirm that also the effect on RGR is caused by this QTL. Furthermore, after genotyping these recombinants with additional markers it is likely that the region can be narrowed down sufficiently to apply a candidate gene approach. An interesting candidate gene in this region is the E3 ubiquitin ligase *BIG BROTHER*. This gene has been described as a negative regulator of organ size, causing phenotypes including wider leaves, thicker stems and larger flowers (Disch *et al.*, 2006). These agreed with the phenotypes that were observed in the NILs. The NILs also displayed reduced apical dominance, which was not

mentioned by Disch *et al*. (2006). However, this could be a phenotype that is related to the Ler genetic background or, alternatively, it could be due to another, closely linked QTL.

An additional QTL with relatively strong effects on RGR, which was not detected by QTL analysis, was identified on the top of chromosome 3 in a NIL with a Sha introgression. The other QTL validated this region, which effects were detected by QTL analysis, was shown to be epistatic to the novel QTL. The close linkage between both QTL is the reason why only the epistatic QTL effects were detected by QTL analysis.

A general observation from the validation experiments was that flowering time effects were always observed in the NILs, although these had not always been detected by QTL analysis. However, these effects were rather weak and, additionally, they did not show the trend to be negatively correlated to the effects on RGR that were also found in these NILs, as was generally found in QTL analysis for QTL that affected both traits. This could be an indication that the effects on flowering time are a secondary effect of a growth specific QTL. Failure to detect these effects by QTL analysis would support this because their effect and contribution to the total genetic variance for this flowering time trait was too small to be detected.

QTL18 described in this thesis (chapter2), detected in the Ler x Kas-2 and the Ler x Sha RIL populations, was fine-mapped using a NIL with a Sha introgression in the genetic background of Ler. This QTL had similar effects in both mapping populations. In the Ler x Sha population, the Ler allele for QTL18 had positive effects on FT, leaf numbers, parameter C, parameter M, and negative effects on PRA, RGR, and parameter B. The effects on PRA could not be validated. The effects on FT, leaf numbers, RGR and CCI were validated and the QTL was fine mapped to the tandem cluster of the four genes *MAF2-MAF5*. The QTL for RGR could not yet be fine-mapped to this resolution but the increased PRA at 42 DAS in short days gave a strong indication that the QTL for this trait is located in a window of 7 genes including the 4 *MAF* genes. BAC sequencing of the Ler and Sha alleles of the *MAF2-MAF5* cluster revealed a gene fusion in place of *MAF2* in the latter accession consisting of the 5' portion of *MAF2* and the 3' portion of *MAF3*. In contrast to similar fusion alleles that were found in 15 other accessions (Caicedo *et al.*, 2009), no intact *MAF3* gene was found in Sha. When comparing the Sha sequence to these published sequences, it was found that its *MAF2/MAF3* fusion allele was identical to those of the accessions Kas-1 and Chi-1. In those two accessions and in Sha, the part of *MAF3* fused to *MAF2* is identical to the complementary part of the retained intact *MAF3* gene in Kas-1 and Chi-1, which indicates a duplication event in the latter two, as suggested by Caicedo *et al.* (2009). However, to delete *MAF3* from the Kas-1 or Chi-1 allele to result in the Sha allele would

require the deletion break-points to occur at homologous base-pairs in both the duplicated and retained parts of *MAF3*, which seems unlikely. An alternative would be to assume an ancestral situation with an intact *MAF2* and *MAF3* gene, the latter being identical to those still retained in Kas-1 and Chi-1. In this case a single deletion could explain the discovered situation of the fused Sha *MAF2/MAF3* allele and the absence of an intact *MAF3*. Given the moderate frequency at which these rearrangements were found among Arabidopsis accessions (Caicedo *et al.*, 2009), it is obvious that these genes are prone to such events. The identical alleles for the fused *MAF2/MAF3* in Kas-1, Chi-1 and Sha suggest a common origin. It could be that both alleles are derived from the same aberrant recombination event through misalignment of homologous pairs during meiosis, which resulted in a heterozygous allele of the Sha, and Kas-1/Chi-1 type rearrangements respectively. These alleles could then have segregated to found the ancestors of the populations that nowadays carry the various alleles in homozygous state.

Genomic complementation with the L*er* alleles of *MAF2-4* showed that the small indel in the 5' intergenic region of *MAF4* is not the functional polymorphism of the QTL. However, transformation with either *MAF2* or *MAF3* showed that also these genes can not explain the QTL effects by themselves. Transformation with *MAF2* led to an increase in flowering time beyond the value of L*er* in the NIL and transformation with *MAF3* did not yield a strong enough increase in flowering time to equal that of L*er*. This has led to the suggestion that both genes might be necessary for proper control of flowering time. Since the *MAF* genes encode proteins of the MADS family of transcription factors there are several mechanism through which this could be realized. These included the formation of homo- or heterodimers, or higher order complexes involving either or both proteins regulating their own expression and/or that of their target genes (de Folter and Angenent, 2006).

The strong effects on flowering time observed in the NIL with the genetic background of L*er* show that in this accession the *MAF2* and, likely also *MAF3*, are important regulators of flowering time. However, the flowering time of the Sha accession is very similar to L*er*, under laboratory conditions (El-Lithy *et al.*, 2004; results not shown). This indicates that in Sha other QTL compensate the loss of flowering time control by *MAF2/MAF3*. A possible candidate for this is *MAF1/FLM* (Werner *et al.*, 2005) on chromosome 1, where a QTL for which the Sha allele increased flowering time could be validated (chapter 3). Furthermore, Sha carries a weak *FLC* allele but a functional allele at *FRI* (Michaels *et al.*, 2003; Shindo *et al.*, 2005), the latter could be sufficient to compensate the loss of *MAF2/MAF3*. These results can be important for molecular ecological studies of flowering time and fitness traits in Arabidopsis, that have mainly focussed on the role of *FRI* and *FLC* (Caicedo *et al.*, 2004; Stinchcombe *et al.*, 2004; Korves *et al.*, 2007). However,

MAF2 and *MAF3* can be important regulators of flowering time, particularly in absence of an active *FRI/FLC* pathway.

Additional studies are underway to evaluate the behaviour of the NILs and transgenic lines under vernalization treatment, which will shed further light on the functional importance of these genes. Furthermore, the transgenic lines have been intercrossed to determine whether the concerted action of both *MAF2* and *MAF3* can complement the phenotypes observed in the NIL by restoring them to the L*er* value. The most discriminative NIL and the transgenic lines will furthermore be analyzed for plant growth, which was not yet possible due to seed batches of insufficient quality. This might allow the positive identification of the control of vegetative plant growth by a plant development pathway in the near future. What is also of interest is to investigate whether the transcripts of the fused *MAF2/MAF3* gene are involved in any of the observed phenotypes. They may act as a dominant negative mutation due to the formation of non-functional dimers.

Overall, this study demonstrates that the genetic basis of the complex trait plant growth can be efficiently elucidated by analyzing a set of growth-related traits at the whole plant, organ, and developmental level. This allowed distinguishing between effects on growth which are correlated to plant development and those that have a distinct genetic basis. In the scope of this task, the multi-trait QTL mapping methodology efficiently yielded comprehensible results. It revealed that the genetic basis of flowering time was largely common between both studied RIL populations and that the same genetic basis controls plant growth rate in a predictable way. A major QTL for flowering time was studied in more detail. Fine-mapping of this QTL and complementation by plant transformation identified *MAF2* and *MAF3* as relevant candidate genes involved in the regulation of flowering time in nature. The fine-mapped QTL will also allow showing whether indeed a flowering time pathway does also control plant growth rate pleiotropically. The plant growth specific genetic architecture was more complex and a distinct genetic basis for this trait in each RIL population was revealed. A major focus of the study was to elucidate the molecular basis of such QTL. This goal has come in sight with the advanced fine-mapping of a plant growth-specific QTL and the development of a set of NILs with which additional QTL have been validated, and that are thus now liable to fine-mapping. With these NILs it could be shown that epistasis likely plays a large role in the control of plant growth. Future work will include testing the NILs in different environments to characterize the QTL but, moreover, to find a contrasting environment for efficient fine-mapping. Once the genes underlying the effects of the QTL are identified it can be investigated whether useful natural variation for them also exists for crop species or related wild species that might allow introgression breeding. Characterisation

of these genes will also allow refining the gene networks that have been built from the analysis of mutants in Arabidopsis.

References

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

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

Anastasiou E., Kenz S., Gerstung M., MacLean D., Timmer J., Fleck C., Lenhard M. (2007). Control of plant organ size by KLUH/CYP78A5-dependent intercellular signaling. *Curr. Biol.* 13: 843-856.

Banerjee S., Yandell B. S., Yi N. (2008). Bayesian quantitative trait loci mapping for multiple traits. *Genetics* 179: 2275-2289.

Beavis W. D., Ed. (1998). QTL analyses: power, precision, and accuracy. In: Molecular Dissection of Complex Traits. New York, CRC Press.

Beemster G. T. S., De Veylder L., Vercruysse S., West G., Rombaut D., Van Hummelen P., Galichet A., Gruissem W., Inze D., Vuylsteke M. (2005). Genome-wide analysis of gene expression profiles associated with cell cycle transitions in growing organs of Arabidopsis. *Plant Physiol.* 138: 734-743.

Beemster G. T. S., de Vusser K., De Tavernier E., De Bock K., Inze D. (2002). Variation in growth rate between Arabidopsis ecotypes is correlated with cell division and A-type cyclin-dependent kinase activity. *Plant Physiol.* 129: 854-864.

Beemster G. T., Mironov V., Inzé D. (2005). Tuning the cell-cycle engine for improved plant performance. *Curr. Opin. Plant Biol.* 16: 142-146.

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

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

Blazquez M. A., Weigel D. (2000). Integration of floral inductive signals in Arabidopsis. *Nature* 404: 889-892.

Borlaug N. (2007). Feeding a hungry world. *Science* 318: 359.

Boss P., Bastow R., Mylne J., Dean C. (2004). Multiple pathways in the decision to flower: enabling, promoting, and resetting. *Plant Cell* 16: S18-S31.

Buse A. (1982). The likelihood ratio, Wald, and Lagrange multiplier tests: an expository note. *Am. Stat.* 36: 153-157.

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

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

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

Chenu K., Chapman S. C., Hammer G. L., McLean G., Salah H. B. H., Tardieu F. (2008). Short-term responses of leaf growth rate to water deficit scale up to whole-plant and crop levels: an integrated modelling approach in maize. *Plant Cell Environ.* 31: 378-391.

Clark R. M., Schweikert G., Toomajian C., Ossowski S., Zeller G., Shinn P., Warthmann N., Hu T. T., Fu G., Hinds D. A., Chen H., Frazer K. A., Huson D. H., Scholkopf B., Nordborg M., Ratsch G., Ecker J. R., Weigel D. (2007). Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* 317: 338-342.

Clerkx E. J. M., El-Lithy M. E., Vierling E., Ruys G. J., Blankestijn-De Vries H., Groot S. P. C., Vreugdenhil D., Koornneef M. (2004). Analysis of natural allelic variation of Arabidopsis seed germination and seed longevity traits between the accessions Landsberg *erecta* and Shakdara, using a new recombinant inbred line population. *Plant Physiol.* 135: 432-443.

Condon A. G., Richards R. A., Rebetzke G. J., Farquhar G. D. (2004). Breeding for high water-use efficiency. *J. Exp. Bot.* 55: 2447-2460.

Cookson S., Chenu K., and Granier C. (2007). Day length affects the dynamics of leaf expansion and cellular development in *Arabidopsis thaliana* partially through floral transition timing. *Ann. Bot.-London* 99: 703-711.

Cooper M., Eeuwijk F. A. v., Hammer G. L., Podlich D. W., Messina a. C. (2009). Modeling QTL for complex traits: detection and context for plant breeding. *Curr. Opin. Plant Biol.* 12: 1-10.

Corbesier L., Vincent C., Jang S., Fornara F., Fan Q., Searle I., Giakountis A., Farrona S., Gissot L., Turnbull C., Coupland G. (2007). FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. *Science* 316: 1030-1033.

de Folter S., Angenent G. C. (2006). *trans* meets *cis* in MADS science. *Trends Plant Sci.* 11: 224-231.

Disch S., Anastasiou E., Sharma V. K., Laux T., Fletcher J. C., Lenhard M. (2006). The E3 ubiquitin ligase *BIG BROTHER* controls Arabidopsis organ size in a dosage-dependent manner. *Curr. Biol.* 16: 272-279.

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

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

Elberse I. A. M., Vanhala T. K., Turin J. H. B., Stam P., Damme J. M. M. v., van Tienderen P. H. (2004). Quantitative trait loci affecting growth-related traits in wild barley (*Hordeum spontaneum*) grown under different levels of nutrient supply. *Heredity* 93: 22-33.

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

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

Flavell R. (2009). Role of model plant species. Plant Genomics**:** 1-18.

Gonzalez N., Beemster G. T., Inzé D. (2008). David and Goliath: what can the tiny weed *Arabidopsis* teach us to improve biomass production in crops? *Curr. Opin. Plant Biol.* 12: 1-8.

Granier C., Aguirrezabal L., Chenu K., Cookson S. J., Dauzat M., Hamard P., Thioux J.-J., Rolland G., Bouchier-Combaud S., Lebaudy A., Muller B., Simonneau T., Tardieu F. (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: 623-635.

Granier C., 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: 1175-1184.

Gray W. M. (2004). Hormonal regulation of plant growth and development. *PLoS Biol.* 2: e311.

Grime J., Hunt R. (1975). Relative growth-rate: its range and adaptive significance in a local flora. *J. Ecol.* 63: 393-422.

Hagenblad J., Tang C., Molitor J., Werner J., Zhao K., Zheng H., Marjoram P., Weigel D., Nordborg M. (2004). Haplotype structure and phenotypic associations in the chromosomal regions surrounding two *Arabidopsis thaliana* flowering time loci. *Genetics* 168: 1627-1638.

Hoffmann M. H. (2005). Evolution of the realized climatic niche in the genus: Arabidopsis (*Brassicaceae*). *Evolution* 59: 1425-1436.

Holland J. B. (2007). Genetic architecture of complex traits in plants. *Curr. Opin. Plant Biol.* 10: 156-161.

Hospital F. (2009). Challenges for effective marker-assisted selection in plants. *Genetica* 136: 303-310.

Hunt R., Cornelissen J. H. C. (1997). Components of relative growth rate and their interrelations in 59 temperate plant species. *New Phytol.* 135.

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

Jansen R. C., Tesson B. M., Fu J., Yang Y., McIntyre L. M. (2009). Defining gene and QTL networks. *Curr. Opin. Plant Biol.* 12: 241-246.

Jansen R., Stam P. (1994). High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* 136: 1447 - 1455.

Jiang C., Zeng Z. B. (1995). Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics* 140: 1111-1127.

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

Juenger T., Pérez-Pérez J. M., Bernal S., Micol J. L. (2005a). Quantitative trait loci mapping of floral and leaf morphology traits in *Arabidopsis thaliana*: evidence for modular genetic architecture. *Evol. Dev.* 7: 259-271.

Juenger T., McKay J., Hausmann N., Keurentjes J., Sen S., Stowe K., Dawson T., Simms E., Richards J. (2005b). Identification and characterization of QTL underlying whole-plant physiology in Arabidopsis thaliana: delta13C, stomatal conductance and transpiration efficiency. *Plant Cell Environ.* 28: 697 - 708.

Keurentjes J., Bentsink L., Alonso-Blanco C., Hanhart C., Blankestijn-De Vries H., Effgen S., Vreugdenhil D., Koornneef M. (2007). Development of a near-isogenic line population of *Arabidopsis thaliana* and comparison of mapping power with a recombinant inbred line population. *Genetics* 175: 891 - 905.

Keurentjes J., Fu J., Terpstra I., Garcia J., Ackerveken G., Snoek L., Peeters A., Vreugdenhil D., Koornneef M., Jansen R. (2007). Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. *Proc. Natl. Acad. Sci. USA* 104: 1708 - 1713.

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

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

Korves T. M., Schmid K. J., Caicedo A. L., Mays C., Stinchcombe J. R., Purugganan M. D., Schmitt J. (2007). Fitness effects associated with the major flowering time gene *FRIGIDA* in *Arabidopsis thaliana* in the field. *Am. Nat.* 169: E141-E157.

Krizek B. A. (2009). Making bigger plants: key regulators of final organ size. *Curr. Opin. Plant Biol.* 12: 17-22.

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

Kusterer B., Muminovic J., Utz H. F., Piepho H.-P., Barth S., Heckenberger M., Meyer R. C., Altmann T., Melchinger A. E. (2007a). Analysis of a triple testcross design with recombinant inbred lines reveals a significant role of epistasis in heterosis for biomass-related traits in Arabidopsis. *Genetics* 175: 2009-2017.

Kusterer B., Piepho H.-P., Utz H. F., Schon C. C., Muminovic J., Meyer R. C., Altmann T., Melchinger A. E. (2007b). Heterosis for biomass-related traits in Arabidopsis investigated by quantitative trait loci analysis of the triple testcross design with recombinant inbred lines. *Genetics* 177: 1839-1850.

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

Langlade N. B., Feng X., Dransfield T., Copsey L., Hanna A. I., Thébaud C., Bangham A., Hudson A., Coen E. (2005). Evolution through genetically controlled allometry space. *Proc. Natl. Acad. Sci. USA* 102: 10221-10226.

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 Physiol. Bioch.* 37: 671-678.

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

Li J., Ji L. (2005). Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* 95: 221-227.

Li Y., Jones L., McQueen-Mason S. (2003). Expansins and cell growth. *Curr. Opin. Plant Biol.* 6: 603-610.

Lisec J., Meyer R., Steinfath M., Redestig H., Becher M., Witucka-Wall H., Fiehn O., Torjek O., Selbig J., Altmann T., Willmitzer L. (2008). Identification of metabolic and biomass QTL in *Arabidopsis thaliana* in a parallel analysis of RIL and IL populations. *Plant J.* 53: 960 - 972.

Mähler M., Most C., Schmidtke S., Sundberg J. P., Li R., Hedrich H. J., Churchill G. A. (2002). Genetics of colitis susceptibility in IL-10-deficient mice: backcross versus F2 results contrasted by principal component analysis. *Genomics* 80: 274-282.

Malmberg R. L., Held S., Waits A., Mauricio R. (2005). Epistasis for fitness-related quantitative traits in *Arabidopsis thaliana* grown in the field and in the greenhouse. *Genetics* 171: 2013-2027.

Maloof J. N. (2002). QTL for plant growth and morphology. *Curr. Opin. Plant Biol.* 6: 85-90.

Malosetti M., Ribaut J., Vargas M., Crossa J., van Eeuwijk F. (2008). A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (*Zea mays* L.). *Euphytica* 161: 241-257.

Melchinger A. E., Piepho H.-P., Utz H. F., Muminovic J., Wegenast T., Torjek O., Altmann T., Kusterer B. (2007). Genetic basis of heterosis for growth-related traits in Arabidopsis investigated by testcross progenies of near-isogenic lines reveals a significant role of epistasis. *Genetics* 177: 1827-1837.

Meyer R., Steinfath M., Lisec J., Becher M., Witucka-Wall H., Torjek O., Fiehn O., Eckardt A., Willmitzer L., Selbig J., Altmann T. (2007). The metabolic signature related to high plant growth rate in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* 104: 4759-4764.

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

Michaels S. D., Bezerra I. C., Amasino R. M. (2004). *FRIGIDA*-related genes are required for the winter-annual habit in Arabidopsis. *Proc. Natl. Acad. Sci. USA* 101: 3281-3285.

Michaels S. D., He Y., Scortecci K. C., Amasino R. M. (2003). Attenuation of *FLOWERING LOCUS C* activity as a mechanism for the evolution of summer-annual flowering behaviour in Arabidopsis. *Proc. Natl. Acad. Sci. USA* 100: 10102-10107.

Mitchell-Olds T., Schmitt J. (2006). Genetic mechanisms and evolutionary significance of natural variation in Arabidopsis. *Nature* 441: 947-52.

Mitchell-Olds T., Willis J. H., Goldstein D. B. (2007). Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nat. Rev. Genet.* 8: 845-856.

Noland M. H., Yuval S., Michael A. L. (2008). The genetic architecture of reproductive isolation in Louisiana irises: pollination syndromes and pollinator preferences. *Evolution* 62: 740-752.

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

Payne R. W., Harding S. A., Murray D. A., Soutar D. B., Baird S. J., J. W. S., Kane A. F., Gilmour A. R., Thompson R., Webster R., Tunnicliffe Wilson G. (2007). GenStat Release 10 Reference Manual, VSN International, Hemel Hempstead.

Perez-Perez J. M., Serrano-Cartagena J., Micol J. L. (2002). Genetic analysis of natural variations in the architecture of *Arabidopsis thaliana* vegetative leaves. *Genetics* 162: 893-915.

Poorter H., 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., Rijn C., Vanhala T., Verhoeven K., Jong Y., Stam P., Lambers H. (2005). A genetic analysis of relative growth rate and underlying components in *Hordeum spontaneum*. *Oecologia* 142: 360-377.

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

Protas M., Conrad M., Gross J. B., Tabin C., Borowsky R. (2007). Regressive evolution in the Mexican Cave Tetra, *Astyanax mexicanus*. *Curr. Biol.* 17: 452-454.

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

Ratcliffe O. J., Kumimoto R. W., Wong B. J., Riechmann J. L. (2003). Analysis of the Arabidopsis *MADS AFFECTING FLOWERING* gene family: *MAF2* prevents vernalization by short periods of cold. *Plant Cell* 15: 1159-1169.

Reif J. C., Kusterer B., Piepho H.-P., Meyer R. C., Altmann T., Schon C. C., Melchinger A. E. (2009). Unraveling epistasis with triple testcross progenies of near-isogenic lines. *Genetics* 181: 247-257.

Röder M., Huang X.-Q., Börner A. (2008). Fine mapping of the region on wheat chromosome 7D controlling grain weight. *Funct. Integr. Genomic.* 8: 79-86.

Sakamoto T., Matsuoka M. (2008). Identifying and exploiting grain yield in rice. *Curr. Opin. Plant Biol.* 11: 209-214.

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

Sang T. (2009). Genes and mutations underlying domestication transitions in grasses. *Plant Physiol.* 149: 63-70.

Semenov M. A., Halford N. G. (2009). Identifying target traits and molecular mechanisms for wheat breeding under a changing climate. *J. Exp. Bot.*: erp164.

Shalit A., Rozman A., Goldshmidt A., Alvarez J. P., Bowman J. L., Eshed Y., Lifschitz E. (2009). The flowering hormone Florigen functions as a general systemic regulator of growth and termination. *Proc. Natl. Acad. Sci. USA* 106: 8392-8397.

Shao H.-B., Chu L.-Y., Jaleel C. A., Manivannan P., Panneerselvam R., Shao M.-A. (2009). Understanding water deficit stress-induced changes in the basic metabolism of higher plants - biotechnologically and sustainably improving agriculture and the ecoenvironment in arid regions of the globe. *Crit. Rev. Biotechnol.* 29: 131.

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

Shindo C., Bernasconi G., Hardtke C. S. (2007). Natural genetic variation in Arabidopsis: tools, traits and prospects for evolutionary ecology. *Ann. Bot.-London* 99: 1043-1054.

Sibout R., Plantegenet S., Hardtke C. S. (2008). Flowering as a condition for xylem expansion in Arabidopsis hypocotyl and root. *Curr. Biol.* 18: 458-463.

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

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

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

Stam P. (1993). Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. *Plant J.* 3: 739-744.

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

Sung S., Schmitz R. J., Amasino R. M. (2006). A PHD finger protein involved in both the vernalization and photoperiod pathways in Arabidopsis. *Genes. Dev.* 20: 3244-3248.

Tardieu F. (2002). Virtual plants: modelling a tool for the genomics of tolerance to water deficit. *Trends Plant Sci.* 8: 9-14.

Tisné S., Reymond M., Vile D., Fabre J., Dauzat M., Koornneef M., 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: 1117-1127.

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

Tsukaya H. (2002). Interpretation of mutants in leaf morphology: genetic evidence for a compensatory system in leaf morphogenesis that provides a new link between cell and organismal theories. *Int. Rev. Cytol.* 217: 1-39.

Tsukaya H. (2008). Controlling size in multicellular organs: focus on the leaf. *PLoS Biol.* 6: e174.

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

Ungerer M. C., Halldorsdottir S. S., Modliszewski J. L., Mackay T. F. C., Purugganan M. D. (2002). Quantitative trait loci for inflorescence development in *Arabidopsis thaliana*. *Genetics* 160: 1133-1151.

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

Vandepoele K., Raes J., De Veylder L., Rouze P., Rombauts S., Inze D. (2002). Genome-wide analysis of core cell cycle genes in Arabidopsis. *Plant Cell* 14: 903-916.

Wang J.-W., Schwab R., Czech B., Mica E., Weigel D. (2008). Dual effects of miR156-targeted *SPL* genes and CYP78A5/KLUH on plastochron kength and organ size in *Arabidopsis thaliana*. *Plant Cell* 20: 1231-1243.

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

Warthmann N., Fitz J., Weigel D. (2007). MSQT for choosing SNP assays from multiple DNA alignments. *Bioinformatics* 23: 2784-2787.

Weigel D., Glazebrook J. (2006a). *In planta* transformation of Arabidopsis. *Cold Spring Harbor Protocols* 2006.

Weigel D., Glazebrook J. (2006b). Transformation of *Agrobacterium* using electroporation. *Cold Spring Harbor Protocols* 2006.

Werner J. D., Borevitz J. O., Warthmann N., Trainer G. T., Ecker J. R., Chory J., Weigel D. (2005a). Quantitative trait locus mapping and DNA array hybridization identify an

FLM deletion as a cause for natural flowering-time variation. *Proc. Natl. Acad. Sci. USA* 102: 2460-2465.

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

Wilczek A. M., Roe J. L., Knapp M. C., Cooper M. D., Lopez-Gallego C., Martin L. J., Muir C. D., Sim S., Walker A., Anderson J., Egan J. F., Moyers B. T., Petipas R., Giakountis A., Charbit E., Coupland G., Welch S. M., Schmitt J. (2009). Effects of genetic perturbation on seasonal life history plasticity. *Science* 323: 930-934.

Wolters H., Jürgens G. (2009). Survival of the flexible: hormonal growth control and adaptation in plant development. *Nat. Rev. Genet.* 10: 305-317.

Yi N., Shriner D. (2007). Advances in Bayesian multiple quantitative trait loci mapping in experimental crosses. *Heredity* 100: 240-252.

Summary

Plant growth is a complex quantitative trait that results from the integration of many physiological processes during growth, including of developmental nature. It is under genetic control but also interacts with the environment. It is an important agronomic trait for being directly involved in crop yield in the form of biomass but also indirectly by providing photo-assimilates, for instance for fruit growth. Plant production needs to keep up with the ever increasing size of the human population and a changing environment. This has been the motivation for many (eco)physiological studies of plant growth that led to a wealth of knowledge concerning this aspect. However, little is known about the genetic and underlying molecular basis of this trait. The study of induced mutants and targeted over-expression lines, mainly in the model species *Arabidopsis thaliana*, did lead to the discovery of many genes involved in plant growth. However, how these genes act in concert remains unclear. Natural variation in Arabidopsis has been an underexploited resource in studies aimed at discovering such genes, yet, it can be a valuable asset. The studies described in this thesis were aimed at elucidating the genetic basis of plant growth in Arabidopsis, and to use this information in paving the way to the discovery of the underlying genes and functional sequence polymorphisms.

A QTL mapping experiment was performed in which the two Arabidopsis RIL populations L*er* x Kas-2 and L*er* x Sha were analyzed simultaneously for a total of 28 growth-related traits. Rosette growth was quantified by digital image analysis of the projected rosette area (PRA) at several time-points from the seedling until a maximum was reached. The change in PRA over the entire expansion phase was modelled with a logistic model and that of the initial exponential phase only with a linear model. The logistic model provided an estimate of the growth rate as well as the maximum rosette size and the inflection point of rosette expansion. The linear model provided an estimate of the relative growth rate (RGR). Furthermore, the area, length and width of the largest leaf, flowering time, leaf number counts, chlorophyll content index and plant height were quantified on each plant. Because all traits were components of plant growth and represented the entire growth phase of the plant, they formed an intricate correlation structure. By using multi-trait methodology based on mixed models, these correlations could be modelled during QTL mapping which provided increased statistical power and a comprehensible analysis. In total, eighteen multi-trait QTL were detected of which some were common to both populations. These QTL were determined to be most important for flowering time but they were also found important in the regulation of plant growth. It was found that these QTL imposed a negative correlation between QTL effects on plant growth and flowering time. However, a

considerable amount of genetic variance was explained by QTL that affected plant growth traits and not the developmental trait flowering time. These QTL were mainly population specific which led to the conclusion that the genetic basis for flowering time and correlated effects on plant growth is common between the studied accessions. They were also found to have a specific genetic basis for plant growth that did not affect flowering time. In addition, it was found that particularly for plant growth traits, two-way interactions between the QTL accounted for relatively large amounts of genetic variance. The plant growth specific QTL were chosen for follow-up studies that aim at the discovery on the genes underlying their effects.

In order to validate the detected QTL a set of 9 near isogenic lines (NILs) was selected by marker assisted selection that carried introgressions in the regions of targeted QTL. These NILs have been analyzed repeatedly for the presence of the QTL effects predicted by QTL analysis. At all 9 loci QTL effects could be validated. These did not always agree exactly with the results of QTL analysis. Effects on RGR were found at loci in the NILs where no QTL effects on this trait had been detected. Furthermore, effects on flowering time were almost exclusively found in the NILs; although these were not detected by QTL analysis. However, these effects were weak and they did not follow the typical trend of being negatively correlated to effects on plant growth rate. This led to the conclusion that in these cases flowering time was a secondary effect of plant growth QTL. Complete inversion of the expected phenotype based on QTL analysis was also observed in NILs. In one case this could directly be explained by two-way interactions that had been detected. For the second case such interaction were hypothesized to occur with un-detected QTL. Indications for the role of epistasis could be identified for two further loci which emphasized the complex genetic architecture of this trait. One QTL was found to cause penetrant morphological differences in a NIL, in which also QTL effects for RGR and leaf size were validated. The effects on morphology allowed the efficient fine-mapping of this QTL, presently to 151 genes likely to be reduced even further soon. Over-all this study demonstrated the accuracy of the multi-trait mapping methodology and the potential of using naturally occurring variation for the elucidation of the molecular bases for plant growth QTL. The now available NILs and others currently under selection will provide the opportunity realize this goal in the near future.

Although QTL for plant growth that also had strong effects on flowering time were not the main interest of this study, no distinction could be made between pleiotropy and linked QTL. Furthermore, the physiological and genetic relation between both traits had been described but it was never proven that they were under the control of the same QTL. Therefore, a QTL that affected flowering time and RGR, among other growth traits, located

on the end of chromosome 5, was fine-mapped and cloned using NILs with Sha introgressions in the L*er* genetic background. Fine-mapping achieved a resolution by which it could be unambiguously proven that the either one or more of the 3 genes *MAF2-MAF4* must be causal to the QTL effects on flowering time. Although these genes had gathered attention for their potential role in flowering time control, this had not yet been proven. The QTL for RGR was fine-mapped to a region containing 4 additional genes but none were better candidates than the *MAF*. Sequence analysis revealed a chimeric fusion between *MAF2* and *MAF3* and the absence of an intact *MAF3* gene in the Sha accession. This was not the first time these genes were found fused together in Arabidopsis but the absence of an intact *MAF3* gene made the Sha allele unique. Complementation studies showed that *MAF4* could not explain the detected QTL effects. However, *MAF3* could not fully complement the effects while *MAF2* increased flowering time beyond the value of the control accession. This led to the hypothesis that both genes are needed to properly regulate flowering time concert. The genetic material generated in this study will allow determining whether this QTL, which clearly belongs in the flowering time pathway, indeed also is involved in the control of plant growth. By doing so, this relation will become formally established.

Samenvatting

De groei van een plant is een komplexe kwantitatieve eigenschap die voortkomt uit de integratie van een veelvoud aan fysiologische processen tijdens de groei, inklusief ontwikkelings aspecten. Deze eigenschap word bepaald door genetische factoren maar word ook beinvloed door het mileu. Het is een belangrijke agronomische eigenschap omdat het de opbrengst van de oogst als biomassa direct kan bepalen. De productie van plantaardige grondstoffen staat continu onder druk om in de navraag van een statig groeiende wereld bevolking te voorzien ondanks tegenwerkende factoren zoals een veranderend klimaat. Dit was de motivatie voor vele (eoc)fysiologische studies die tot een grote hoeveelheid informatie hebben geleid. Er is echter relatief weinig bekend over de genetische basis en de onderliggende molekulaire basis van plantengroei. De studie van kunstmatig geinduceerde mutanten en doelgerichte over-expressie lijnen, voornamelijk in de model plant *Arabidopsis thaliana* (Arabidopsis), heeft tot de ontdekking van vele genen betrokken bij plantengroei geleid. Hoe deze genen gezamelijk bij de groei betrokken zijn is echter nog grotendeels onduidelijk. De in de natuur voorkomende genetische variatie in Arabidopsis is tot zover nog maar weinig benut voor de ontdekking van zulke genen. Deze variatie kan echter een waardevolle aanvulling betekenen voor de kennis die al opgedaan is. De studies beschreven in dit proefschrift hadden als doel om de genetische en molekulaire basis van in de natuur voorkomende variatie voor plantengroei in Arabidopsis the ontdekken.

Om te bepalen waar zich de kwantitatieve loci, de QTL, die plantengroei beinvloeden liggen is een QTL-karterings experiment uitgevoerd. Twee recombinante inteelt lijn (RIL) populaties zijn gefenotypeerd voor 28 groei-gerelateerde eigenschappen. Het geprojecteerde oppervlak van de Arabidopsis rozet is herhaaldelijk gemeten tijdens de groei aan de hand van digitale afbeeldingen. De toename van dit oppervlak over de tijd is gemodelleerd met een logistisch model tijdens de komplete groeicurve en een lineair model tijdens alleen de initiele exponentiele groei fase. Het logistische model leverde een schatting van de groeisnelheid, het maximale PRA en het inflektie punt van de groeicurve. Het lineaire model leverde een schatting van de relatieve groeisnelheid (RGR). Bovendien is het oppervlak, de lengte en de breedte van het grootste blad, de bloeitijd, het aantal bladeren en de chlorofyl gehalte index (CCI) bepaald van iedere plant. Omdat alle eigenschappen componenten van plantengroei waren, verspreid over de gehele groei fase, had deze data een uitgebreide correlatie structuur. Door gebruik te maken van multi-eigenschap methodologie gebaseerd op gemengde modellen konden deze correlaties uitgebuit worden bij het karteren van de QTL. In totaal zijn 18 multi-eigenschap QTL ontdekt. Sommige bevonden zich op overeenkomstige lokaties in beide RIL populaties en deze waren vooral belangrijk voor het

bepalen van de bloeitijd. Ze speelden echter ook een grote rol in het bepalen van de plantengroei en de effecten op beide klassen van eigenschappen was negatief gecorreleerd. Desalniettemin werd een aanzienlijke hoeveelheid genetische variantie voor plantengroei verklaard door QTL die geen effect hadden op de bloeitijd. Deze QTL waren vooral populatie specifiek en daarmee een idicatie voor het bestaan van een unieke genetische basis voor plantengroei in de bestudeerde Arabidopsis accessies. Bovendien werd vooral voor groei eigenschappen een aanzienlijke hoeveelheid genetische variantie door twee-weg interacties tussen de ontdekte QTL verklaard. De plantengroei specifieke QTL zijn gekozen voor vervolgonderzoek met als doel het ophelderen van de onderliggende molekulaire variatie.

Om de ontdekte QTL te valideren is een set van 9 vrijwel isogene lijnen (NILs), met introgressies van bepaalde accessies in de genetische achtergrond van een kontrole accessie, geselecteerd met merker geassisteerde selectie. Deze NILs zijn herhaaldelijk geanalyseerd en de aanwezigheid van QTL effecten kon in iedere lijn aangetoond worden. De effecten in the NILs kwamen echter niet altijd volkomen overeen met de resultaten van QTL kartering. Effecten op RGR werden gevonden op loci waar geen QTL effecten gedetecteerd waren. Bovendien werden in praktisch iedere NIL effecten op bloeitijd gevonden welke ook niet altijd met QTL kartering ontdekt waren. Deze effecten waren echter over het algemeen zwak zodat geconcludeerd kon worden dat het hier om sekundaire effecten van plantengroei QTL zou kunnen gaan. Complete omkering van fenotypisce effecten op plantengroei in de NILs, vergeleken met de gedetecteerde QTL effecten is ook waargenomen. In een enkel geval kon dit door de ontdekte twee-weg interacties tussen QTL verklaard worden. Waar dit niet het geval was zouden interacties met niet ontdekte QTL hiervoor verantwoordelijk kunnen zijn geweest. Indicaties voor epistatische interacties werden voor nog twee loci gevonden die de nadruk op de komplexe genetische architectuur van plantengroei legden. Een NIL voor een bepaalde QTL vertoonde sterke effecten op de gehele morphologie van de plant. Deze werden begeleid door effecten op RGR en de afmetingen van het grootste blad. De effecten op de morphologie van de plant maakten het mogelijk deze QTL efficient fijn te karteren tot een gebied dat 151 genen bevatte. Over het algemeen tonen deze resultaten de nauwkeurigheid van de toegepaste QTL kartering methodologie aan. Bovendien demonstreren ze de potentie van natuurlijk voorkomende genetische variatie voor het ophelderen van de molekulaire basis van plantengroei in Arabidopsis. De NILs die voor deze studie ontwikkeld zijn maken het mogelijk dit doel in de nabije toekomst te realiseren.

Ondanks dat QTL met effecten op plantengroei en tevens op bloeitijd niet direct interessant waren voor deze studie, kon geen onderscheid gemaakt worden tussen pleitropie

end genetisch gekoppelde QTL. De genetische en fysiologische relatie tussen beide eigenschappen beschreven, er was echter nog geen bewijs geleverd dat ze inderdaad onder de invloed van dezelfde QTL lagen. Een QTL met effecten op bloeitijd en, onder andere, RGR is daarom fijn-gekarteerd en gekloneerd. Hiervoor zijn NILs met introgressies van de accessie Sha in de genetische achtergrond van L*er* gebruikt. De resolutie behaald met fijn-kartering stond toe om aan te tonen dat 1 of meerdere van de 3 genen *MAF2-MAF4* verantwoordelijk moesten zijn voor de QTL effecten op bloeitijd. Alhoewel interesse voor deze genen gewekt was vanwege hun potentiele rol in het bepalen van bloeitijd in de natuur, was dit niet eerder direct aangetoond. De QTL voor RGR is fijn-gekarteerd tot een gebied dat nog 4 extra genen bevat die echter geen betere kandidaten zijn dan de *MAF*. De analyse van de DNA sequentie van de *MAF* genen leidde tot de ontdekking van een fusie tussen *MAF2* en *MAF3* waarbij een intact *MAF3* gen volledig afwezig was in de Sha accessie. Dergelijke fusie allelen van beide genen waren recentelijk beschreven voor andere accessies maar de afwezigheid van een intact *MAF3* gen maakte het Sha allel uniek. Komplementatie van het Sha allel in de NIL met het L*er* allel van *MAF4* kon de gedetecteerde QTL effecten niet verklaren. L*er MAF3* kon echter de effecten van het Sha allel ook niet volkomen komplementeren terwijl *MAF2* bloeitijd zo sterk liet toenemen dat het die van de L*er* accessie overtrof. Dit leidde tot de hypothese dat zowel *MAF2* als *MAF3* nodig zijn voor de regulatie van bloeitijd zoals het in L*er* aangetroffen wordt. Het genetische materiaal dat in deze studie is ontwikkeld voor deze experimenten zal het mogelijk maken om aan te tonen of deze QTL, die duidelijk tot het bloetijd regulatie netwerk behoord, inderdaad ook bij de regulatie van plantengroei betrokken is. Wanneer did zo blijkt te zijn zou deze relatie voor het eerst formeel aangetoond zijn.

Acknowledgements

Looking back at the last four years I can not help from scratching the back of my head wondering how fast this time has passed. I have come to the realization that this has been the most exciting and constructive time of life. During my time in Cologne many people have played their part in this experience either professionally or otherwise. However small this might have been it was significant. My gratitude goes out to all of you. I would like to use this opportunity thank some persons in particular.

Firstly, I would like to thank my promoter Maarten Koornneef. By offering me the position as PhD student, back in 2005, you have made it all happen. Being a director obviously takes a lot of time and effort. However, you have always dropped by the office spontaneously to see how the work was coming along. Furthermore, I could always drop by your office also whenever there was the need, which I much appreciated. Your friendly, positive nature seems to radiate out into the department of Plant Breeding and Genetics and generally makes it a pleasant place to work.

I would further like to thank my co-promotor and daily supervisor Matthieu Reymond. Right from the beginning until the present day I consider myself lucky to have ended up in your group. In part this is due to the enjoyable atmosphere you manage to maintain in the group. However, your continuous availability and the frequent animated discussions have also played a major role for me. Nevertheless you gave me a lot of freedom and that made me feel right 'at home'. Not to forget the fact that you were always willing to help out in the lab or the greenhouse as well.

In the early days of AG Reymond I had the pleasure to share the office and lab with the entire 4 person group. It is a pity that the remaining two group members had to move on eventually. Hugues, right from day one I knew we would be having a lot of fun and my suspicion proved right. Ania, I must admit that I genuinely miss having someone around (you) who talks to her computer when nobody else has got the time or mood to listen. Friends, you take a special place in my memories. Luckily it was not for long though until I found myself sharing the office with Nadine, Aina and Samija. I would like to thank you, and the rest of the group, Rubén, Navot and Inga for the all the good times we have and the help with scientific issues.

Sometimes my experiments involved planting thousands of seeds and taking care of over a thousand plants at a time. This would be impossible without you, Regina. Thank you for taking great care of my plants and for all the help in the greenhouse with starting and conducting experiments. Of course all those thousands of plant needed to be genotyped. To this end I would like to thank Barbara and Nele for all their help in the lab.

I would also like to thank Fred van Eeuwijk and Marcos Malosetti for having me at Biometris for two weeks to teach me how to use the mixed models. This was most instructive and I consider it a very valuable experience.

A special thanks to Maria. You play an important role in my life as my best friend and even more and I hope it will stay like this for ever. Your continuous support and care during the writing of the thesis make you part of it.

To do my PhD in Cologne meant that I had to move away from my old friends. However, the resulting distance proved to be only of geographical nature. You saw and heard me a lot less than you used to but our friendship did not suffer. For this I thank you, Maarten, Eddy, Martijn and Marc. I would also like to thank my new friends that I made here in Cologne: Thanos, Lorna, Mandi and Seth. You genuinely make life more interesting over here.

Last but not least, a big thanks to my mother and Dick for their unconditional support during and preceding my PhD.

Education Statement of the Graduate School

Experimental Plant Sciences
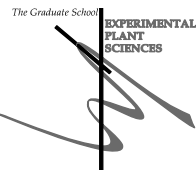
1) Start-up phase	*date*
► **First presentation of your project**	
"Characterization of QTL for plant growth and related traits in Arabidopsis"	Jun 16, 2005
► **Writing or rewriting a project proposal**	
► **Writing a review or book chapter**	
Genetic and molecular analysis of growth responses to environmental factors using *Arabidopsis thaliana* natural variation	2006
► **MSc courses**	
► **Laboratory use of isotopes**	
Subtotal Start-up Phase	*3.0 credits**

2) Scientific Exposure	*date*
► **EPS PhD student days**	
MPIZ PhD days	2005
MPIZ PhD days	2006
PhD student day, Wageningen University	Sep 13, 2007
MPIZ PhD days	2007
MPIZ PhD days	2008
► **EPS theme symposia**	
MPIZ Institute days	2005
MPIZ Institute days	2006
MPIZ Institute days	2007
MPIZ Institute days	2008
► **NWO Lunteren days and other National Platforms**	
NWO Lunteren	Apr 04-05, 2004
NWO Lunteren	Apr 07-08, 2008
► **Seminars (series), workshops and symposia**	
Seminars by invited speakers at MPIZ	2005-2009
Workshop on growth phenotyping and imaging in plants	Jul 17-19, 2007
► **Seminar plus**	
Carlos Alonso-Blanco	2005
Olivier Loudet	2006
Michael Tomasello	Dec 05, 2007
Murray Grant	Nov 12, 2008
Kirsten Bomblies	Mar 18, 2009
► **International symposia and congresses**	
Symposium on Plant Evolution and Domestication	Oct 25-26, 2006
6. Plant GEM	Oct 03-06, 2007
► **Presentations**	
Scheduled progress reports before the department of Plant Breeding and Genetics at the MPIZ	Dec 07, 2005
idem	Sep 06, 2006
idem	May 16, 2007
idem	Apr 24, 2008
idem	Apr 08, 2009
Poster presentation at annual PhD-day	Sep 14, 2006
Workshop on growth phenotyping and imaging in plants	Jul 17, 2007
PhD Workshop 'Natural Variation in Plants'	Sep 26, 2008
Annual PhD-day 2008 (MPIZ-Koeln)	Oct 21, 2008
► **IAB interview**	
► **Excursions**	
Subtotal Scientific Exposure	*17.4 credits**

3) In-Depth Studies	*date*
► **EPS courses or other PhD courses**	
Plant Genetics-from mutants to quantitative genetics (IMPRS course)	Apr 21-22, 2005
Graduate Course "Environmental signalling: Arabidopsis as a model	Aug 22-24, 2005
Mathematics in Plant Biology	Feb 19-24, 2006
Graduate Course "Environmental signalling: Arabidopsis as a model	Aug 27-29, 2007
SISG - Quantitative genetics	Sep 03-05, 2007
SISG - QTL mapping	Sep 05-08, 2007
SISG - Association mapping	Sep 10-12, 2007
EPS PhD Workshop 'Natural Variation in Plants'	Aug 26-29, 2008
► **Journal club**	
Literature discussion in inter-group meetings	2005-2009
► **Individual research training**	
Personal training in statistical analysis of my data	Oct 15-26, 2007
Subtotal In-Depth Studies	*13.4 credits**

4) Personal development	*date*
► **Skill training courses**	
Career Symposium organized at MPIZ	Feb 01, 2008
► **Organisation of PhD students day, course or conference**	
Involved in establishment and organisation of monthly interdepartmental PhD-student meetings at MPIZ	2006-2008
► **Membership of Board, Committee or PhD council**	
Member of PhD-student organisation at MPIZ	2006-2008
Subtotal Personal Development	*4.7 credits**

TOTAL NUMBER OF CREDIT POINTS*	**38.5**

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the
Educational Committee of EPS which comprises of a minimum total of 30 credits

* A credit represents a normative study load of 28 hours of study

Curriculum vitae

Bjorn Pieper was born on 15 August 1976 in Enschede, The Netherlands. In 1996 he completed the KMBO School in Laboratory Technology applied to chemistry. Following this, his education was interrupted for 3 years while he was working in a business not related to science. In 1999 he resumed his education in Laboratory Technology at the MBO Rijn IJssel College in Arnhem, specifying in botany, from which he graduated in 2002. This included 6 month of trainee-ship performed at Plant Research International, Wageningen, The Netherlands, working on *Agrobacterium*-mediated transformation of ornamental rose. The graduation thesis encompassed another 6 months during which he was involved in genetic studies on flax at the Laboratory of Plant Breeding of the Wageningen University. During the same year he started an HBO education in Laboratory Technology at Larenstein, Arnhem, The Netherlands, specializing in Plant Biotechnology. The graduation thesis was conducted at the Laboratory of Genetics of the Wageningen University, working on the development of genetic tools in the metal hyperaccumulator *Thlaspi caerulescens*. In 2005 Bjorn acquired the degree of BSc in completion of these studies. The same year he started his PhD research at the Max Planck Institute for Plant Breeding Research in Cologne, Germany of which this thesis is the result.