Natural variation in shade- and ethylene-induced differential growth and transcription in Arabidopsis

Natuurlijke variatie in schaduw- en ethyleen-geïnduceerde differentiële groei en transcriptie in Arabidopsis
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. J.C. Stoof, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op dinsdag 9 juni 2009 des middags te 12.45 uur

door

Laurens Bastian Snoek
geboren op 10 juli 1977, te Almelo

Promotor
Prof. dr. L.A.C.J. Voesenek

Co-promotor
Dr. A.J.M. Peeters

The research for this thesis was funded by The Netherlands Organization for Scientific Research (NWO), Program Genomics Grant 050-10-029.

Table of contents

ISBN
9789064643422

Chapter 1

Introduction

L. Basten Snoek, Laurentius A.C.J. Voesenek, Anton J.M. Peeters.

Flooding stress

Since plants are in general confined to a limited space and are unable to escape this environment, it is of the utmost importance that they sense and respond appropriately to any change in that environment. Because of this sessile lifestyle plants are continuously threatened by biotic attackers and abiotic stress. To cope with environmental changes plants must adjust to improve their chances for survival and to control the use of limiting resources. Plants have evolved various strategies to deal with environmental variability demonstrating the enormous ability to adjust phenotypically to changes in the environment (phenotypic plasticity). These, usually heritable, acclimations include physiological processes, anatomical structures, morphological changes and behavioral traits, or combinations of these.

Flooding is a severe abiotic stress that plants may encounter. Any excess of water, whether it is waterlogging of the roots only or complete submergence of the entire plant, is very damaging and nearly always leads to considerable losses in plant productivity. It has been estimated that world wide around 10% of the farmland faces regular and incidental flooding, leading to a loss of approximately 20% crop yield (Voesenek et al., 2006 and references therein). In the USA alone crop losses over the last 25 years due to flooding have been estimated to be around 50 billion dollar (Mittler, 2006).

Terrestrial plants are usually not very well adapted to flooding and will die rapidly when such an event occurs. Flooding tolerant species are characterized by a range of metabolic and morphological adaptations (Jackson, 1985; Crawford, 1992; Perata and Alpi, 1993; Armstrong et al., 1994; Drew et al., 2000; Sauter, 2000; Colmer, 2003; Gibbs and Greenway, 2003; Voesenek et al., 2006). Semi-aquatic species inhabit areas that are regularly flooded and these species have evolved two strategies to cope with these conditions. Flooding leads to rapid oxygen shortage in non-photosynthesizing tissues which induces energy production via a fermentative pathway. This can be combined with a down regulation of redundant metabolic processes to limit the use of resources until the water retreats (Voesenek et al., 2006 and references therein). Another adaptation stimulates aeration of submerged plant parts via enhanced (differential) growth to bring leaf tips above the water surface (Voesenek et al., 1990; Sauter, 2000). The various traits associated with this adaptation are collectively called the “low oxygen escape syndrome” (Bailey-Serres and Voesenek, 2008). These traits are studied for many years to unravel the molecular, cellular, physiological and anatomical mechanisms induced upon submergence (Rijnders et al., 2000; Vriezen et al., 2000; Cox et al., 2004; Benschop et al., 2005; Vreeburg et al., 2005).

Flooding tolerance of *Rumex palustris*

Rumex palustris is a flooding tolerant member of the Polygonaceae family. The flooding tolerance of this species is characterized by several traits related to escape from the submerged environment.

When plants undergo full submergence gas exchange is severely hampered by the fact that gas diffusion in water is 10,000 times slower in water than in air (Armstrong, 1979). *R. palustris* senses the change in environment and one of the pivotal factors for that is the rapid build-up (Banga et al., 1996) of ethylene inside plant tissues due to continuous production and slow diffusion into the water layer. Furthermore, oxygen levels decrease rapidly, mainly in non-photosynthesizing tissues (Mommer et al., 2006).

One of first responses in *R. palustris* induced by ethylene accumulation is a change of the angle of some of the petioles by differential growth thus lifting the tip of the leaf towards the water surface. This differential growth process is called hyponastic growth. When a certain threshold angle is reached (approximately 50 degrees) the petiole, and to a much lesser extent the blade, starts to elongate to bring the leaf tip above the water level thus restoring gas exchange (Cox et al., 2003; Cox et al., 2004).

Next to ethylene, also the plant hormones Abscisic acid (ABA), Gibberellic acid (GA) and auxin are involved in these growth responses to submergence. Ethylene induces a fast decrease in the ratio between ABA and GA (Cox et al., 2004; Benschop et al., 2005). In both rice and *R. palustris* a strong and rapid reduction of the endogenous ABA concentration was observed (Hoffmann-Benning and Kende, 1992; Benschop et al., 2005). In *R. palustris* this decline is as in rice (Saika et al., 2007), at least partly, brought about by a down regulation of several 9-cis-epoxycarotenoid dioxygenases (NCED), enzymes involved in the rate limiting step in ABA biosynthesis, and enhanced ABA catabolism (Benschop et al., 2005; Saika et al., 2007). It is clear that in both Rumex and rice, ABA acts as a negative regulator of underwater elongation. This negative effect of ABA is related to its inhibition of enhanced GA biosynthesis (Benschop et al., 2005). Recently, Benschop et al., 2006 demonstrated that the presence of ABA actively down regulated the GA biosynthesis gene *RpGA3ox1*. The role of auxin is less clear, it seems necessary for the hyponastic response but when auxin transport is inhibited the response is only delayed (Cox et al., 2006). More downstream, the involvement of cell wall loosening proteins, expansins, was demonstrated (Vriezen et al., 2000; Vreeburg et al., 2005).

Use of Arabidopsis in flooding research

Rumex palustris is an ideal plant to study submergence-induced growth responses from a physiological perspective. From a genetic and molecular point of view it suffers some major drawbacks. The species is hexaploid and there is little genomic information beside the fact that it has a large genome. Moreover, the species is very recalcitrant to transformation, which together with the hexaploidy, makes generating and testing transgenics a difficult task.

Arabidopsis thaliana meets all the criteria mentioned above with respect to genetics and molecular biology, but it lacks many traits (e.g. enhanced shoot elongation, aerenchyma development) associated with flooding tolerance. Interestingly, Arabidopsis does show hyponastic growth when exposed to submergence or elevated ethylene levels. Furthermore, similar differential growth responses are induced in this species by spectral

neutral shade (Millenaar et al., 2005) and elevated temperatures (van Zanten *et al.*, unpublished results). Another important feature was that Arabidopsis showed ample natural genetic variation among Arabidopsis accessions with respect to hyponastic growth (Millenaar et al., 2005; Pierik et al., 2005). Taken together, this makes Arabidopsis a suitable species to study part of the submergence responses as observed in *Rumex palustris*.

Natural variation in Arabidopsis

In nature variation in phenotypes is present among individuals of most organisms, including Arabidopsis. This variation is influenced by two factors, the environment and the genetic composition of the individuals. The interaction of these two factors determines the plants phenotypes. When Arabidopsis accessions are grown in the same environment it becomes possible to study the genetics of a certain phenotype (Nordborg et al., 2005). The observed variation between different phenotypes that can be attributed to the genotype is expressed as the broad sense heritability (H^2) of a phenotype. When the H^2 is high the phenotypes are highly determined by the genetic differences between the tested accessions.

Phenotypes determined by one or only a few genes often have such a high heritability and can be observed as distinct classes during segregation. These phenotypes are called qualitative and are relatively easy genetically dissected. In contrast, phenotypes which show a more continues distribution of phenotypes are called quantitative. The observation that the phenotypic classes show a large overlap is mostly due to the involvement of many genes and possible epistatic interactions (Koornneef et al., 2006). Also a high interaction with the environment can further complicate the genetic dissection of the variation within a trait (Scarcelli et al., 2007). All of these factors (allelic effects and interaction with the environment) can add positively or negatively to a phenotype when it is classified and therefore not be clearly associated with a genotype. However, in natural populations many traits are found to have a quantitative distribution, like, flowering time, circadian rhythm, light and hormone responses, leaf architecture, seed-dormancy and phytate and phosphate accumulation (Alonso-Blanco et al., 1998; Swarup et al., 1999; Borevitz et al., 2002; Perez-Perez et al., 2002; Alonso-Blanco et al., 2003; Bentsink et al., 2003). This quantitative variation provides these natural populations to cope with ever changing environmental conditions and so ensure the survival of the species (Gailing et al., 2004; Botto and Coluccio, 2007).

Quantitative trait analysis

The development and growing ease at which molecular markers can be generated has led to an increased use of natural variation to study quantitative traits. Even though the genetic regulation of such traits is complex, much progress has been made in their genetic dissection (Alonso-Blanco and Koornneef, 2000; Koornneef et al., 2004), due to the combination of the molecular markers and sophisticated mapping methods (Jansen, 1993; Zeng, 1994; Kao et al., 1999; Zeng et al., 1999). The development of immortal mapping populations has further aided the study of the genetic mechanisms behind natural variation. Especially Recombinant Inbred Line (RIL) populations have been developed and used increasingly in the past decades (Alonso-Blanco et al., 1998; Loudet et al., 2003; El-Lithy et al., 2004). In contrast to mutant screens, in which the variation of one gene is studied to unravel its position and function, a RIL screen consists of variations of multiple genes.

These RIL populations can therefore be used to study many different traits associated with many different environments. The ease at which these populations are generated in the model species Arabidopsis has made it one of the favorite tools to study plant genetics.

A RIL population is made by a cross between two genetically different, preferably homozygous, Arabidopsis accessions. By recombination the individuals of such a population have a different genotypic composition consisting of the parental genotypes. By selfing the F2 to at least the F8 a near homozygous, immortal population is created which can be tested time after time. By genotyping the individual RILs and the availability of molecular markers and statistics, Quantitative Trait Loci (QTLs) can be found. Even though precision of QTL mapping depends on factors like population size, genetic map density and heritability of the studied trait it has proven its use in many cases (Swarup et al., 1999; El-Assal et al., 2001; Paran and Zamir, 2003; Kroymann and Mitchell-Olds, 2005; Price, 2006).

QTLs often contain many genes and still need conformation and further fine mapping up to the level of the single gene. For this purpose Near Isogenic Lines (NILs) are frequently used (Keurentjes et al., 2007). These lines contain a small genetic introgression of one of the parental backgrounds in the genetic background of the other parent. In these NILs most traits have been mendelized (to segregate according to Mendel's laws) and can be studied as single gene mutants for which the exact position of the gene can be found by mapping of the progeny of the repeated back-crosses.

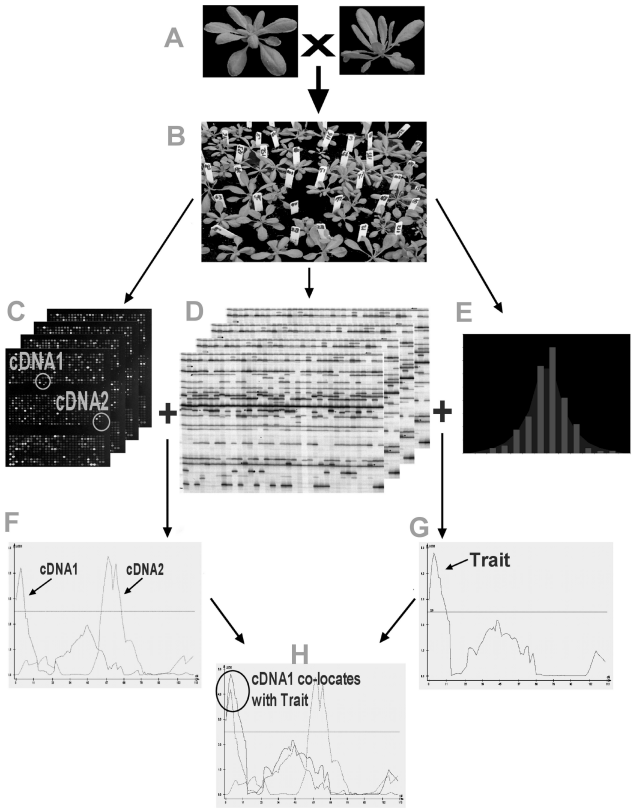


Figure 1.1 Genetical genomics as proposed in Jansen & Nap (2001). Two accessions are crossed (a) to create a segregating population (b). From each individual of this population three types of data are obtained: Transcript levels (c), Genotype (d) and Phenotype (e). Combining (c) and (d) will give QTLS for expression levels (eQTLs) (f). By combining (d) and (e) QTLs for phenotypes are obtained (g). By comparing (f) and (g) for co-locating QTLs (h), a transcriptional regulation network underlying that trait can be constructed.

Natural variation in transcript abundance; genetical genomics

Nowadays, high-throughput measuring of transcript levels, often genome-wide, can give highly detailed information on transcript regulation. In these experiments most often environmental, time or single-gene perturbations are used as a method for the construction of transcript networks. Recently, several studies have successfully used a multi-factorial genetic perturbations method to gain insight in genetics of transcript regulation (Schadt et al., 2003; Bing and Hoeschele, 2005; Li et al., 2006; Keurentjes et al., 2007; West et al., 2007). This method, called "genetical genomics" (Jansen and Nap, 2001) makes use of the combination of expression profiling and QTL mapping in a segregating population (Fig. 1.1).

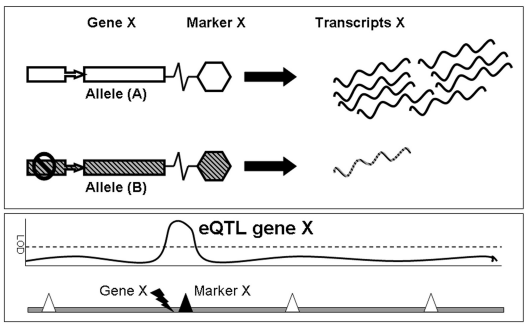


Figure 1.2 eQTL for local regulation. When two alleles (A & B) of a gene (X) are differently transcribed due to a polymorphism in or nearby the gene, the variation in transcript abundance will map to the physical position of that gene.

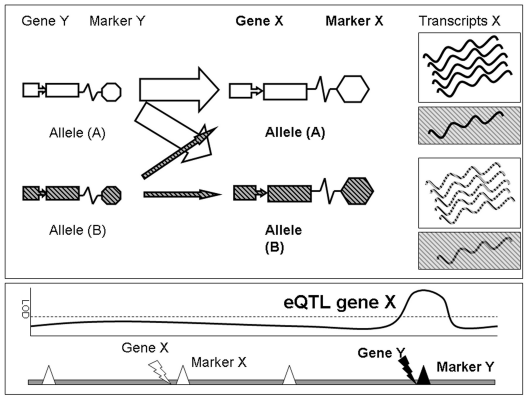


Figure 1.3 eQTL for distant regulation. When the transcription of a gene (X) is determined by the allele type of another gene (Y), the variation in transcript abundance will map to the physical position of the gene (Y) which allelic difference causes the difference in transcript abundance. In a grey Y genetic background gene X is hardly transcribed whereas in a white Y genetic background gene X is highly transcribed.

Natural variation among several different Arabidopsis accessions can also be observed for transcript abundance. When the transcript levels of the individuals of a RIL population are determined, QTLs for expression differences (eQTL) between alleles of genes can be calculated. The eQTL(s) found for an individual gene contain its putative regulator(s). The eQTL is determined for local regulation if a gene is physically located on its own eQTL (Fig. 1.2). Because this may be caused by auto-regulation (a gene regulating its own expression) the eQTLs can also be called putative cis regulatory. When the eQTL maps to a different position than the physical position of the gene, the eQTL is determined for distant regulation (Fig. 1.3). This eQTL contains a putative regulator for the gene (target) under study. By correlating the biological features or transcript abundance of target and putative regulator putative interactions can be found. When subsequent steps are investigated a putative transcript regulatory network can be constructed.

Recently, this novel approach has successfully been tested in Arabidopsis (Keurentjes et al., 2007; West et al., 2007). The

highly complex architecture of transcript abundance was revealed in both the Bay-0 x Sha (West et al., 2007) and L*er* x Cvi RIL populations (Keurentjes et al., 2007). Of the Bay-0 x Sha population the genome-wide expression profiles of 211, 6 weeks old RILs was characterized with single channel micro-arrays (Affymetrix microarrays). The expression profiles of 160, 1 week old RILs of the L*er* x Cvi population were determined by two channel spotted arrays (Allemeersch et al., 2005). Despite the differences in experimental conditions and plant age, general patterns of transcript level controlled by the genetic background were discovered. In both experiments many eQTLs were identified and distant regulation was more abundant than local regulation even though for local regulation stronger QTLs were found.

Another common feature was the discovery of so called hotspots of transcript regulation (HTR). These are genomic loci at which more eQTLs were found than can be expected by chance and therefore may contain putative "master regulators". Although much more eQTLs were found and a more detailed description was made in the Bay-0 x Sha study, the use of the L*er* x Cvi data was extended to putative transcript regulation network generation. In the L*er* x Cvi population a positive correlation between eQTL and SNP frequency was found, especially for local regulation. Also the amount of eQTLs per gene ontology class was studied. While most classes contained roughly similar relative amounts of eQTLs a significant under representation was found for genes involved in regulatory processes. This may have biological significance but could also be caused by the relative low abundance of the transcripts of those genes and hence their more difficult detection. Finally, a putative regulatory network was presented (Keurentjes et al., 2007), in which previously unknown regulators of genes involved in flowering time were detected by combining expression trait profiling with eQTL mapping, gene annotation and extended by Iterative Group Analysis (IGA) (Breitling et al., 2004).

Genetical genomics experiments on a number of different species, like *Homo sapiens, Zea mays, Saccharomyces cerevisiae*, *Mus musculus, Caenorhabditis elegans* and *Arabidopsis thaliana* (Bing and Hoeschele, 2005; Bystrykh et al., 2005; Li et al., 2006; Keurentjes et al., 2007; West et al., 2007) have been shown to extend the knowledge of gene interaction networks. Two experiments in Arabidopsis indicate that there are promising possibilities for future genetic research in Arabidopsis and demonstrate that "genetical genomics" is feasible in this species too.

Scope of this thesis

Natural variation exists for many traits, including those that enable plants to cope with environmental stress. The genetic mechanisms behind coping with these stresses have been studied extensively and some regulatory networks have been unraveled. However, due to the complex polygenic nature of many stress-related traits a lot remains unknown. A promising new technique called "genetical genomics" proofed to be feasible in Arabidopsis during this thesis project. The data generated through genetical genomics enables the construction of putative regulatory networks. The general research question of this thesis is therefore to unravel regulatory networks operating during ethylene- and low light-induced hyponastic growth in Arabidopsis. To this end QTL analyses of phenotypic leaf angle traits and global transcription analysis (genetical genomics) will be used to explain and dissect natural variation in hyponastic growth and rosette compactness.

Chapter two describes hyponastic growth in detail and has a focus on a QTL study to find large effect controlling loci for leaf angle related traits. Two leaf regions of differential growth were identified and could be separated as genetically (partly) independent. Furthermore, a difference in response to ethylene compared to low-light was found. A QTL involved in the response to both ethylene and low-light was found to be caused by the allelic difference of the *ERECTA* gene. Some other QTLs could also be narrowed down to a single gene.

In Chapter three the light influenced growth traits; rosette area, rosette diameter (fermax), rosette compactness and Relative Growth Rate (RGR) are investigated by QTL analysis. Several QTLs were found and for some the underlying gene could be identified. ERECTA again proved to play an important role in these traits, but also PHYB had a functional allelic difference.

In chapter four a method is described to construct a network for transcript regulation by combining genetical genomics data with single gene perturbation transcript profiling. By using the difference in transcript abundance between the *ERECTA* mutant L*er* and its wild type Lan as a start we constructed (part) of the downstream signaling cascade.

In chapter five we concentrate on the comparison of two genetical genomics experiments using the L*er* x Cvi RIL population. The results of a experiment, where the RIL population is treated with 3 hours of low light, are compared with a previous experiment on the same population (Keurentjes et al., 2007). Apart from network construction, the genetical genomics data of low light treated L*er* x Cvi RILs is used to explain variations in light affected traits measured in chapters two and three.

In chapter six a NIL is used to study the possibilities to extend the method of network construction described in chapter four. Although in chapter four the effects of a single gene were used, often the gene underlying a QTL is not known and the whole locus needs to be studied. By the transcript profiling of a NIL the effects of a single “small” introgression on transcript levels are known and can be used as a start for constructing the regulatory network.

Chapter seven contains a general discussion about the research described in this thesis.

Chapter 2

Hyponastic growth of leaves in *Arabidopsis thaliana* acts via two genetically separable differential growth regions.

L. Basten Snoek, Frank F. Millenaar, Laurentius A.C.J. Voesenek, Anton J.M. Peeters.

Abstract

The ability of plants to react appropriately to changes in the environment is key for their survival. Plants exposed to natural shade or submergence often shows resource-directed growth. In dicots this is achieved via two distinct processes: upward movement of the leaves (hyponastic growth) and elongation of the entire petiole. *Arabidopsis thaliana* accessions show hyponastic leaf growth when exposed to elevated ethylene and low-light. In this study we used two RIL populations, Columbia x Landsberg *erecta* (Col x L*er*) and Landsberg *erecta* x Cape Verde islands (L*er* x Cvi), to analyze natural variation in hyponastic growth in response to low-light (spectral neutral shade) and ethylene. Through a detailed description of the traits underlying hyponastic growth two genetically separable regions of differential growth were identified. Natural variation was present for all of those traits and QTLs were found. The genes, *CRYPTOCHROME2* and *ERECTA* known polymorphic between L*er* and Cvi and *NON-PHOTOTROPIC HYPOCOTYL 4* were shown to be important for hyponastic growth phenotypes.

Introduction

The ability of plants to react appropriately to changes in the environment is key for their survival. In this respect re-orientation of plant growth optimizes the location of plant organs towards limiting natural resources, such as light, oxygen and carbon dioxide and thus may improve plant fitness. An example of re-orientation of growth is the phototropic response, in which differential growth directs plant-organs towards the light source (Firn and Digby, 1980; Hangarter, 1997; Ahmad et al., 1998; Ballare, 1999; Friml et al., 2002). Re-orientation of leaf growth is also clearly demonstrated during shade- and submergence avoidance (Voesenek et al., 1989; Clua et al., 1996; Gautier et al., 1997; Ballaré, 1999; Cox et al., 2003; Pierik et al., 2003; Pierik et al., 2004; Millenaar et al., 2005). During these acclimations, growth is directed towards the resources light and air (e.g. CO_2 and O_2), respectively.

When dicots are exposed to natural shade or submergence, resource-directed growth is achieved via two distinct processes: upward movement of the leaves (hyponastic growth) and elongation of the entire petiole (Cox et al., 2003; Pierik et al., 2003). Although differential growth responses to shade and submergence show phenotypic similarity, the induction is triggered by different cues. For shade-induced hyponasty and petiole elongation a decrease in the R to FR ratio and a reduction in blue-light are crucial factors (Smith and Whitelam, 1997; Ballaré, 1999; Pierik et al., 2004). Recently, it was shown for *Nicotiana tabacum* that the gaseous hormone ethylene too is essential for well-timed shade avoidance responses (Pierik et al., 2003). During submergence-induced hyponastic growth, however, ethylene is the pivotal factor inducing growth and petiole elongation, such as in the semi-aquatic plant *Rumex palustris* (Voesenek et al., 2006).

Ethylene, however, is not the only hormone involved in this elongation response. Extensive work done on *R. palustris* shows that other hormones like abscisic acid (ABA), gibberellins (GA) and auxin play important roles in this process too (Cox et al., 2004; Benschop et al., 2005; Vreeburg et al., 2005; Benschop et al., 2006; Cox et al., 2006; Benschop et al., 2007). More downstream in the signal transduction cascade cell wall loosening proteins, such as expansins (Vriezen et al., 2000; Vreeburg et al., 2005), are involved. The dependency on various hormones, signal-transduction cascades and consequently a large number of genes and proteins make hyponastic growth a complex polygenic quantitative plant trait.

Interestingly, *Arabidopsis thaliana* accessions also show hyponastic leaf growth when exposed to elevated ethylene concentrations and low-light levels (Millenaar et al., 2005). Furthermore, as for many other traits (Koornneef et al., 2004), considerable natural genetic variation for hyponastic growth exists in Arabidopsis accessions (Millenaar et al., 2005). Variation was found for initial angle, for the response kinetics to increased ethylene levels or low-light and for the final angle due to these treatments. Millenaar et al. (2005) also demonstrated that hyponastic growth in Arabidopsis leaves is characterized by two locations of differential growth. One is located at the basal end of the petiole; the other at the basal region of the leaf blade. To escape unfavorable conditions, via hyponastic growth, it is essential that the plant raises it leaf-tips as high as possible. This can be achieved by the reaction of the petiole and extended by the reaction of the leaf-blade (Millenaar et al., 2005).

In plant populations with enough genetic and phenotypic variation, Quantitative Trait Loci (QTL) analysis is a powerful tool to study complex polygenic traits to identify genomic loci, genes and alleles (El-Assal et al., 2001; Koornneef et al., 2004; ter Steege et al., 2005). This technique requires the combined study of physiological characteristics and molecular genetics (Koornneef et al., 2004). In Arabidopsis, the generation of Recombinant Inbred Lines (RIL) (Alonso-Blanco et al., 1998; Alonso-Blanco and Koornneef, 2000), resulting from a cross between two contrasting accessions, greatly facilitated QTL analysis (Koornneef et al., 2004).

In this study we used two of these RIL populations, Columbia x Landsberg *erecta* (Col x L*er*) (Alonso-Blanco et al., 1998) and Landsberg *erecta* x Cape verde islands (L*er* x Cvi) (Alonso-Blanco et al., 1998), to analyze natural variation in hyponastic growth in response to low-light (spectral neutral shade) and elevated ethylene. The QTLs (and candidate genes) that are specific for and/or shared between the leaf blades and petioles are discussed. Natural variation in the pleiotropic functions of the well studied genes *CRYPTOCHROME2* (and *ERECTA* and *PHYTOCHROMEB*) were found. Moreover, genes like *GIGANTEA* and *NON-PHOTOTROPIC HYPOCOTYL4* (NPH4) were found to function in hyponastic growth. The role of NPH4 was identified to be most important for the basal part of the leaf blade.

Results

Hyponastic growth phenotypes

Hyponastic growth in Arabidopsis is the change in angle due to differential growth of the petiole, the leaf-blade or a combination of both. Typical phenotypes of leaf movement (Fig. 2.1 A - I) show that a response can take place either via A, D, G or H to I, when the lamina is the first organ to respond or via A, B, C or F to I when the petiole is the first responding organ. More extreme, intermediate and combinations of phenotypes of initial angle and response are possible and were observed.

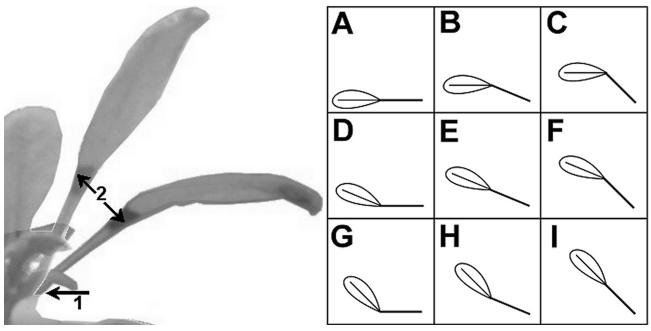
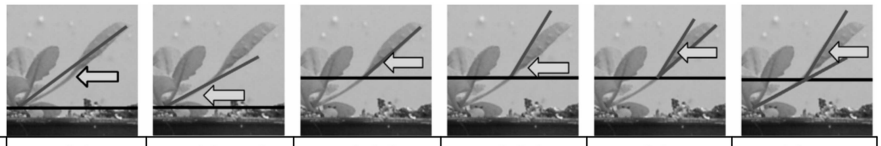


Figure 2.1 Left: Compilation of Cvi before (lower (darker) leaf) and after (higher (lighter) leaf) 6 hours low-light treatment showing two distinct regions (arrows 1 and 2) where differential growth of the petiole occurs. Right: Schematic overview of different ways of hyponastic growth response of the leaf.

Various components that determine the final leaf angle form the basis for the QTL mapping. The leaf-tip angle (LTA) (Fig. 2.2A) was measured to describe the total chance in angle of the leaf tip due to hyponastic growth. To determine their specific contribution to hyponastic growth, petiole angle (PA, Fig. 2.2B) and petiole bending angle (PBA, Fig. 2.2F) were measured. The PA describes the differential growth at the base the petiole. This

pushes the leaf upwards since most of the remaining part of the petiole stays (nearly) straight. The PBA describes the activity of the second region of differential growth near the base of the leaf blade. It increases the angle of the tip of the leaf even further.

The tip of a leaf with a downward curvature exhibits a constitutively lower leaf tip angle and therefore there was a need to distinguish the leaf shape from the leaf-blade response. By combining the leaf-blade basal angle (LBBA) (Fig. 2.2C) and leaf-blade top angle (LBTA) (Fig. 2.2D) the leaf shape angle (LSA, Fig. 2.2E) could be calculated.



Angles:	**A** Leaf Tip Angle	**B** Petiole Angle	**C** Leaf Blade Basal Angle	**D** Leaf Blade Top Angle	**E** Leaf Shape Angle	**F** Petiole Bending Angle
Initial	LTA	PA	LBBA	LBTA	LSA	PBA
Final	LTA	PA	LBBA	LBTA	LSA	PBA
Response	RLTA	RPA	RLBBA	RLBTA	RLSA	RPBA

Figure 2.2 Leaf angle traits measured/calculated that were used for the QTL analysis. Initial angles were measured at the start of the treatment. Final angles were measured after 6 hours of treatment for both low light and ethylene. The difference between Initial and Final angle is described as the response (R). Trait E (leaf shape, LSA) was calculated based on traits C and D, trait F (petiole bending, PBA) was calculated based on traits B and D.

Phenotypes of parental accessions

The three accessions used as parents for the two RIL populations (L*er* x Cvi and Col x L*er*) all showed the two distinct regions of differential growth of the leaf in response to both low-light and ethylene. They showed a significant (2 sided t-test; $p < 0.05$) difference for most of the traits at the start of the experiments (Fig. 2.3). In general, Col has the most prostate growth form and has the lowest PA, PBA and LSA. L*er* can be identified by its curved leaves which results in the highest LSA. Cvi on the other hand displays flat leaf blades without a curvature with an LSA value comparable to Col. Nevertheless, Cvi can be clearly distinguished from the other two accessions by the relative upright position of its petioles. These phenotypes and combinations of them are also visible among the individual RILs of the two populations.

The parental accessions were shown to have very different amplitudes of response to low-light and ethylene treatment and confirm the data of (Millenaar et al., 2005). This includes the observation that Col responds to both low-light and ethylene (Fig. 2.3) treatment, whereas L*er* petioles hardly respond to ethylene but do respond to low-light. Interestingly, some remarkable differences were found for the responses of leaf blades when compared to the responses of the petioles. The position of the leaf blade can be influenced by differential growth of the most basal part of the petiole (PA) and through bending of the petiole/lamina border (PBA). This is most noticeable in Col and Cvi, which exhibit the largest increase in angle of the leaf blade through bending (ResponsePBA) in low-light treatments. The ResponsePBA however, is achieved differently in these two accessions. The laminae of Col plants have a lower initial angle (like Fig. 2.1B), while Cvi laminae display a higher angle after 6 hours of low-light treatment (like Fig. 2.1H), both

resulting in a larger response in ResponsePBA. L*er* laminae can in contrast to its petioles, respond to both treatments, albeit to a lesser extend than Col and Cvi (Fig. 2.3).

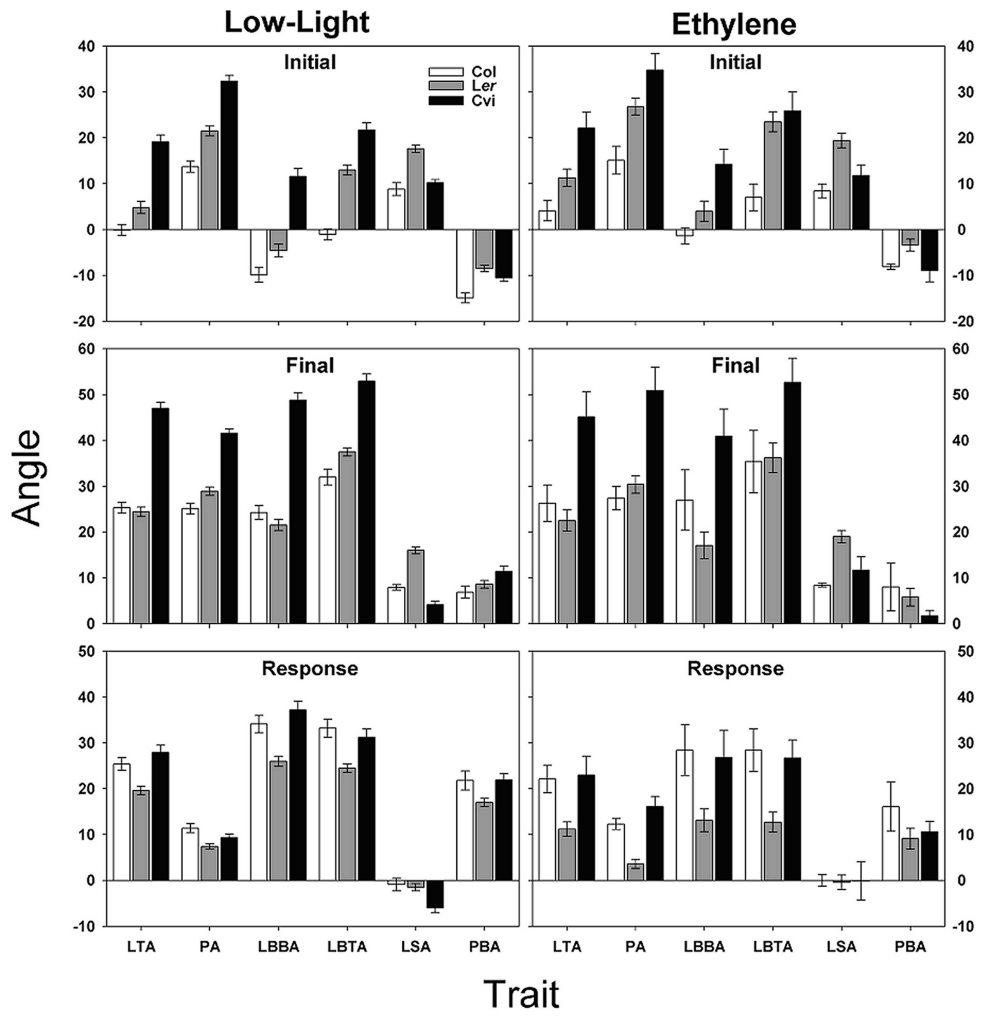


Figure 2.3 Initial, final and response angles in degrees of the parental accessions Col (white bars), L*er* (grey bars) and Cvi (black bars). Graphs on the left show low-light treatment (Col, n=15; L*er,* n=52; Cvi, n=39). Graphs on the right show ethylene treatment (Col, n=14); L*er*; n=12); Cvi, n=16). Error-bars show standard error.

Transgression and heritability

The variation in phenotypes found in the RIL populations (SI Fig. 2.1 to 2.4) extends beyond the variation observed in the parental genotypes in most cases, thus demonstrating transgression. To estimate the genetic effect of the within-trait variation, broad sense heritability (H^2) was determined. The H^2 was different for the traits, between

and among treatments and populations (Table. 2.1). The traits for which the absolute value was used (at the start and after 6 h of treatment) had in general a higher H^2 when compared to the traits that describe the response. Traits in which the leaf shape was present had the highest H^2 (Table. 2.1) indicating that this morphological trait is highly controlled by the genetic background.

Table 2.1: Broad sense heritability (H^2) for various initial, response and final angle traits, calculated for two RIL populations and two treatments.

		L*er* x Cvi		Col x L*er*	
Heritability		Low- Light	Ethylene	Low- Light	Ethylene
Initial	LTA	0.66	0.59	0.48	0.38
	PA	0.60	0.43	0.64	0.32
	LBBA	0.66	0.63	0.41	0.39
	LBTA	0.56	0.53	0.55	0.44
	LSA	0.68	0.58	0.60	0.61
	PBA	0.32	0.35	0.39	0.28
Response	LTA	0.28	0.48	0.45	0.07
	PA	0.43	0.50	0.46	ND
	LBBA	0.29	0.42	0.42	0.08
	LBTA	0.30	0.48	0.40	0.11
	LSA	0.28	0.19	0.21	0.06
	PBA	0.28	0.35	0.34	0.16
Final	LTA	0.69	0.55	0.48	ND
	PA	0.49	0.45	0.30	ND
	LBBA	0.73	0.59	0.54	ND
	LBTA	0.58	0.48	0.42	ND
	LSA	0.72	0.59	0.73	0.22
	PBA	0.48	0.44	0.32	0.13

The relative role of specific angle traits determining final leaf angle

To identify the importance of the specific traits (Fig. 2.2) determining the final leaf tip angle we used a linear model. In this model we used the initial angles and the response of the traits, PA, LSA and PBA as explaining factors for final LTA. The role of most of these six traits determining the final LTA is equal for both treatments, as shown by their adjusted regression coefficients (Table. 2.2). A remarkable difference, however, was found between the response of the two regions of differential growth (ResponsePA and ResponsePBA) which enable the lift of the lamina during hyponastic growth. For instance, the ResponsePBA shows a much more important role in the response to low-light than the response to ethylene. In Fig. 2.4 the ResponsePA (x-axis) is plotted against the ResponsePBA (y-axis) when ethylene (white) or low-light (black) were used to induce hyponastic growth. The two treatments are separated along the y-axis which indicates that the distal part of the petiole nearest to the lamina (PBA) reacts much more to shade then to the ethylene treatment.

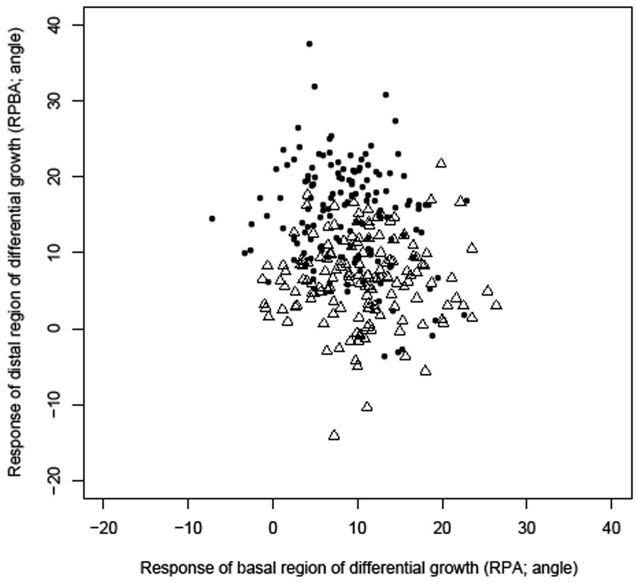


Figure 2.4 Scatter-plot of the responses (in degrees) of the two regions of differential growth (ResponsePA: x-axis; ResponsePBA: y-axis) of the RILs of the L*er* x Cvi population during the hyponastic response to low-light (black dots) or ethylene (white triangles).

The leaf shape has a negative influence on the final LTA because increased curvature results in a lower leaf-tip angle (LTA)(data not shown). Since the leaf shape hardly changes during both treatments, it only has a constitutive effect on the final LTA. In the L*er* x Cvi population the initial PA plays the most important role, whereas in the Col x L*er* population this is shared between initial PA and ResponsePBA, especially when treated with low-light (Table. 2.2).

Table 2.2: Standardized (adjusted) linear regression coefficients of underlying components of final leaf tip angle (LTA) showing their relative importance in the two RIL populations after different treatments. All estimated coefficients were significant ($p<10^{-5}$).

	Ler x Cvi		Col x Ler		
Trait	Low-Light	Ethylene	Low-Light	Ethylene	
PA	0.81	0.94	0.64	0.63	
LSA	-0.24	-0.27	-0.35	-0.17	LTA at end of experiment
PBA	0.17	0.14	0.27	0.09	
RPA	0.28	0.34	0.39	0.35	
RLSA	-0.07	-0.08	-0.09	-0.05	
RPBA	0.30	0.17	0.62	0.32	

Global QTL distribution shows extensive natural variation

To characterize natural variation in hyponastic growth and identify major effect loci causing variation we used two treatments, ethylene and spectral neutral shade, and two RIL populations, L*er* x Cvi and Col x L*er* (likelihood of odds; LOD thresholds: supplemental information (SI) Table 2.1; all QTLs can be found in SI Table. 2.2 to 2.5). In total 9 QTL rich regions in which QTLs for multiple traits co-locate are detected (Fig. 2.5A and 2.5B). These QTL rich regions are more or less evenly spread across the chromosomes and the distribution in both populations was quite similar. The low-light experiments in the L*er* x Cvi population resulted in one (ResponseLBBA) to eight (LBBA after low-light treatment) QTLs with an average of 4.8 QTLs for all traits. The ethylene experiment with this population resulted in a lower average number of QTLs (3.2) with a range of zero (ResponseLSA) to eight (LTA after ethylene treatment) QTLs. In general, less QTLs were found in both treatments with the Col x L*er* population compared to the L*er* x Cvi population. The difference in number of QTLs found between initial, final and response QTL is dependent on the trait, but in general, response QTLs were less abundant. Most of the highly significant QTLs for initial angle traits were also found for the final angle traits in the L*er* x Cvi population. The Col x L*er* population, on the other hand, showed much more inducible variation.

While some of the QTL rich regions harbored (nearly) all measured traits, some could be attributed to specific traits. A QTL rich region that was identified in both populations was located on chromosome **2,** (L*er* x Cvi **2**;40-60 cM, Col x L*er* **2**;30-50 cM). It is involved in almost all final and response traits for both treatments. In this QTL region the L*er* allele has in general a negative effect on most angle traits; a positive effect was only detected for LSA. More specifically, in the L*er* x Cvi population the QTL regions **1**;0-35 cM, **1**;70-126 cM, **2**;60-71cM, **3**;0-10 cM and **3**; 70-80 cM are involved in PA and as a consequence in LBBA, LBTA and LTA. In the Col x L*er* population, QTLs in three comparable regions **1**;0-30 cM, **1**;70-90 cM, **3**;50-60 cM displayed much weaker effects or were none existing. Two other comparable regions, **2**;55-65 cM and **3**;0-10 cM, however, did also contain QTLs in this population. The allelic effect of L*er* in **3**;0-10 cM, however, having a positive effect in a Cvi background, had the opposite effect in a Col background. Regions with more QTLs in the Col x L*er* population were found on chromosome 4 and 5 (**4**;10-25 cM, **4**;45-55 cM, **5**;0-25 cM, **5**;70-80 cM). The two QTL containing regions on chromosome 5 are involved in the response to low-light. The first of these QTL rich regions located on the top of chromosome 5 has a large role in the response of the Col x L*er* population. In the L*er* x Cvi population only one large effect QTL for initial PBA was found on this position. In the L*er* x Cvi population the bottom of chromosome 5 (**5**;85-105 cM) contained a QTL for the response of the petiole and the response of lamina to low-light. In the Col x L*er* population this locus had a much larger effect on the response to low-light. Although, also different QTLs were found between treatments and populations, in general many QTLs were found on the same positions. We showed that extensive natural variation exists in both initial leaf orientation and hyponastic growth in response to ethylene or low-light.

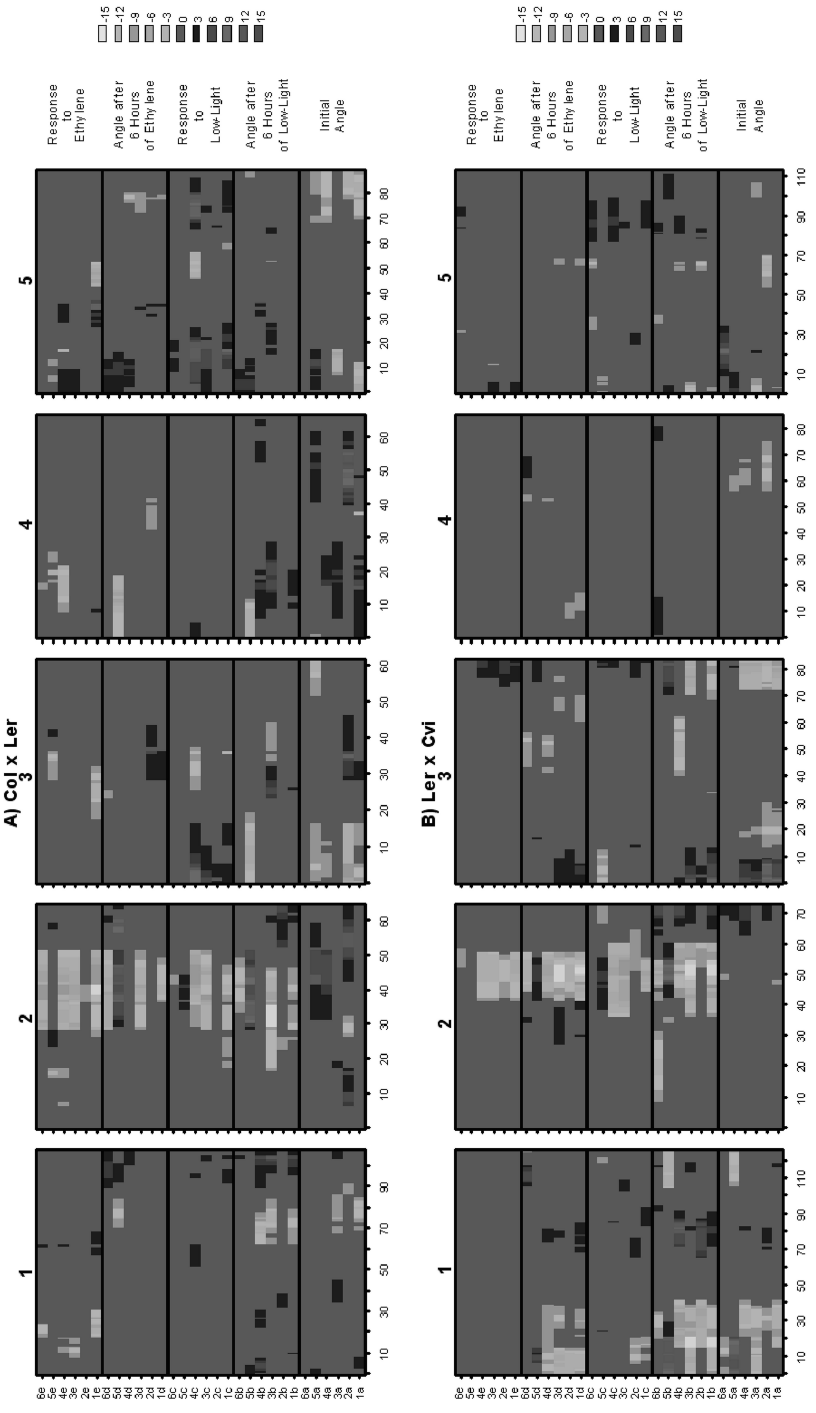


Figure 2.5 LOD profiles, multiplied by the sign of the additive effect of the L*er* allele, of six traits (Y-axis; Traits LTA, PA, LBBA, LBTA, LSA and PBA are numbered 1 to 6) initial (a), after (b & d) and response (c and e) to low-light and ethylene (Y-axis; a to e). The 5 chromosomes of Arabidopsis are numbered 1 to 5 (top of graph) and the accumulated genetic distance in cM is depicted on the X-axis. Upper graph (A) shows Col x L*er*, bottom graph (B) shows L*er* x Cvi.

QTLs for initial angles

Light quantity is one of the determinants of leaf angle under standardized growing conditions (Pierik et al., 2005); data not shown). Although all RILs were grown in a similar light quantity, different initial angles were measured for individual RILs. The clear difference between the parents for initial angle traits has a genetic base, as shown by a high H^2, and segregated in the RIL population (Table. 2.1).

Within the L*er* x Cvi population the variation in initial angle traits could largely be explained (36 to 50 %) by a number of QTLs. Five QTLs were calculated for LTA which is determined by the PA, LSA and PBA. The QTLs found for LTA are most similar to the QTLs found for PA and LSA, although LSA plays a smaller role than PA in determining the LTA. For the PBA some QTLs were found together with those of LTA, but the strongest QTL for PBA which was found on chromosome 5 was not detected for LTA.

At the top of chromosome 3, in the L*er* x Cvi population, the effect of the L*er* allele gives rise to higher angles while in the Col x L*er* population it had the opposite effect. A closely linked locus with the same L*er* effect, however, could also be the Col x L*er* counterpart.

QTLs for the initial LTA co-locate with QTLs for initial PA and LSA, but not with initial PBA. This supports our finding that the LTA is largely determined by the PA. Genetic variation for specific components determining the orientation of the leaf is present under standardized growing conditions. Major effect QTLs for a single specific leaf angle trait are present, as well as QTLs that influence the initial angles of more than one leaf angle trait.

QTLs for Response and Final angles

Two treatments, ethylene and low-light, were used to initiate hyponastic growth. Although in both treatments the response traits had a lower H^2 than the absolute angle traits, we did find QTLs for the responses. QTLs, mapping on chromosome 2 between 43 and 53 cM were present in both populations for each of the two treatments and had a similar effect. The hyponastic response was lower in lines carrying the L*er* (*erecta*) allele. Nevertheless, some differences in the magnitude of the response to the different treatments and between the two populations were present. The effect of this locus on chromosome 2 on hyponastic response is most severe in the Col x L*er* population when treated with ethylene and explained most of the variation. In the L*er* x Cvi population other QTLs with smaller effects, for the response of specific parts were found (Fig. 2.5). The region at the top of chromosome 3 contained a QTL for initial angles in both populations. In the L*er* x Cvi population these were still present as QTLs for absolute angles after six hours of low-light or ethylene. In the Col x L*er* population the QTL could only be related to the response upon low-light. The top of chromosome 5 contained many QTLs in the Col x L*er* population, both for response and final angle. In the L*er* x Cvi population only scattered QTLs for some of the traits co-located. For both populations this locus was involved in the response to low-light, but the effect was much stronger in the Col x L*er* population. The L*er* allele of this locus increased the response, mainly through bending of the petiole (PBA, ResponsePBA).

In the L*er* x Cvi population a specific locus for the response of the petiole to low-light was found on top of chromosome 1. The L*er* allele had a negative effect on the

measured traits. On a nearby region also very clear QTLs for initial and even stronger for final angle of all measured traits of the leaf were found. This and the fact that also QTLs for final angle after ethylene treatment, but not for the response to it located here suggest that this locus also has an influence on normal as opposed to induced leaf movement. Only a small, but still significant QTL was detected in the Col x L*er* population (**1**;15-25 cM) for the response of the leaf to ethylene.

QTLs for absolute angles and some of the response QTLs in the two populations were found on chromosome 5 around 90 cM (L*er* x Cvi) and 80 cM (Col x L*er*). This locus was involved in the response to low-light, through ResponsePA or ResponsePBA. In the L*er* x Cvi population some co-locating QTLs for final angle were observed, whereas in the Col x L*er* population the QTLs for initial angle were located at this position.

QTLs for response and final angles were, in both populations, located on chromosomes 1 (response in L*er* x Cvi; final angle in Col x L*er*), 3 (final angles) and 5 (response QTL). The Erecta locus or a locus very closely linked to it on chromosome 2 was found to affect the responses in both populations and treatments where it affects the response to ethylene the most. In most cases the same loci are involved in determining the initial angle as well as in determining the response.

QTLs for Leaf Shape Angle

As the petiole and lamina respond to low-light or ethylene the leaf-tip is the most distal part that changes in angle. Both petiole and laminae can cause changes in the leaf angle. A downward curved leaf, however, will have a constitutive lower leaf-tip angle than a leaf with a more flat morphology. To distinguish between QTLs for the shape- and position of the lamina we calculated the shape of the leaf (LSA). At least some of the variation for LSA was expected to have a genetic base.

In the L*er* x Cvi population, four QTLs for the LSA trait could be mapped before application of the treatments, explaining 48% of the variation found. Three QTLs increased the LSA as expected by the more curved leaves of L*er,* whereas one decreased LSA. Still the total effects of all QTLs would predict the parents to show a difference of 11 degrees in their LSA which is within the support interval of the measured difference between the parents (Fig. 2.3).

Because the mutant allele of *erecta*, carried by L*er*, makes the lamina more curved (Torii et al., 1996), the only QTL that could be expected for LSA would be on the Erecta locus. Indeed after the treatments, especially low-light, a large effect QTL was found on the Erecta locus. After the treatments the effect of the QTL showed that the L*er* allele results in ~7.5 degrees more curved lamina than its Cvi counterpart. The total effect of the QTLs, which could be mapped after treatment with low-light, resulted in a calculated difference of ~14 degrees between the parents. This was somewhat higher than the difference measured directly between them (Fig. 2.3).

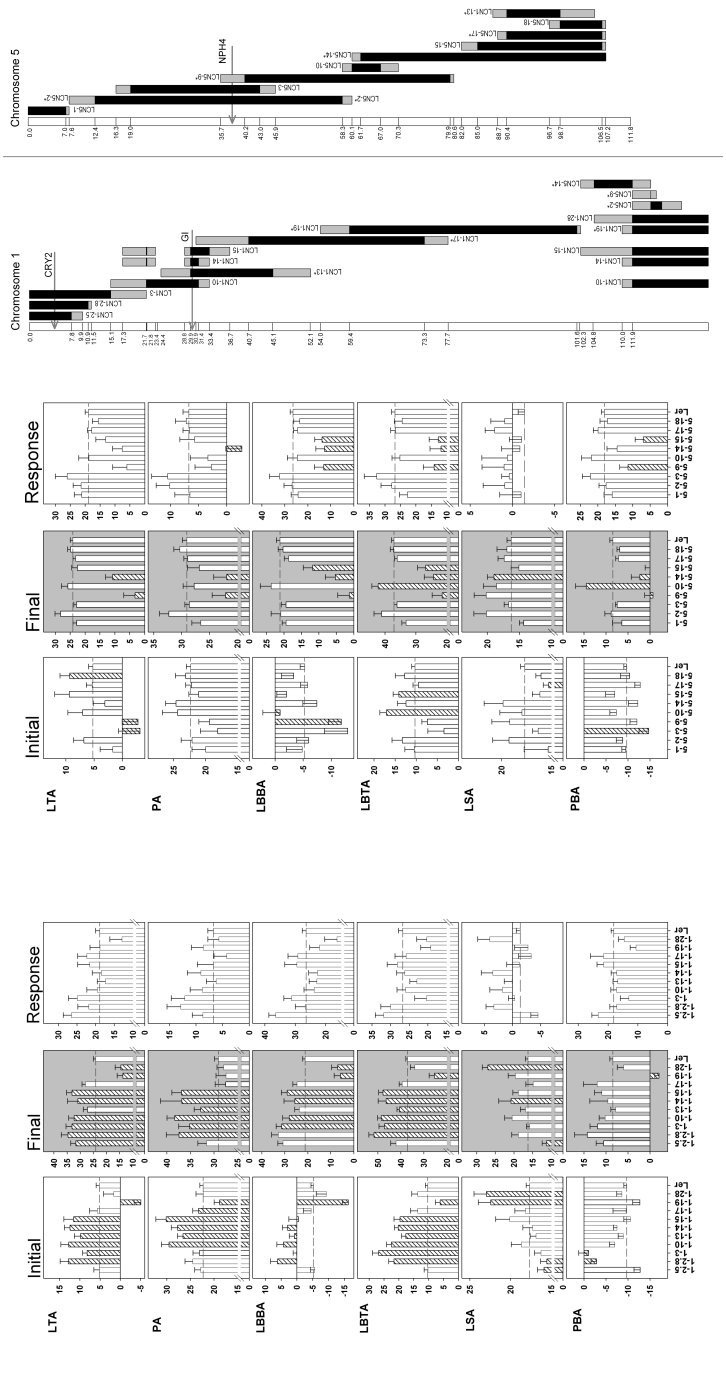


Figure 2.6 Description of QTL containing segments on chromosome **1** (left bar graphs and left segment figure) and chromosome **5** (right bar graph and right segment figure) (segment lengths in centiMorgan). Next to the chromosome NILs containing a Cvi introgression (depicted in black bars, grey bar indicate the region were the cross-over has occurred) in a L*er* background. NILs with multiple introgressions are indicated with an asterisk. Candidate genes are noted with name and arrow. The bargraphs; phenotypic values (y-axis: angles in degrees; initial angle: left; final angle: middle and grey).

Near Isogenic Lines confirming QTLs

The availability of a large number of Near Isogenic Lines (NIL) (Keurentjes et al., 2007), possessing a Cvi introgression in a L*er* background, enabled us to confirm several QTL positions that were detected with the L*er* x Cvi RIL. Most major effect QTLs were supported by a NIL. Furthermore, NILs with partially overlapping introgressions were used to decrease the QTL size and thus the number of candidate genes. This could be done for two QTLs, one on chromosome 1 and one on chromosome 5 (Fig. 2.6).

NILs confirm QTLs on chromosome 1 for Leaf shape (LSA) and absolute angles (initial and final)

The QTLs detected in the L*er* x Cvi population on the top of chromosome 1 (**1;**0-35), could be split into two groups. Traits that describe leaf shape (LSA) and those that determine absolute angles of leaf, lamina and petiole. The phenotypes, before treatment, of the partially overlapping NIL showed that these two groups of QTLs are indeed caused by two separate loci. The NIL LCN1-2.5, 1-2.8 and 1-3 showed a lower LSA, while the flanking NIL LCN1-10, 13, 14 and 15 showed a LSA similar to L*er*. PA on the other hand was higher in NIL LCN1-10, 13, 14 and 15 than L*er* (Fig. 2.6). The response QTLs could not be confirmed because none of the NILs showed a significantly different response, although a tendency was found to a larger response in LCN1-2.5 and 1-2.8.

The QTL for LSA (**1**;0-20cM) co-located with a QTL for flowering time, described by Alonso-Blanco *et al*. and El-Assal *et al*. El-Assal et al., 2001) of L*er* and Cvi . They showed that this is caused by an allelic difference of *CRY2*, by testing a transgenic line containing the Cvi *CRY2* allele in a L*er* background (L*er*T4-CRY2-Cvi; (El-Assal et al., 2001)). When we tested this transgenic line a significantly lower LSA was found when compared to L*er*. This shows that CRY2 is also involved in determining LSA, where the L*er* allele is responsible for more curved leaf-blades. The involvement of CRY2 and light intensity on LSA became even clearer when plants were grown under a higher light intensity (400 µmol m^{-2} s^{-1}); the leaf-blades of the transgenic line had an almost completely flat appearance whereas the leaf-blades of L*er* remained curved (Fig. 2.7).

Figure 2.7 Leaf phenotype of L*er* (right) and L*er*T4-CRY2-Cvi (L*er* transformed with the *CRY2* allele of Cvi; left) when grown at 400 µmol m^{-2} s^{-1} light.

Indirectly CRY2 is also involved in LBBA and LTA. QTLs of the traits which, together, determine LSA, LBBA and LTA, have allelic effects in the opposite direction.

The QTLs for absolute leaf-, lamina- and petiole angles (**1**;20-35cM) could be confirmed by several NILs (LCN1-10, 13, 14, 15) (Fig. 2.6). Together these NILs indicate that a small introgression positioned around 8.0 Mbp from the top of the chromosome must

contain the gene causal to these QTLs, resulting in a much more upright position of the leaves of the NILs containing a Cvi introgression at this position. A second locus involved in the absolute angles could be located around 21 cM from the top, but smaller and single introgression NILs are necessary to investigate this, since the flanking QTLs for other traits can influence the QTLs under study.

QTL region **1**;70-126 contained a QTL for LSA. This QTL could be confirmed by a NIL with a single introgression at the end of chromosome 1 LCN1-28 which had a higher LSA than L*er*. Other NILs with an introgression at this region showed a tendency to a higher LSA. Only, LCN1-19 before treatment and LCN1-14 and 5-14 after treatment had a significantly higher LSA. This could possibly be caused by the second or third introgression of all these lines except LCN1-28. The QTLs for LTA and PA was confirmed by the lower angle of LCN1-19.

By comparing the phenotypes of the NILs on chromosome 1 with L*er* we could narrow down loci of several QTLs. The phenotypic difference and resulting QTLs for LSA were caused by the *CRY2* polymorphism between L*er* and Cvi. The QTL for the initial angle of traits influenced by PA could be placed on a small ~5cM locus. Furthermore, QTLs for initial angle (increased by the L*er* allele) and LSA (decreased by the L*er* allele) were confirmed by LCN1-19 and 1-28, respectively.

NILs confirm QTLs on chromosome 5 involved in the response to low-light

Chromosome 5 contained a large effect QTL (**5**; 85-100) for the response of petiole and lamina to low-light (Fig. 2.5). The QTL and direction of the additive effect could be confirmed by LCN 5-14 and 5-15 (Fig. 2.6). In both these NILs the petioles did not respond, whereas a small response was observed for the leaf blades. NILs flanking LCN5-14 (LCN5-9, 5-10, 5-17 and 5-18) all showed a hyponastic response to low-light. A QTL for PBA that was only found at the initial measurement was confirmed by LCN5-3 which had a significantly lower initial PBA. In LCN5-9 a difference in the final and response PBA was observed when compared to L*er*, probably caused by the same QTL. A candidate gene for this PBA QTL, *NPH4* (Stowe-Evans et al., 1998) was located on the Cvi introgression of LCN5-2, 5-3 and between the markers identifying the upper border of the Cvi introgression in LCN5-9. When testing the mutant, *nph4*, we observed a small difference in initial PBA when compared to the wildtype Col. Interestingly, when we studied the response to low-light of the *nph4* mutant it responded to a much lesser extent through PBA compared to Col (Fig. 2.8). The petioles, however, responded to a similar extend. The difference in PBA response between *nph4* and Col was larger than the QTL study would predict. Even though the exact genetic difference causing the QTL remains unknown, *NPH4* is a good candidate and shows that the two locations of differential growth are at least in part controlled by different genes.

By testing NILs on chromosome 5 we could narrow down a QTL for PBA to the genetic locus of 35.7 to 45.9 cM (=10.2 cM) and a QTL for the response to low-light could be placed in the region 82.2 to 90.4 cM (=8.2 cM).

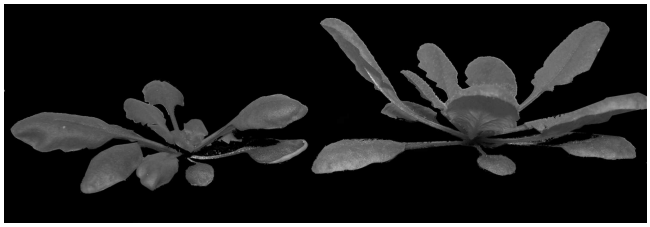


Figure 2.8 Phenotypic difference after 6 hours of low-light treatment between, *nph4* (left) and wild type Columbia-0 (right), showing the more responding laminas (distal region of differential growth) of Columbia-0.

Other NILs confirming QTLs

Several more QTLs could be confirmed by single NILs (Table. 2.3). The QTLs in region **2**;40-60 were validated by LCN2-7 (~9.0 to ~ 12.0 Mbp.). It harbors the *ERECTA* gene which is known to be different between L*er* and Cvi and is therefore a likely candidate for the gene causing the allelic difference. *ERECTA*, non-functional in L*er*, is substituted in this NIL by the functional copy of Cvi. No significant differences with L*er* for any initial angle traits where found thereby supporting the absence of a QTL at this locus. After treatment with neutral shade, QTLs were found for all traits with high LOD scores and large effects. These QTLs where confirmed by the phenotype of LCN2-7 compared with L*er*.

The phenotype of LCN 3-1 confirms two QTL rich regions, **2**;60-71 and **3**;0-10. This NIL has two introgressions, at the bottom of chromosome 2 (~17.2 to ~19.2 Mbp) and the top of chromosome 3 (0 to ~0.70 Mbp) that both contain QTLs with similar effect. This results in a NIL significantly different for all measured traits before and after treatment, having more curved leaves as well as petioles and laminae with a lower initial and final angle than L*er*.

At chromosome 4 the QTL for LTA, PA, LBBA and LBTA were confirmed by LCN4-6 (~10.4 to ~12.8 Mbp) which has higher angles for these traits independent of the treatment.

Table 2.3: NILs confirming QTLs, Traits significantly different (2-sided t-test; $p<0.05$) from L*er* are in bold. Cvi introgressions in a L*er* background: LCN2-7 (Chr2: ~9.0 to 12.0 Mbp.), LCN3-1 (Chr2: ~17.2 to ~19.2 Mbp.) & (Chr2: ~0.00 to ~0.70 Mbp.), LCN4-6 (Chr4: ~10.4 to ~12.8 Mbp.)

	Initial					
NIL / Trait	LTA	PA	LBBA	LBTA	LSA	PBA
L*er*	5.21	22.31	-5.26	10.31	**17.67**	-9.64
LCN2-7	7.96	22.63	-2.50	13.02	15.52	-9.61
LCN3-1	**-8.93**	**11.90**	**-19.59**	**2.56**	**24.14**	**-6.64**
LCN4-6	**12.60**	**31.23**	**2.72**	**23.35**	**20.63**	-7.88
	Lowlight					
L*er*	**24.11**	29.05	**21.16**	37.08	**16.25**	8.44
LCN2-7	**34.65**	**32.42**	**33.38**	**45.54**	**12.15**	**13.12**
LCN3-1	**0.56**	**16.00**	**-5.92**	**19.42**	**25.33**	**3.42**
LCN4-6	**34.25**	**38.15**	**31.23**	**48.40**	17.17	10.25

Discussion

Hyponastic growth

We studied the process of hyponastic growth of the whole leaf as induced by ethylene or low-light. A detailed description of this differential growth process was developed to study the underlying genetics. Natural variation for hyponastic growth of the petiole upon ethylene or low-light treatment is present among Arabidopsis accessions as shown by Millenaar et al. (2005). We used two different RIL populations, Col x L*er* and L*er* x Cvi and found natural variation for the initial angle and the response of the petiole to the treatments of the parental accessions of the two RIL populations thus confirming the result of Millenaar et al. (2005). By studying hyponastic growth of the whole leaf we could elucidate a strategy to escape unfavorable environmental conditions by hyponastic growth in more detail. In Arabidopsis, hyponastic leaf growth results from differential growth in two regions: the basal part of the petiole and the basal part of the leaf blade (lamina). Both petiole and lamina respond, partially independent, to ethylene and low-light. Together, the differential growth responses of these growth regions lift the tip of the lamina. We distinguished 6 leaf angle traits: leaf tip angle, petiole angle, leaf blade basal angle, leaf blade top angle, leaf shape angle and petiole bending angle. By a linear regression analysis we showed the relative role of these traits in determining the final orientation of the leaf (LTA). The initial PA and the response traits ResponsePA and ResponsePBA were explaining most of the variation.

Genetic regulation of two regions exhibiting differential growth

A remarkable difference in the nature of response was found between low-light and ethylene treatments (Fig. 2.4). In both RIL populations the relative importance of the ResponsePBA was less during ethylene treatment, compared to the low-light treatment. This indicates that the involvement of cell elongation in the most basal and distal parts of the petiole are regulated differently. It was shown that lamina and petiole can be regarded as different organs with respect to growth (elongation) (Tsukaya et al., 2002; Millenaar et al., 2005). Millenaar et al. (2005) shows that there is no significant correlation between the effect of ethylene on the elongation of the petiole and leaf-blades. For the response to light this difference was shown by Tsukaya et al., 2002. Apart from their description of the difference in response of lamina and petiole to blue light, red/far-red ratio and darkness, they also describe the mutant gene *acaulis2* (*acl2*) that affects cell size exclusively in the petiole, confirming this difference between petiole and lamina on a genetic level. Furthermore, a mutation in *PHYB* was described to have an opposite effect on the cell sizes of the leaf-blade (smaller) and the petiole (longer). They conclude, from a genetic perspective, that petioles and leaf blades can be regarded as different organs (Tsukaya et al., 2002). In Arabidopsis we also found a difference between the basal and distal part (the part of the leaf where the petiole gradually turns into the lamina) of the petiole in hyponastic growth and QTLs found for the different parts. From this perspective the basal- and distal part of the petiole can also be regarded as (partially) different organs. These provide specific locations for differential growth leading to hyponasty. The involvement of AUX/IAA factor NPH4 (ARF7) in differential growth of the distal region of the petiole as a response to low-light suggests that auxin involvement may be different between the two

regions of differential growth leading to hyponasty. Either auxin is much less important for the basal part of differential growth or the auxin signal is gated via other AUX/IAA factors than NPH4. The inability of the laminae of *nph4* to respond to low-light resembles the wild-type (Col) response to ethylene in which the laminae also hardly respond. It is therefore tempting to speculate that ethylene has a negative influence on the function of NPH4. Enforcing this speculation are the observed influence of ethylene on NPH4 by Harper *et al.* (Harper et al., 2000) who found that the reduced gravitropic curvature of *nph4* could be recovered by ACC application.

Identity of LSA QTLs; *ERECTA* and *CRY2*

A large influence of the Erecta locus on leaf shape (LSA) was found and is very likely to be caused by ERECTA. This gene is different between L*er* and Cvi because of the mutation in the L*er* allele (Torii et al., 1996). They describe that the *erecta* mutation causes a defect in the elongation of plant organs, including the petioles and leaf blades. A NIL (LCN2-7, (Keurentjes et al., 2007) with an introgression that harbors the Cvi allele of *ERECTA* in a L*er* background has leaves with much less curvature than L*er* and confirms that *ERECTA* is a very likely candidate for this difference in LSA.

A QTL study on natural variation in flowering time of the L*er* x Cvi RIL, lead to the identification of an allelic difference in the *CRY2* gene between L*er* and Cvi (El-Assal et al., 2001). The Cvi allele of *CRY2* causes earlier flowering especially noticeable under short days. When L*er* plants where transformed with the *CRY2*-Cvi allele they flowered much earlier. Because of this known allelic difference between the parents and the co-location of a QTL we investigated the possible role of CRY2 in LSA. Co-location of a LSA QTL with a cloned flowering time QTL revealed a new phenotypic effect of the allelic difference between the CRY2 gene of L*er* and Cvi. When we tested the same transgenic L*er*::CRY2-Cvi plants for LSA we found that these plants had laminae with much less curvature when compared to L*er*. Furthermore, the effect on flowering time could still be observed as well. The flat leaf-blade phenotype due to the Cvi CRY2 allele was visible even better when plants were grown under 400 µmol m^{-2} s^{-1} (Fig. 2.7), showing the involvement of light intensity on LSA. Moreover, the laminae became almost totally flat; pointing out that the CRY2-Cvi allele can almost completely complement the effect on LSA of the L*er* type Erecta locus, which makes the laminae more curved.

Co-locations with QTLs for other traits

In Arabidopsis, QTLs have been detected for a wide range of traits. Allelic variation was found for many traits ranging from morphological to life history traits (Koornneef et al., 2004). This enabled us to compare these to our own study for co-location of QTLs. By doing this we could place hyponastic growth in a better biological framework and identify or predict associated processes. From all studied traits studied, we choose the obvious ones to compare with our study. These include movement of leaf-blades and petioles or treatment with various light (quality or quantity) conditions and/or hormone treatments.

Co-location with circadian rhythm QTLs

Circadian rhythm in leaf movement is observed for many Arabidopsis accessions (Michael et al., 2003). The change in angle of the leaves, petioles and laminae that are observed in hyponastic growth could very well share similar genetic components (QTLs) since it is a phenocopy of the circadian movement, suggesting a (at least partly) similar mechanism. Several QTL studies have been performed on a number of aspects of circadian rhythm in the Col x L*er* or L*er* x Cvi population through rhythmic movement of the leaves (Swarup et al., 1999; Michael et al., 2003; Edwards et al., 2005) or rhythmic expression of a reporter gene (Darrah et al., 2006).

Four hyponastic growth QTL rich regions (L*er* x Cvi: **1**;0-35 cM, **2**;40-60 cM, **3**;0-10 cM, **5**;85-105 cM) showed considerable co-location with QTLs identified for amplitude, period or phase of the circadian rhythm. Especially the co-location of QTLs on chromosome 1 (**1**;0-35 cM) and 2 (**2**;40-60 cM) were of particular interest because of the candidate genes thought to cause the natural variation in circadian rhythm are also candidates for the variation in hyponastic growth.

The co-location of circadian rhythm QTLs with QTLs for hyponastic growth could be caused by genes involved in hyponastic growth normally, rhythmically expressed by the clock or components of the circadian clock through which the hyponastic response is gated or modulated. On chromosome 1 the co-location on the *GIGANTEA* locus suggests at least the latter, since it is involved in the rhythm of leaf movement (Fowler et al., 1999) and rhythmically modulated itself (David et al., 2006). An independent role of the clock, however, cannot be ruled out; especially since recently multiple functions of GIGANTEA have been found acting independently of the circadian clock (Cao et al., 2005; Martin-Tryon et al., 2007; Oliverio et al., 2007).

On chromosome 2 Swarup et al. (Swarup et al., 1999) showed that the NOTROPPO (NOT) QTL was caused by an allelic difference in the ERECTA gene, where the L*er* allele lengthened the period. However, a QTL for phase of CAB:LUC expression was found by Darrah *et al*. (Darrah et al., 2006) but they could not confirmed it by either a NIL or Wt (Landsberg, Lan-0). Michael et al. (Michael et al., 2003) identified a QTL on chromosome 2 in the Col x L*er* population for the amplitude of the circadian movement of the leaves. No clue was given by the authors whether the difference in the *ERECTA* gene could have caused this.

Altogether, the allelic variation of *ERECTA*, the co-location with the hyponastic response QTLs and identification of *ERECTA* to be involved in the period of the circadian rhythm, makes *ERECTA* a very likely candidate for the variation between the RILs observed during hyponastic growth.

Co-location with light- and hormone response QTLs

Light and hormone-action play and important role in hyponastic growth (Pierik et al., 2005). Genetic variation in light and hormone sensitivity could result in differences in hyponastic growth between RILs. An extensive study on variation in hypocotyl growth as response to the hormones gibberellic acid (GA) and brassinosteroids (BR) as well as different light treatments was performed by Borevitz et al., 2002 using the L*er* x Cvi population.

QTL region **1**;0-35, co-located with Light 1. Borevitz et al. (2002) found this to be a major locus for variation in hypocotyl length in response to light quality or gibberellic acid treatment in the L*er* x Cvi population. This indicates that growth responses of the hypocotyl as a result of the light quality act through the plant hormone GA. For the other plant hormone (BR) they tested, no QTL was found for the absolute hypocotyl length on this position. When we re-analyzed the data, however, and subsequently calculated the QTL for the difference in hypocotyl length between white light the BR treatment (data not shown) a highly significant QTL was found on this position. This indicated that BR may also play a modulating role in light-controlled growth responses, like hyponastic growth.

Hyponastic growth QTL region **2**;40-60 co-located with the QTL Hypocotyl 2 (HYP2) (Borevitz et al., 2002). They show that the variation in hypocotyls length is caused by the different alleles of *ERECTA* for BLUE, DARK and BR treatments. Therefore *ERECTA* is also a very likely candidate gene for the hyponastic growth QTL at this locus. Our re-analysis of the difference in hypocotyls length between white light treated and far-red or GA treated hypocotyls identified two other QTLs at this locus. QTLs for the length difference of far-red treated hypocotyls were found on the left border of this locus involved in hyponastic growth. Interestingly, another highly significant QTL for the far-red difference was found near the bottom of chromosome 2. Variation in length of hypocotyls treated with GA lead to a QTL which shared the entire locus with the QTL for hyponastic growth on this position. LIGHT 2, the QTL rich region found in the study of Borevitz et al. (2002) is on the edge of QTL region **2**;40-60 and contains *PHYB*. This is a well known gene coding for an important light-receptor and is allelic between L*er* and Cvi. Borevitz *et al.* (2002) sequenced *PHYB* from L*er* and Cvi and considerable nucleotide variation in the promoter region as well as synonymous and replacement changes in the coding region were found. Quote “It is surprising that a photoreceptor like PHYB could be a candidate for a QTL because null mutants have deleterious and enormous effects throughout their development. If this allele of *PHYB* is the LIGHT 2 QTL it must only affect a subset of downstream processes controlled by PHYB” (Borevitz et al., 2002). The hyponastic phenotype and elongated petioles of the several PHYB knockouts, however, make this a good candidate gene for our hyponastic growth QTL.

Further co-location was found for QTLs for individual treatments like, FARRED2 at the end of chromosome 2, RED3 at the end of chromosome 3 and possibly BLUE4 at ¾ of chromosome 4.

Conclusions

Natural variation was found for hyponastic leaf movement in response to independent low-light and ethylene treatments. This hyponastic growth of the leaf is the result of differential growth in two regions of the petiole, the most basal part and the basal part of the lamina where the petiole gradually converts into the lamina. Several QTLs were shared by the two regions of differential growth, but not all. These unshared QTLs, show that basal and distal part of the petiole can be regarded as different regions for differential growth both contributing to hyponasty. More genetic evidence for two different locations of differential growth in the petiole came from the *nph4* (Stowe-Evans et al., 1998) mutant. The petioles of this mutant respond normally with respect to the basal part, but hardly bend

when subjected to low-light. The altered response of the *nph4* mutant also suggests the involvement of auxin in low-light-induced hyponastic growth. Furthermore, it strengthens the evidence for the genetically separable regions of differential growth at the basal and distal part of the petiole.

The Erecta locus is the most important locus shared between the hyponastic response to ethylene and low-light. The large effect of this locus resulted in less QTLs for the response to ethylene than for the response to low-light. Not only determined the Erecta locus the severity of the differential growth response, also leaf shape was influenced., The Erecta locus was, however, not the only locus influencing the leaf shape. We present evidence that the QTL for LSA at the top of chromosome 1 is caused by the known allelic difference of *CRY2* between L*er* and Cvi.

The QTLs found for the low-light treatment for initial, final and response angles were compared to other QTL studies. Co-locating QTLs show that hyponastic growth shares QTLs (genetic components) with the circadian rhythm and its temperature compensation mechanism as well as the light sensory mechanism of hypocotyls growth.

The diversity of processes in which genes, and very likely candidates, identified in this study like, CRY2, NPH4, ERECTA, GIGANTEA and PHYB are involved show that the complex nature of hyponastic growth as shown by hormone interaction studies extents to gene regulation. The large amount of natural variation in phenotypes (and alleles) indicates that hyponastic growth is an important adaptive trait of Arabidopsis.

Materials and Methods

Plant material and Growth conditions

Two *Arabidopsis thaliana* Recombinant Inbred Line (RIL) populations were used, L*er*-1 x Cvi-0 (Alonso-Blanco et al., 1998) and Col-0 x L*er*-1 (Alonso-Blanco et al., 1998). Plants were germinated on wet filter paper, grown on a mixture of soil and perlite at 20°C, 70% (v/v) relative humidity, 9-h photoperiod (short day photoperiod, 200 µmol m^{-2} s^{-1} photosynthetically activated radiation photon flux density) as described in Millenaar *et al.* (Millenaar et al., 2005), with the following modifications. Because of the large variation in flowering time in the L*er* x Cvi RILs (Alonso-Blanco et al., 1998; Alonso-Blanco et al., 1998) in short days, due to the presence of the Cvi *CRY2* allele (El-Assal et al., 2001), the experiments started 21 days after transplantation just before the earliest flowering RIL started to bolt. Water was given manually 3 times per week. Since hyponastic growth occurs to a similar extent over an extended period of the plants life (Millenaar et al., 2005) Near Isogenic Lines (NILs) could be tested 28 days after transplantation.

Experimental setup

The low-light experiment on the L*er* x Cvi population (156 RILs of 162 RILs) was split into 4 groups of approximately 50 RILs with 10 RILs overlapping between experiments. This experiment was done twice and contained two to four plants per RIL. In this way data of 3 to 12 individuals per RIL were collected. The low-light experiment on the Col x L*er* (88 RILs) population was split into 3 groups of approximately 40 RILs with 10 RILs overlap between experiments. Each group contained 2 to 6 plants per RIL. In this way data of 3 to 11 individuals per RIL were obtained.

For the ethylene application experiment on the L*er* x Cvi population, 4 groups of 50 semi-randomly chosen RIL were made in which 10 RILs were overlapping between experiments. This experiment contained 3 to 4 plants per RIL. Two to six individual measurements per line were obtained. The ethylene experiments on the Col x L*er* population had 3 separate groups of about 30 RILs per experiment. Of these groups at least 15 RILs were overlapping between experiments. These experiments contained 2 plants per RIL, 2 to 4 individual measurements per line were obtained.
From all experiments means of the individual RILs were used in QTL analysis. Of the parental lines 1 to 6 individuals were measured in every experiment. NILs were tested using the same conditions as used for the RILs.

Plant trait quantification

The angles of the third and fourth youngest petiole and leaf blade (6-8 leaves < 5mm total) were measured before and after 6 hours of treatment. This was realized by photographing plants with a digital camera (Casio QV-3500EX). The angles were measured with the line tool of ImageJ (http://rsb.info.nih.gov/ij/). The angle of specific points with the horizontal (left and right side of the plant; described in Fig. 2.2) were measured and the average was used as the value for an individual plant. These directly measured angles (petiole angle (PA), leaf tip angle (LTA), leaf blade basal angle (LBBA) and leaf blade top angle (LBTA)(Fig. 2.2 A - D) were used to calculate the traits: leaf shape angle (LSA) and petiole bending angle (PBA) (Fig. 2.2 E,F and results).

Plant treatments

Shading treatment was as described by Millenaar *et al.* (Millenaar et al., 2005), with the following modification: the reduction in light-quantity was achieved by shading cloth only. Plants were subjected to a 6 h spectrally neutral shade resulting in a light intensity of 15 ± 5 μmol m^{-2} s^{-1} (~10% of the growing conditions).

Treatment with 5 μL L^{-1} ethylene was done as described in Millenaar *et al.* (Millenaar et al., 2005), with the following changes: per cuvette (30 x 20 x 30 cm, 18 L) 10 plants were treated. The ethylene application was performed in a flow through system with a flow rate of 10 L min^{-1}. The ethylene concentration was checked regularly by gas-chromatography.

Data analysis

All traits were found to have a normal or near to normal distribution (see supplemental Fig. 2.1 t/m 2.4) and were used in QTL analyses without mathematical transformation. None of the traits show a distinct binary distribution and therefore multiple controlling loci are expected for each trait. The Arabidopsis genetic map for the L*er*-Cvi and the Col-L*er* RIL populations was obtained from the NATURAL website (http://www.dpw.wau.nl/natural/). The core map of both populations was used in initial QTL mapping. For fine mapping available markers where added up to 1 per cM. Composite interval mapping (CIM) in QTL Cartographer (Wang et al., 2007) was utilized for all QTL mapping. As settings for CIM, the forward and backward regression method and a window size of 7 cM were used. Genome-wide thresholds per trait were estimated in QTL Cartographer using 1000 permutations (supplemental Table. 1). For clarity, all thresholds

(which were very similar) were averaged per population per treatment to determine the QTLs. When QTLs were overlapping they are called co-locating and they were treated as located in nearby-region when the QTL borders are within 5 cM of each other.

Statistical analyses were performed in SPSS (Release 12.0.1; 11Nov 2003, SPSS inc.), and Excel 2002 (Microsoft). Broad sense heritability (H^2) as the proportion of the total variance, explained by the between line effects was calculated using the general linear model function of SPSS. SPSS was also used to calculate the adjusted linear regressions coefficients in determining the relative role of specific components determining final leaf-tip angle. In the linear model used the final leaf-tip angle was dependent on six factors. These, initial petiole-, leaf shape- and petiole bending angle (PA, LSA and PBA) and their response (ResponsePA, ResponseLSA and ResponsePBA) were compared by their adjusted regression coefficients.

Chapter 3

Genetic architecture of *Arabidopsis* rosette compactness reveals a functional allelic difference in *PHYTOCHROME B*.

L. Basten Snoek, Martijn van Zanten, Hendrik Poorter, Laurentius A.C.J. Voesenek, Anton J.M. Peeters

Abstract

Rosette plants like Arabidopsis have to prevent shading of leaves by its own new developing leaves because of its flat radial growth. One of the main features of a developing rosette with respect to "self-shading" of leaves is the compactness of the rosette. This involves the length of the petiole and the shape of the laminae for which natural variation has been found. However, the way in which these traits determine the architecture of the whole rosette and the influence of the rosette shape on petiole length and leaf-shape remains largely unknown. By QTL analysis using the Landsberg *erecta* (L*er*) x Cape Verde Islands (Cvi) RIL population, we present the genetic framework of rosette compactness and rosette growth. Moreover, a functional allelic difference between the *PHYTOCHROME B* allele of L*er* and Cvi was identified. The use of rosette compactness to describe plant morphology is validated by the identification of the well-known polymorphic gene *ERECTA* to be involved in rosette compactness. Furthermore, we show evidence for feedback regulation of the genetically determined compactness on petiole length and leaf-shape.

Introduction

To escape unfavorable and find optimal light conditions plants use elongation of specific organs like stems, petioles and to a lesser extent laminae (Franklin and Whitelam, 2005). Rosette plants like Arabidopsis, missing an elongated stem at the vegetative phase, are mainly dependent on petiole elongation for shade avoidance. Because of its flat radial growth a rosette plant tends to prevent shading of leaves by its own new developing leaves. This involves the petiole that by increase in length and/or differential growth leading to hyponasty can prevent much overlap. Obviously, the shape of the laminae is an important factor too, since broad laminae overlap much earlier than small ones. Natural variation in leaf shape and petiole length has been found (Perez-Perez et al., 2002; Juenger et al., 2005) as well as total leaf number until bolting (Clarke et al., 1995; Alonso-Blanco et al., 1998). However, the way in which these traits determine the architecture of the whole rosette and the influence of the rosette shape through self-shading on petiole length and leaf shape remains largely unknown.

The development of the whole rosette can be described as a function of its underlying traits. One of the main features of a developing rosette with respect to "self-shading" of leaves is the compactness of the rosette. Two traits whose interaction makes up rosette compactness are petiole length and laminae shape. These two traits can be measured by the maximum feret diameter (fermax). The fermax is the longest strait line that can be drawn between two points on the edge of a (non-circular) object. Together, the area covered by the rosette and fermax determine compactness. For example a plant that covers a small area and has a large fermax becomes less compact. Fermax and rosette area and therefore compactness are influenced by light and have a clear genetic base. This is reflected in the phenotypes of a mutant of the well studied photoreceptor PHYTOCHROME B (Reed et al., 1993) and the variation found in leaf shape in an Arabidopsis recombinant inbred line population (Perez-Perez et al., 2002; Juenger et al., 2005). The involvement of a photoreceptor shows that light (perception) is an important factor in determining rosette compactness. Furthermore, it can also be altered by changes in hormone production/signaling (Tsukaya et al., 2002) or the interaction of hormones and light (Tsukaya et al., 2002; Hisamatsu et al., 2005).

Variation in rosette compactness may involve some genetic components determining hypocotyl length. For hypocotyl length of seedlings grown under light of different wavelengths, natural variation was found in *Arabidopsis* (Borevitz et al., 2002). However, evidence exists that growth and elongation of petioles and laminae is governed by their own, largely specific, genetic program (Tsukaya et al., 2002). Components at the start of a signal transduction pathway, for example the photoreceptors, will influence the length of the hypocotyls as well as that of the petiole and even the shape of the laminae. Further downstream, genes will probably play a more specific role, for example by determining specifically petiole length but not the laminae shape or hypocotyl-length.

The identified natural variation for light sensitivity in the Landsberg *erecta* (L*er*) x Cape Verde Islands (Cvi) RIL population (Borevitz et al., 2002) suggested that in this population natural variation for rosette compactness may also exist. Furthermore, it provides an excellent tool to identify alleles involved in petiole length and laminae shape, both underlying rosette compactness.

By QTL analysis using the L*er* x Cvi RIL population (Alonso-Blanco et al., 1998), we identified the genetic framework of rosette compactness, (relative) rosette growth and evidence for feedback regulation of the genetically determined compactness on petiole length and leaf shape. Moreover, a functional allelic difference between the *PHYB* allele of L*er* and Cvi was found. Our results are validated by the identification of the well-known polymorphic gene *ERECTA* which is known to influence leaf shape (Torii et al., 1996; Xu et al., 2003; Shpak et al., 2004) and therefore is involved in rosette compactness. A mutant (NON PHOTOTROPIC HYPOCOTYL 4) and an introgression line of gene MDB9 (At1g01460) with altered rosette compactness are identified and described further specifying the underlying components that make up rosette compactness.

Results

Natural variation in rosette compactness

Parental phenotypes

To determine the extent of natural variation in rosette compactness in the segregating Landsberg *erecta* (L*er*) x Cape Verde islands (Cvi) population, the fermax and area of the rosette were measured at two different time points, three days apart, to measure the expansion of the rosette fermax by the same petiole(s) in time. We compared the phenotypic distribution of the Recombinant Inbred Lines (RILs) and their parents (Fig. 3.1). Variation in phenotypes was found for maximum feret diameter (fermax), rosette area and compactness. The underlying components of compactness, fermax and rosette area, showed transgression at both time points. Of the parents, L*er* had a bigger rosette area, whereas Cvi had a larger fermax (a larger diameter). This resulted in different compactness of the two genotypes. A plant that covers a small area and has a large fermax becomes less compact. In contrast to the transgression observed for the underlying traits, no obvious transgression was found for the resulting compactness. This indicates that Cvi has the majority of the alleles with negative effect (and L*er* the positive) of genes (loci) that play a role in determining the compactness of the rosette. The difference between L*er*, closed rosette (high compactness) and Cvi, open rosette (low compactness) becomes especially clear when the leaf-blades of L*er* start to overlap. This occurs much earlier and more severely in L*er* than in Cvi but somewhat later than the two observation points of the RIL study (personal observation).

Heritability

To assess the contribution of the genetic background we calculated the broad sense heritability (H^2) (table 3.1). At both time-points rosette area and fermax showed a H^2 of around 50%, but this was always higher at the second time point. This was also found for compactness with approximately 50% at time-point one (t=18d) and increases to almost 65% at time point two (t=21d). This suggests that the traits, especially rosette compactness, are under increasing control of the genotype. The fact that the rosette compactness has the highest H^2 suggests the adjustment of the petiole- and laminae morphology to a (partly) predefined rosette compactness.

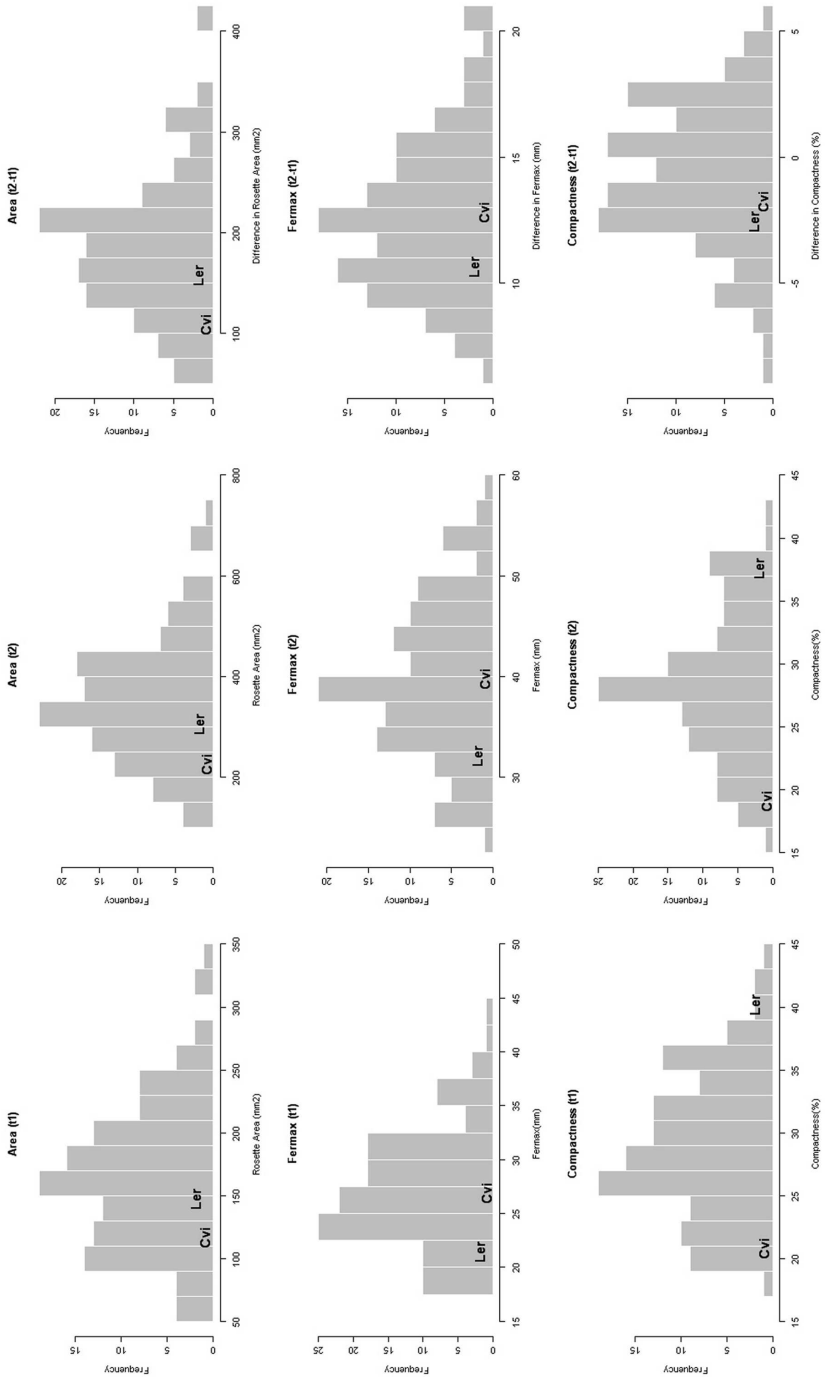


Figure 3.1 Phenotype distribution within the recombinant inbred lines (RILs) and parents of the L*er* x Cvi population. Rosette area, fermax and compactness at both time-points (TI= 18 days after transfer to soil and TII = 21 days after transfer to soil) are displayed as well as the difference between the two time-points (t2-t1). Approximate parental phenotypic values are depicted by their names.

Table 3.1 Broad sense heritability (H^2), as percentage of total variation in phenotypes that can be explained by the genotypes within the recombinant inbred line population.

Time point	I	II
AREA	**50**	**53**
FERMAX	**46**	**51**
COMPACTNESS	**53**	**64**

Genetic architecture of rosette compactness

To isolate important loci involved in rosette compactness we correlated the phenotypes with the genotypes by Quantitative Trait Locus (QTL) mapping (Fig. 3.2). We used both time points to investigate the influence of rosette growth on rosette compactness. Because of the differences in phenotypes between the parents as well as the distribution of phenotypes in the RILs and the high H^2, QTLs were expected and indeed identified for all traits. In general, strong QTLs were found on chromosome 5 for all traits and on chromosome 2 for compactness specifically.

Rosette area and fermax

We first compared the underlying components of rosette compactness, i.e. rosette area and fermax. These were measured at two time points, three days apart, to determine rosette growth. For rosette area two major (I;**5**;35-48 (I = time point one=18 days after transfer to soil; II = 21 days after transfer to soil), **5** = chromosome 5, 35-48 = 38 to 48 centiMorgan), II;**5**;36-49 and I;**5**;77-83, II;**5**;77-83) and two minor QTLs (I;**4**;49-55, II;**4**;48-58 and I;**5**;58-68, II;**5**;59-69) were found at similar positions at both time points (Fig. 3.2; Table 3.2). The two major QTLs identified on chromosome 5 had opposing effects of similar size. For the fermax at both time points three major (I;**3**;72-80, II;**3**;78-80, I;**5**;42-52, II;**5**;38-45 and I;**5**;76-86, II;**5**;84-94) and two minor QTLs (I;**2**;43-53, II;**2**;44-53 and I;**3**;0-8, II;**3**;0-12) were found at similar positions (Fig. 3.2; table 3.3). Similar to the QTLs found for rosette area the two major QTLs for fermax on chromosome 5 had opposing effects of similar size. The other major QTL (I;**3**;72-80, II;**3**;70-80), with positive effect, is located at the bottom of chromosome 3.

Two QTLs (and both effects) were shared between the rosette area and fermax. Both are located on chromosome 5 and had opposite allelic effects. Apparently no single locus will give a plant a larger rosette area and a smaller fermax or vice versa. Hence, no co-locating QTLs with opposite effects (trait wise) were found for rosette area and fermax.

Rosette compactness

Rosette compactness is a compound trait and can be calculated from the rosette area and fermax. Compared to the QTLs identified for the determinants of compactness, rosette area and fermax, several new (at loci not identified for area and fermax) QTLs for compactness could be identified (I;**2**;35-42, II;**2**;35-39, I;**2**;48-52, II;**2**;49-59, I;**2**;55-63, I;**2**;63-71, I;**3**;0-2, II;**3**;0-5, I;**3**;40-46, II;**3**;45-55, I;**3**;78-80, II;**3**;76-80 and I;**5**;68-76, II;**5**;54-73) (Fig. 3.2; Table 3.4). As for the QTLs identified for rosette area and fermax most QTLs for compactness at both time points were detected at similar positions.

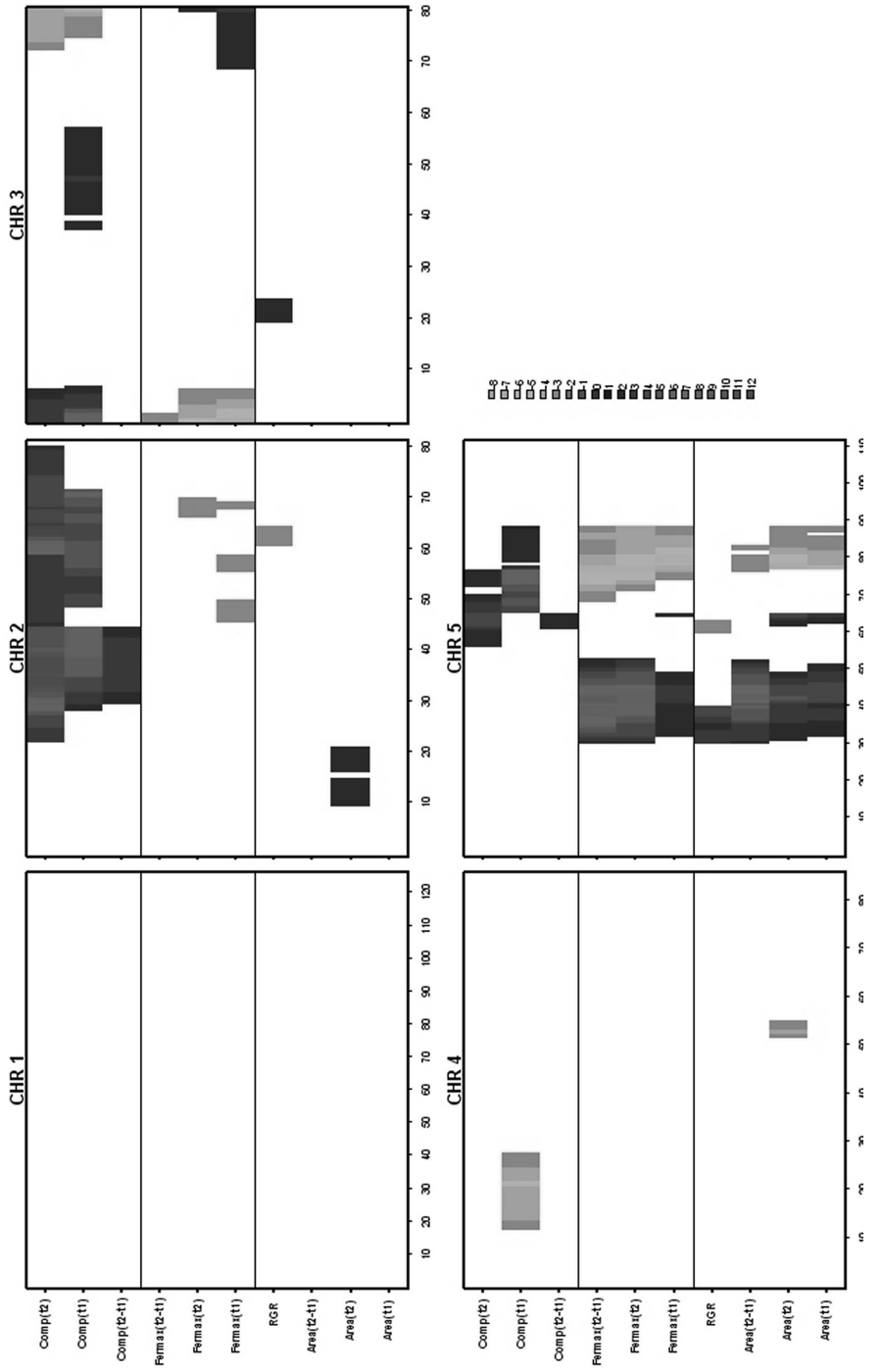


Figure 3.2 QTLs found for Area, Fermax, Compactness and Relative growth rate. Colors indicate the LOD score and the sign of the L*er* allelic effect, see legend next to chromosome 5. Y-axis: Traits (TI = 18 days after transfer to soil, TII = 21 days after transfer to soil) X-axis: Position on the chromosome expressed in centiMorgan.

Table 3.2 QTLs for rosette area, chromosome (CHR), effect of the L*er* allele compared to the average of all RILs (L*er*EFF), interval of the QTL in cM (INTER), top of the QTL (TOP) and LOD score of the top (TOP) at both time-points (left = time-point I, right = time-point II).

TRAIT	L*er*EFF (mm^2)	CHR	INTER	TOP	LOD	TRAIT	L*er*EFF (mm^2)	CHR	INTER	TOP	LOD
Area I	-18.8	4	48-55	51.9	2.9	Area II	-29.9	4	47-58	53.0	3.1
	+21.7	5	35-48	43.1	5.2		+48.7	5	36-49	43.9	5.7
	+18.2	5	58-68	65.0	3.4		+32.2	5	58-68	66.0	2.4
	-23.2	5	77-83	80.0	5.4		-45.2	5	76-82	80.0	5.0

Table 3.3 QTLs for maximum feret diameter (fermax), for details see legend Table 3.2.

TRAIT	L*er*EFF (mm)	CHR	INTER	TOP	LOD	TRAIT	L*er*EFF (mm)	CHR	INTER	TOP	LOD
Fermax I	-1.1	2	43-53	47.0	1.9	Fermax II	-2.1	2	44-53	50.0	3.5
	-1.2	3	0-8	0.0	2.8		-1.6	3	0-12	0.0	2.1
	+1.6	3	72-80	79.0	4.3		+2.6	3	69-80	79.0	5.6
	+2.0	5	41-52	46.8	5.2		+3.3	5	38-45	43.1	8.4
	-2.0	5	76-86	78.0	5.0		-2.5	5	83-89	87.0	4.7

Table 3.4 QTLs for rosette compactness, for details see legend Table 3.2.

TRAIT	L*er*EFF (%)	CHR	INTER	TOP	LOD	TRAIT	L*er*EFF (%)	CHR	INTER	TOP	LOD
Comp I	+1.0	2	34-42	35.0	8.8	Comp II	+1.0	2	34-39	35.0	8.7
	+1.0	2	48-52	49.0	6.9		+1.0	2	49-58	58.0	7.8
	+0.9	2	54-62	60.1	5.2						
	+0.7	2	63-71	70.0	4.7						
	+0.9	3	0-2	0.0	7.0		+0.8	3	0-5	0.0	7.0
	+0.8	3	40-45	42.0	6.7		+0.6	3	45-53	53.0	3.3
	-0.7	3	77-80	79.0	5.5		-1.0	3	76-80	79.1	8.9
	+1.0	5	67-76	72.6	8.4		+0.8	5	53-73	63.8	5.0

Table 3.5 QTLs for change in area, fermax and compactness and RGR, for details see legend Table 3.2.

TRAIT	L*er*EFF	CHR	INTER	TOP	LOD	TRAIT	L*er*EFF	CHR	INTER	TOP	LOD
Area (mm^2)	+5.8	2	23-43	35	2.0	RGR	+3.4	2	24-43	35	2.6
	+32.3	5	30-48	42	5.9		+7.0	5	30-48	42	7.7
	-28.7	5	76-81	77	4.1		-5.2	5	75-81	77	4.2
Fermax (mm)	+1.5	5	30-49	42	6.8						
	-1.1	5	71-87	77	4.2	Comp (%)					
							+0.72	4	13-26	22	2.0

From the phenotypic distribution of the RILs (Fig. 3.1), we expected the majority of the QTLs to show a positive effect of the L*er* allele, leading to a more compact rosette. Indeed this was found as the only QTL with a negative effect of the L*er* allele on compactness was located at the bottom of chromosome 3 (I;**3**;78-80, II;**3**;76-80).

Since mutation in the *ERECTA* gene in L*er* influences leaf shape (Torii et al., 1996; Xu et al., 2003; Shpak et al., 2004), making the rosette more compact, a QTL on the Erecta locus could be identified as expected (I;**2**;48-52, II;**2**;49-59). However, other QTLs, flanking Erecta, were also found (I;**2**;35-42, II;**2**;35-39, I;**2**;48-52, II;**2**;49-59, I;**2**;55-63, I;**2**;63-71). The bottom half of chromosome 2 explained about half of the variation found for time point I with four separate QTLs (I;**2**;35-42, I;**2**;48-52, I;**2**;55-63, I;**2**;63-71). Only two separate QTLs (II;**2**;35-39, II;**2**;49-59) were found for time point II at chromosome 2, still explaining about a considerable amount, one third, of the variation.

In general, compared to the QTLs found for rosette area and fermax, the QTLs for compactness had higher LOD scores. The QTLs for compactness co-located with those of fermax on chromosome 3 with opposite effects for the two traits. No QTLs for compactness were found on those positions where fermax co-located with the major QTLs for rosette area.

Absolute and relative growth

Measurements at two time-points enabled us to determine the relative growth rate (RGR) as well as the absolute differences between rosette area, fermax and compactness. These parameters describe part of plant growth and can be used to study the role of the genetic background in these specific aspects of plant growth. The QTLs for absolute differences in rosette area and fermax all co-located with their QTLs found for them at the single time points (Fig. 3.2; Table 3.5). This means that a plant that has a larger area at time point I increases its rosette area correspondingly resulting in a still larger area at time point II. This effect was also found for fermax. The change in compactness describes how the rosette grows and expands by sequential leaves that cover the potential area during time. For compactness the difference between the two time points, however, shows only one major QTL, located at chromosome two. In contrast to the change in rosette area and fermax (multiple QTLs) the change in compactness in time is affected by much less QTLs.

To study growth in more detail we used the Relative Growth Rate (RGR). This trait describes the increase in area relative to the plant area already present (Evans 1976; Poorter 2001). A few QTLs could be detected for this trait. A couple of minor QTLs co-located with compactness QTLs indicating that the possible overlap of laminae has an effect on how accurate the RGR could be determined. However, the most significant QTL for RGR co-located with the most significant QTLs for rosette area and fermax, which shows that the allelic difference of this locus really affects the pace at which a plant grows.

Near Isogenic Lines

For confirmation and more exact location of the QTLs we tested several of the near isogenic lines (NILs) described by Keurentjes et al. (2007). These NILs contain a Cvi introgression in a L*er* background. Most QTLs were confirmed and the locus size could be decreased in this way. Equally important, testing the NILs on chromosome 2 enabled us to more specifically determine the underlying traits causing the difference in compactness.

NILs confirming QTLs on chromosome 5

On chromosome 5 QTLs for rosette area and fermax were identified at both time points as well as the difference between the two. To confirm these QTLs we measured the NILs LCN5-1 to 5-6 (Fig. 3.3). These NILs contain an introgression harboring or flanking the QTLs for fermax and rosette area (Fig. 3.4). Some, but not all, NILs confirmed the QTLs for fermax and rosette area.

The fermax QTL at time point I, for which the L*er* allele shows a positive effect, was only confirmed by LCN5-3 (p=0.05) although the p-values of LCN5-2 and LCN5-6 were close to significant (p=0.08 and 0.09 resp.). No NILs were significantly different from L*er* for the second time point. However, the three NILs (LCN5-2, LCN5-3 and LCN5-6) with the lowest p-values at the first time point showed again the lowest p-values at the second time point (0.11, 0.09 and 0.12 resp) supporting the QTL effect. The difference in fermax between the two time points could not be confirmed and no indication of a QTL effect was found since all tested NILs were statistically similar to L*er* (Fig. 3.3).

For rosette area the QTLs found for both time points also had a positive L*er* effect and could be confirmed by LCN5-2 and LCN5-3 (**I**; p=0.01 and 0.05; **II**; p=0.01 and 0.04). LCN5-6 which also should confirm the QTL, based on the introgressed region, was only indicative of a QTL effect. It had a smaller average area than L*er* but not significant (**I**; p=0.10; **II**; p=0.09). The difference in area was confirmed again by LCN5-2 (p=0.01). The increase in area of LCN5-3 and LCN5-6 was smaller than that of L*er* but not-significant (p=0.11 and 0.09 respectively).

Together, these results show that the most likely locus of the QTLs for fermax, rosette area and the difference in area on chromosome 5 is between 35.7 and 45.9 cM.

NILs confirming QTLs on chromosome 3

On both ends of chromosome 3 QTLs for fermax and compactness were identified. On the top of this chromosome a QTL with a positive L*er* effect on compactness but negative on fermax was partially confirmed by LCN3-1 (Fig. 3.5). The compactness of this NIL was significantly lower than that of L*er* and although this NIL also had a larger fermax as the QTL would predict, it was not significant. The L*er* effect of the QTLs on the bottom of the chromosome was positive for fermax and negative for compactness. The QTLs for these two traits could both be confirmed by LCN3-17 which had a significantly lower fermax and a higher compactness. These two NILs show that the fermax by itself has a large influence on rosette compactness and might harbor natural alleles influencing petiole length.

Determining the number of loci on chromosome 2 involved in rosette compactness

For compactness at time point I a broad QTL consisting of several peaks was found at the bottom half of chromosome 2. The three independent peaks suggest flanking multiple individual QTLs. This compactness QTL on chromosome 2 was further investigated by assigning each NIL of that chromosome to their specific genetic bin(s) (as in Keurentjes et al., 2007) and tested against the observed L*er* phenotype(s). These bins are predefined genomic loci and a NIL is assigned to a specific bin if that NIL contains a Cvi introgression at that locus.

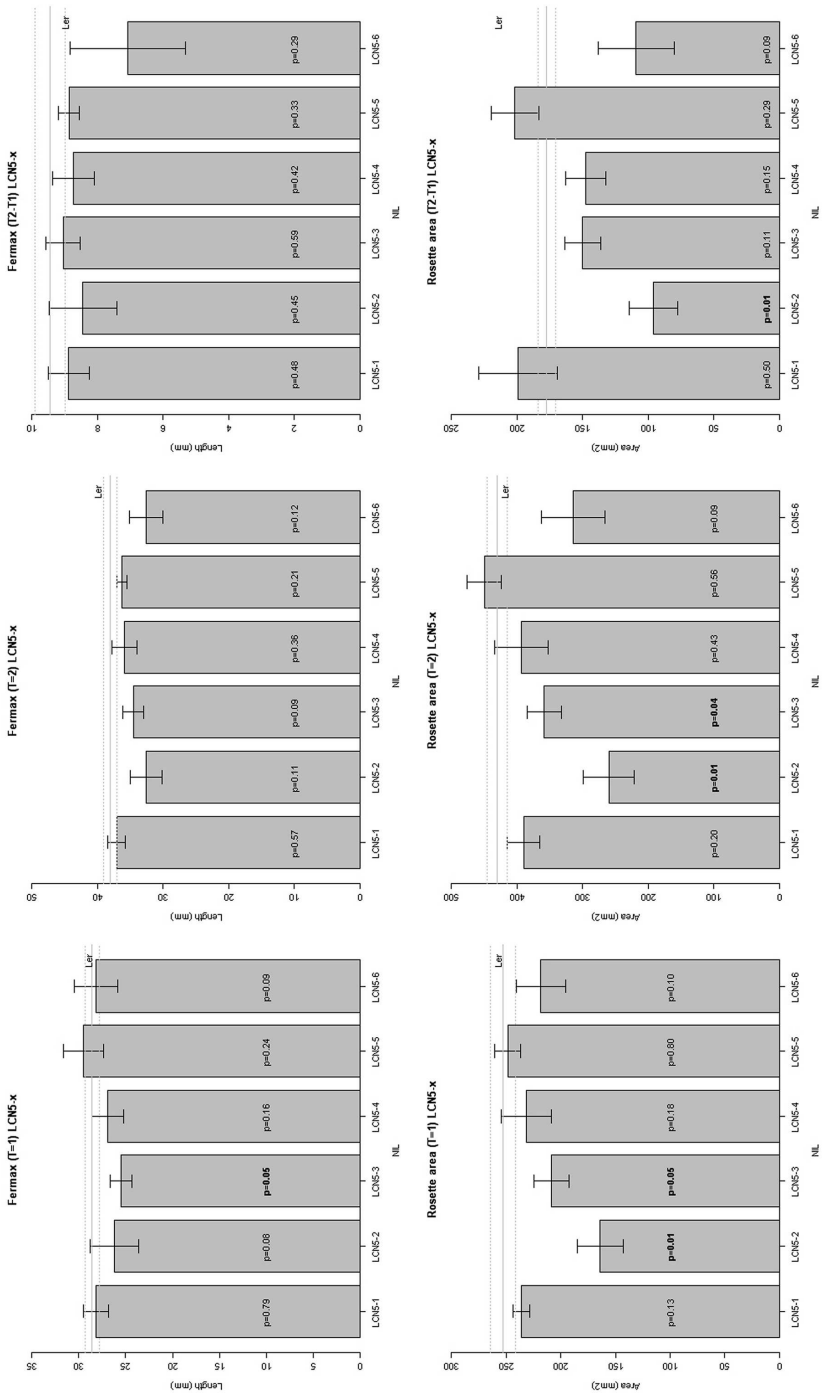


Figure 3.3 Near Isogenic Lines (NILs) with a Cvi introgression (in a L*er* background) harboring or flanking the QTLs for rosette area (I;5;35-48, II;5;36-49) anc fermax (I;5;42-52, II;5;38-45). L*er* phenotypic value and standard errors are indicated by the grey horizontal lines. NIL phenotypic values significantly differen from L*er* are indicated with their p value in bold (tI = 20 days, tII = 23 days).

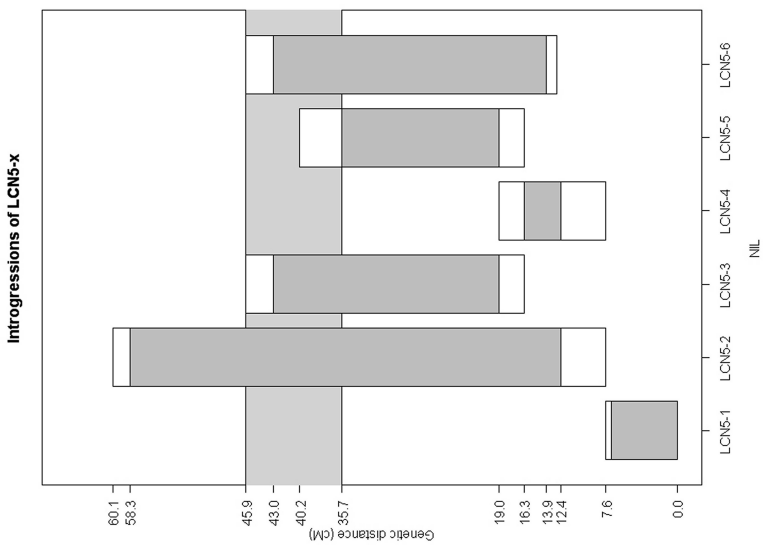


Figure 3.4 Introgressions of LCN5-x. Light-grey area (35.7-45.9 cM) is most likely locus containing the fermax and area QTLs. The grey areas of the Cvi introgressions into L*er* are flanked by two Cvi markers. The white parts are flanked by a Cvi marker on one side and a L*er* marker on the other.

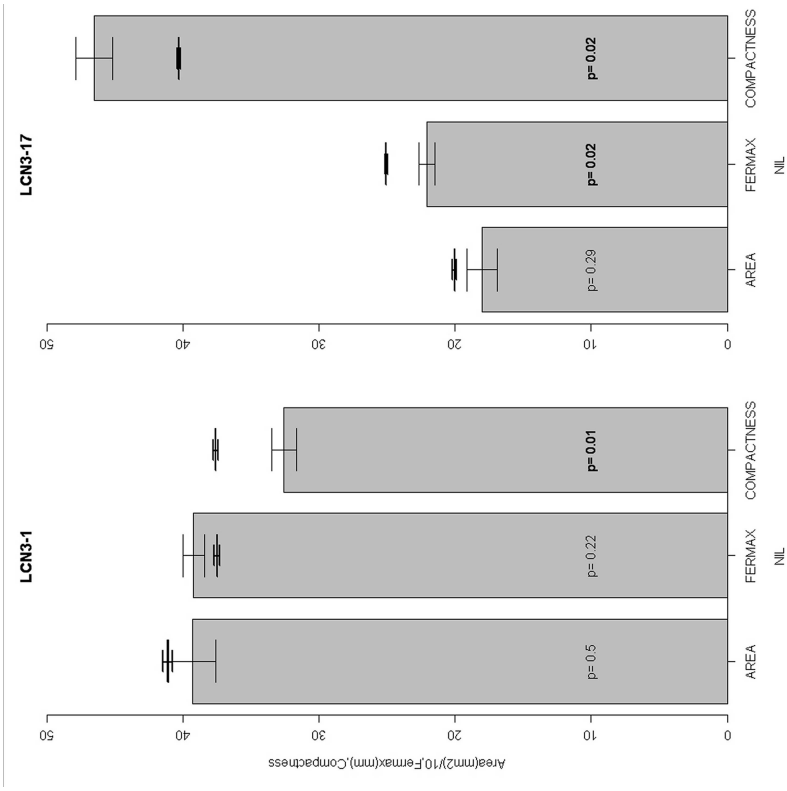


Figure 3.5 Rosette area, fermax and compactness of LCN3-1 (t = 23 days) and LCN3-17 (t = 19 days) compared to L*er*. L*er* phenotypic values are indicated in thick black lines. For both the standard error is indicated per trait. P-values of t-test are indicated in the bars; significant values are in bold.

By testing the NILs containing an introgression spanning part of this broad QTL region we could identify three independent loci involved in rosette compactness (Fig. 3.6). Two of these contain genes known to affect rosette compactness. The first locus (I;**2**;35-42, II;**2**;35-39) with a peak, most to the left of the broad QTL, contained *PHYTOCHROME B* (*PHYB*). This QTL peak was confirmed by LCN2-7 which has smaller laminae and relatively longer petioles than L*er* (Fig. 3.7) The second peak, in the middle, (I;**2**;48-52, II;**2**;49-59) contained the extensively studied gene *ERECTA* (*ER*). This QTL peak was confirmed by LCN2-8, 2-9, 2-11 and 2-15 which are of similar size to L*er* but have longer petioles (Fig. 3.7). The third locus (I;**2**;63-71), near the end of the chromosome was confirmed by LCN2-17. This NIL has much narrower laminae than L*er* (Fig. 3.7) but does not contain a Cvi *PHYB* or *ER* allele and so must be the result of a third not yet identified locus influencing rosette compactness.

For two of the three identified loci influencing compactness candidate genes likely to be causal for the QTL(s) were found. This enabled us to test whether the alleles of these genes, *PHYB* and *ER* are functionally different between L*er* and Cvi.

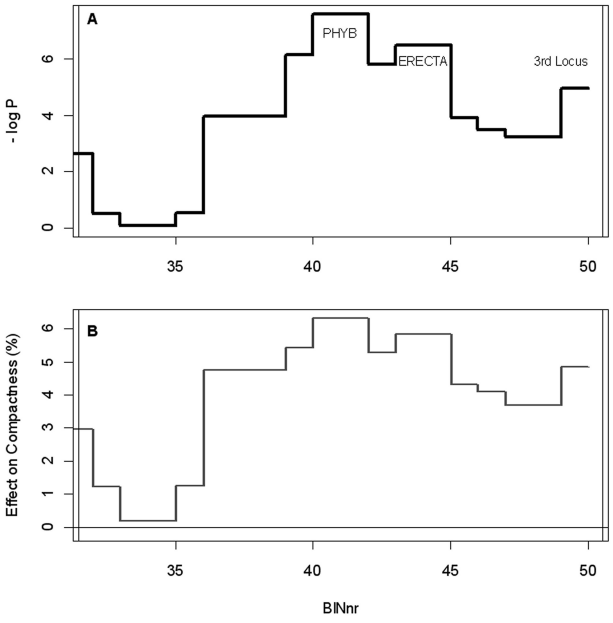


Figure 3.6 Chromosome 2: NIL bin mapping; –log p score of phenotypic values of compactness compared to L*er* (A) and effect difference between L*er* and Cvi bin (allele) (B). The position of genes (*PHYTOCHROME B* (*PHYB*) and *ERECTA*) identified as allelic different between L*er* and Cvi are indicated in panel A as well as the 3rd identified locus influencing compactness.

Identification of genes causal for QTLs

Candidate genes ERECTA and PHYB

For compactness two of the large effect QTLs on chromosome 2 contained a likely candidate gene. Within the support interval of the first QTL *PHYB* was located and the second was on top of *ERECTA*. Both genes are known to influence petiole length and leaf development (Reed et al., 1993; Xu et al., 2003) and as a result rosette compactness.

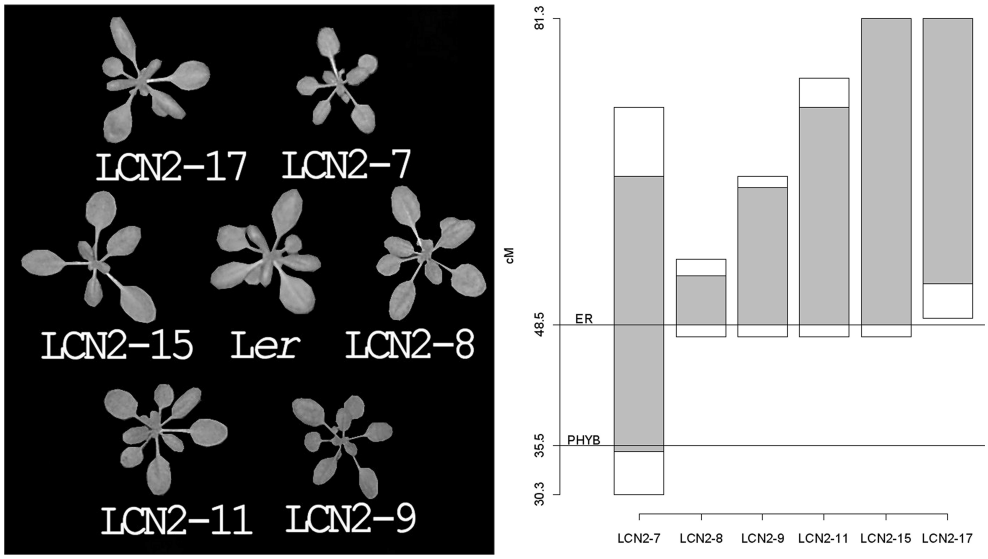


Figure: 3.7 Phenotypes of NILs with a Cvi introgression at chromosome 2. Upper left panel: Phenotypes of 6 NILs with different introgressions compared to L*er*. Lower panel: Phenotypes of L*er* compared to NIL 2-17, containing a small Cvi introgression (~12 to ~20 Mbp), 30 days after germination, showing an altered leaf phenotype. Upper right panel: introgression regions of the NILs shown in the upper left panel. Position of *PHYTOCHROME B* (*PHYB*) and *ERECTA* (*ER*) are indicated by the horizontal lines.

Moreover, allelic differences (polymorphisms in the nucleotide sequence) in these genes between L*er* and Cvi have been described. The *PHYB* alleles from Cvi and L*er* have been sequenced and found to have considerable nucleotide variation in the promoter as well as synonymous and replacement changes in the coding region (Borevitz et al., 2002). The *ERECTA* gene makes a good candidate since L*er* contains a well known mutation in this gene (Torii et al., 1996; Shpak et al., 2004) affecting its functionality, whereas the Cvi allele is assumed to be fully functional. We tested if the QTLs could be caused by the allelic polymorphisms of these genes between L*er* and Cvi.

To test if ERECTA causes a difference in compactness we compared L*er* and its wild-type Landsberg (Lan) for compactness (Fig. 3.8). This showed that indeed ERECTA influences compactness causing L*er* to be much more compact than Lan. Further observation shows that the difference in compactness is mainly due to the shorter petioles of L*er* (data not shown). This is also shown by the QTL for fermax in which the L*er* allele has a negative effect on fermax, indicating a shorter petiole. In chapter 2 a QTL for curved

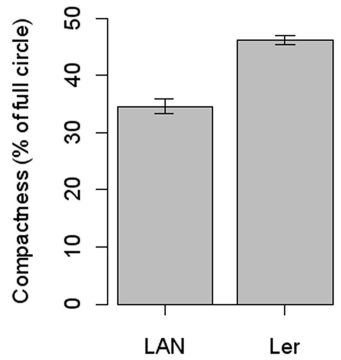


Figure 3.8 Difference in compactness between L*er* and its wildtype Lan, confirming that the QTL on chromosome 2 (I;**2**;48-52, II;**2**;49-59) is caused by the allelic difference of *ERECTA*.

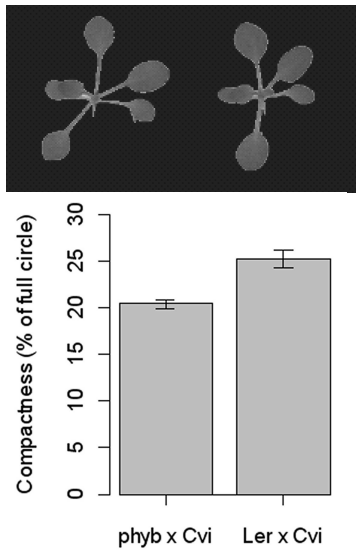


Figure 3.9 Complementation of a *phyb* mutant with the Cvi *PHYTOCHROME B* allele in relation to rosette compactness. Lower panel: difference in compactness between the F_1-offspring of a cross between a *phyb* mutant in a L*er* background and Cvi and a cross between L*er* and Cvi. Upper panel: top view of the rosettes of those F_1-offspring, showing the longer petioles and the smaller leaves of the *phyb* x Cvi F_1 offspring. Left: *phyb* x Cvi F_1 Right: L*er* x Cvi F_1.

leaves was found to be caused by an allelic difference between L*er* and Cvi in *erecta*. L*er* contains the mutant allele of *erecta* which causes the leaves to become curved and the petioles to remain shorter (Torii et al., 1996; Shpak et al., 2004). These curved leaves should also result in a lower compactness since the area that can be measured is smaller. But the L*er* allele results in a higher compactness, thus the curved leaf phenotype could not substitute for the effect of the shorter petioles.

A candidate gene for the QTL (peak) located above *ERECTA*, affecting rosette compactness on chromosome 2 (I;**2**;35-42, II;**2**;35-39) is *PHYB*. To detect whether a functional allelic difference exists between the L*er* and Cvi allele, we made two crosses to compare to each other. Cvi was crossed, reciprocal, with both L*er* and a *phyB* knockout mutant (PHYB5; (Reed et al., 1994), in the L*er* background (*phyB*L*er*). The only genetic difference in the completely heterozygous F_1's of both crosses was in only one gene, *PHYB*. Both F_1's are heterozygous for *ERECTA*/*erecta* and will show a wildtype *ERECTA* phenotype. Moreover, both L*er* x Cvi F_1's will have the *PHYB*Cvi allele combined with either a functional *PHYB*L*er* allele or a mutant *phyB*L*er* allele. A phenotypical difference between the F_1 plants of the two crosses implies that not only the *PHYB*Cvi allele is recessive compared to *PHYB*L*er* but also that the alleles exhibit a functional difference. Indeed the rosette compactness was different; the *PHYB*Cvi allele results in a less compact, Cvi-type rosette, confirming the functional allelic difference of *PHYB* between L*er* and Cvi (Fig. 3.9). To rule out the effect of the difference in copy number of the functional *PHYB* genes, e.g. the L*er* x Cvi F_1's contain two functional alleles and the *PHYB*L*er* x Cvi F_1's contain only one, we also compared L*er* with a L*er* x *PHYB*(L*er*) cross. No difference was found in rosette compactness or any of the other measured traits. This shows that there is no large dose effect for functional *PHYB* alleles.

Taken together, these results provide strong evidence that the QTL for rosette compactness on chromosome 2 is caused by a functional allelic difference of PHYB between L*er* and Cvi.

Candidate genes; mutants and T-DNA introgression lines

Several other candidate genes for rosette compactness were tested by analyzing mutant or T-DNA insertion lines. The specific roles of the tested candidate genes on rosette area and fermax making up compactness were also determined.

To test which gene at the top of chromosome 3 may cause the QTL for rosette compactness the phenotype of a T-DNA insertion lines was studied. This line containing an insertion in gene At3g01460 (showing similarity to the human METHYL-CPG-BINDING DOMAIN 9) exhibited a significant difference in rosette compactness with its wild-type Col-0 (Fig. 3.10, Fig. 3.11) that remained stable over time. To test whether a change in laminae shape or petiole length or a combination caused the rosette compactness to be lower in the T-DNA insertion line we measured the length of the lamina, petiole and whole leaf (Fig. 3.12). The similar area and small, but significant difference in rosette feret already suggested the length of the petiole to be the cause of the less compact rosette. The introgression line had significantly longer petioles (circa. 0.4 cm) whereas the laminae length was not different.

The QTL rich region (**5**;30-45) on chromosome 5 may be identical to the one identified in Chapter 2. In that study a candidate gene, *NPH4/ARF7*, was tested and found to influence a specific part of hyponastic growth. To test whether *NPH4/ARF7* is also a likely candidate gene for the area QTL on chromosome 5 we tested this mutant (Harper et al., 2000) against its wild type Col (Fig. 3.11; Fig. 3.12). A difference in rosette area was observed caused by the smaller leaves of *nph4*. The smaller leaves, but similar petiole length therefore results in a smaller compactness. Although this makes a good candidate for the QTL for rosette area as well as RGR the physical position of this gene falls outside the QTL position determined by the NILs.

Together, these results show that through one of the underlying components (petiole length and lamina size/shape) rosette compactness can be influenced separately at the single gene level.

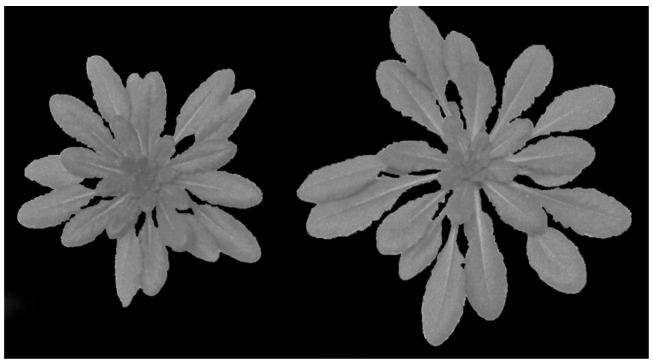


Figure 3.10 Top view of Col-0 (left) and the SALK T-DNA insertion line (N554659) of gene At3g01460 (right) showing that this gene functions in determining the rosette compactness by controlling petiole length.

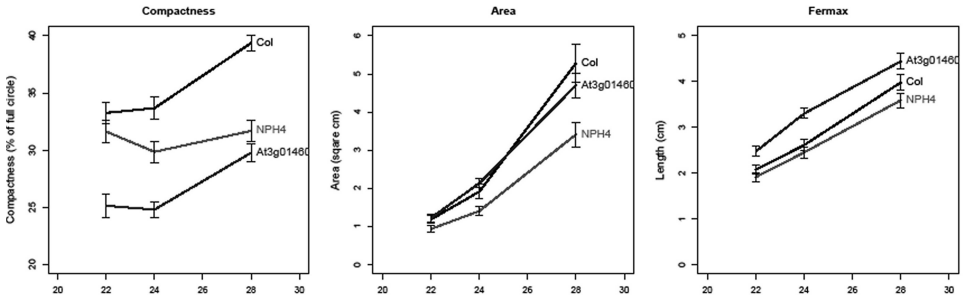


Figure 3.11 Rosette area, fermax and compactness 22, 24 and 28 days after germination of Col-0, *non phototropic hypocotyl 4* (*nph4*/*arf7*) and a SALK T-DNA insertion line of gene At3g01460.

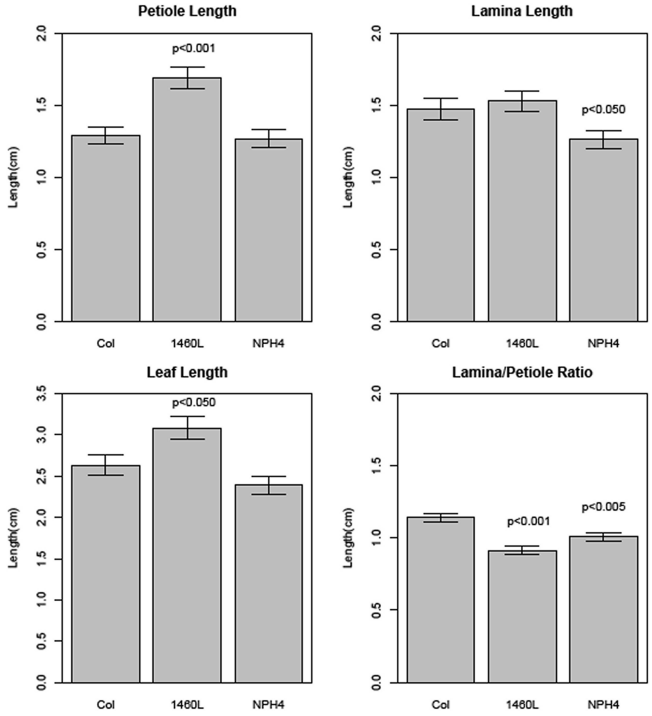


Figure 3.12 Length of the petiole (upper left), lamina (upper right) and leaf (lower left) and lamina petiole ratio (lower right) of Col-0, *non phototropic hypocotyl 4* (*nph4/arf7*) and a SALK T-DNA insertion line of gene At3g01460 (indicated by 1460L). P-values are given when the difference with Col-0 is significant.

Discussion

Compactness as a measure for light sensitivity

We used rosette compactness as measure for light-sensitivity as it describes two of the main effects of the light adaptation response, petiole elongation and decreased rate of leaf expansion (Tsukaya et al., 2002; Hisamatsu et al., 2005; Kim et al., 2005; Kozuka et al., 2005). By identifying and confirming QTLs, it was shown that rosette compactness is a genetically determined plant trait brought about by a limited number of loci in the L*er* x Cvi Recombinant Inbred Line (RIL) population. Confirmation of QTLs with NILs showed a more detailed effect of the QTLs on components that constitute rosette compactness. The functional allelic difference of *PHYB* between L*er* and Cvi identified in this study shows that (at least partially) compactness is a measure for light sensitivity.

QTLs identified for the different traits

The three traits, fermax, rosette area and rosette compactness can be seen as three genetically partially separate groups since not all QTLs are shared. The major QTLs for traits, measured at time point I, II and even the difference between the two time-points are most similar to each other. Since compactness is a composite trait calculated from the other two traits, fermax and area, some QTLs for compactness can be expected to be enhanced at one locus, but absent on another. This is because an opposite effect of the two traits enhances the effect of compactness while a similar effect of the two decreases the allelic effect on compactness. For instance, the first QTL rich region on chromosome 5 (around 40 cM) contains QTLs for both the fermax and rosette area group, that both have a negative effect of the L*er* allele. Since the L*er* allele has a similar negative effect for both traits, this would result in a total reduction in plant size. The compactness, however, is not affected by this locus. The QTL rich region on chromosome 5 (around 80 cM) shows QTLs for all three groups. Here, both rosette area and fermax are positively influenced by the L*er* allele but the compactness is decreased by it. Together, this shows that even if the fermax and rosette area are affected by the same QTL, the effect on compactness is dependent on the interaction of effect sizes of fermax and rosette area.

The QTLs for compactness on chromosome 3 can be explained by just one trait. The fermax has QTLs at both top and bottom of the chromosome with an opposite effect to that for compactness. On the top of chromosome 3 a QTL for compactness with a negative effect of the L*er* allele was found. The effect of the QTL for fermax at this locus was positive for the L*er* allele, hence making the diameter of the rosette larger while the rosette area is not affected, thus decreasing the compactness. The other way around, a L*er* allele resulting in a smaller fermax while the rosette area is not affected by it and thus resulting in a more compact plant, can be found at the end of chromosome 3.

Most striking are the QTLs for compactness on chromosome 2. These very strong QTLs mainly affect compactness and hardly fermax or rosette area. Although some (minor) QTLs for fermax are found to co-locate with QTLs for compactness, their effect is not enough to explain the observed compactness QTLs. Therefore, the combinations of loci with (opposite) effects on fermax and rosette area are expected to underlie these compactness QTLs. These effects are probably too small to be detected as QTLs but

together they have an influence on the compactness for which a QTL was found. Because of the small effects of the QTLs on fermax and rosette area and the strong effect on compactness the effect of the QTLs on fermax is opposite to the effect on rosette area. In this way compactness is double affected and results in strong QTLs with a large effect.

Functional allelic difference of ERECTA and PHYB

One gene in particular, *ERECTA*, was expected to explain a QTL because of its known influence on leaf shape and petiole length (Torii et al., 1996; Shpak et al., 2004) and its allelic difference between L*er* and Cvi. Although the QTL found on the Erecta locus was very strong it was also very broad and consisted of several independent peaks. Nevertheless, the QTL and the effect of *ERECTA* were confirmed by comparing L*er* and its wild-type Lan (Fig. 3.8).

Through NIL analysis it became clear that at least two other loci (genes) influencing compactness are flanking *ERECTA* (~49 cM) and together these three independent loci make up this broad QTL. More positioned to the top of the chromosome relative to *ERECTA* is *PHYB* located. This gene has a known influence on both petiole length and leaf expansion. The fact that no QTLs were found for fermax (petiole length) and rosette area (leaf expansion) on the position of PHYB (chr 2; ~35 cM) can be explained by the effects of PHYB on those traits. A mutation in *PHYB* results in longer petioles and a reduced leaf expansion. This has a strong effect on the rosette compactness; less area is spread out further. Even if the allelic effect between the L*er* and Cvi PHYB type would be small for petiole length and leaf expansion the compactness could be significantly altered.

Despite being candidate gene in several QTL studies (Borevitz et al., 2002) a functional allelic difference between the L*er* and Cvi type *PHYB* was not known. The phenotypic effects of a *phyB* mutant in a L*er* background (*phyB*L*er*) could not be complemented with a *PHYB*Cvi allele in a *phyB*L*er* x Cvi F_1. The L*er* x Cvi F_1 showed a compacter phenotype than the *phyB*L*er* x Cvi F_1. We therefore conclude that Cvi has a natural functional allelic difference of *PHYB* compared to L*er*. Several, but not all functions of the Cvi-PHYB allele are impaired or altered leading to a more light sensitive phenotype.

Candidate genes NPH4/ARF7 and At3g01460

A candidate gene approach showed that petiole length and leaf area, making up rosette compactness, can be separated on a genetic level. The SALK line of At3g01460 and NPH4 both showed a decreased compactness. In the case of At3g01460 this is because of longer petioles while in the *NPH4/ARF7* mutant the leaf area is smaller compared to wild-type.

While At3g01460 (which shows similarity to the human *METHYL-CPG-BINDING DOMAIN* 9 gene) is physically located on the QTL at the top of chromosome 3 and therefore a likely candidate, *NPH4/ARF7* is at the border of the QTL on the first half of chromosome 5 and therefore not a very likely candidate. However, phenotypic analysis of knockouts/mutants of both of these genes gives further details of the genetic regulation of rosette compactness. The insertion line of gene At3g01460 shows that this gene is mainly influencing petiole length in contrast to *nph4/arf7* in which the laminae size and shape are affected.

Candidate genes involved in processes determining rosette compactness

The two components, fermax and rosette area, used to calculate rosette compactness are determined by three traits, petiole length, leaf area and overlapping leaves. Fermax is dependent on petiole length and the shape of the laminae; longer laminae mean a higher fermax. These also determine the rosette area but in a different way, they both determine how much the individual leaves overlap each other. Therefore the genetic components affecting petiole length and leaf-shape will influence rosette compactness as well and are important candidate genes for the fine-mapping of the unconfirmed QTLs for compactness, fermax and rosette area.

Despite *PHYB* and *ERECTA* being responsible for the very broad compactness QTL on chromosome two, a third locus was identified through NIL/bin analysis. This locus near the bottom of the chromosome contains at least one gene with a functional allelic difference between L*er* and Cvi. Several genes affecting the morphology of the leaf are positioned at this locus. Genes determining the adaxial side of the leaf like *PHABULOSA* (*PHB*; At2g34710) (McConnell and Barton, 1998; Mallory et al., 2004) or the abaxial side like, the *YABBY* class, *YAB1* (At2g45190) and *YAB5* (AT2g26580) (Siegfried et al., 1999) are candidate genes. Genes involved in leaf primordia determination like *ASYMMETRICLEAVES1* (*AS1*; At2g37630) (Byrne et al., 2000), *SERRATE* (*SE*; At2g27100) (Clarke et al., 1999; Grigg et al., 2005) and *PICKLE* (*PKL*; At2g25170) (Ogas et al., 1999) are also good candidates. Furthermore, genes determining leaf length like *ROTUNDIFOLIA4* (*ROT4*; At2g36985) (Narita et al., 2004) and genes involved in the amount of cells like *LOWCELLDENSITY1* (*LCD1*; At2g37860) (Barth and Conklin, 2003) could be polymorphic between L*er* and Cvi as well.

Conclusions

Leaf shape, petiole length and overall "growth plan"

From our study we conclude that the overall "growth plan" of the rosette is a genetically regulated trait on which leaf shape and petiole length are (partially) dependent. This can be observed by the higher heritability of the compactness trait. Moreover, all the traits, rosette area, fermax and compactness are under the influence of the environment. This results in an interesting feedback mechanism of traits influencing each other. Fermax and rosette area are determining the compactness, while compactness creates the micro-environment in which both traits operate. So, although within their genetic constraints, the area and petiole length (fermax) are determined by the micro-environment of the "growth plan" of the rosette. Especially when the leaves start to overlap, the overall "growth plan" starts to give feedback on the other traits that are determined by a genotype*environment interaction.

Functionally different alleles between Ler and Cvi

Multiple QTLs were found for all measured traits. For two of the compactness QTLs we could determine the causal genes. The L*er* and Cvi allele of these genes cause a difference in phenotype. The allelic difference between L*er* and Cvi for the well known gene *ERECTA* caused a difference in compactness. Furthermore, we showed that the alleles of *PHYTOCHROME B* of L*er* and Cvi are functionally different for rosette compactness formation.

Materials and Methods

Growing conditions and experimental setup

Plants were grown under short days (9h light, 15 dark) and a light intensity of 200 µmol m^{-2} s^{-1}. Seeds were imbibed on wet filter paper for 3 days after which the germinated seedlings were transferred to soil. Further details are described in Millenaar et al., 2005. To identify natural variation and QTL mapping, 120 recombinant inbred lines (RILs) of the L*er* x Cvi population (Alonso-Blanco et al., 1998) were grown at once in four replicate batches. Each batch was divided in 10 trays containing 3 plants of 14 separate genotypes, parental genotypes included. In this way we obtained measurements of 12 plants per RIL. The parental genotypes were measured 35 times. For conformation of the QTLs the near isogenic lines (NILs) were grown in 3 batches, at the same time and in the same growth chamber as the RILs. Per NIL 6 to 9 individual plants were measured. Tests on other genotypes have been performed under similar conditions.

Measurements

Photos of the trays of plant were taken at two time points 21 and 24 days after the start of germination (18 and 21 days after transfer to soil) for the RILs, the other genotypes are indicated in the figure legends as days after transfer to soil. These photos were processed with KS400 imaging software using a custom designed script. Rosette areas as well as the maximum feret diameter (fermax), the longest line that can be drawn between two points on the edge of a (non-circular) object, were determined for individual plants. These two shape parameters, fermax and rosette area are defined so that they describe the overall shape of the rosette by the compactness trait. Further validation of the use of fermax as a leaf length estimate came from the measurements of the length of the petiole and laminae of the candidate genes.

Compactness was calculated by dividing the measured area by, half fermax squared, multiplied by π (pi). This results in the percentage of coverage by the rosette of a full circle with a diameter of Fermax.

$$Compactness = \frac{Area}{(\frac{1}{2} Fer\max)^2 \times Pi}$$

QTL mapping

QTL mapping was performed in QTL cartographer (Wang et al., 2007). LOD thresholds were determined by 1000 permutations. Composite interval mapping (CIM) with the back and forward control method (p=0.05), was used to calculate the QTLs. The markers and genetic map were used as in Keurentjes et al., 2007.

NIL bin linkage

As described in Keurentjes et al., 2007. We tested if the mean phenotypic value of each bin of the NILs of chromosome 2 was different from the mean phenotypic value of L*er* by a linear model.

Chapter 4

Identification of a regulatory network downstream of ERECTA by means of Genetical Genomics.

Inez R. Terpstra & L. Basten Snoek, Joost J. B. Keurentjes, Anton J.M. Peeters & Guido van den Ackerveken.

Abstract

Many factors affect the transcript levels of genes, including their genetic background. By application of genetical genomics the latter can be studied in a multifactorial way. In this paper we present a new way to utilize genetical genomics and discuss the resulting new insights in regulatory networks. To this end we combined genetical genomics on L*er*/Cvi RILs with expression profiling of the wild type versus *erecta* mutant. This resulted in three major findings. 1) Confirmation of the presumed regulation of a significant number of genes by the main hotspot for transcript regulation (HTR) in our previous study. 2) Extension of single gene mutant analysis to identify in detail interactions in its downstream regulatory processes. 3) Integration into the network of genes that have no significant eQTLs themselves, through their effect on transcript levels measured in single gene mutation experiments. Applied to the extremely pleiotropic Arabidopsis gene *ERECTA*, we confirm that this is the causal gene for the major HTR in the Arabidopsis L*er*/Cvi RIL population. We present evidence for the presence of multiple allelic differences between L*er* and Cvi in other regulatory loci of the ERECTA signaling pathway and present candidate genes for these loci. Analysis of other publicly available single gene mutation experiments, allowed us to place ERECTA upstream of a MAPK signaling cascade.

Introduction

Transcript levels of genes are influenced by many factors including their genetic background. The latter can be studied in a multifactorial way by application of genetical genomics (Jansen, 2003). In this concept, the level of expression of single genes in segregating populations is treated as a quantitative trait, for which quantitative trait loci (eQTLs) can be mapped. Using DNA microarrays for expression profiling, these eQTLs can be determined at a genome wide scale. This approach, introduced by Jansen and Nap, 2001, has been successfully applied to several model organisms like maize, mouse, man, yeast (Brem et al., 2002; Schadt et al., 2003; Yvert et al., 2003; Monks et al., 2004) and recently also to *Arabidopsis thaliana* (Vuylsteke et al., 2005; DeCook et al., 2006; Keurentjes et al., 2007; West et al., 2007). In general, QTL analysis results in the identification of genome regions, responsible for the observed variation in the trait under study. As these regions contain hundreds of genes it is often difficult to identify the gene(s) responsible for that specific trait. Genetical genomics offers the possibility to limit the number of candidate genes for expression and phenotypic QTLs.

In mouse, for example, Mehrabian et al., 2005 pursued the gene underlying the QTLs for traits related to cardiovascular and metabolic diseases. They validated the role of their prime candidate gene, *ALOX5*, by comparing the gene expression profile of the *alox5* mutant to the eQTLs mapped in the same segregating population. The overlap between the genes with changed transcript levels in the mutant and the genes with an eQTL on *ALOX5*, confirmed that this gene is a susceptibility gene for obesity and bone traits. Another strong feature of genetical genomics is the possibility to construct genetic regulatory networks (Zhu et al., 2004; Bing and Hoeschele, 2005; Li et al., 2005; Mehrabian et al., 2005; Keurentjes et al., 2007). Based on the assumption that changes in transcript levels of regulatory (transcription) factors, underlying the eQTLs of the regulated genes, are co-regulated with those genes (Bing and Hoeschele, 2005) networks can be constructed. They can be further refined in combination with SNP information (Li et al., 2005; Mehrabian et al., 2005), Bayesian network modeling (Zhu et al., 2004; Li et al., 2005) or extended by Iterative Group Analysis (IGA) (Breitling et al., 2004; Keurentjes et al., 2007).

IGA was applied in a recent experiment on a segregating population of Arabidopsis recombinant inbred lines (RILs), derived from the parental accessions Landsberg *erecta* (L*er*) and Cape Verde Islands (Cvi) (Keurentjes et al., 2007). The most influential locus in this RIL population was the Erecta locus. More eQTLs mapped to this locus than can be expected by chance and it is therefore a Hotspot for Transcript Regulation (HTR). In L*er* this locus contains the mutated *ERECTA* gene, whereas Cvi has a functional *ERECTA* gene, making this gene a prime candidate for causing the HTR. The ERECTA protein is a membrane bound leucine-rich repeat receptor-like Ser/Thr kinase (LRR-RLK) (Torii et al., 1996), known to regulate developmental, hormone and defense processes (Godiard et al., 2003; Shpak et al., 2004; Llorente et al., 2005). Despite or maybe due to its extreme pleiotropic effects, direct molecular targets of ERECTA have not been identified so far.

In this chapter we present a novel way to utilize genetical genomics and discuss the resulting new insights in regulatory networks. To this end we combined genetical genomics on L*er*/Cvi RILs with expression profiling of the wild type versus *erecta* mutant.

This resulted in three major findings: i) It confirmed the presumed regulation of a significant number of genes by the main HTR in our previous study (Keurentjes et al., 2007), ii) It enables the extension of single gene mutant analysis to identify in detail interactions in its downstream regulatory processes, and iii) It allows the integration into the network of genes that have no significant eQTLs themselves, through their effect on transcript levels measured in single gene mutation experiments.

Applied to the extremely pleiotropic Arabidopsis gene *ERECTA*, we confirm that this is the causal gene for the major HTR in the Arabidopsis L*er*/Cvi RIL population. We present evidence for the existence of multiple allelic differences between L*er* and Cvi in other regulatory loci of the ERECTA signaling pathway and present candidate genes for these loci. Extending the analysis to other publicly available single gene mutation experiments allowed us to link ERECTA to a downstream MAPK signaling cascade.

Results

Strategy to analyze the ERECTA controlled transcription network

We determined the differently abundant transcripts (DATs) between the ERECTA mutant Landsberg *erecta* (L*er*) and its wild-type Landsberg (Lan) parent. As the Cape Verde islands (Cvi) accession also has a functional *ERECTA* allele, in theory, genes of which the transcript levels are controlled by ERECTA, are expected to be differentially abundant between L*er* and Lan (Fig. 4.1a) and to show an eQTL on the ERECTA locus in the L*er*/Cvi population (Fig. 4.1b).

Not all genes, however, will be directly regulated by the ERECTA protein, but are regulated via downstream genes. If all downstream genes are identical in L*er* and Cvi, all L*er*Lan DATs will have an eQTL that maps to *ERECTA* only. If one or more of the downstream genes are polymorphic between L*er* and Cvi, and cause differences in gene expression levels of their target genes, part of the L*er*Lan DATs will have an eQTL on this downstream regulator (Fig. 4.1c). Combination of these single gene mutation and multifactorial (Ler/Cvi RILs) experiments therefore, can be used to identify regulatory factors downstream of ERECTA and to start the construction of a transcriptional regulatory network.

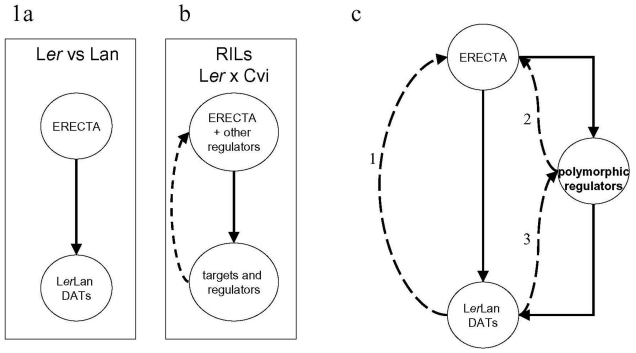


Figure 4.1 Microarray analysis combined with genetical genomics can reveal part of the signal transduction pathway downstream of a gene with a functionally different transcript abundance, in this case ERECTA.

a. Microarray analysis of L*er* vs. Lan results in differentially abundant transcripts, caused by the *erecta* mutation (solid arrow).
b. In the L*er*/Cvi RILs ERECTA is polymorphic and will affect transcript abundance, resulting in eQTLs on the Erecta locus. Besides ERECTA, many other regulators are polymorphic and will affect transcript abundance resulting in eQTLs on other loci too. The effects of both are depicted with the dashed arrow.
c. The eQTLs for the DATs will identify downstream regulators that are polymorphic between L*er* and Cvi. If the transcript abundance of regulators downstream of ERECTA is affected by the polymorphism in the ERECTA gene (solid arrow) than those regulators will have eQTLs mapping to the Erecta locus (dashed arrow). The transcripts affected by those regulators (solid arrow) will have eQTLs mapping to the Erecta locus (dashed arrow). Moreover in case those regulators are polymorphic between L*er* and Cvi, the affected genes (L*er*/Lan DATs) will also have an eQTL mapping to the physical position of that polymorphic regulator (dashed arrow).

Estimating the effect of *erecta* in the L*er*/Cvi RIL population

Seedlings (7 d) of Lan and the *erecta* mutant L*er* were analyzed on DNA microarrays. Of the 24,065 genes represented on the arrays, 2049 transcripts (8.5%) were found to be differentially abundant (L*er*Lan DATs; false discovery rate (fdr) at 0.05). Of these L*er*Lan DATs, 1378 transcripts were less and 671 were more abundant in L*er* compared to Lan. In another experiment we used this same DNA microarray platform to measure the transcript levels of 7 d old seedlings grown under similar conditions in 159 L*er*/Cvi RILs (Keurentjes et al., 2007). In this study, eQTLs were identified for 4066 genes (fdr at 0.01), which is 17% of the total number of profiled genes. So, a random subset of genes will, on average, contain 17% of genes with at least one eQTL. Of the 2049 L*er*Lan DATs, 728 genes had at least one eQTL, with a total of 863 eQTLs. This is 36% of the DATs which is more than twice the genome-wide percentage of genes with significant linkage. Thus, the effect of the allelic difference of ERECTA between L*er* and both Lan and Cvi is strong enough to observe its effects in a multi-factorial perturbation experiment on the L*er*/Cvi RIL population.

Once the general overrepresentation of genes with an eQTL in the L*er*Lan DATs was discovered, we investigated if the eQTLs of these genes mapped to the Erecta locus. The experiments on the L*er*/Cvi RIL population showed that the Erecta locus contained most eQTLs. It was therefore designated a Hotspot for Transcript Regulation (HTR) (Keurentjes et al., 2007) (Fig. 4.2a). For eQTLs that map to this HTR, the gene *ERECTA* itself is a likely controlling element, due to its known polymorphism. However, no estimate could be given for the number of eQTLs at this locus that are actually influenced by the *ERECTA* gene. Of the 4066 genes with an eQTL a total of 176 genes (4.4%) map to the Erecta locus. Of these, 142 are distantly and 34 locally regulated. In the set of 728 DATs with an eQTL, 98 genes (13.4%) map to *ERECTA*, of which 87 are distantly and 11 locally regulated. So, almost 55% (98 of 176) of the genes with an eQTL at the Erecta locus have a direct link to the *erecta* mutation in L*er*. This is 61% for the distantly regulated genes and only 31% of the genes regulated locally.

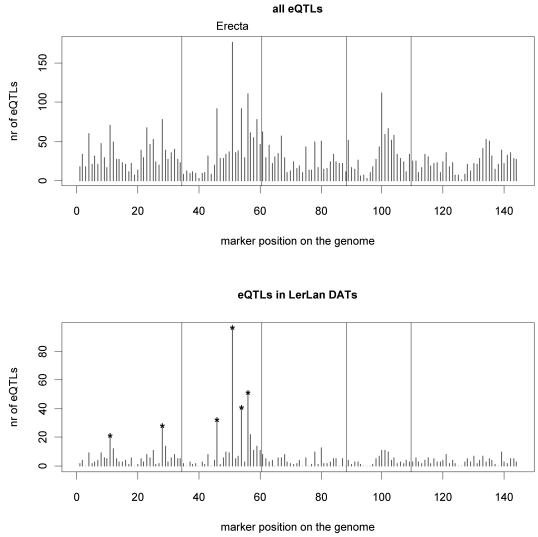


Figure 4.2 eQTL distributions of all profiled genes (a) compared to the genes with differently abundant transcripts (DATs) between Lan and L*er* (b).

a. For each marker position on the genome the number of eQTLs in the L*er*/Cvi RILs is plotted.
b. The distribution of eQTLs in the L*er*Lan DATs. Significantly over-represented markers are marked with an asterisk ($p < 0.01$).

Polymorphic regulators downstream of ERECTA

Of all the 728 L*er*Lan DATs with an eQTL, only 98 (13%) map to the Erecta locus. The remaining 630 (87%) eQTLs must therefore have arisen from nucleotide polymorphisms between L*er* and Cvi that have resulted in changes in gene expression downstream in the ERECTA pathway. We examined the distribution of eQTLs in the L*er*Lan DATs (Fig. 4.1b) to identify the candidate intermediate regulators. This resulted in 6 HTRs, including *ERECTA*, in the L*er*Lan DATs collection.

Overlap in eQTL profiles reveals shared regulation

To identify the polymorphic intermediate regulators, that influence the transcript levels of the many genes that map to the different HTRs, we searched for common eQTL profiles in the L*er*Lan DATs. When the regulatory function of these regulators is largely dependent on their transcript levels, a high correlation between the eQTL profiles of these regulators and their targets (L*er*Lan DATs) can be expected. By clustering the eQTL profiles of the L*er*Lan DATs in a 3 by 3 Self Organizing Map (SOM), a specific cluster containing 643 genes was found (Fig. 4.3a). The mean eQTL profile of the genes in this co-regulated cluster (Fig. 4.3b) has a broad eQTL peak on chromosome 2. This peak contains *ERECTA* and the 3 other HTRs on the same chromosome. Another locus, at the bottom of chromosome 1, also co-locates with an HTR. For this latter peak, a -log (P) ± 1.5 interval could be defined ranging from marker 18 (71.9 cM) to 31 (117.4 cM), with the peak at marker 28 (101.6 cM).

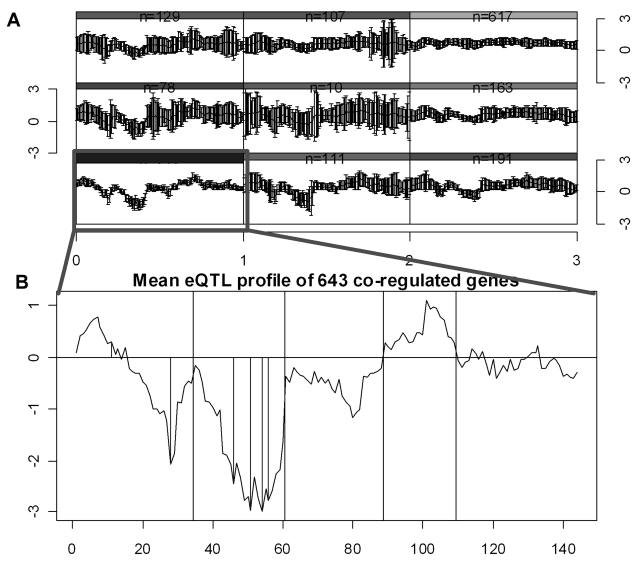


Figure 4.3 Clustering eQTL profiles
a) Self organizing map (SOM) based on PCA of eQTL profiles of L*er*Lan DATs, which shows common regions of genetic regulation of transcript abundance for 643 genes in bottom left cluster.
b) Mean eQTL profile of co-regulated cluster. Average significance of linkage (y-axis) is plotted at each marker position on the genome (x-axis). The sign is indicative for the additive effect of the L*er* allele on the log-ratio. HTRs are indicated with vertical bars.

Transcription factors (TFs) regulated by ERECTA

The 643 genes with the shared eQTL profile (Fig. 4.3) are expected to share the same regulatory control, which could be one or more Transcription Factors (TFs). To investigate this, we examined this cluster of genes for overrepresentation of TF binding sites. In the 3000 bp regions upstream of the start codon of the 643 genes, two significantly over represented binding sites were found. These are the CARCGW8GAT motif in 497 genes and the W-box in 550 genes (both $p < 10^{-10}$). The CARCGW8GAT motif is the specific binding site for the MADS-BOX gene AGL15 (Tang and Perry, 2003). The gene AGL15 does not co-localize with any of the HTRs. The W-box motif is a binding site for WRKY (named after their conserved amino acids) class TFs. In Arabidopsis, this family of TFs has 74 members (Eulgem et al., 2000; Yamasaki et al., 2005). Several WRKY genes are located near identified HTRs and we further analyzed the involvement of WRKY TFs in the co-regulation of the 643 genes in the selected cluster.

The L*er*Lan DATs contained 115 TFs, of which 13 belong to the WRKY class of TFs. The WRKYs underlying the peaks of the shared eQTL profile are considered the primary candidates. Of these WRKYs, WRKY6 is physically located in the –log (P) ± 1.5 eQTL interval (71.9 – 117.4 cM) on chromosome 1 defined for the mean QTL profile. On chromosome 2, the WRKY TFs 15, 25, 33, 44, 46 and 59 are located within the broad eQTL interval. Together, this reduced the number of candidates TFs to seven.

Co-expression analysis of selected TFs

To determine if any of these 7 genes are responsible for affecting transcript levels downstream of ERECTA, a co-expression analysis was performed. Of each of the 7 selected WRKYs, the 1000 highest correlated genes were compared to each other and the L*er*Lan DATs. The WRKYs that showed the highest overlap between these gene sets are the most likely candidates for the regulation of the majority of the 643 clustered genes. These are WRKY6, WRKY15, WRKY25, WRKY33 and WRKY46. For each of these WRKYs almost 400 genes of their corresponding top 1000 gene sets overlap with the L*er*Lan DATs. Their gene sets show even higher overlap between each other (table 4.1). These WRKYs are themselves part of the co-regulated cluster. WRKY44 and WRKY59, on the other hand, do not show a clear correlation.

Table 1 Genes sets of the 1000 most positively correlated genes to the candidate WRKY TFs and their overlap with the L*er* vs. Lan DATs and each other. LANLERDE: stands for the genes with differently abundant transcripts between L*er* and Lan. (website: www.arabidopsis.leeds.ac.uk/act/coexpanalyzer.php)

		WRKYs						
	LANLERDE	**(6)**	**(15)**	**(25)**	**(33)**	**(44)**	**(46)**	**(59)**
WRKY6	392	X	628	594	584	3	614	5
WRKY15	392	628	X	892	842	1	833	6
WRKY25	387	594	892	X	793	1	839	4
WRKY33	362	584	842	793	X	4	807	6
WRKY44	29	3	1	1	4	X	3	6
WRKY46	380	614	833	839	807	3	X	3
WRKY59	23	5	6	4	6	6	3	X

WRKY6 is polymorph between Ler and Cvi

WRKY6 belongs to group II of the WRKY family of TFs, containing one WRKY domain. The gene contains 6 exons, translated into a protein of 553 amino acids. The region spanning amino acids 251 – 294 in the 4th exon contains a nuclear targeting signal (Robatzek and Somssich, 2001). Sequencing of WRKY6 revealed several polymorphisms between the L*er* and Cvi allele, one resulting in the substitution of Threonine at position 286 by Alanine in L*er* (data not shown).

Downstream targets of ERECTA that regulate the WRKYs

In order to find the components in the signal transduction cascade between ERECTA and the WRKYs, we studied, based on literature and our own data, WRKY regulation in more detail. WRKY TFs are associated with Mitogen Activated Protein Kinase (MAPK) signaling cascades that are induced during plant defense (Asai et al., 2002). WRKY25 and WRKY33 proteins are *in vitro* substrates of MPK4 (Andreasson et al., 2005) and interact in a yeast two-hybrid screen with MKS1, a substrate of MPK4. Furthermore MPK4 is part of a MAPK cascade consisting of the MAP3K MEKK1 and the MAP2Ks MKK1 and MKK2 (Asai et al., 2002; Teige et al., 2004; Andreasson et al., 2005; Meszaros et al., 2006; Suarez-Rodriguez et al., 2007). Moreover, MPK4 itself is one of the genes contained in the co-regulated cluster of genes sharing the common eQTL profile, along with the selected WRKY TFs. Although MAPK transcript levels in general do not correlate with the activity of the protein (Ichimura et al., 2006), this is an indication of feedback regulation on the MAPK cascade through WRKY gene expression. In order to investigate the involvement of MPK4 downstream of ERECTA we utilized publicly available microarray data to overlay our genetical genomics experiments.

Common eQTL profiles in MAPK cascade mutants

Transcript profiles of several mutants or over-expressor lines of upstream components of the MPK4 signaling cascade have been determined and analyzed in previous studies (Teige et al., 2004; Andreasson et al., 2005; Meszaros et al., 2006; Suarez-Rodriguez et al., 2007). We re-analyzed the profiles in the knockout mutants *mpk4*, *mkk1* and *mekk1*, as well as in the over-expressors of *MKK2* and *MKS1* in a similar manner used to analyze the L*er*Lan DATs. Our analysis showed a strong enrichment of eQTLs in the *mekk1*, *mkk1* and *MKK2* DATs (29%, 29 % and 33% resp., compared to 17% genome wide). In particular, the eQTL distribution of *mekk1* and *MKK2* was very similar in profile and HTRs, as seen for *erecta* (Fig. 4.4b and 4.4c). Together, this provides strong evidence to position *MEKK1* and *MKK2* downstream of *ERECTA*.

A double mutant er/mpk4 shows an additional cascade converging downstream of erecta

Surprisingly, the knockout mutant of *MPK4,* a downstream target of MKK2, had much less enrichment for eQTL (23%), but a very similar eQTL distribution as the genes downstream of MEKK1, MKK2 and ERECTA (Fig. 4.4e). This similar distribution was not expected, because the *mpk4* mutation is in the *erecta* background. Genes completely dependent on ERECTA could never be identified in a comparison between an *erecta* mutant (L*er*) and a double mutant (*er/mpk4*). Genes with DATs in *er/mpk4* compared to L*er* do show overlap

with genes downstream of ERECTA, indicating that these genes are also dependent on ERECTA. However, this dependency has to be partial, since the MPK4 comparison was made in the absence of functional ERECTA. A likely explanation can be the involvement of an additional regulator, acting through a common part of the pathway downstream of ERECTA, comprising MPK4 (Fig. 4.4a). This would indicate that at the level of MPK4, or upstream, different signaling pathways converge with the ERECTA signaling cascade. Compared to the L*er*Lan DATs, the eQTL distribution of the genes with DATs between L*er* vs. *er/mpk4* showed an enrichment of eQTL on chromosome 4 (Fig. 4.4e). This enrichment of eQTL disappeared in the eQTL distribution of the genes with DATs in both, L*er* vs. Lan and L*er* vs. *er/mpk4* (Fig. 4.4f). It became more pronounced in the L*er* vs. *er/mpk4* exclusive set of DATs (Fig. 4.4g). This locus probably harbors the additional regulator. In that case, due to its position on chromosome 4, MKK1 is a likely candidate. Its transcript levels are locally regulated and this gene is known to interact with MPK4 (Meszaros et al., 2006). As *MKK1* is not differentially abundant in L*er* vs. Lan, we expect it to act independent of ERECTA. We obtained more evidence for this idea when we investigated the transcripts that are differentially abundant between L*er* and *er*/*mpk4,* but not between L*er* and Lan. The HTRs on chromosomes 1 and 2, containing the WRKY TFs, are still present in the eQTL distribution for this subset of genes, confirming that the WRKY TFs can be placed downstream of MPK4 (Fig. 4.4g). Most importantly, the HTR on *ERECTA* disappeared from the eQTL distribution (Fig. 4.4f compared to 4g), consistent with the positioning of ERECTA as one of the upstream regulators of MPK4. On chromosome 4, the enrichment of eQTL was more profound, reinforcing the hypothesis that the additional upstream regulator at this locus acts independent of ERECTA.

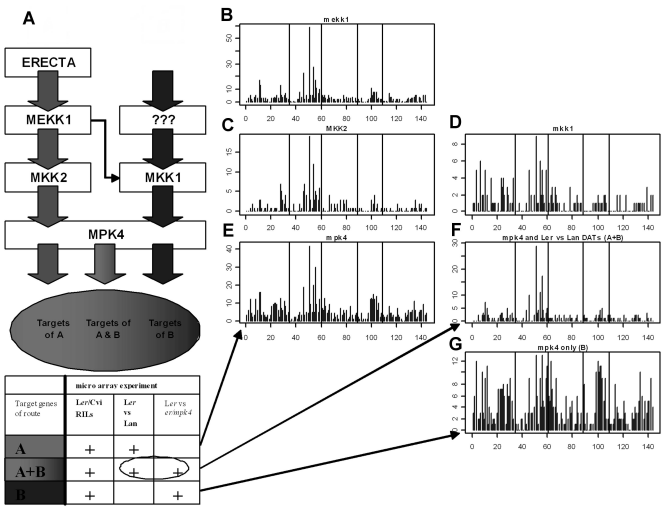


Figure 4.4 Model explaining the likelihood of an additional regulator upstream of MPK4 that acts independently of ERECTA. The scheme shows the types of genes affected by ERECTA (A), the additional regulator (MKK1) (B) or both (A&B). b) to g) eQTL distributions caused by genes downstream of ERECTA, markers on x-axis and number of eQTL on y-axis. The mutants in the MPK4 cascade result in similar eQTL distributions found for the L*er*Lan DATs as shown for mekk1 (b), MKK2 (c), mkk1 (d) and mpk4 (e) mutants. The A&B group show similar eQTL distribution as the A genes as shown in the overlap between L*er*Lan DATs and mpk4 affected genes (f). The genes in B, still display the eQTL in the HTR harboring the WRKY TFs (f-g), but lack the regulation by ERECTA (g).

Discussion

Network analysis based on eQTL and monogenic mutant analysis

The identification of genes underlying eQTLs is an important first step in the construction of genetic regulatory networks. The chromosomal regions that underlie the expression Quantitative Trait Loci (eQTLs) often span several megabases and may contain several hundreds of candidate genes responsible for the trait. To identify more detailed genetic or regulatory interactions, different methods have to be employed, e.g. Bayesian approaches (Zhu et al., 2004), co-expression analysis (Bing and Hoeschele, 2005), and protein-protein interaction data (Franke et al., 2006). In this study we used single gene mutant analysis to validate the genetic interactions downstream of *ERECTA*, the prime candidate for the largest Hotspot for Transcript Regulation (HTR) identified in the L*er*/Cvi Recombinant Inbred Line (RIL) population and the genes with eQTLs that map to this HTR. Besides this validation, we were able to deduce other regulatory steps downstream of *ERECTA*, by combining clustering, Transcription Factor Binding Site (TFBS) analysis, co-expression analysis and additional single gene mutant experiments (publicly available). Especially useful proved the eQTL distribution analysis of groups of Differentially Abundant Transcripts (DATs) from several mutants. If the mutated gene (or either an up- or downstream component of its signaling cascade) is polymorphic between L*er* and Cvi the specific HTR can be used in constructing a putative transcription regulatory network.

ERECTA is the major HTR in L*er*/Cvi RILs

The largest collection of eQTLs at a single marker found by Keurentjes et al. (2007) was on the *ERECTA* locus. Because of the possibility of other polymorphic genes in this area with a strong influence on gene expression, this cannot be solely attributed to *ERECTA*. By analyzing the differences between L*er* and Lan we were able to show that almost 60 percent of the genes with a distant eQTL mapping to the Erecta locus are truly regulated by the *ERECTA* gene. This is most likely an underestimate as it is known that expression of the *ERECTA* gene shows temporal and spatial regulation during plant development (Yokoyama et al., 1998). In this study we could only identify those transcripts that are influenced by ERECTA specifically at the developmental stage (7 d old seedlings) used under our experimental conditions.

MAPK signaling and WRKY Transcription Factors

Other Hotspots for Transcript Regulation (HTRs) were identified in the group of DATs between L*er* and Lan. These contain genes that operate, at least partially, downstream of *ERECTA*. The genes that mapped to these HTRs were enriched for the W-box TFBS. Co-expression analysis of relevant regulatory WRKY TFs resulted in the selection of 5 candidate WRKY TFs that may be downstream regulators in the ERECTA pathway (Fig. 4.5). Moreover, it also indicates that they are polymorphic between L*er* and Cvi, like WRKY 6. We found strong evidence that the Leucine Rich Repeat – Receptor Like Kinase (LRR-RLK) ERECTA protein is regulating WRKY type transcription factors 6, 15, 25, 33 and 46. Other LRR-RLKs are also capable of activating WRKY TF expression. It is known that LRR-RLK FLS2 regulates the expression of WRKYs 22 and

29 via the MEKK1-MKK4/5-MPK3/6 kinase cascade after recognition of bacterial flagellin (Asai et al., 2002). This co-expression and possible redundancy of WRKYs has been described in several other studies. Besides WRKY22 and 29, WRKY11 and 17 are found to act partially redundant as negative regulators of basal resistance (Journot-Catalino et al., 2006). Similarly, WRKYs 18, 40 and 60 show partial redundancy in response to different microbial pathogens (Xu et al., 2006). In our experiment we found multiple WRKYs (6, 15, 25, 33 and 46) acting together as likely regulators downstream of ERECTA.

The redundancy in signaling is also present at the level of the LRR-RLKs, where ERECTA and its homologues ERL1 and ERL2 act synergistically (Shpak et al., 2004). These LRR-RLKs have also been connected to another MAP3K, YODA (YDA), in a study of Ingram *et al*. (2005)(Ingram, 2005), where they were shown to be involved in stomatal development (Shpak et al., 2005; MacAlister et al., 2007; Pillitteri et al., 2007).

Map kinase cascades are described as a general activating mechanism for WRKY type transcription factors (Asai et al., 2002). Although MEKK1 can signal through multiple MAP2Ks and possibly be activated by multiple LRR-RLKs, we found evidence to place the MEKK1-MKK1/2-MPK4 signaling cascade between ERECTA and the selected WRKY TFs (Fig. 4.5). With the mutant *mpk4*, that was generated in L*er* we were able to identify a set of transcripts that share part of the ERECTA-induced MPK4 signaling cascade, but are also controlled by a regulator other than ERECTA. Regulation of transcript level abundance of this set of genes is likely mediated by MKK1.

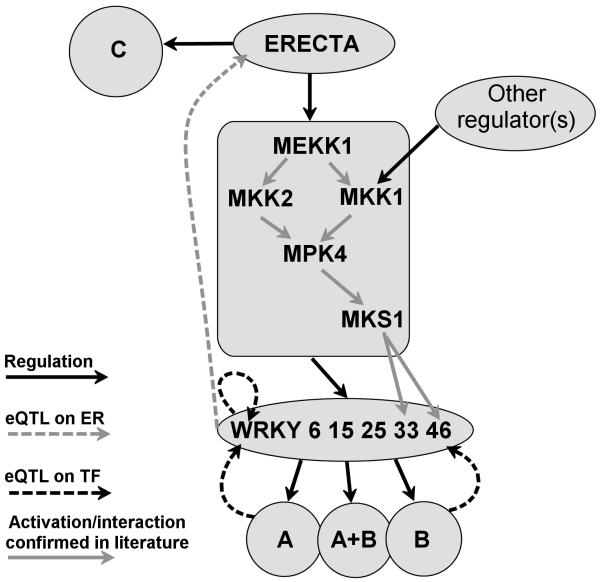


Figure 4.5 Model of ERECTA controlled transcription network
ERECTA is regulating transcript abundance through the MEKK1-MKK1/2-MPK4 cascade. This cascade can also be activated independent of ERECTA through MKK1. The candidate WRKY TF regulators are placed downstream of ERECTA, regulated through the MPK4 signaling cascade. The signaling steps are by black arrows, in grey the interactions are shown that could be confirmed in literature and could be confirmed by our data, based on eQTL enrichment and distribution profiles. The dashed lines indicate eQTL. The WRKY TFs are the regulators downstream of ERECTA, as deduced from the TFBS enrichment analysis of the co-regulated genes in the L*er* vs. Lan DATs.

Targets of the WRKYs

WRKY TFs are known for their ability to influence their own expression. In *Petroselinum crispum* (parsley) PcWRKY1 regulates its own expression by binding to 3 W-box elements in its own promotor (Turck et al., 2004). WRKY33 is an orthologue of PcWRKY1, and based on our eQTL analysis, this WRKY might very well regulate its own expression, although other WRKY TFs are also likely involved in its regulation (Lippok et al., 2007). Like WRKY33, MPK4 shares a common (regulatory) eQTL profile, indicating that MPK4 might be under the same genetic regulatory control and as such contributing to the feedback mechanism. Furthermore Robatzek and Somssich (2002) (Robatzek and Somssich, 2002) describe auto regulation for WRKY 6. Most of the targets they identified were confirmed in our study. MPK4 is known as a negative regulator of Salicylic Acid (SA) and Jasmonic Acid (JA)/Ethylene (ET) dependent responses, especially via EDS1 and PAD4. These marker genes for SA responses act independently of ERECTA. Other marker genes regulated downstream of an MPK-signaling cascade were in addition influenced by ERECTA, like *CHIB, PR1, PR4* and *VSP1*. We could not determine involvement of MPK signaling for *NPR1*, *PAL1* and *LOX1* and only detected regulation by ERECTA.

Phenotypes of the downstream targets of ERECTA

Arabidopsis plants with mutated ERECTA, including L*er*, all have a small compact rosette (Torii et al., 1996). This difference in growth is, partly, also found when genes downstream of ERECTA are altered. The most extreme example is the *mpk4* mutant in a L*er* background which has a severe dwarf phenotype (Andreasson et al., 2005). A substrate of MPK4, MKS1, has less effect on the phenotype when over-expressed. The plants with 35S::MKS1 in a L*er* background are compact and a little bit smaller than L*er*, although not as small as the *mpk4*/*er* mutant. The RNAi line of MKS1, also in a L*er* background shows a less compact rosette than L*er*, due to its longer petioles. When the RNAi construct of MKS1 is introduced in an *er*/*mpk4* mutant background the plants are phenotypically very similar to L*er* confirming its opposite effect on growth. The mutants of MKK1 and MKK2, of which MPK4 is a direct target, where only available in a Col background, nevertheless *mkk2* shows a growth phenotype (Teige et al., 2004). When grown under normal conditions it is a little bit smaller than Col, however, when treated with cold or salt this difference is much more pronounced. One step further upstream functions MEKK1. Like *er*/*mpk4*, *mekk1* shows a dwarf phenotype when grown at 24°C, but this was much less severe at 32°C (Suarez-Rodriguez et al., 2007). Of the WRKY TFs placed just beneath this map-kinase cascade some mutants were described. When WRKY6 was over-expressed in a Col background this resulted in small shunted plants (Robatzek and Somssich, 2002). Smaller rosettes were also found when WRKY33 was mutated (Zheng et al., 2006). Over expressing WRKY33 did not result in an altered rosette growth although the leaves seem somewhat smaller (Andreasson et al., 2005). The fact that mutants of genes that possibly operate downstream of ERECTA, show a similar or stronger effect on growth than the *erecta* mutation, is consistent with their role in the ERECTA pathway.

Conclusion

Strong evidence was found that ERECTA is regulating a group of WRKY transcription factors. By a combination of genetical genomics, single mutant experiments and literature research we showed that the MEKK1-MKK1/2-MPK4 signaling cascade constitutes the likely steps between ERECTA and these WRKYs.

Methods and Materials

Plant material

The accessions L*er* and Lan were grown for 7 days and the aerial parts harvested as described in Keurentjes et al. (2006).

Microarray procedures

Sample preparation, hybridization and analysis were performed as described in Keurentjes et al. (2007). Arabidopsis DNA microarrays were provided by the Galbraith laboratory (University of Arizona, Tucson, AZ) and were produced from a set of 70-mer oligonucleotides, representing 24065 unique genes (Array-Ready Oligo Set, version 1.0, Qiagen-Operon). DNA probes immobilization and hybridization was performed according to instructions from the Galbraith laboratory. Arrays were scanned by using a ScanArray Express HT (PerkinElmer, Wellesley, MA) and quantified by using Imagene 6.0 (BioDiscovery, El Segundo, CA). Microarray analysis was performed using the R-package Limma (Smyth et al., 2005; Ritchie et al., 2006; Ritchie et al., 2007).

Statistics

We used the eQTLs as calculated by Keurentjes (2007). To determine significant overrepresentation of eQTLs at markers we used 1000 permutations. One permutation was a randomly drawn subset, out of the total set of genes profiled. This random subset had the same size as the specific subset to be tested. After all the permutations, confidence intervals were determined assuming a normal distribution of eQTLs per marker. TFBS overrepresentation was determined with the Athena database (O'Connor et al., 2005; http://www.bioinformatics2.wsu.edu/cgi-bin/Athena/cgi/analysis_select.pl). Co-expression analysis was performed using the Arabidopsis Co-expression Tool (ACT) (Manfield et al., 2006; http://www.arabidopsis.leeds.ac.uk/act/coexpanalyzer.php). The SOM function in the statistical software program R was used to find correlating sets of QTL profiles within a specific group of genes.

Chapter 5

Genetical genomics reveals natural variation in constitutive and environmentally modulated transcript regulation.

L. Basten Snoek & Inez R. Terpstra, Rene Deckter, Laurentius A.C.J. Voesenek, Guido van den Ackerveken & Anton J.M. Peeters

Abstract

Genome*environment interactions are the main determinants of gene transcription levels. Both environment and genome are highly variable. Unraveling how natural genomic variation affects gene expression in reaction on changing environmental conditions and subsequent variation in phenotypes is one of the major goals of genetics. We compare a genetical genomics experiment in which Recombinant Inbred Lines of *Arabidopsis thaliana* are treated with neutral shade, to a previous experiment, both using the Landsberg *erecta* x Cape Verde islands population. The differences in treatment and stage of development between the two experiments are used to gain insight in the influence of variation of both factors on gene transcription and regulation. The natural genetic variation within these individual experiments is identical whereas environmental factors are different between the two experiments. Genetic parameters like heritability, transgression and Single Nucleotide Polymorphism linkage are evaluated and insight in the biological processes affected by the natural variation in gene expression due to the genome or genome*environment interaction are presented. We found evidence for both constitutive and plastic regulation of gene expression.

Introduction

Many developmental processes in organisms and the response to the environment are directed and modulated by genetic components. Therefore most of the variation present in those responses is the result of complex interactions among gene products specific for an environment or developmental stage.

The static genotype of an organism is transcribed into a dynamic pool of transcripts. Most of these transcripts are translated into proteins, which in turn perform most of the cellular functions and so determine the functioning of the whole organism. One of the important factors that determine which functions are performed and to what extent, is the level at which a protein is present. The levels of many proteins are, at least in part, determined by the abundance of their RNA templates. Given the genetic make-up of an individual organism, transcript abundance is the first level at which variation in development and response to environmental changes can occur.

The complexity and net-like structure of signaling and response cascades puts a limit to the possibility to unravel molecular mechanisms underlying developmental and environmental responses through single mutant analyses. By studying the combination of natural variation and high-throughput transcript profiling, these net-like signaling cascades can be unraveled (visualized) through multi-factorial experiments. Variation in genotype leading to differences in transcript levels can be used to find Quantitative Trait Loci (QTL) for these transcript levels (e(xpression)QTLs). These eQTLs can be used to construct transcription regulatory networks, since the locus of the component (or gene) influencing transcript abundance is known, due to the location of the eQTL for that transcript.

This technique “Genetical Genomics” (GG) has been applied successfully in a number of model organisms (Brem et al., 2002; Bing and Hoeschele, 2005; Brem and Kruglyak, 2005; Bystrykh et al., 2005; Li et al., 2006; Keurentjes et al., 2007; West et al., 2007) uncovering a vast amount of (natural) variation in transcript abundance, induced by the recombination of genetic backgrounds. Subsequently, transcription regulatory networks of different complexity, revealed by phenotypes for which natural variation was found, have been constructed through genetical genomics (Jansen and Nap, 2001; Bing and Hoeschele, 2005; Keurentjes et al., 2007). In *Arabidopsis thaliana*, two genetical genomics experiments have been performed (Keurentjes et al., 2007; West et al., 2007). In Keurentjes et al. (2007, Genetical Genomics 1; GG1) the genetical genomics data was further used as a proof of concept to construct a transcription regulatory network of genes involved in flowering time.

Because the individual genotypes were grown in one static environment, only developmental responses, for which transcript levels are settled, can be studied. In a *Caenorhabditis elegans* RIL population the comparison of gene expression when grown in two different environments show the highly plastic nature of gene expression regulation (Li et al., 2006). Many responses to a change in the environment are relatively fast and the result of transient changes in transcript abundance. Therefore, much of that particular natural variation will only manifest itself after induction. This made the previous studies less suitable for studying transcript regulation of a fast inducible response like hyponastic growth. To study the differences (and similarities) in transcript regulation, when Arabidopsis deals with changes in the environment, we treated the Arabidopsis

Recombinant Inbred Line (RIL) population derived from the accessions Landsberg *erecta* (L*er*) and Cape Verde Islands (Cvi) (Alonso-Blanco et al., 1998) with 3 hours of neutral shade. This treatment is known to affect transcript levels (Millenaar et al., 2006) as well as phenotypic traits like hyponastic growth for which natural variation was identified in this population ((Millenaar et al., 2005; Chapter 2, this thesis).

We measured genome-wide transcript levels of 120 L*er* x Cvi RILs (Alonso-Blanco et al., 1998). To investigate the influence of genetic background and environment on transcript levels, we compared the genetical genomics experiment described in this paper (GG2) with our previous experiment using the same population (GG1) (Keurentjes et al., 2007). The micro-array experimental design (Fu and Jansen, 2006) used in the two experiments was identical. Observed similarities in the influence of genotype on gene expression denote the constitutive influence of genetic background on transcript abundance. As growth conditions, developmental stage and treatment were different between the two experiments, additional influence of genetic background dependent on these conditions could be observed. In GG1, the aerial parts of 7-day old seedlings, grown on agar under long-day conditions were used. In GG2, wholly developed leaves of 3-week old plants, grown on soil under short-day conditions and treated with 3 hours of shade were used. Expression profiling in GG1 was performed on the 70-mer oligonucleotide spotted Arabidopsis Oligonucleotide Microarrays (http://ag.arizona.edu/microarray) representing 24065 genes. In GG2, the expression profiles were measured on 150-450-mer spotted arrays, CATMA version 2 (Allemeersch et al., 2005), representing 20833 genes. The number of genes represented on both arrays was 19205. Altogether, this comparison is an excellent opportunity to investigate robust constitutive influence of genetic background on transcript abundance. Furthermore, plasticity eQTL, specific for a developmental stage or treatment can be studied, unraveling further the interaction between genotype and environment in relation to transcript abundance.

Heritability, transgression and constitutive eQTLs are compared and discussed. Furthermore, plastic eQTLs, specific for a developmental stage or treatment were found. Groups of genes over-represented with eQTLs, showed which biological processes are affected by genotype*environment interactions. This enables further unraveling of the interaction between genotype and environment in relation to transcript abundance.

Results

Genetic analysis of gene expression in two environments/developmental stages reveals a moderate constitutive influence of the genetic background

Only a small fraction of the differentially abundant transcripts (DATs) between the parents is constitutively different

To estimate the parental influence on transcript abundance, we measured the DATs between L*er* and Cvi. Between them, 3228 genes had DATs (false discovery rate (fdr) at 0.05) after 3 hours of low light treatment (GG2). This number of DATs is considerably higher than in the previous experiment (GG1) in which only 808 DATs were observed (fdr at 0.05) (Keurentjes et al., 2007). The overlap contains a group of 262 genes, representing DATs independent of the two tested developmental stages or environments. The transcript levels of the genes with DATs between the parents, in one or both experiments, are expected to be largely determined by genotype. So they are expected to have an eQTL, unless this difference between the parents is caused by many alleles with small effects or environmental noise.

Heritability; the influence of the genotype is increased during treatment and development

By comparing the broad sense heritability (H^2) of the transcript levels in the two experiments of both parents and RILs, we could study the influence of genetic background on variation in transcript levels between the RILs compared to the total variation in transcript abundance. The median H^2 calculated for the RILs, was higher in both experiments than the median H^2 of the parents (Fig. 5.1 and SI Table 5.1). Of the parents, the median H^2 in GG2 was higher than the median H^2 in GG1, whereas GG1 showed a little higher median H^2 for the RILs (Fig. 5.1).

To investigate whether the influence of the genotype on transcript levels is similar in the parents and RILs, we compared the H^2 of individual transcripts in the parents versus the H^2 in the RILs (Fig. 5.2A, B, SI Table 5.2). In GG1, ~8% of the genes show a H^2 >0.5 in both parents and RILs (Fig. 5.2A). In GG2, this was almost twice as high with ~15% of genes with a H^2 >0.5 (Fig. 5.2B). In GG2 a distinct group of genes with a H^2 >0.75 in both parents and RILs could be identified (Fig. 5.2B). The transcript levels of these genes are highly controlled by genotype under the conditions of GG2 in both parents and RILs.

We also compared the influence of genetic background in the parents and the RILs between the two genetical genomics experiments. When the transcript levels of the parents in GG1 were compared to those in GG2, ~4% of genes had a H^2 >0.5 in both (Fig. 5.2C). For the RILs in GG1 and GG2, ~14% of genes showed a H^2 >0.5 in both experiments (Fig. 5.2D). Taken together, this shows that only a moderate number of transcripts are under constitutive genetic control and are thus hardly influenced by the environment. Furthermore, the increased heritability of transcript levels in GG2 shows an increased influence of the genotype during treatment/development.

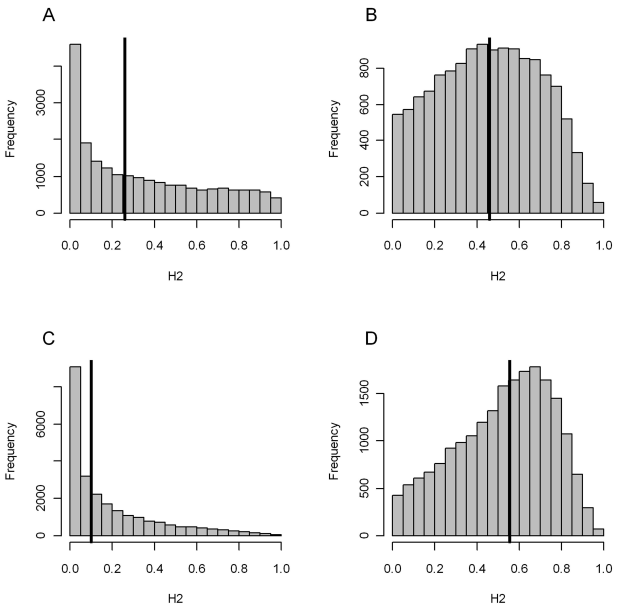


Figure 5.1 Frequency distributions of the heritability (H^2) of the transcript measured in GG2 (this chapter) in the parental lines (L*er* and Cvi) (A) and in the recombinant inbred lines (RILs) (B), in GG1 (Keurentjes et al., 2007) in the parents (C) and in the RILs of GG1 (D). Median is indicated by the black vertical bar.

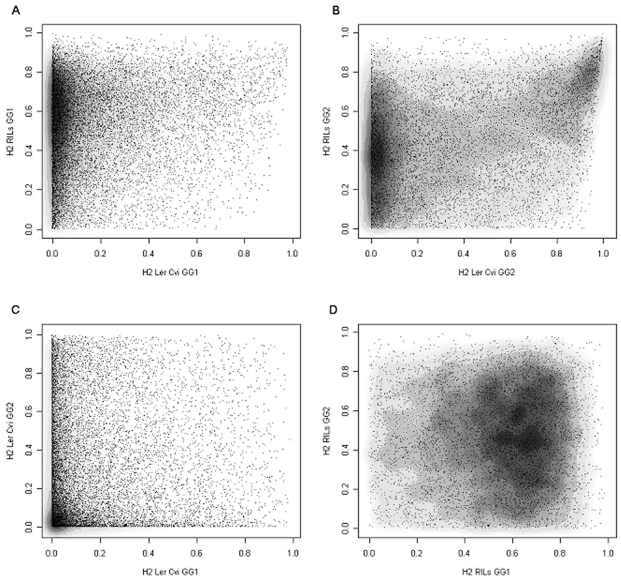


Figure 5.2 Broad sense heritability (H^2) of all transcripts compared between experiments and parents (L*er* and Cvi) and the recombinant inbred lines (RILs).
A) H^2 distribution of the parents (x-axis) and RILs (y-axis) of GG1 compared
B) H^2 distribution of the parents (x-axis) and RILs (y-axis) of GG2 compared
C) H^2 distribution of the parents of GG1 (x-axis) and GG2 (y-axis) compared
D) H^2 distribution of the RILs of GG1 (x-axis) and GG2 (y-axis) compared

The number of genes showing transgression is decreased during treatment and development and is correlated to heritability

Transcript levels in the RILs can extend beyond the parental levels. This transgression is the result of the influence of multiple alleles that may have a balancing effect on transcript levels in the parents. In a segregating population, however, new combinations of these alleles in the individuals may alter this balance and result in an overall increased positive or negative influence on gene transcript levels. By comparing both experiments we were able to estimate the number of genes showing constitutive transgression.

We determined transgression by calculating the number of RILs exhibiting transcript levels beyond the means of the parental lines (Fig. 5.3A and SI Table 5.3). In GG1 ~31% of the genes showed a higher abundance, whereas ~22% of the genes showed lower transcript abundance in >30 RILs (threshold used in GG1 (Keurentjes et al., 2007) to determine transgression; ~19% of the lines). This was much lower in GG2 were ~5% of genes show higher and ~7% showed lower transcript abundance in >20 RILs (~17% of the lines; comparable to the GG1 threshold). Hardly any genes showed transgression to both sides (<1% in both experiments; SI Table 5.3). Only a small percentage of transcripts showed transgression in both GG1 and GG2, 1.6% showed higher and 1.4% showed lower transcript abundance than the extreme parent.

Given the H^2 and transgression, this enabled us to study the relation between these parameters (Fig. 5.3B-D and SI Table 5.4, 5.5) and gain insight in the factors driving transgression. No specific correlation was found between the H^2 and transgression in the parents (Fig. 5.3D). In the RILs transgression was positively correlated with H^2, more so in GG2 than in GG1 (Fig. 5.3B-C). This positive correlation is likely due to the increase in the range of transcript levels by transgression, which reduces the relative contribution of the experimental noise and environmental variation to the observed variation. This indicates that genetic variation drives transgression in general.

Genetic regulation of transcript abundance

Global eQTL distribution; the number of eQTLs per gene decreases exponentially

To assess linkage of variation in transcript abundance we applied the Multiple QTL Mapping (MQM) method on gene expression profiles of the RIL population, previously also used in GG1. In GG2 we mapped 7868 eQTLs for 6676 genes (~32% of total measured transcripts; Fig. 5.4). Applying the same multiple testing correction to the data obtained in GG1, we mapped 5912 eQTLs in GG1 for 5126 genes (~21% of total measured transcripts). For the majority of genes, a single eQTL could be mapped. The numbers of genes with multiple eQTLs decreases exponentially with increasing number of eQTLs mapped per gene (Fig. 5.5) in both experiments.

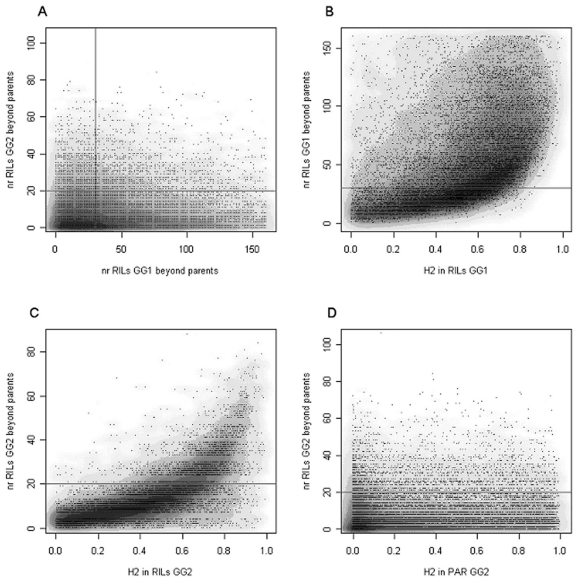


Figure 5.3 Transgression, number of recombinant inbred lines (RILs) with transcript levels beyond 2* times the standard deviation (SD) of the extreme parent per gene. Thresholds are in red.
A. in GG1 (Keurentjes et al., 2007) and GG2 (this chapter)
B. transgression compared to H^2 in GG1
C. transgression compared to H^2 in GG2
D. transgression compared to the parental (L*er* and Cvi) heritability (H^2) in GG2

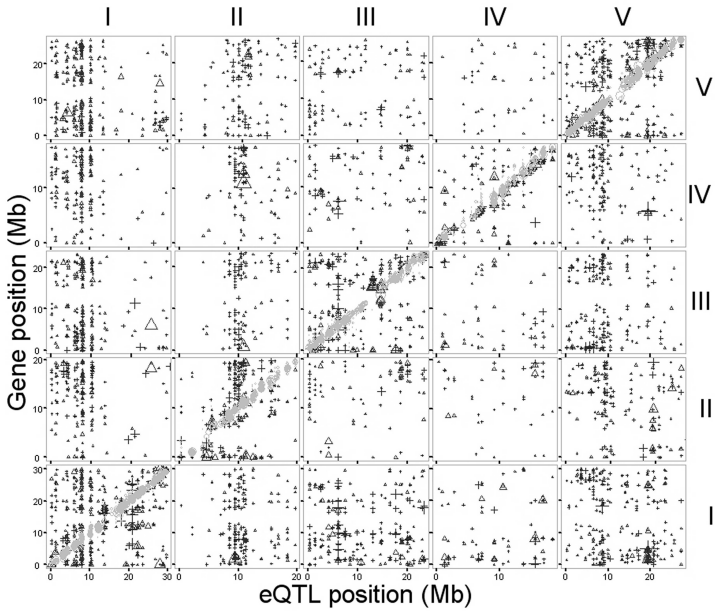


Figure 5.4 Expression quantitative trait loci (eQTL) distribution in GG2 (this chapter). Gene position (in mega basepairs (Mb)) expressed against the position of the eQTL(s) (in Mb). Chromosomes 1, 2, 3, 4 and 5 from left to right and bottom to top, as indicated. eQTLs for local regulation are indicated by the grey diagonal. eQTLS with a positive L*er* effect in triangles and with a positive Cvi effect as "plus" signs.

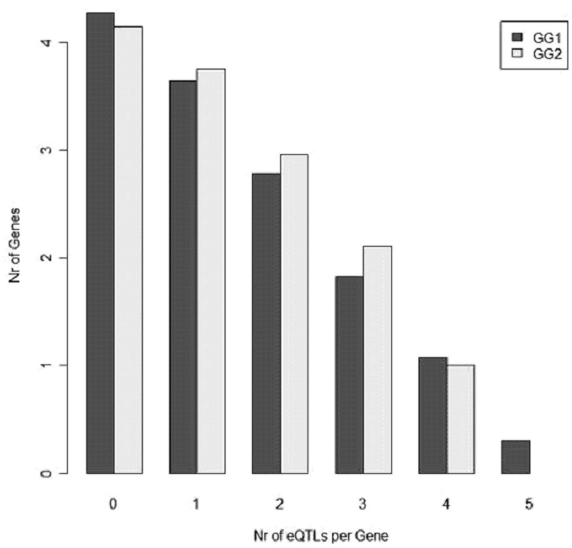


Figure 5.5 Frequency distribution of the number of expression quantitative trait loci (eQTLs) per gene (y axis in logarithmic scale).

Distant regulation occurs more frequent than local regulation

A higher percentage of eQTLs was found in GG2 in general, but the ratio between local and distant regulation was highly similar in GG1 and GG2 (SI Table 5.6). The fraction of genes with either local, distant regulation or both was in both experiments ~40%, ~50 % and ~10%, respectively. This shows that the genes with the polymorphisms causing the initial variation in transcript abundance ultimately cause the variation in transcript abundance of many other genes.

DATs between Ler and Cvi are caused by the genetic background

Most of the genes with DATs between the parents are expected to have an eQTL, since the only variable is the genotype and therefore the DAT must be caused by a genetic difference between the parents. An eQTL was found for ~74% of the parental DATs (SI Table 5.6, 5.7) of the genes in GG2. This is much higher than the percentage of genes with an eQTL of all genes on the array (~32%). Of the L*er*/Cvi DATs with an eQTL, ~51% was locally regulated; ~40% was distantly regulated and ~12 % both. This is in contrast to the fractions of local and distant regulation of all genes with an eQTL, for which ~40, ~50 and ~10% were found, respectively. A similar observation was made in GG1 (SI Table 5.6, 5.7), where in the whole set ~60% of the genes show distant regulation, whereas in the DATs this fraction is only ~40%. Local regulation increases from ~40% to more than ~50% in the DATs with eQTL. So, local regulation is more common in the genes with DATs between L*er* and Cvi, strengthening the conclusion that the majority of the DATs between L*er* and Cvi are caused by the genetic background.

The genes with an eQTL, but without DATs between L*er* and Cvi are expected to show transgression. Because an eQTL means that a difference in transcript abundance excists between groups of RILs, transgression should be observed if the gene with that

eQTL does show DATs between L*er* and Cvi. To test this idea we plotted the fraction of eQTLs against transgression. (Fig. 5.6). The relative number of genes with eQTL shows a positive correlation with transgression for the genes without L*er* Cvi DATs, as opposed to the slightly negative correlation for the genes with L*er* Cvi DATs supporting this idea. However, the majority of the genes with an eQTL do not show transgression. This means that most eQTL effects lie within the 2*SD of the extreme parent.

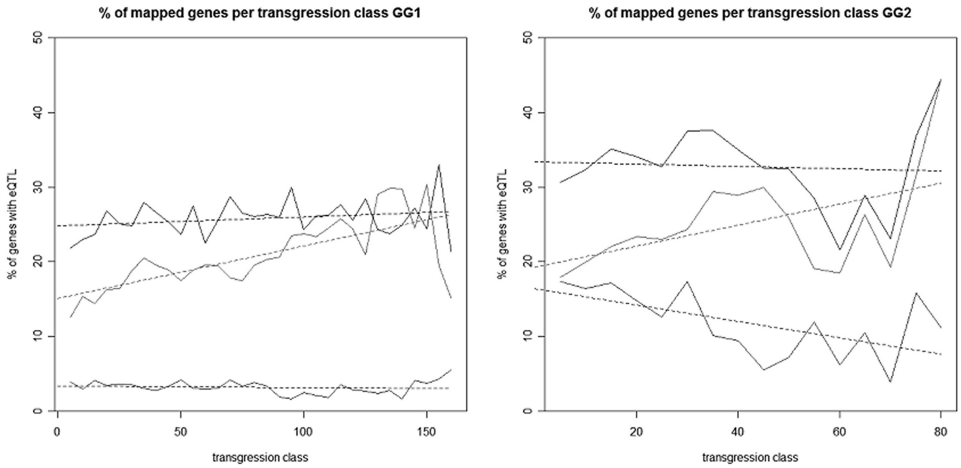


Figure 5.6 Number of mapped genes per transgression class (upper line), divided in genes with L*er*Cvi DATs (lower line) and without DATs (middle line). Left panel: GG1, right panel: GG2.

The chance of finding an eQTL increases with transgression and a high heritability

eQTLs are generally more frequently found for genes with expression levels that are largely determined by genetic background (high H^2). Such a relation between the H^2 and the number of genes with an eQTL was found in both the parents and RILs (Fig. 5.7). Especially in the parents the relative number of eQTLs show a sharp increase with higher H^2 (>0.75). For the RILs in GG2, this sharp increase corresponds to an increase in the fraction of genes with an eQTL per H^2 class. In the GG1 RILs no correlation between H^2 and fraction of genes with an eQTL was found, hence the horizontal line of this relative number of genes with an eQTL per H^2 group. The correlation previously found between transgression and H^2 also shows that, on average, the genes showing transgression are under stronger influence of the genome when recombination has created new genotypes. Another indication for this is the higher number of eQTLs in the genes without DATs in the parents, but with an eQTL (Fig5.6, SI Table 5.8, 5.9).

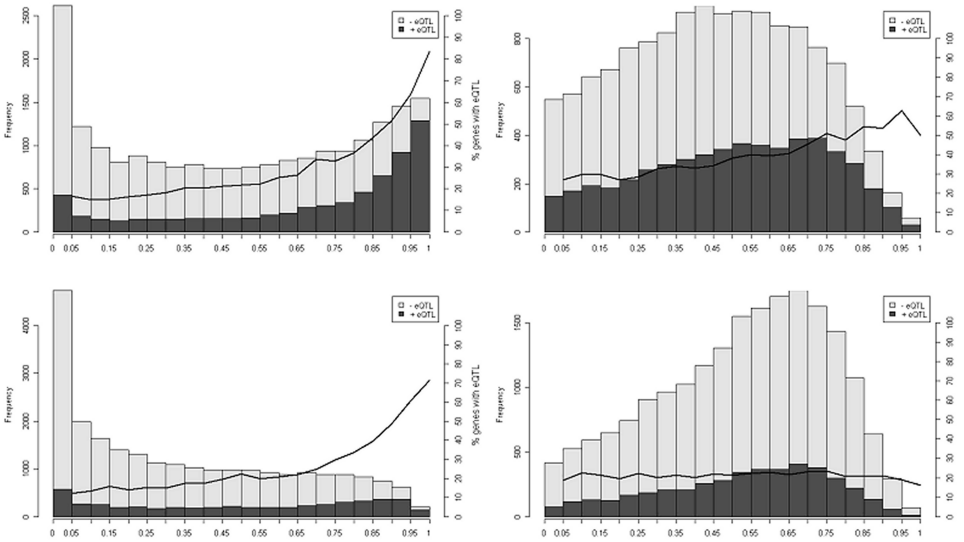


Figure 5.7 Number of mapped genes per heritability (H^2) class (x-axis) in GG2 (this paper; top) compared to GG1 (Keurentjes et al., 2007; bottom) for parents (L*er* and Cvi; left) and recombinant inbred lines (RILs; right).

Correlation between eQTL distribution, gene density and SNP frequency

We hypothesize that the major sources of differences in transcript abundance are single nucleotide polymorphisms (SNPs). Therefore the relation between eQTL distribution, gene density and SNP frequency was studied by their occurrence per marker. A correlation between SNP frequency and gene density could not be detected, thus enabling the study of their individual correlation with eQTL distribution.

eQTL distribution for local regulation is highly correlated with gene density

Between the number of eQTLs per marker and the genes located on the marker under study (marker locus) a medium (0.48) but highly significant ($p<10^{-8}$) correlation was found genome-wide. Per chromosome this correlation varied, with a strong correlation of 0.89 ($p<10^{-7}$) on chromosome 4 on one hand and with no significant correlation (0.26; p=~0.14) on chromosome 5 on the other. Genome-wide, some variation was also found between the two specific groups of eQTLs, local and distant. The local eQTL distribution was with 0.70 highly correlated ($p<10^{-15}$) to gene density, whereas the distribution of the distant eQTLs had only a correlation of 0.30 ($p<0.0005$). The correlations per chromosome were slightly different. The local eQTL distribution showed significant correlation of >0.60 at each chromosome. For the distant eQTL distribution no correlation was found on chromosomes 2 and 5 and the correlation per chromosome was lower.

The eQTL distribution for local regulation was highly correlated with gene density. In contrast, the eQTL distribution for distant regulation, depending on the chromosomal region under study, was much less or not correlated with gene density.

eQTL distribution for distant regulation is correlated to SNP frequency

We corrected the eQTL distribution for gene density because of the correlation between the two. The SNP frequency was corrected for the sampling frequency because the studied loci are of different sizes. A correlation was found between the eQTL distribution (corrected for gene density) and SNP frequency (corrected for sampling frequency). A correlation of 0.32 ($p<1*10^{-4}$) was found between the genome-wide distribution of all eQTLs and the SNP frequency. Differences were found between the different chromosomes. The distribution of the distant eQTLs showed a stronger correlation (0.32; $p<1*10^{-4}$) with SNP frequency than the distribution of the local eQTL (0.23; p=0.006). Since a higher frequency increases the chance to contain a SNP causal for multiple eQTL for distant regulation more than for local regulation these results support the fact that the eQTLs are mapping to the physical position of the causal polymorphism.

eQTL distribution for local regulation is correlated to genes containing SNPs

Furthermore, we investigated the correlation between SNPs and eQTLs by determining which genes harbored the sampled SNPs. The total eQTL distribution and position of these genes were better correlated (0.40; $p<5*10^{-7}$) than the genome-wide distribution of all eQTLs and the SNP frequency. The difference in correlation between local and distant eQTL distribution was also larger. Local eQTLs were much more correlated (0.64; $p<10^{-15}$) with the position of the SNP containing genes than the distant eQTLs (0.23; $p<0.01$).

In summary this means that the correlation of SNPs and eQTLs for local regulation is stronger if not only the single SNPs, but also the genes containing them are considered. However, the distant eQTLs were stronger correlated to the SNP frequency. This suggests that not the number of SNPs, but the gene in which the SNP is located is important for local (possibly cis) regulation and that the frequency at which SNPs are found is correlated with distant (possibly trans) regulation.

Constitutive and plastic natural variation of transcript abundance

Hotspots for Transcript Regulation are largely dependent on Treatment and Developmental Stage

Hotspots for transcript regulation (HTR) are identified in most genetical genomic experiments (Brem et al., 2002; Schadt et al., 2003; Bystrykh et al., 2005; Li et al., 2006; Keurentjes et al., 2007; West et al., 2007). These HTRs are genomic regions containing significantly more eQTLs than can be expected by chance and may either point to master regulators or indicate very gene dense genomic regions. To determine the HTRs in GG2 and their possible co-occurrence in GG1 we analyzed the eQTL distribution in both experiments (Fig. 5.8A, B). The eQTL distributions for local regulation (plotted in green) show a stronger correlation (~0.64) than the eQTL distribution for distant regulation (~0.21). This indicates that the eQTLs for distant regulation are much more dynamic and more dependent on environment/developmental stage whereas the eQTLs for local regulation better reflect the constitutive differences in transcript levels between the RILs.

When the total numbers of eQTLs per marker were considered, HTRs were found in both GG1 and GG2 (Fig. 5.8). The main HTR (>300 eQTLs) identified in GG2 was located at the top of chromosome one (around marker 11). Much less regulation by this

locus was found in GG1. In GG1, the HTR with most eQTLs (>200) was the Erecta locus (marker 51). In GG2 however, the number of eQTLs on Erecta, was severely reduced, but still well above the threshold. Several other HTRs, mostly specific for either GG1 or GG2, were found. The cold-spots, genomic regions containing few eQTLs, on the other hand were found at similar positions. These positions are roughly located on and nearby the centromeres and identify gene-sparse regions.

As the distant eQTLs seemed to be most influenced by the specific experimental conditions, we determined the fraction of eQTLs for distant regulation as a percentage of the total number of eQTLs per marker (Fig. 5.9). A relatively higher fraction of eQTLs was distantly regulated by loci on the top of chromosome 1 and chromosome 5 in GG2 (3weeks old; SD; low-light treated), and the bottom of chromosome 1 in GG1 (1 week old; LD; no treatment). Chromosome 3 showed a considerably lower fraction of distant eQTLs in GG1. Together, this shows that the position and especially the number of eQTLs on the HTRs are largely dependent on the treatment and/or developmental stage of the RIL population. If the a single gene underlying the HTR directs many different other genes in different pathways, it can be regarded as a master switch. The data show that at a given developmental stage and environment, different HTRs are predominant, indicating that master switches can change in time.

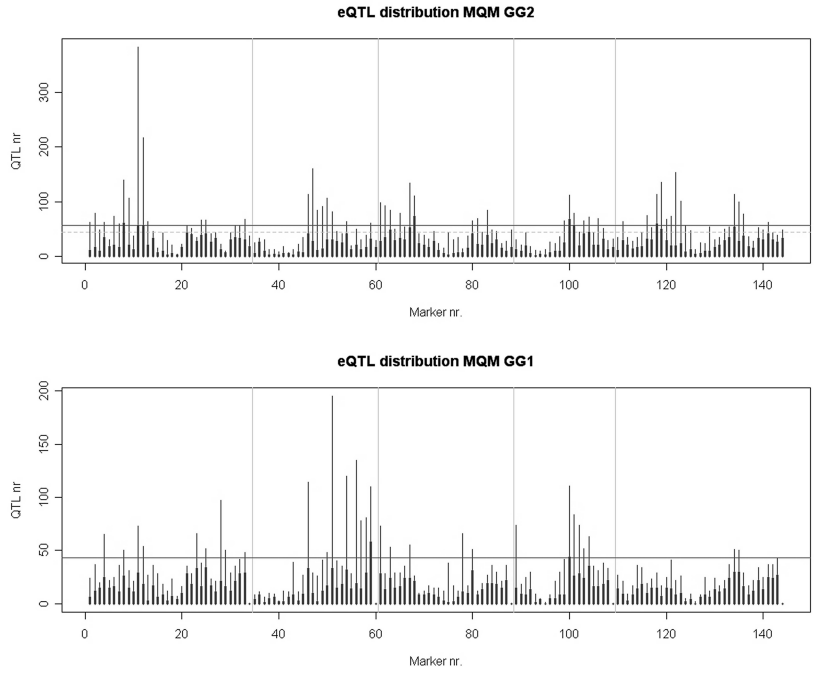


Figure 5.8 Expression quantitative trait loci (eQTL) distributions in GG1 (Keurentjes et al., 2007; bottom) and GG2 (this chapter; top). eQTLs for local regulation are the thick dark grey lines, eQTLs for distant regulation are the thin light grey lines, total height of bar is all eQTLs. Chromosome border are in vertical grey lines.

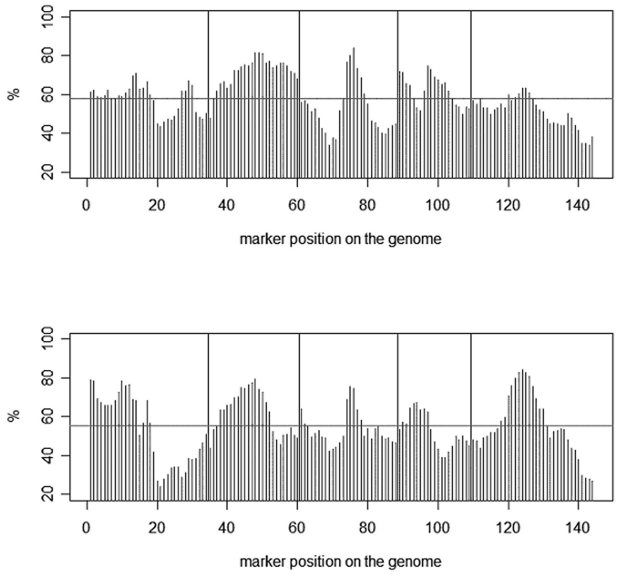


Figure 5.9 Fraction of distant expression quantitative trait loci (eQTLs) as percentage of total number of eQTLs per marker. GG1 (Keurentjes et al., 2007; upper panel), GG2 (this chapter; lower panel). The horizontal line indicates the mean percentage over all markers.

Comparison of the eQTL profiles of both experiments show constitutive and plastic genetic regulation of transcript abundance.

In the two genetical genomics experiments the same population was used. The differences, stage of development and treatment of the RILs, gave us the opportunity to investigate constitutive regulation and induced/stage dependent eQTLs. First, we calculated the correlation of the –log(P) profiles found in both experiments (Fig. 5.10). For the genes without eQTLs (< determined threshold) in either experiment, or with an eQTL in 1 experiment only, we observed that the correlation between the – log(P) profiles of the gene pairs from both experiments were normally distributed around zero (Fig. 5.10A-C). Genes with an eQTL just below the threshold in one of the experiments may cause the positively skewed distribution in figure 5.10B-C.

Most of the genes with an eQTL in both experiments show a strong positive correlation between their –log(P) profiles (~50% Fig. 5.10D, lower right panel); the genetic background has a similar influence on the transcript levels of these genes. Surprisingly, almost 6% of the genes with an eQTL in both experiments show a strong negative correlation between their – log(P) profiles. The transcript levels of these genes are affected by the same loci in both experiments but in an opposite manner.

Considering the strong constitutive influence of the genetic background on the transcript levels of these positively correlated genes, they are expected to have DATs between L*er* and Cvi. Indeed, of the 262 genes with DATs between the parents in both experiments 60% show highly positively correlated –log(P) profiles, whereas less than 1% is negatively correlated.

Of the genes with a strong positively correlated –log (P) profile 93% contained an eQTL in the same interval (88% for the negatively correlated profiles). This was only ~17%

for the genes with non- or weakly correlated –log (P) profiles. When the eQTLs per gene per experiment were compared we observed that 1183 eQTLs (for 1124 genes) were located on the same interval (is ~20% of total eQTLs in GG1 and ~15% of total eQTLs in GG2). Of these eQTLs ~79% were locally regulated (~21% distant). The directional effect of the eQTLs was the same for 1037 eQTLs (489 L*er* pos. and 558 L*er* neg.). So 146 eQTLs were found that are in overlapping regions but have an opposite effect in the two experiments (~70% local). Apparently 15-20% of all eQTLs are under strong control of the genetic background and therefore identified as constitutively under genetic control (at least in both experiments). Of these constitutive eQTLs ~12% have an opposite effect between the two experiments. Even though some eQTLs were observed on similar positions in both experiments, the effect of it can be under (partial) environmental control.

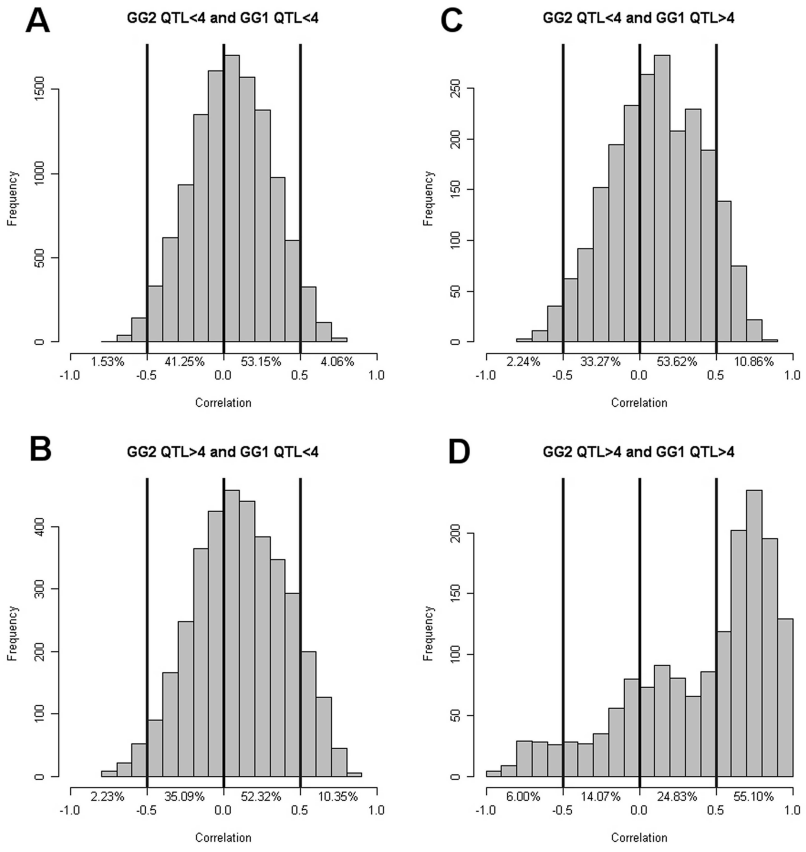


Figure 5.10 Histogram of the correlation of –log p profiles per gene. Correlation between the –log p profiles of gene with no expression quantitative trait loci (eQTL) in both experiments A), genes with a eQTL in one of the experiments, resp GG1 (Keurentjes et al., 2007) B), GG2 (this chapter) C) and genes with an eQTL in both experiments (GG1 and GG2) D).

Constitutively influence of the genetic background on biological processes

The over-represented annotation groups in the ~1000 genes with a constitutive eQTL represent the biological processes under (partly) constitutive influence of the genetic background (Table 5.1).

For *cellular component* there was an over-representation of genes targeted to the plastid, chloroplast and more specifically the thylakoid. From the other organelles, the mitochondria and nucleus were also found to be over-represented. For *molecular function*, genes showing transferase, hydrolase, oxidoreductase and kinase or transcription factor activity were found to be over-represented. Also observed were genes involved in "*nucleic acid, nucleotide or protein binding"*. For "*biological process"* an over-representation of genes involved in cellular, primary or macro molecule-metabolic processes were found. Also found to be over-represented were "*biosynthetic process"*, "*establishment of localization"*, "*cell organization and biogenesis"*, "*multi cellular organization"*, "*regulation of biological process"* and "*cell communication"*. Moreover, genes involved in the "*response to chemical, abiotic stress or endogenous stimulus"* were shown to be over-represented. The genes in these groups play a key role in regulation of vital processes in plants and are very likely to be causal for the constitutive phenotypic differences between L*er* and Cvi. Furthermore, it indicates in which processes a phenotypic difference may be expected between L*er* and Cvi.

Processes with constitutive eQTL but with opposite effect

A small number of genes (112) had eQTLs at the same loci in both experiments, but with an effect in opposite direction. Some over-represented annotation groups were also observed in this group of genes (Table 5.2).

For "*biological process"* most genes were found in the cellular, primary or macro molecule metabolic process. Furthermore, in the biological process group, "*response to chemical stimulus*", "*establishment of localization*" and "*regulation of biological process*" were found. For "*cellular component*" major groups were "*plastid*", "*mitochondrion*" and "*nucleus*". These genes and processes respond differently in L*er* compared to Cvi depending on the environmental/developmental condition. The annotation groups mentioned are the most likely groups in which natural variation in gene*environment interaction between L*er* and Cvi can be found.

Transcript regulatory network leading to flowering (Flowering time genes)

In GG1 we were able to construct a putative regulatory network for the genes involved in flowering time (Keurentjes et al., 2007). We did a similar calculation using the same genes and statistical method, iGA (Breitling et al., 2004), to determine which genes in the current experiment (GG2) are the main regulators for the genes involved in flowering time. Candidate regulators in both experiments were listed with the calculated PC values (measure of how likely the gene is a regulator of multiple genes involved in the same process, in this case flowering time).

It is clear that in GG2, completely different candidate regulators are selected. There is no overlap in the selected regulators between the two experiments. Only 9 genes, mostly of unknown function, overlap between the sets of all candidate regulators with significant PC values.

Table 5.1 Over-represented annotation groups in the group of genes with constitutive eQTLs. Number of genes with a constitutive eQTL per annotation group are given as the percentage of genes with a constitutive eQTL in the parent annotation group (Cellular Component, Molecular Function or Biological Process).

Constitutive eQTLs			985	
	Cellular Component		605	
		Plastid	170	28.10%
		Chloroplast	167	27.60%
		Thylakoid	24	3.97%
		Nucleus	58	9.59%
		Mitochondria	118	19.50%
	Molecular Function		571	
		Transferase act.	118	20.67%
		Hydrolase act.	109	19.09%
		Nucleic acid binding	92	16.11%
		Oxidoreductase act.	61	10.68%
		Protein binding	56	9.81%
		Nucleotide binding	46	8.06%
		Kinase act.	44	7.71%
		Transcription factor act.	43	7.53%
	Biological Process		492	
		Cellular metabolic proc.	297	60.37%
		Primary metabolic proc.	268	54.47%
		Macro-molecule metabolic proc.	194	40.04%
		Biosynthetic proc.	87	17.68%
		Establishment of localization	62	12.60%
		Cell organization and biogenesis	44	8.94%
		Multi cellular organization	32	6.50%
		Cell communication	31	6.30%
		Regulation of biological process	50	10.16%
		Response to chemical stimulus	43	8.74%
		Response to abiotic stimulus	40	8.13%
		Response to stress	38	7.72%
		Response to endogenous stimulus	34	6.91%

Table 5.2 Over-represented annotation groups in the groups of genes with constitutive eQTLs with opposite effect. Number of genes with a constitutive eQTL with opposite effect per annotation group are given as the percentage of genes with a constitutive eQTL with opposite effect in the parent annotation group (Cellular Component, Molecular Function or Biological Process).

Constitutive eQTLs with opposite regulation			112	
	Cellular Component		75	
		Plastid	22	29.33%
		Nucleus	7	9.33%
		Mitochondria	7	9.33%
	Molecular Function		---	
	Biological Process		62	
		Cellular metabolic proc.	39	62.90%
		Primary metabolic proc.	30	48.39%
		Macro-molecule metabolic proc.	22	35.48%
		Establishment of localization	12	19.35%
		Regulation of biological process	9	14.52%
		Response to chemical stimulus	12	19.35%

The genotype influences transcript levels involved in biological processes specific for treatment and development stage

The genotype may affect biological processes differently when plants are in a different environments or developmental stages. Some of these genotypic differences will be of influence in specific of in multiple environmental or developmental stages. This can be studied by determining the annotation groups that contain more genes with an eQTL than can be expected by chance. These processes are under strong control of the genotype at the specific experimental conditions of GG1 or GG2. The larger annotation groups are of interest because of the more general processes they describe, since they contain more genes, whereas the small groups highlight specific parts of the general processes.

Biological processes under genetic control specific for GG1; shown by over-represented annotation groups

In GG1 four main functional groups under influence of the genotype were found (Fig. 5.11 and Table 5.3). Group 1 consisted of annotation groups involved in defense pathways. Group 2 contained annotation groups involved in hormone processes and group 3 was represented by groups functioning in stress response/signal transduction pathways. Finally, group 4 includes sulfur and nitrogen metabolism.

The largest annotation group overrepresented for genes with an eQTL is "*defense response*". This general process is further specified by smaller annotation groups all over-represented for genes with eQTLs. This shows that under the experimental conditions of GG1 (seedling stage, no treatment) the defense against pathogens is very important and

under control of the genotype. Natural variation in transcript levels of genes involved in these processes was observed.

Several hormone processes are also controlled by the genotype in the seedling stage. The groups related to the two hormones with a clear role in the defense to pathogens, jasmonic and salicylic acid are overrepresented with eQTLs. Other groups related to the hormones, abscisic acid and especially auxin were also found to be overrepresented with eQTLs. These hormones are (partially) controlled at a transcript abundance level by the genotype in the seedling stage.

The third group, stress responses, can be divided in two, by their type of function. First water house keeping was found to be under genetic control specifically in the seedling stage. Secondly, the annotation groups linked to other stress responses are under genotypic control in the seedling stage.

Together, these results show that these important processes in seedling development are under genotypic control and natural variation is observed. If one considers the natural variation and eQTL found for the annotation groups as a measure for their importance for processes in the developing seedlings, than the defense, auxin and water state are very important and thus under high control of the genotype.

Biological processes under genetic control specific for GG2; shown by over-represented annotation groups

In GG2 three main functional groups under influence of the genotype were found (Fig. 5.12 and Table 5.4). Group 1 consists of annotation groups involved in "*energy/metabolic pathways"*. Group 2 contains annotation groups involved in "*hormone processes"* and group 3 is represented by groups functioning in the "*response/signal transduction pathways"*.

The largest annotation group over-represented for genes with an eQTL is "*metabolic process*". This general process is further specified by smaller annotation groups all over-represented for genes with eQTLs. These annotation groups show that under the experimental conditions of GG2 (3h of neutral shade treatment) the energy reserves and the anaerobic generation of energy are under control of the genotype and natural variation in these processes was observed.

Several hormone processes are also controlled by the genotype when treated with 3 hours of neutral shade. Annotation groups over-represented with eQTLs, together with literature evidence (Millenaar et al., 2005; Pierik et al., 2005; Benschop et al., 2007), show that the processes involving the hormones gibberellic acid, brassinosteroids, abscisic acid and ethylene are probably important for the response to neutral shade. They are involved at the transcript abundance level controlled by the genotype when treated with 3 hours of neutral shade.

The third group can be divided in two by their type of function. The group 3a facilitates all kinds of responses but clearly indicate an induction of gene expression, protein biogenesis and cellular processes. In group 3b the annotation groups are linked to more specific processes like for example, "*response to light stimulus*". Together, these results show that on all levels of the response to reduced light conditions natural variation in genotypic control can be found. Furthermore, the transcriptome of plants is indicative for the environment they were in and is dependent on their genetic background. Even more

striking is that the process of transcription on translation itself is reflected in the transcriptome. Clearly, these plants were actively reacting to the reduced light conditions by regulated expression of genes involved in all major processes involved in cellular change (transcription, translation, protein folding etc.). The combination of genetical genomics and gene annotation has identified which hormones are important, with which processes the response to reduced light intensity genes are shared and how the energy, necessary for this process, is managed.

Biological processes under genotypic control dependent on the environment

Clear differences are found between the overrepresented annotation groups in GG1 and GG2. These differences could be of constitutive nature (many genes within an annotation group have eQTLs but in only one of the experiments the group is above the threshold) but could also be largely dependent on the environment (developmental stage). To test this we compared the difference in relative amount of eQTL per annotation group per experiment (Fig. 5.13 and Table 5.5). These groups largely represent the differences indicated by directly comparing the annotation groups specific for each experiment. Some new annotation groups were found that are significantly different between the two experiments. New annotation groups specific for GG1 were "*proteolysis*", "*DNA replication*"," *response to molecules of bacterial origin*" and "*flower morphogenesis*". New groups specific for GG2 were "*transport*", "*ovule development*", "*chromatin modification*" and "*trichome differentiation*". Like the groups from the single experiments, these groups also reflect the stage or environment the tested RILs were in.

Table 5.3 Specific processes under genetic control in GG1 (Keurentjes et al., 2007), represented by Gene Ontology annotation term over-represented with eQTLs.

Group	GO term	Number of genes in group	Number of genes with eQTL	Percentage
1) Defense pathways				
	Defense response	272	90	33.1
	Response to nematode	45	16	35.6
	Defense response to fungus	30	12	40.0
	Defense response to bacterium	28	12	42.9
	Defense response, incompatible interaction	9	5	55.6
	Systemic acquired resistance	8	5	62.5
	Regulation of defense response	5	4	80.0
	Detection of bacterium	3	3	100.0
	Negative regulation of programmed cell death	5	4	80.0
	Regulation of hypersensitive response	3	3	100.0
	Putrescine biosynthetic process	2	2	100.0
	Spread of virus within host	2	2	100.0
	Hypersensitive response	17	8	47.1
2) Hormone processes				
	Response to salicylic acid stimulus	81	29	35.8
	Jasmonic acid biosynthetic process	17	8	47.1
	Salicylic acid biosynthetic process	2	2	100.0

	Response to abscisic acid stimulus	138	52	37.7
	Auxin homeostasis	9	6	66.7
	Indoleacetic acid biosynthetic process	8	8	100.0
	Auxin metabolic process	6	4	66.7
	Indoleacetic acid biosynthetic process	8	8	100.0
	Response to jasmonic acid stimulus	90	29	32.2
3) stress response/signal transduction pathways				
	Response to salt stress	141	45	31.9
	Response to water deprivation	78	26	33.3
	Response to osmotic stress	44	19	43.2
	Hyper-osmotic salinity response	26	11	42.3
	Response to desiccation	14	10	71.4
	Response to water	8	5	62.5
	Response to stress	72	27	37.5
	Response to wounding	70	28	40.0
	Response to temperature stimulus	11	6	54.5
	Response to external stimulus	2	2	100.0
	Response to cold	119	45	37.8
4) Sulfur and nitrogen metabolism				
	Regulation of nitrogen utilization	13	7	53.8
	Sulfate assimilation	13	9	69.2
	Sulfur metabolic process	13	9	69.2
Other overrepresented annotation groups				
	Electron transport	367	104	28.3
	DNA endoreduplication	14	7	50.0
	Lactose catabolic process via UDP-galactose	14	7	50.0
	Adventitious root development	4	3	75.0
	Cysteine biosynthetic process	7	4	57.1
	Histidine biosynthetic process	6	5	83.3
	Proline biosynthetic process	5	3	60.0
	Pyrimidine nucleotide biosynthetic process	5	3	60.0
	Camalexin biosynthetic process	3	3	100.0

Table 5.4 Specific processes under genetic control in GG2 (this chapter), represented by Gene Ontology annotation term over-represented with eQTLs.

Group	GO term	Number of genes in group	Number of genes with eQTL	Percentage
1) energy/metabolic pathways				
	Metabolic process	401	106	26.4
	Fatty acid biosynthetic process	44	22	50.0
	Glycerol metabolic process	14	8	57.1
	Acetate fermentation	12	8	66.7
	Reductive pentose-phosphate cycle	11	7	63.6
	Starch catabolic process	10	7	70.0
	Glucose catabolic process to lactate and acetate	10	8	80.0
	Starch metabolic process	7	5	71.4
	Pentose-phosphate shunt, non-oxidative branch	8	7	87.5

	Sulfate assimilation	13	8	61.5
	D-ribose catabolic process	6	5	83.3
	5-phosphoribose 1-diphosphate biosynthetic process	5	5	100.0
	Branched chain family amino acid biosynthetic process	5	4	80.0
	Carotene biosynthetic process	5	4	80.0
	Peptidoglycan biosynthetic process	3	3	100.0
	Pyrimidine ribonucleotide biosynthetic process	3	3	100.0
	Viral replication complex formation and maintenance	3	3	100.0
	Catechol catabolic process, ortho-cleavage	2	2	100.0
	Deoxyribose phosphate metabolic process	2	2	100.0
	Protocatechuate catabolic process, ortho-cleavage	2	2	100.0
2) Hormone processes				
	Response to gibberellin stimulus	62	27	43.5
	Positive regulation of gibberellic acid mediated signaling	4	4	100.0
	Brassinosteroid homeostasis	5	4	80.0
	Abscisic acid catabolic process	2	2	100.0
	Regulation of salicylic acid metabolic process	2	2	100.0
	Response to 1-aminocyclopropane-1-carboxylic acid	2	2	100.0
	Detection of ethylene stimulus	2	2	100.0
3) Response/signal transduction pathways				
	Translation	250	97	38.8
	Protein folding	188	80	42.6
	RNA processing	69	30	43.5
	Ribosome biogenesis and assembly	63	29	46.0
	Cell redox homeostasis	54	28	51.9
	Transcription	48	22	45.8
	Phosphorylation	22	12	54.5
	Chloroplast organization and biogenesis	27	18	66.7
	DNA endoreduplication	14	8	57.1
	Microtubule-based process	12	8	66.7
	RNA interference, production of ta-siRNAs	7	5	71.4
	RNA interference	2	2	100.0
	Positive regulation of cell proliferation	5	5	100.0
	Mitochondrion organization and biogenesis	4	4	100.0
	Negative gravitropism	7	6	85.7
	Embryonic pattern specification	3	3	100.0
	Cellular respiration	2	2	100.0
	Inflorescence development	2	2	100.0
	Apoptosis	16	9	56.3
	Response to light stimulus	90	37	41.1
	Response to UV-B	25	14	56.0
	Gravitropism	12	7	58.3
	Response to temperature stimulus	11	7	63.6
	Response to hypoxia	7	5	71.4
	Negative gravitropism	7	6	85.7
	Response to super oxide	2	2	100.0

Table 5.5 Processes under genetic control, specifically for GG1 (Keurentjes et al., 2007) or GG2 (this chapter), represented by Gene Ontology annotation term over-represented with eQTLs.

Group	GO term	Number of genes in group	nr genes with eQTL in GG1	nr genes with eQTL in GG2	Percentage GG1	Percentage GG2
GG1						
	Proteolysis	291	77	79	26.5	27.1
	Defense response	251	81	74	32.3	29.5
	Response to abscisic acid stimulus	138	52	43	37.7	31.2
	Response to wounding	70	28	22	40.0	31.4
	DNA replication	38	8	4	21.1	10.5
	Defense response to fungus	30	12	8	40.0	26.7
	Defense response to bacterium	28	12	6	42.9	21.4
	Response to fungus	16	7	2	43.8	12.5
	Response to desiccation	14	10	4	71.4	28.6
	Sulfur metabolic process	13	9	3	69.2	23.1
	Salicylic acid mediated signaling pathway	9	3	0	33.3	0.0
	Auxin homeostasis	9	6	1	66.7	11.1
	Response to water	8	5	1	62.5	12.5
	Indoleacetic acid biosynthetic process	8	8	3	100.0	37.5
	Response to molecule of bacterial origin	7	3	0	42.9	0.0
	Histidine biosynthetic process	6	5	2	83.3	33.3
	Regulation of defense response	5	4	1	80.0	20.0
	Negative regulation of programmed cell death	5	4	1	80.0	20.0
	Pyrimidine nucleotide biosynthetic process	5	3	0	60.0	0.0
	Indole glucosinolate biosynthetic process	3	3	1	100.0	33.3
	Regulation of hypersensitive response	3	3	1	100.0	33.3
	Flower morphogenesis	3	2	0	66.7	0.0
	Camalexin biosynthetic process	3	3	0	100.0	0.0
	Salicylic acid biosynthetic process	2	2	0	100.0	0.0
	Detection of bacterium	2	2	0	100.0	0.0
GG2						
	Transport	321	63	113	19.6	35.2
	Translation	250	57	97	22.8	38.8
	Ribosome biogenesis and assembly	63	13	29	20.6	46.0
	Transcription	48	10	22	20.8	45.8
	Fatty acid biosynthetic process	44	10	22	22.7	50.0
	Chloroplast organization and biogenesis	27	9	18	33.3	66.7
	Transcription initiation	24	1	9	4.2	37.5
	Ovule development	20	0	10	0.0	50.0
	Gravitropism	12	1	7	8.3	58.3
	Microtubule-based process	12	2	8	16.7	66.7
	Acetate fermentation	12	2	8	16.7	66.7
	Reductive pentose-phosphate cycle	11	2	7	18.2	63.6
	Regulation of gene expression, epigenetic	11	1	6	9.1	54.5
	Chromatin modification	11	0	6	0.0	54.5

	Trichome differentiation	10	0	5	0.0	50.0
	Starch catabolic process	10	1	7	10.0	70.0
	Glucose catabolic process to lactate and acetate	10	2	8	20.0	80.0
	Glucose catabolic process to D-lactate and ethanol	8	0	5	0.0	62.5
	Pentose-phosphate shunt, non-oxidative branch	8	2	7	25.0	87.5
	Purine base biosynthetic process	7	0	4	0.0	57.1
	Plastid organization and biogenesis	7	0	4	0.0	57.1
	Positive regulation of cell proliferation	5	1	5	20.0	100.0
	Carotene biosynthetic process	5	0	4	0.0	80.0
	Brassinosteroid homeostasis	5	0	4	0.0	80.0
	Mitochondrion organization and biogenesis	4	1	4	25.0	100.0
	L-serine biosynthetic process	4	0	3	0.0	75.0
	Aspartate family amino acid biosynthetic process	4	0	3	0.0	75.0
	Aerobic, glycerol catabolic process	4	0	3	0.0	75.0
	Glyceraldehyde-3-phosphate catabolic process	4	0	3	0.0	75.0
	Positive regulation of gibberellic acid mediated signaling	4	0	4	0.0	100.0
	Response to super oxide	2	0	2	0.0	100.0
	Protocatechuate catabolic process, ortho-cleavage	2	0	2	0.0	100.0
	Catechol catabolic process, ortho-cleavage	2	0	2	0.0	100.0
	Inflorescence development	2	0	2	0.0	100.0
	Regulation of salicylic metabolic process	2	0	2	0.0	100.0
	Response to 1-aminocyclopropane-1-carboxylic acid	2	0	2	0.0	100.0

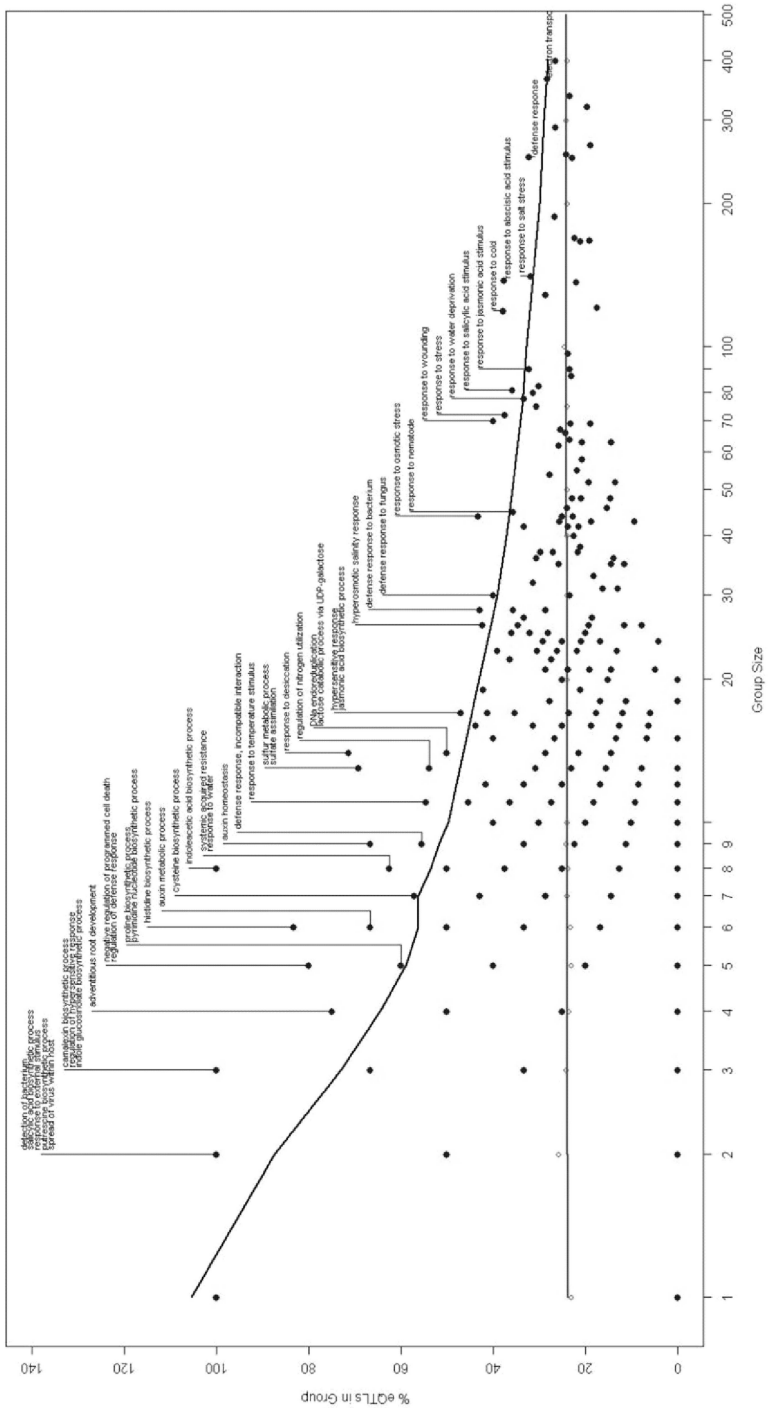


Figure 5.11 GG1 (Keurentjes et al., 2007), Gene Ontology annotation groups overrepresented with gene with eQTLs. 1000 permutations on groups sizes show the following (random groups of (1,2,3,4,5,6,7,8,9,10,20,30,40,50,75,100,200,300,400) genes have been picked 1000x and the average number of genes with eQTLs was determined as well as the standard deviation. Horizontal line: is the average percentage of genes with an eQTL of a randomly picked group of genes of a certain size. Black decresing line: is the 2 sd (97.8%) probability line. Dots are the different annotation groups; the dots above the black decresing line are annotation groups with more genes with eQTLs than can be expected at random. Group names are shown in graph.

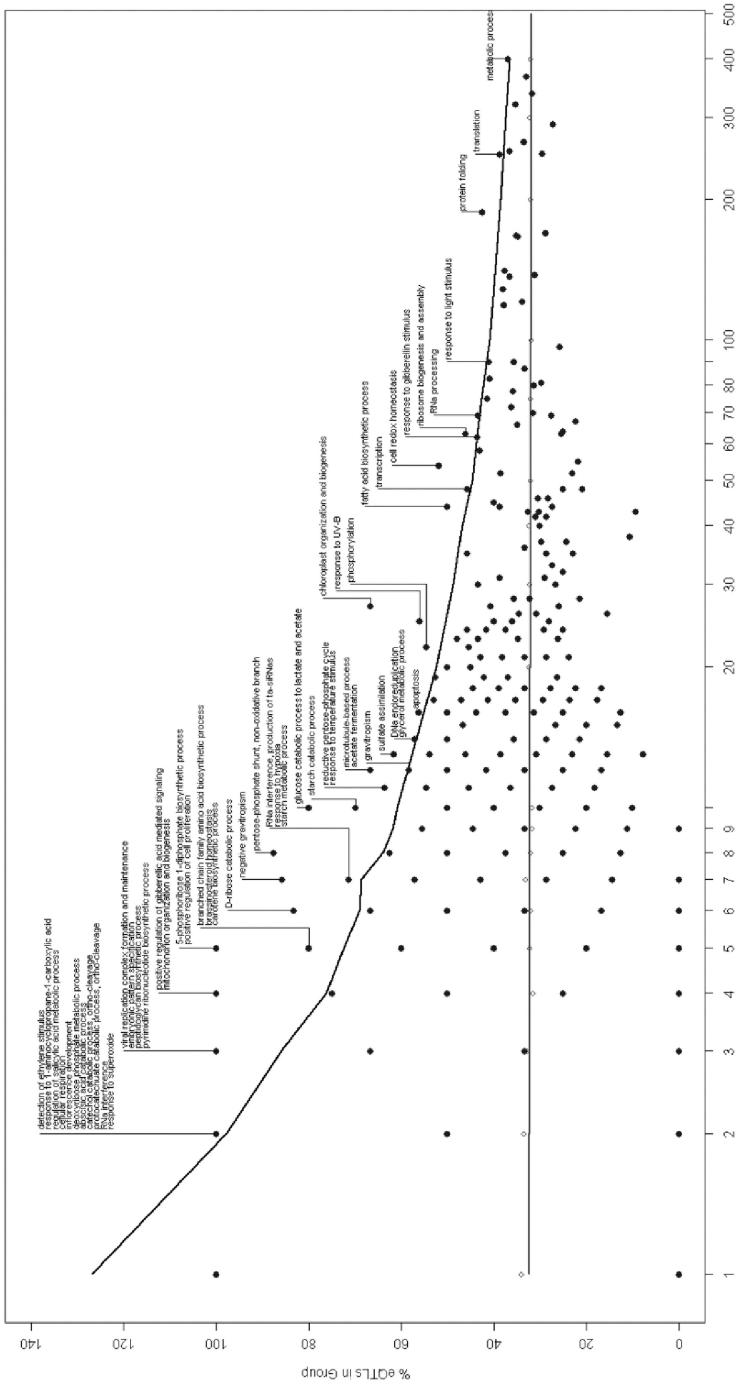


Figure 5.12 GG2, Annotation groups overrepresented with gene with eQTLs. For legend see figure 5.11

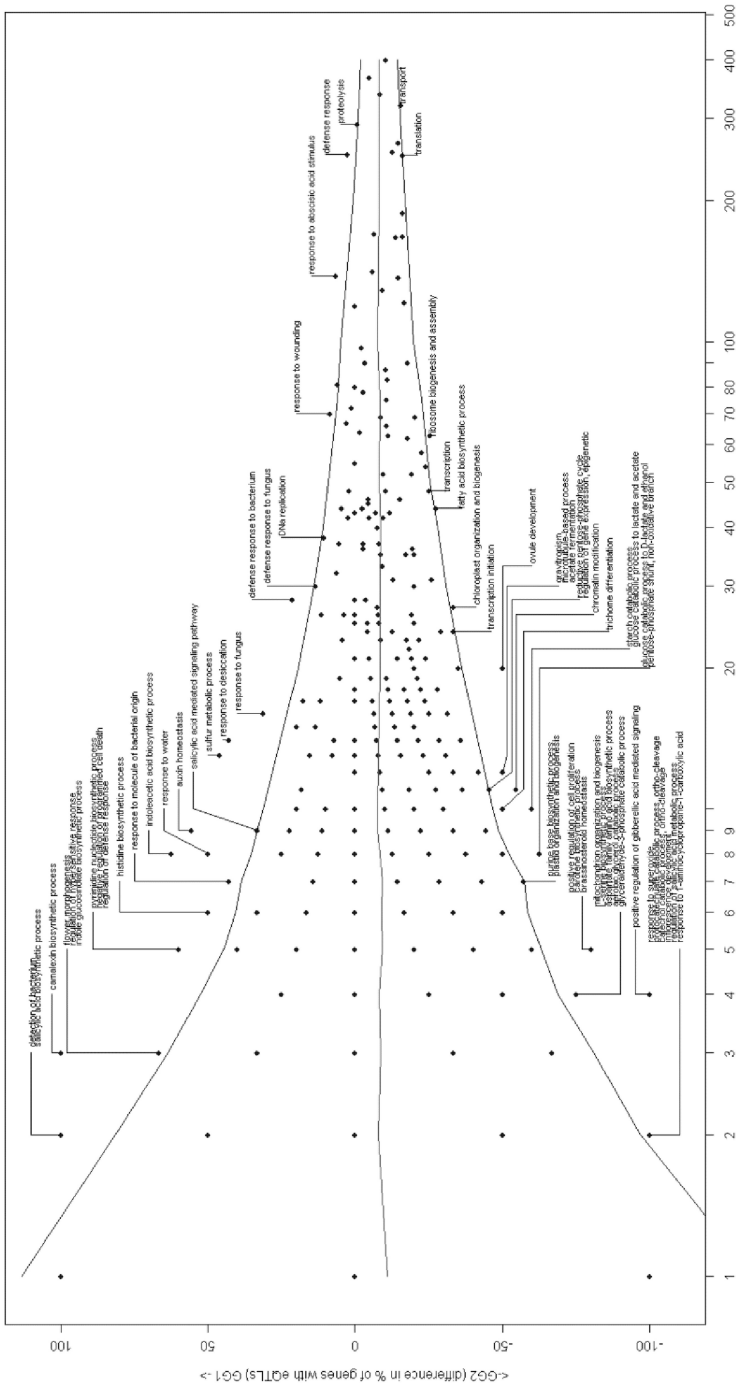


Figure 5.13: Annotation groups specific for GG1 or GG2. The lower groups are specifically overrepresented with eQTLs in GG2 and the upper groups in GG1. For legend see figure 5.11

Discussion

Only a small fraction of the differentially abundant transcripts (DATs) between the parents is constitutive

Differences in transcript levels were detected between the parental lines in GG1 (808) and GG2 (3228). The number of DATs found between the parental lines in GG2 is remarkable similar to those found between the parental lines of the Bay x Sha population (3351) (West et al., 2007). These plants were profiled when they were in their rosette stage (6weeks post-germination). More DATs were found between parental lines tested in GG2 than in GG1 indicating that the global transcript levels become increasingly different during development or as a response to changing environmental conditions. The differences in DATs found between the parents in the two experiments may be caused by two factors, time and treatment. The time in which the genotype could influence transcript levels is 3 times longer in GG2 than in GG1 and if we assume that certain transcripts accumulate in time, even initial small expression differences can lead to large enough transcript abundance variation to be picked up by eQTL mapping at later stages. Furthermore, it could be that the seedling stage of Arabidopsis is a much more conserved process, less influenced by the environment, than development at a later stage. Furthermore, we expect that the induced or reduced expression of genes as a response to the neutral shade treatment did not settle to a new balanced state yet and thus causes more variation leading to an increased amount of DATs in the GG2 experiment. Moreover, a difference in sensitivity of the two micro-array platforms used may also give rise to some differences in the number of observed DATs. In other species, like mice and yeast, a remarkable difference was found with some of the parental lines used in a genetical genomics experiment. In yeast 25% ($p<0.005$) to 50% ($p<0.15$) of the parental genes showed a difference in abundance (Brem et al., 2002). In mice the parental lines showed 33% DATs between them (Schadt et al., 2003). This is much higher than the observed variation between the Arabidopsis parental lines used in GG1 (~5%), GG2 (~15%) and the Bay x Sha population (~15%) (West et al., 2007). It could either be that the mice and yeast parental lines are more genetically divergent than the Arabidopsis parental lines or that transcript levels are more balanced in Arabidopsis.

Heritability at three levels: parents, RILs & genes

The larger number of DATs between the parents tested in GG2 compared to GG1, also resulted in a larger number of genes under strong control of the genotype (e.g. a high H^2) in GG2. This is probably caused by similar reasons as for the differences in DATs between the parents tested in the two experiments (parental lines tested in GG1 vs. parental lines tested in GG2). Furthermore, it shows that more genes are under strong control of the genotype when the parental lines are fully developed, shade treated plants.

In contrast to the parental lines the heritability determined with the RILs shows a larger group of genes under influence of the genotype in the seedling stage (GG1). Indeed if the expression levels in the seedling stage are much more conserved in the parental lines, this must be the result of balancing effect of different alleles which together result in comparable transcript levels in those parental lines. In contrast to the parental lines, of which the genotypes are products of natural selection, the RILs have not been subjected to

any selection pressure (except for being able to reproduce). Therefore the RILs will also contain less favorable combinations of alleles leading to more variation, which in turn increases the chance of detecting a high control of the genetic background. Since the between-spot variations of the parental slides are used to calculate the H^2 of transcript abundance in the RILs the differences between the parental lines also play a role in the difference between the H^2 of the RILs in the two experiments. This cannot explain, however, the difference by itself and a more general "biological" process as discussed before is expected to play a role.

When we plot the H^2 of the individual genes in both experiments we can determine whether those genes are affected in a similar way or specific for one condition (Fig. 5.2). The comparison of the H^2 of the parental lines with the RILs in both GG1 (Fig. 5.2A) and GG2 (Fig. 5.2B) shows that the influence of the genotype is larger for most genes in the RILs. However, we would expect a large number of genes to have transcript level variations in the parental lines to also have DATs between the RILs (since they are "genetic mosaics" of their parents). Especially in GG2 this group was observed (Fig. 5.2B, upper right corner). In GG1 only a trend was visible, but no clustered sets of genes as in GG2 could be identified. The fact that some genes with a high H^2 in the parental lines had a low H^2 in the RILs could mean that some genes are controlled by combinations of alleles which are randomly mixed in the RILs causing a lack of genotypic control.

The comparison of the H^2 between experiments shows that most genes in the parental lines (Fig. 5.2C) are under genotypic control under specific conditions. However, for the RILs, in which the genotypic influence is stronger, most genes are found to be under genetic control in both experiments (Fig. 5.2D). This means that independent of the experiment the genotype plays a role in transcript abundance of many genes. However, this does not mean that the environment does not play a role since the graph shows a cloud of genes with a certain H^2 under both conditions and not a strait diagonal line from the lower left corner to the upper right. It is very well possible that the genotype/environment interaction is playing a large role here.

Transgression vs. heritability

The comparison of transgression and heritability (Fig. 5.3B, C, and D) show a correlation between the number of RILs with expression values beyond the parental range and heritability in both GG1 and GG2. In GG1 (Fig. 5.3B) genes with a high H^2 always show transgression. This is in contrast with GG2 (Fig. 5.3C) were not all genes with a high H^2 show transgression. On the other hand the genes with a very low H^2 hardly show transgression in GG2 whereas in GG1 even some of those genes show transgression. Again, like in determining the H^2, the difference between the parental transcripts levels plays a consistent role in these results too. Still the positive relationship between transgression and H^2 in the RILs is clear, especially when compared with the relationship of the H^2 of the parents and transgression (Fig. 5.3D) which shows no correlation at all.

Global eQTL distribution

The relative amount of genes with an eQTL increased from ~20% in GG1 to ~30% in GG2. Like the increased number of DATs between the parents in GG2, this can be caused by either the older stage of development of the plants in GG2 or the transcription

inducing/reducing effect of the low-light treatment in GG2. The ~20% and ~30% genes with an eQTL for GG1 and GG2 respectively, are in contrast to the relative number of genes with an eQTL (~69%) found in the Bay x Sha population (West et al., 2007). Because it is very unlikely that this difference is only caused by the populations used it could be caused by the difference in the type of platforms that were used, Affymetrix arrays for the Bay x Sha and spotted arrays for GG1 and GG2. The platforms and experimental setup affect the power to detect eQTLs significantly. In GG1 and GG2, 160 and 120 RILS were profiled respectively, from which 80/60 data points per gene were obtained for eQTL determination. In the Bay x Sha population more RILs (211) and the use of the Affymetrix arrays enabled them to use 211 data points per gene to calculate eQTLs. This is also reflected in the larger total number of eQTLs detected and the number of eQTLs per gene.

In a yeast population of 40 segregants 9 (Brem et al., 2002) to 34% (Brem and Kruglyak, 005) of the genes showed at least one eQTL. In a population of a 111 F2 mice, 16 (LOD > 4.3) to 47% (LOD>3) of the genes showed at least one eQTL. When we combine the percentage of eQTLs found in the different organisms, it is clear that population size, micro-array platform and experimental setup are the main determinants of the number of eQTLs that can be detected.

Distant vs. Local regulation

In both experiments relatively similar percentages of genes are locally, distant or both regulated (40, 50 and 10%, respectively). Combined with the relative amount of all genes with an eQTL, this indicates that globally both local en distant regulations are affected to the same extent independent by the developmental stage or treatment. In the Bay x Sha population however 33% of genes show local regulation (West et al., 2007) which was similar to the percentage of locally regulated genes in yeast (Brem et al., 2002). Due to the increased detection power in their experiments they probably observed more eQTLs for distant regulation of which the effects in general are smaller.

eQTLs of the DATs between Ler and Cvi

Most of the genes with DATs between the parents are expected to have an eQTL. This was found for the parental DATs in both experiments. Furthermore, local regulation was found to be more common among the genes with DATs between L*er* and Cvi, which indicates that most of the differently expressed genes between L*er* and Cvi are caused by a polymorphism, either in the regulatory sequence or one that may affect the stability of the transcript. The genes with an eQTL, but without DATs between L*er* and Cvi show a positive correlation with transgression. However, the majority of the genes with an eQTL do not show transgression, which means that most eQTL effects lie within 2 SD of the extreme parent. Together, this shows that much of the transcription level differences are caused by the combination of small effect regulators. This was also reflected in the Bay x Sha population, 63% of the genes with an eQTL showed for at least 10% of the RILs transcript levels 1 SD beyond the parental range. Although this is a little more relaxed setting of determining transgression, the authors conclude that most effects must be very small (West et al., 2007). In yeast only 20% of the DATs between the parental lines were found to have an eQTL. This is much less than expected and may be due to a too simple mapping method (Brem et al., 2002).

Relation of eQTLs with heritability and transgression

A positive relation between the H^2 and the frequency of genes for which an eQTL was detected was found in GG2 for both parental an RIL determined H^2. This was like expected, however in GG1 only a trend was observed. This could mean that although the H^2 in GG1 was high for many genes, the effect of the individual loci affecting the transcripts levels of those gene were small and therefore undetectable by QTL analysis. In GG2 this must mean that at least some of the loci did have an effect large enough to be identified by QTL mapping. Strong transcript regulation by a light responsive pathway could cause this, when allelic variation of components of this pathway exists between L*er* and Cvi. These results indicate that a treatment will increase the chance of finding (strong) eQTLs for more genes than using a balanced state. Furthermore, this could indicate that regulation of transcript levels as a response to a treatment involves fewer components than balancing a gene's transcript level when in a constant environment. Considering transgression it shows that a treatment also maximizes the differences in some transcript levels between the RILs as well as the parents, increasing the correlation of transgression and the frequency of genes with an eQTL.

eQTL distribution for local regulation is highly correlated with gene density

The eQTL distribution for local regulation was highly correlated with gene density, as was also found in GG1 (Keurentjes et al., 2007). In contrast, the eQTL distribution for distant regulation, depending on the chromosomal region under study, was much less or uncorrelated with gene density in our study. This demonstrates that a gene harboring the causal polymorphism for its own transcript level differences and thus contain an eQTL for local regulation can affect the levels of multiple other transcripts which can cause many eQTLs for distant regulation. This weakens the correlation with gene density for distantly regulated genes, especially when the causal gene is in a gene sparse region.

eQTL distribution for distant regulation is correlated to SNP frequency

In a previous study, in GG1 (Keurentjes et al., 2007), a correlation was observed between the eQTL distribution (corrected for gene density) and SNP frequency (corrected for sampling frequency). This is in contrast with result from experiments with the Bay x Sha population were it was found that most genes with local eQTLs (91%) did not harbor a single feature polymorphism (SFP) (West et al., 2007). In our study, the distributions of the distant eQTLs showed a stronger correlation with SNP frequency than the distribution of the local eQTL. Since a higher SNP frequency at a locus makes it more likely that that locus harbors SNP(s) causal for multiple eQTL for distant regulation more than for local regulation these results support the fact that the eQTLs must map to the physical position of the causal polymorphism. Furthermore, a gene harboring the causal polymorphism causing its own transcript level difference can affect the levels of multiple other transcripts, as found in the correlation between eQTL distribution and gene density.

eQTL distribution for local regulation is correlated to genes containing SNPs

The total eQTL distribution and position of these genes were better correlated than the genome-wide distribution of all eQTLs and the SNP frequency. The difference in correlation between local and distant eQTL distribution was also larger. Again a remarkable

contrast with the Bay x Sha population were most genes with local-eQTLs (91%) did not harbor a single feature polymorphism (SFP) (West et al., 2007). The results from our study, however, show that the correlation of SNPs and eQTLs for local regulation is stronger if not only the single SNPs, but also the genes containing them are considered. However, the distant eQTLs were stronger correlated to the SNP frequency. This is a strong indication that not the number of SNPs, but the gene in which the SNP is located is important for local (possibly cis) regulation and that the frequency at which SNPs are found is more important for distant regulation.

Distribution of the number of eQTLs per gene

The number of eQTLs per gene reduces exponentially (Fig. 5.5). The chance of finding additional eQTLs per gene becomes increasingly smaller. This would make random co-location of multiple eQTLs not very likely. Therefore, co-location of multiple eQTLs increases the likelihood of shared regulation, making them especially useful as initial steps in constructing putative transcription regulatory networks. In both GG1 and GG2 we found 0-5 eQTLs per gene, whereas in the Bay x Sha population 0-11 eQTLs per gene could be observed (West et al., 2007). In mouse 0-4 eQTLs per gene (Schadt et al., 2003) and in yeast 0-10 eQTLs per gene (Brem and Kruglyak, 2005) were found. Like the total number of eQTLs detected we suggest that this increase is due to a difference in detection power due to experimental design.

Comparison of the eQTL profiles of both experiments show constitutive and plastic genetic regulation of transcript abundance

For the genes without eQTLs in either experiment (GG1 & GG2), or with an eQTL in one experiment only, we observed that the correlation between the –log(P) values of both experiments follows a normal distribution around zero (Fig. 5.10A-C). Genes with an eQTL just below the threshold in one of the experiments probably caused the positively skewed distribution in figure 5.10B-C. The genes with an eQTL in only one of the GG experiments are very likely to be regulated by a plasticity QTL (pQTL) as was found in *C. elegans* (Li et al., 2006). Although in *C. elegans* this was specifically tested and it was shown that temperature was the variable that was causal for this plasticity we speculate that in Arabidopsis many of the genes with an eQTL specific for a treatment will also be regulated by a similar pQTL. Furthermore, the HTR on chromosome I (Fig. 5.4 and 5.10) may be one of such plasticity hotspots as was found for temperature in *C. elegans* (Li et al., 2006). This HTR in Arabidopsis could be a major switch in the transcription responses to reduced light. Most of the genes with an eQTL in both experiments have a strong positive correlation between their – log(P) profiles (~50% Fig. 5.10D, lower right panel). The genetic background has a constitutive influence on the transcript levels of these genes. The 262 genes with DATs between the parents in both experiments 60% have highly positively correlated – log(P) profiles, whereas less than 1% is negatively correlated.

Surprisingly, of all genes with an eQTL in both experiments almost 6% have a strong negative correlation between their – log(P) profiles. The transcript levels of these genes are affected by the same loci in both experiments but in an opposite manner. These loci probably harbor a gene which has one allele that is regulated whereas the other allele is constitutively expressed at the same level or regulated less strong.

Of the genes with a strong positively correlated –log (P) profile 93% contained an eQTL in the same interval (88% for the negatively correlated profiles). This was only ~17% for the genes with non- or weakly correlated –log (P) profiles. Not only shows this the constitutive regulation of the genetic background, but also that the information of genetic regulation is contained in the whole – log P profile, instead of just the significant eQTLs.

The eQTLs per gene per experiment show that ~20% of total eQTLs in GG1 and ~15% of total eQTLs in GG2 are the result of constitutive genetic regulation. Of these eQTLs ~79% were locally regulated and ~21% distant. The directional effect was the same for almost all of those eQTLs. The distribution of negative and positive effects was equal between the parental alleles. (~47% L*er* pos. and ~53% L*er* neg.). Only ~12% of the constitutive eQTLs had an opposite effect in the two experiments (~70% local). These eQTLs are perfect examples that the effect of even a constitutive eQTL can be under (partial) environmental control.

Most of these constitutive eQTLs must be at the basis of many of the constitutive (phenotypic) differences between L*er* and Cvi. In addition, the constitutive eQTLs with opposite effect probably cause a difference in response of L*er* and Cvi to the differences in treatment/development between the two experiments. The largest over-represented annotation groups were those involved in metabolism. This could mean that the metabolite QTLs measured in the L*er* x Cvi population (Keurentjes et al., 2006) could be more constitutive over environments than the eQTLs in this population. The other large annotation group over-represented in the constitutive eQTL group was chloroplast localization. This specifies that many constitutive differences between Ler and Cvi could be derived from a difference in chloroplast function. Many metabolites are synthesized in this organelle suggesting a link between the constitutive eQTLs for metabolite genes and cell organelle too.

Flowering time, transcriptional network

We were able to construct a putative network of transcript regulation leading to flowering as in GG1 (Keurentjes et al., 2007). None of the putative regulators, however, were observed in both networks. Only 9 genes, mostly of unknown function show overlap between the sets of all genes with significant PC values. (Only the gene with the most significant PC value is selected as regulator for each regulated gene) This is most likely due to the different (number of) eQTLs for each gene and the difference in growth conditions (short day in GG2 compared to long day conditions in GG1). Flowering time is regulated through integration of different environmental signals like photoperiod, temperature, hormone signals and/or stress. Plants grown in short day conditions drive gene expression through different parts of this signaling network, than plants grown under long day conditions. But the main reason will probably be that genes regulating flowering time are predominantly expressed in the apical meristems, and not so much in the leaves. The fraction of RNA derived from the meristems is much higher in seedlings as compared to the fraction in extracts derived from mature leaves. In addition, the low-light treatment drives gene expression to LL responsive pathways, increasing the effect of co-locating cis-eQTLs that are not related to flowering time, and thus increase the noise, on the correlation analysis applied in iGA.

Biological processes affected by the genetic background under specific conditions

The overrepresented GO annotations groups per GG experiment show which processes are affected by the genetic background at the specific conditions of those experiments. In the GG1 auxin related processes and defense against pathogens is under genetic control. These two processes are important for the seedling stage plants used in the GG1 experiment. Auxin was found to be very important for hypocotyl elongation (Cheng et al., 2007). The high level of control of defense related genes in seedlings is likely due to the high selection pressure of pathogens on small seedlings.

In GG2 the older plants (3 weeks), treated with 3 hours of neutral shade, show very different processes under genetic control. Most of these processes are linked to the cellular response to a changing environment. All of the processes making a change within a cell or plant organ possible where found to be under genetic control at the transcript level. It is striking that natural variation in transcript abundance can be found for many genes involved in basic processes like transcription, translation, chromatin remodeling, protein modification and even the assembly of the ribosomes. The transcriptomes of a RIL population reflect which processes are important and for which variation can be found under the specific conditions (treatment/developmental stage) of the experiment.

The combination of the important biological processes for the two experiments show that for flowering time the genes for flowering and reproduction are expressed before any classical phenotypical variation can be observed in the RILs.

The number of processes and genes found to be controlled by the genetic background under specific conditions show that it is very likely that the amount of functional natural alleles is much larger than can be estimated from phenotypical QTLs alone.

Conclusions

The constitutive influence of the genetic background is reflected in the transcriptome. But more of the transcript levels are under genotype*environment influence as shown by the comparison of heritability, transgression and eQTL profile correlation.

The eQTLs for local regulation are best correlated to the genes harboring a SNP, indicating that most local-eQTLs are caused by SNPs. More evidence was found by the eQTLs for distant regulation which were better correlated to SNP frequency. This shows that indeed most eQTLs must have a causal SNP.

The flowering time network constructed in GG1 (Keurentjes et al., 2007) could not be replicated with the data from our study (GG2) because, indications were found that most of the RILs were in a different transcriptional stage regarding flowering as indicated by the over-represented annotation group "*flower morphogenesis*" in GG1 and "*ovule development*" in GG2.

The difference in differently abundant transcripts between the parental lines and the number of genes for which an eQTL could be detected in all three experiments (Keurentjes et al., 2007; West et al., 2007); this study) show that the "old" parental genomes must have been under natural selection for a general balance in transcript abundance. This indicates that many genes have an optimum expression level with regard to fitness regardless of the genetic background.

The transcriptome is indicative for the processes for which variation can be observed under specific conditions and thus shows which processes are important for responding to those environmental conditions. Auxin and defense responses are important in the hypocotyl/seedling stage. The whole cellular machinery making the response to low-light possible is regulated during this response. Shown by the groups describing transcription, translation, protein modification, energy generation, hormone processes and chromatin modification are all processes for which natural variation in transcript abundance can be found specifically in low-light treated plants.

Methods and materials

Plant growth

To identify natural variation in transcript abundance, 120 RILs of the Ler x Cvi population (Alonso-Blanco et al., 1998) were grown at the same time, divided in 10 trays containing 3 plants of 14 individual genotypes, parents included. In this way we obtained 3 plants per RIL and 9 per parent. Plants were grown under short days (9h light, 15 dark), a light intensity of 200 µmol m^{-2} s^{-1} and watered 3 times per week.

Shade Treatment and harvested material

All RILs and parents were treated with 3 hours of neutral shade (ca 10% of growth conditions) 21 days after transfer to soil (24 days after germination). The three most responsive leaves (petiole and lamina) per plant were harvested and pooled per genotype for RNA isolation and transcription profiling.

Microarray analysis

In short, RNA was extracted and purified using the RNeasy kit (Qiagen, Valencia, U.S.A). Amplification and labeling was performed with the MessageAmp aRNA kit (Ambion, Austin, U.S.A). Amplified RNA was used to generate labeled cDNA with incorporation of 5-(3-aminoallyl)-dUTP and labeled with either Cy3 or Cy5 mono-reactive dye (Amersham, Piscataway, U.S.A.). All cDNA products were purified using the RNeasy kit (Qiagen, Valencia, U.S.A). The CATMA (Allemeersch et al., 2005) Arabidopsis DNA microarrays were provided by Utrecht University (the Netherlands) and were produced from a set of (150-450)-mers representing 20833 unique genes. Hybridizations performed on CATMA arrays and arrays were scanned using ScanArray Express HT (PerkinElmer, Wellesley, U.S.A.) and quantified with Imagene 6.0 (BioDiscovery, El Segundo, U.S.A.).
The Limma package (Smyth et al., 2005; Ritchie et al., 2006; Ritchie et al., 2007) for the statistical work environment R was used for normalization. All array data is submitted to ArrayExpress.

Statistical analysis

To determine differential abundance of transcripts between the two parents, we applied a linear model using the Limma package (Ritchie et al., 2006; Ritchie et al., 2007) for the statistical work environment R. For each gene the *p* value and the corresponding *q* values (Storey and Tibshirani, 2003) were computed.
The following adaptations on the procedures developed in GG1 were used to calculate the broad sense heritability (H^2) in the parents and RILs. Log-signal intensities of gene expression were used to test for genetic variance of expression traits. We compared this genetic variance with and without removing spot effects, by treating them as a random effect in a linear mixed model. Heritability of expression in the parental accessions was calculated as follows (Hegmann and Possidente, 1981):

$$H_P^2 = \frac{0.5 \times Vg}{0.5 \times Vg + Ve}$$

where *Vg* and *Ve* represent the components of variance among and within accessions respectively. The factor 0.5 was applied to adjust for the 2 times over-estimation of additive genetic variance among inbred strains.
Heritability of expression within the RIL population was calculated using the pooled variance of the parents as an estimate of the within line variance.

$$H^2_{RIL} = \frac{V_{RIL} - V_e}{V_{RIL}}$$

Where V_{RIL} and V_e are the variance among adjusted expression intensities in the segregants and the pooled variance within parental measurements, respectively. To prevent overestimation we removed outlier's more than three standard deviations away from the mean values. We discarded negative heritability values.
Transgressive segregation was determined in terms of the standard deviation of the individual parents (Brem and Kruglyak, 2005). We calculated the number of RILs, *n*, whose expression level lay above $\mu_{max}+2*SD_{max}$ or below $\mu_{min}-2*SD_{min}$; where μ and *SD* are the mean and the standard deviation of the parental phenotypic values, respectively, and max indicates the parent with the highest value and min indicates the parent with the lowest value.

Multiple QTL analysis

eQTL were mapped using the procedures based on MQM mapping, as developed in GG1. A genome-wide p value threshold of $2.23x10^{-3}$ at α=0.05 for a single trait was estimated by a 10,000 permutation test (Churchill and Doerge, 1994). But for a study with 24,065 gene transcripts, we controlled the false discovery rate (FDR) based on the pool of p values for all markers and all transcripts. Because the p values are correlated when markers are linked, the FDR increases depending on the number of markers on a chromosome (Benjamini and Yekutieli, 2005). In our experiment the maximum number of markers reached 35 (chromosome 5) and a simulation analysis (not shown) using Storey's algorithm to control the FDR (Storey, 2002) at a desired level indeed showed a 4.4 fold increase of the actual FDR. To account for this we corrected the FDR by a factor 5 and calculated the genome wide p value threshold at Storey's FDR of 0.01 for all gene-marker p values to make sure that the real FDR rate is below 0.05 (corrected FDR=0.05). The estimated p-value threshold then corresponded to $5.29x10^{-5}$. As neighboring markers stand for partially/largely similar tests, we decided to leave them out, resulting in significance thresholds of –log(P) = 3.83 in GG1and 3.62 and GG2. and this threshold was used as significance threshold for the detection of eQTL.
Explained variance of detected eQTLs was estimated by fitting expression ratios of all detected eQTLs and their interactions in a linear model. We used ANOVA to estimate the fraction of variance explained by each eQTL and eQTL interactions.

Threshold determination

To decide which genetic factors have a significant effect on the regulation of gene expression, we needed to establish a significance threshold. The False Discovery Rate (FDR) was used to correct the threshold for the large number of simultaneous tests. To determine the FDR in this experimental setup, we used a specific method, which takes into

account, that neighboring markers stand for partially/largely similar tests and are thus left out. (For each eQTL, we only considered one gene/marker combination representing the eQTL, and discarded the other marker/gene combination within the 1.5 log(P) interval of the eQTL in the multiple testing procedure.) Without correction for this redundancy, the threshold will be too strict, due to the increased number of "significantly linked" markers in the whole set. With our method, we determined the FDR threshold in both GG1 and GG2 to be 3.83 –log(P) and 3.62 –log(P) respectively.

Correlation of SNP frequency, gene density and eQTL distribution.

The SNP set published in (Nordborg et al., 2005) was used. The SNP frequency and gene density were determined for each marker by counting the number of SNP or genes between the flanking markers of the marker under study (marker locus). Indels were treated as a single polymorphism and the number of SNPs was corrected for the number of sampled sequences per marker locus. The number of eQTLs at the marker locus, were determined by the number of eQTLs above the established genome-wide threshold.

Determination of eQTL overrepresentation in annotation groups

We determined the overrepresentation of eQTLs in annotation groups in both experiments and in the comparison of the two experiments (GG1 and GG2). The overrepresentation per experiment was determined by comparing the amount of genes with an eQTL in a specific annotation group with the amount of genes with an eQTL in a random group of genes of similar size. A random group of genes will contain a similar percentage of genes with an eQTL as the whole set average. However the standard deviation (sd) is dependent on the group size that is drawn from the whole set (e.g. a small group has a relatively higher sd that a larger group). For the group sizes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 75, 100, 200, 300 and 400 genes the mean and sd of genes with an eQTL was determined by 1000 permutations. With the mean and the sd of these random groups of genes the confidence interval of every group size could be determined and hence the annotation groups that contain more eQTLs than can be expected by chance.

To compare the two experiments the means were subtracted and the confidence intervals determined by the combination of the standard deviations of both experiments. In this way annotation groups specifically under genetic control in one of both experiments are farthest from zero.

Chapter 6

Genetic variation in shade-affected transcript abundance is underlying the difference in phenotype between L*er* and a Cvi introgression line.

L. Basten Snoek & Inez R. Terpstra, Martijn van Zanten, Hendrik Poorter, Dirk Blom, Guido van den Ackerveken, Laurentius A.C.J. Voesenek & Anton J.M. Peeters

Abstract

Recent advancements in especially transcript profiling opened the possibility to measure many transcript levels (variables) under different environmental conditions and for many different genotypes. To obtain biologically relevant answers from these colossal datasets, statistical methods are required to establish relevant relations between genes and between transcript levels and phenotypes. This has been applied successfully in a number of experiments, particularly in genetical genomics. One of the characteristics of building networks is the vast number of possible starting points (genes) when all the natural variation in transcript abundance is considered. First, we reduced the number of possible starting points in unraveling a transcript regulation network, biologically. Natural variation can be limited by using near isogenic lines (NIL, LCN1-10), in which only the introgressed locus is different when compared to the wildtype. When for this genotype the expression profile is measured only that part of the natural variation caused by the introgression is revealed instead of all the variation of two recombined genotypes. Secondly, we used several methods, like local association mapping, trait transcript correlation, eQTL position and the difference in transcript abundance in LCN1-10, each of which can act as candidate gene selection factors. The results from these analyses where combined by rank product testing. The best candidate gene putatively causal for the difference in phenotype between L*er* and LCN1-10 was used to generate a putative transcript regulation network. Furthermore, genes in this network were used to enlighten the biological processes putatively leading to the differences in shade-affected phenotypes.

Introduction

One of the major goals of quantitative genetic research is the understanding of the complex interaction between molecular genetic factors leading to variation in phenotypes. To enable this, natural variation has been exploited extensively in genetic research the last decade by quantitative trait loci (QTL) mapping (Alonso-Blanco and Koornneef, 2000; Borevitz and Nordborg, 2003; Koornneef et al., 2006; Keurentjes et al., 2008). In *Arabidopsis thaliana* several naturally functional different alleles have been identified by phenotypic screens of recombinant inbred line (RIL) populations and subsequent QTL mapping (Alonso-Blanco et al., 1998; El-Assal et al., 2001; Kliebenstein et al., 2001; Alonso-Blanco et al., 2003). For many traits, QTLs have been identified, but in most cases the causal genes have not been cloned. The selection of candidate genes and subsequent fine-mapping, cloning and complementation of alleles is laborious and still is a bottleneck in the process of pinpointing QTLs to quantitative gene(s) and quantitative trait nucleotide polymorphism(s) (QTN).

QTN containing gene(s) were successfully identified by applying local association mapping (Ehrenreich et al., 2007; Zhao et al., 2007). By sequencing (part) of the QTL in multiple accessions sequence polymorphisms can be identified. Subsequently, polymorphisms associated with the variation in the trait of interest can be found. These associated polymorphisms result in a reduction of candidate genes or even the cloning of the gene causal for the quantitative variation. Nowadays, the ever increasing number of published polymorphisms between accessions enables local association-mapping for nearly all chromosomal positions (Kim et al., 2006).

The recent development of a population of near isogenic lines (NILs) containing the segmented Cape Verde Island (Cvi) genome in a Landsberg *erecta* (L*er*) background (Keurentjes et al., 2007) can speed up the confirmation and fine-mapping of QTLs segregating between these two accessions. It further provides a way to find extra QTL for certain phenotypes as NILs can be phenotypically different from L*er*, even when no (clear) QTL was found at the NIL specific loci in the recombinant inbred lines (RILs) (Keurentjes et al., 2007). Extensive effects of Cvi introgressions in a L*er* background on transcript levels have been observed by Juenger et al. (2006) who compared transcription profiles of L*er*-2 and a NIL containing two small Cvi introgressions on chromosome 2 and 3 in a L*er*-2 background (Juenger et al., 2006). Many transcript levels were influenced by these introgressions also of genes physically located outside the introgressed Cvi locus. It is likely that those variations in transcript levels are causal to the variation in phenotypes between L*er* and a NIL.

Recently, extensive evidence for this idea was found by applying genetical genomics to Arabidopsis (Keurentjes et al., 2007; chapter 5). With the successful establishment of genetical genomics in Arabidopsis it is possible to link the genome wide variation in transcript levels to the genotype. Especially the chromosomal positions of eQTLs for local regulation can be used to decrease the number of candidate genes further. When variations in phenotypes are caused by variations in transcript abundance they will be correlated. By searching for correlation between the phenotypes and the transcript levels in the RILs, candidate genes can be selected more precisely. Using this method, candidate genes for several traits have been identified in *Mus musculus* (mouse) (Bystrykh et al.,

2005) and *Saccharomyces cerevisiae* (yeast) (Bing and Hoeschele, 2005; Brem and Kruglyak, 2005)

Each of these four methods reveals genes most likely associated with the variation in phenotypes observed in the individuals used in the experiments. A phenotypic QTL can thus be studied in more detail by applying these four selection criteria, i) a trait associated SNP, ii) transcript level differences between L*er* and a NIL, iii) eQTLs and iv) the correlation of transcript levels and traits in a RIL population. This will result in four lists of candidate genes. We hypothesize that the genes that are identified as top candidates in all of the lists are, by rank product testing (Breitling et al., 2004), the most likely genes causal for the variation in phenotypes.

We measured the genome wide transcription profile of a NIL compared to L*er*. This NIL, LCN1-10 has an introgression harboring some of the QTLs found in chapter 2. Furthermore, LCN1-10 showed also some differences for some of the phenotypes measured in chapter 3. To extend our selection criteria we compared the expression profiles with a genetical genomics experiment (chapter 5). In that experiment, Genetical Genomics 2 (GG2), Arabidopsis plants were treated with 3 hours of spectral natural shade and major natural variation in shade-induced transcript levels was uncovered. To enable comparison between the experiments, L*er* and LCN1-10 were subjected to the same neutral shade treatment prior to transcript profiling. The transcript level differences of LCN1-10 and L*er* can be combined with the eQTLs from GG2 to select candidate genes. This will be an interesting comparison by itself because many eQTLs can be found on the introgression of LCN1-10. This locus was previously identified as a hotspot for transcript regulation (HTR) in GG2 (Chapter 5) and the expression profile of LCN1-10 should reflect this. The last refinement in our list of candidate genes comes from association mapping with the single nucleotide polymorphisms (SNPs) located on the introgression of LCN1-10. The ~20 accessions profiled with Perlegen (www.perlegen.com/**)** SNP-arrays accessible through TAIR were used for this (ftp://ftp.arabidopsis.org/home/tair/Sequences/Perlegen_Array_Resequencing_Data_2007/).

In this chapter we present our findings for each of the four selection criteria and the candidate genes selected by combining the criteria. The phenotypes for which LCN1-10 differs from L*er* (Chapter 2 & 3) are initial petiole, leaf and petiole bending angle and rosette area, compactness and fermax (a measure for rosette diameter). By correlating variation in transcript levels and phenotypes we identified *PHYB* as a candidate gene. The alleles of this gene were previously identified as functional different between L*er* and Cvi (chapter 3). By combining the difference in transcript levels of LCN1-10 and eQTLs from GG2 we found many putative regulators. For local association mapping we showed that it can result in further refinement of the candidate gene list. Finally, we present a putative transcriptional regulatory network for the best candidate gene and used the members of this network to enlighten the biological processes underlying the phenotypic difference between LCN1-10 and L*er* with respect to the traits under study.

Results

LCN1-10 genotype and phenotype

Trait phenotype

Near isogenic line (NIL) LCN1-10 was identified as phenotypically different from L*er* for traits describing the position of the leaf: petiole angle, lamina angle and leaf angle before and after a neutral shade treatment (chapter 2). Furthermore, we measured the spectral neutral shade-influenced traits plant area, fermax and rosette compactness (chapter 3). In general LCN1-10 is a larger plant with a less compact rosette and more upright (hyponastic) leaves (Fig. 6.1). This NIL has two small Cvi introgressions of each of 3 to 4 Mbp in length, ~5 – ~8.2 Mbp (~950 genes; located at approximately 1/6th from the top of chromosome 1) and ~26.5 – ~30.4 (~1150 genes; end of chromosome 1). We assumed that one or more genes on these introgressions must be causal for the differences in these phenotypes.

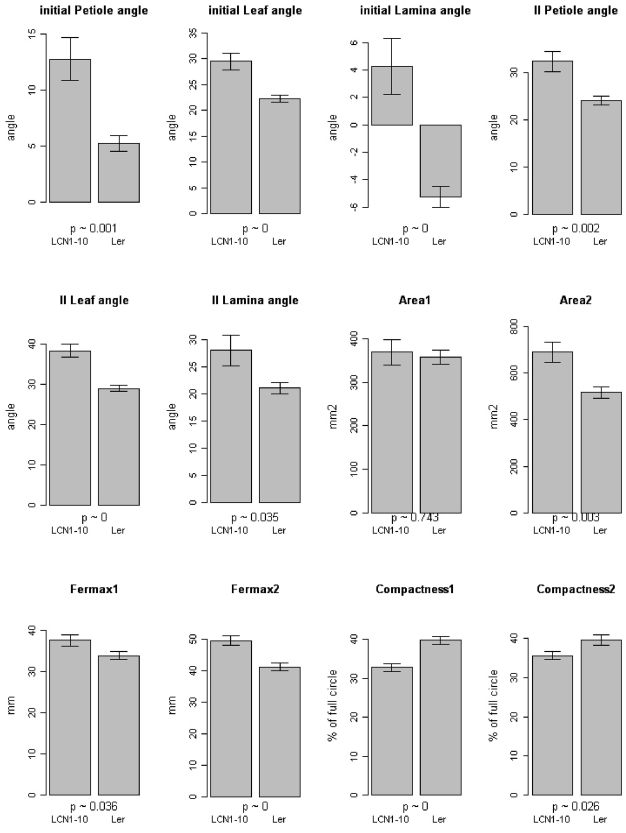


Figure 6.1 Phenotypes of LCN1-10 compared to L*er*. Initial and low-light (ll) treated Petiole, Lamina and Leaf angles; the measurements are explained in chapter 2 of this thesis. Area, Fermax and Compactness at two time points; the measurements are explained in chapter 3 of this thesis. Standard errors are indicated. n =12.

Expression phenotype

To identify genes which transcript levels were influenced by the physical Cvi introgression of LCN1-10 we compared the transcript levels with those of L*er*. We used a 3 h shade treatment to enable comparison of transcript abundance with the genetical genomics experiment described in chapter 5 (called GG2). This shade treatment was found to attenuate the phenotypes caused by a difference in response to the light condition during growth (chapter 5). We identified 2146 genes (false discovery rate, FDR=0.05) with differently abundant transcripts (DATs) between LCN1-10 and L*er*. Of these 2146 genes 166 are physically located on the first introgression and 155 on the second. This means that the majority of the genes (1825 ~85%) with DATs are physically located outside of the introgressions and thus must result from trans-regulation. This is also reflected by the relatively limited overlap with the parental DATs (SI data 6.1). The genes with DATs between LCN1-10 and L*er* therefore, form a selection criterion. Not only must the candidate gene be physically located on the introgression, it is very likely that its transcript levels are also different.

Trait correlation with transcript abundance variation in the L*er* x Cvi recombinant inbred line population

Some of the natural variation in transcript abundance in the shade treated recombinant inbred lines (RILs) is assumed to be causal for the natural variation in traits like growth and position of the leaves.

To identify genes for which variation in transcript abundance may be underlying the variation in these light-influenced traits, the correlation between the phenotypic values and transcript levels of each RIL was calculated. We found 641 genes showing a significant correlation ($p<0.001$) with at least one of the studied traits. These genes, as well as the traits, were clustered based on trait/transcript level correlation and compared to identify similar regulated groups of traits. Several groups of genes specifically correlated to one or a group of traits were found (Fig. 6.2; Table 6.1 and SI Data 6.2). The correlation of those genes to the specific trait groups serve as a selection criterion to select the gene(s) causal for the phenotypes of LCN1-10 when compared to L*er*. Several QTLs for these traits were also found at the positions of the introgressions of LCN1-10 (chapter 2).

Table 6.1 Clusters of traits that could be identified after clustering correlated gene transcripts. See also figure 6.1. For description of the traits see chapter 2 and 3.

Group	Traits
1	Petiole angle (PA) before and after low-light treatment; Leaf shape angle (LSA) before and after low-light treatment; Leaf tip angle (LTA) after low-light treatment.
2	Response to low-light of: petiole bending angle (PBA); leaf blade bottom angle (LBBA) ; leaf tip angle (LTA) and leaf blade top angle (LBTA).
3	Leaf mass area (LMA) of low-light and high-light grown plants.
4	Compactness (both time points).
5	Petiole bending angle (PBA) before and after treatment with low-light.
6	Growth traits: Fermax (both time points and difference between the two); Area (both time points and difference between the two); Relative growth rate (RGR)

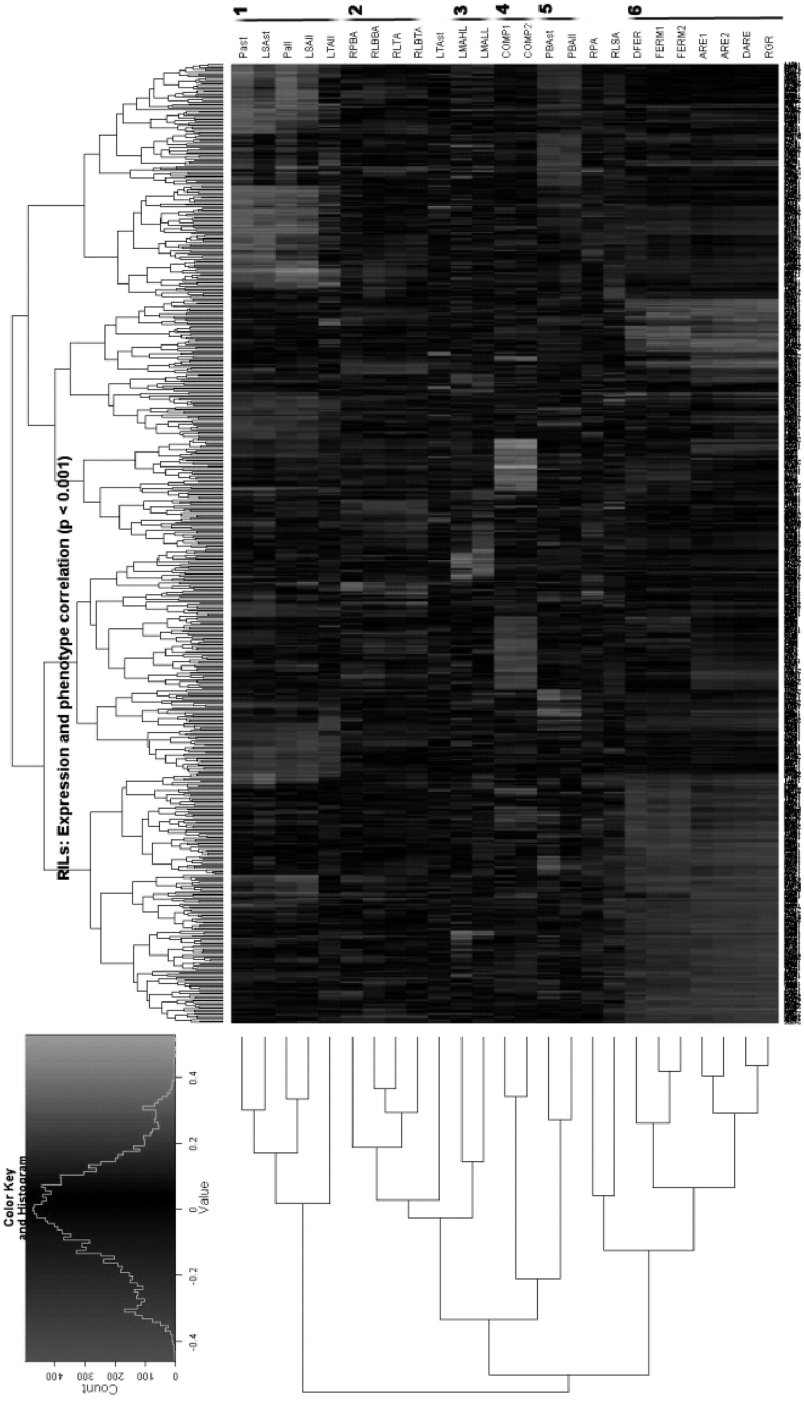


Figure 6.2 RILs; expression and phenotype correlation (p<0.001). Traits and genes clustered by the correlation between the variation in traits and the variation in transcript levels in the RILs. Six clusters of traits were identified based on the similarity of genes to which they were correlated (described in table 6.1).

eQTLs on the introgression of LCN1-10

In chapter 5 we measured the genome wide transcript levels of the L*er* x Cvi recombinant inbred lines (RILs) for genetical genomics. Many expression quantitative trait loci (eQTL) could be calculated. These eQTL showed a hotspot for transcript regulation on the first introgression of LCN1-10. Also on the second introgression multiple eQTLs were found. The genes with DATs between LCN1-10 and L*er* are therefore very likely to have an eQTL. Indeed, most eQTLs mapped to one of the introgressions of LCN1-10 (Fig. 6.3). In total we found that ~47% of the 2147genes with DATs between LCN1-10 and L*er* had an eQTL (Fig. 6.3). Of those genes ~25% had an eQTL on one of the introgressions. The eQTLs mapping to the introgressions were mostly local eQTLs (~54% total). More local eQTLs, relative to all eQTLs, mapping to the second introgression were found (~69%), whereas less local eQTLs that mapped to the first introgression were found (~40%). The rest of the DATs with eQTLs did not map to one of the introgressions (~75% of DATs with eQTLs). These are probably regulated indirectly by a gene on one of the introgressions. Again, this shows that distant regulation occurs more frequently. Furthermore, the eQTLs on the introgressions are the third selection criterion to find the gene(s) causal for the differences in phenotype between LCN1-10 and L*er*.

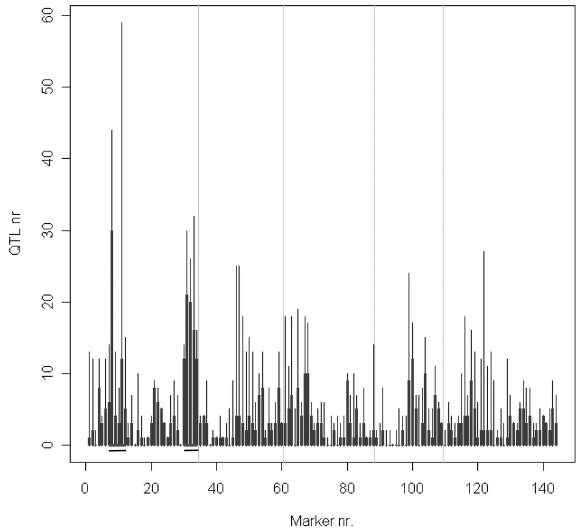


Figure 6.3 Expression quantitative trait loci (eQTL) distribution of the genes with differently abundant transcripts (DATs) between LCN1-10 and L*er*. eQTLs for distant regulation are in thin lines and eQTLs for local regulation are in thick lines, total height of line stands for all eQTL. Chromosome 1 to 5 are indicated left to right separated by the grey lines. Marker numbers are cumulative. Cvi introgressions of LCN1-10 are indicated by the horizontal black lines (on chromosome 1).

SNP correlation with Phenotypes (local association mapping)

To identify single nucleotide polymorphisms (SNPs) correlated to the variation in the traits we used the "Perlegen" SNP set available though TAIR and measured the six traits of the accessions of that set for which L*er* and LCN1-10 exhibits different phenotypes. The SNP must be located on one of the introgressions of the NIL to be causal for the difference in phenotypes. We therefore limited our association mapping to the introgressions of LCN1-10 and the SNPs for which L*er* and Cvi were different. Although only 20 accessions were used this can be very useful when combined with other data. We

found many associated SNPs, for some traits with p-values as low as 10^{-19}. For each gene we determined the SNP with the lowest p-value and used that as a fourth selection criterion.

Combining all information to find genes most likely to be causal for the variation in phenotype.

We have defined four selection criteria; i) Variation in transcript levels correlated to the variation in phenotypes, ii) Genes physically located on the introgression of LCN1-10, especially those with DATs between LCN1-10 and L*er*, iii) Genes with an eQTL on one of the introgressions of LCN1-10 and iv) Genes containing a SNP associated with the variation in one of the traits. The four selection criteria are combined by rank ordering them according to their best linked members (which are then the genes on top of the list) and calculating their likelihood for the individual position in the four lists. Hereafter, the 4 individual likelihood values are multiplied to obtain a combined estimate by rank product (Breitling et al., 2004). This estimate is used to find those genes most likely to be causal for the different phenotypes of LCN1-10 compared to L*er* under the defined environmental conditions.

Candidate genes

The top 2% of the genes of the combined criteria lists for each trait were selected as initial candidate genes (Table 6.2 to 6.7). Interestingly, some genes appeared in multiple lists and might therefore be causal to some of the variation in all of these traits. Many of these genes were also present in the lists of combined traits (Table 6.8).

Table 6.2 Top 2% of genes with a score linked to initial petiole angle. Gene: Agi number, NILDAT: difference in transcript level (log ratio) between L*er* and LCN1-10 (negative values are transcripts less abundant in L*er*), BH: Benjamini-Hoghberg adjusted p-values for NILDAT, eQTL: - log p score multiplied by the sign of the effect, Traitcorrel: Correlation of the variation in trait with the variation in transcript level in the RILs, p-val: p-value of the correlation between the trait and transcript level, The rank scores for Trait, NILDAT, eQTL, SNP and the rank product adjusted p-value are in the last five columns.

Gene	NIL DAT	BH	eQTL	Trait correl	p-val	Trait	NIL DAT	eQTL	SNP	Rank product
At1g15730	-0.78	0.0325	-11.63	0.38	0.0000	1	89	21	10	<0.0001
At1g19090	-0.20	0.2295	NA	0.34	0.0002	2	283	748	2	0.0006
At1g16880	-0.98	0.0188	-17.48	0.19	0.0474	108	72	5	34	0.001
At1g23780	-1.04	0.0004	-14.09	0.18	0.0480	109	2	11	748	0.0011
At1g16920	-0.52	0.0375	-18.49	0.13	0.1564	240	98	3	31	0.0012
At1g16080	-0.48	0.0266	-14.13	0.29	0.0019	8	83	10	748	0.0031
At1g16460	-1.43	0.0007	-17.08	0.17	0.0699	148	8	6	748	0.0032
At1g17250	-1.12	0.0326	-13.06	0.30	0.0013	6	90	14	748	0.0034
At1g19050	-0.61	0.4530	-20.94	0.26	0.0055	18	431	1	748	0.0036
At1g21270	-1.56	0.0011	-5.51	0.08	0.3815	417	13	86	14	0.0043
At1g16445	0.82	0.0094	12.36	-0.21	0.0260	69	54	18	105	0.0044
At1g16840	0.26	0.3208	17.92	-0.15	0.1084	201	336	4	30	0.0051
At1g15880	0.43	0.0447	10.10	-0.05	0.5737	542	113	31	6	0.0068
At1g17190	1.62	0.0027	14.29	-0.07	0.4762	479	27	9	117	0.0076
At1g15900	-1.35	0.0004	-8.64	0.18	0.0591	136	3	51	748	0.0088

Table 6.3 Top 2% of genes with a score linked to initial PBA, for legend see table 6.2

Gene	NIL DAT	BH	eQTL	Trait correl	p-val	Trait	NIL DAT	eQTL	SNP	Rank product
At1g15730	-0.78	0.0325	-11.63	-0.33	0.0003	6	89	21	2	<0.0001
At1g23780	-1.04	0.0004	-14.09	-0.23	0.0132	44	2	11	26	<0.0001
At1g16840	0.26	0.3208	17.92	0.38	0.0000	1	336	4	748	0.0006
At1g23050	-0.72	0.0040	-11.09	-0.23	0.0131	43	38	23	33	0.0009
At1g17160	0.43	0.0559	14.81	0.13	0.1670	204	124	8	8	0.0011
At1g16460	-1.43	0.0007	-17.08	-0.23	0.0150	47	8	6	748	0.0011
At1g15900	-1.35	0.0004	-8.64	-0.28	0.0027	15	3	51	748	0.0011
At1g16880	-0.98	0.0188	-17.48	-0.30	0.0010	9	72	5	748	0.0013
At1g14660	-0.88	0.0059	-10.58	-0.36	0.0001	3	45	27	748	0.0017
At1g15960	2.07	0.0012	12.60	0.27	0.0033	18	15	16	748	0.0021
At1g21270	-1.56	0.0011	-5.51	-0.14	0.1379	180	13	86	21	0.0027
At1g14890	1.15	0.0029	10.30	0.33	0.0003	7	29	30	748	0.0029
At1g16920	-0.52	0.0375	-18.49	-0.26	0.0049	23	98	3	748	0.0032
At1g19050	-0.61	0.4530	-20.94	-0.28	0.0031	16	431	1	748	0.0032
At1g22240	0.02	0.9580	6.95	0.15	0.1160	153	715	66	1	0.0044

Table 6.4 Top 2% of genes with a score linked to initial leaf angle, for legend see table 6.2

Gene	NIL DAT	BH	eQTL	Trait correl	p-val	Trait	NIL DAT	eQTL	SNP	Rank product
At1g16460	-1.43	0.0007	-17.08	-0.12	0.2085	155	8	6	230	0.0011
At1g23205	1.64	0.0012	9.04	-0.17	0.0570	35	16	44	77	0.0012
At1g14680	-1.26	0.0008	-9.11	0.17	0.0707	44	10	43	103	0.0012
At1g16920	-0.52	0.0375	-18.49	0.07	0.4157	309	98	3	22	0.0012
At1g15730	-0.78	0.0325	-11.63	0.06	0.5261	411	89	21	3	0.0012
At1g23780	-1.04	0.0004	-14.09	0.11	0.2486	184	2	11	734	0.0018
At1g21270	-1.56	0.0011	-5.51	0.10	0.2543	188	13	86	17	0.0023
At1g23050	-0.72	0.0040	-11.09	0.15	0.0931	60	38	23	75	0.0026
At1g17190	1.62	0.0027	14.29	-0.15	0.1012	68	27	9	277	0.003
At1g16880	-0.98	0.0188	-17.48	-0.03	0.7393	558	72	5	28	0.0034
At1g15880	0.43	0.0447	10.10	-0.15	0.0938	61	113	31	32	0.0044
At1g15900	-1.35	0.0004	-8.64	0.07	0.4372	330	3	51	190	0.0055
At1g14900	0.78	0.0039	5.43	-0.19	0.0416	25	36	90	123	0.0056
At1g14660	-0.88	0.0059	-10.58	0.14	0.1352	94	45	27	101	0.0068
At1g16840	0.26	0.3208	17.92	-0.05	0.5636	441	336	4	21	0.007

Table 6.5 Top 2% of genes with a score linked to rosette area, for legend see table 6.2

Gene	NIL DAT	BH	eQTL	Trait correl	p-val	Trait	NIL DAT	eQTL	SNP	Rank product
At1g15900	-1.35	0.0004	-8.64	0.01	0.9151	709	3	51	4	0.0001
At1g21270	-1.56	0.0011	-5.51	0.03	0.7064	575	13	86	1	0.0002
At1g23040	-0.90	0.0013	-10.08	-0.15	0.1090	111	18	32	13	0.0006
At1g23780	-1.04	0.0004	-14.09	-0.15	0.0930	96	2	11	748	0.001
At1g23070	-0.63	0.0031	NA	-0.23	0.0119	16	31	748	7	0.0016
At1g16460	-1.43	0.0007	-17.08	-0.11	0.2449	238	8	6	748	0.0052
At1g20130	-1.25	0.0011	NA	0.20	0.0269	30	14	748	33	0.0059
At1g14660	-0.88	0.0059	-10.58	0.23	0.0114	14	45	27	748	0.0073
At1g15880	0.43	0.0447	10.10	-0.02	0.7894	621	113	31	6	0.0073
At1g17160	0.43	0.0559	14.81	0.00	0.9577	737	124	8	20	0.008
At1g15960	2.07	0.0012	12.60	-0.16	0.0788	83	15	16	748	0.0085
At1g16880	-0.98	0.0188	-17.48	0.17	0.0616	66	72	5	748	0.0092
At1g17190	1.62	0.0027	14.29	-0.15	0.1062	110	27	9	748	0.0101
At1g20120	-0.65	0.0515	-4.04	0.09	0.3495	331	122	139	5	0.0142
At1g14680	-1.26	0.0008	-9.11	0.15	0.1053	108	10	43	748	0.0161

Table 6.6 Top 2% of genes with a score linked to fermax, for legend see table 6.2

Gene	NIL DAT	BH	eQTL	Trait correl	p-val	Trait	NIL DAT	eQTL	SNP	Rank product
At1g23780	-1.04	0.0004	-14.09	-0.15	0.0952	110	2	11	26	<0.0001
At1g21270	-1.56	0.0011	-5.51	0.17	0.0569	61	13	86	2	<0.0001
At1g16460	-1.43	0.0007	-17.08	-0.13	0.1426	144	8	6	748	0.0032
At1g15900	-1.35	0.0004	-8.64	0.00	0.9865	753	3	51	45	0.0032
At1g14880	-3.88	0.0009	-8.95	0.13	0.1484	146	12	47	73	0.0038
At1g23040	-0.90	0.0013	-10.08	-0.12	0.2096	202	18	32	59	0.0044
At1g20130	-1.25	0.0011	NA	0.19	0.0403	44	14	748	19	0.0052
At1g14660	-0.88	0.0059	-10.58	0.22	0.0147	16	45	27	748	0.008
At1g17160	0.43	0.0559	14.81	-0.04	0.7033	586	124	8	31	0.0095
At1g16880	-0.98	0.0188	-17.48	0.16	0.0775	87	72	5	748	0.0121
At1g15730	-0.78	0.0325	-11.63	0.13	0.1595	162	89	21	85	0.0132
At1g15380	-1.92	0.0005	-2.59	0.02	0.8587	679	4	199	49	0.0134
At1g17890	0.26	0.0764	12.38	-0.12	0.2098	203	152	17	53	0.0139
At1g20330	0.43	0.1495	4.80	0.06	0.5137	448	224	113	3	0.016
At1g19850	0.67	0.0240	NA	-0.02	0.8708	684	79	748	1	0.0178

Table 6.7 Top 2% of genes with a score linked to rosette compactness, for legend see table 6.2

Gene	NIL DAT	BH	eQTL	Trait correl	p-val	Trait	NIL DAT	eQTL	SNP	Rank product
At1g16920	-0.52	0.0375	-18.49	0.15	0.0947	50	98	3	24	<0.0001
At1g16460	-1.43	0.0007	-17.08	0.14	0.1285	71	8	6	125	0.0001
At1g23050	-0.72	0.0040	-11.09	0.20	0.0294	14	38	23	112	0.001
At1g23780	-1.04	0.0004	-14.09	0.05	0.6065	425	2	11	158	0.001
At1g17190	1.62	0.0027	14.29	-0.13	0.1491	81	27	9	76	0.001
At1g21270	-1.56	0.0011	-5.51	-0.21	0.0205	9	13	86	155	0.001
At1g15960	2.07	0.0012	12.60	-0.20	0.0286	13	15	16	748	0.0012
At1g16880	-0.98	0.0188	-17.48	0.05	0.5803	409	72	5	35	0.0032
At1g15730	-0.78	0.0325	-11.63	-0.03	0.7830	573	89	21	6	0.0043
At1g16840	0.26	0.3208	17.92	-0.08	0.4013	269	336	4	23	0.0052
At1g14680	-1.26	0.0008	-9.11	0.18	0.0530	26	10	43	748	0.0052
At1g23070	-0.63	0.0031	NA	-0.27	0.0031	4	31	748	92	0.0052
At1g16445	0.82	0.0094	12.36	-0.14	0.1161	63	54	18	156	0.0055
At1g15380	-1.92	0.0005	-2.59	-0.18	0.0441	21	4	199	748	0.0071
At1g20270	-0.50	0.0155	2.46	-0.19	0.0359	17	66	212	66	0.0088

Table 6.8 Top 2% of genes with combined adjusted rank product scores. Gene: Agi code of selected genes, Scores are given for eQTL, difference in transcript level between L*er* and LCN1-10 (NILexpr), SNP associated with compactness, fermax, area and petiole, petiole bending and leaf angle, the correlation of transcript levels with those traits as well as the combined scores.

Gene	At1g21270	At1g15730	At1g20120	At1g23560	At1g15880	At1g15550	At1g15760	At1g17330	At1g23050	At1g23780	At1g16445	At1g23070	At1g16880	At1g16920	At1g20330	At1g16840	At1g22920	At1g22170	At1g17190	At1g20990
eQTL	86	21	139	175	31	748	748	33	23	11	18	748	5	3	113	4	158	189	9	165
NIL expr	13	89	122	445	113	504	167	703	38	2	54	31	72	98	224	336	471	121	27	413
Comp SNP	155	6	111	42	121	11	21	163	112	158	156	92	35	24	51	23	10	219	76	98
Comp Trait	9	573	386	47	320	7	744	465	14	425	63	4	409	50	304	269	402	251	81	381
Ferm SNP	2	85	9	134	67	112	76	37	748	26	748	68	748	748	3	748	132	36	748	748
Ferm Trait	61	162	408	446	427	288	453	235	298	110	286	470	87	506	448	182	149	52	343	223
Area SNP	1	748	5	748	6	748	27	41	748	748	748	7	748	748	39	748	748	748	748	748
Area Trait	575	287	331	203	621	212	399	249	740	96	443	16	66	163	457	142	69	15	110	112
Score rosette	0	0.0106	0.0037	0.1266	0.009	0.0406	0.0903	0.0418	0.0437	0.0008	0.0878	0.0001	0.0095	0.0094	0.0062	0	0.0106	0.0037	0.1266	0.009
Pet SNP	14	10	124	43	6	16	7	102	125	748	105	114	34	31	748	30	101	45	117	46
Pet Trait	417	1	313	11	542	27	424	21	350	109	69	701	108	240	251	201	304	306	479	161
PBA SNP	21	2	748	10	748	4	3	748	33	26	32	748	748	748	748	748	748	748	748	12
PBA Trait	180	6	11	52	70	495	125	68	43	44	317	360	9	23	246	1	138	571	151	223
Leaf SNP	17	3	74	19	32	10	9	289	75	734	229	68	28	22	78	21	62	23	277	550
Leaf Trait	188	411	24	644	61	85	104	214	60	184	333	198	558	309	83	441	431	614	68	118
Score angle	0.0002	0	0.0063	0.0005	0.0018	0.0008	0.0004	0.0418	0.0006	0.0008	0.0045	0.2552	0.0004	0.0008	0.2375	0.0002	0	0.0063	0.0005	0.0018
Score All	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

The best candidate gene was At1g21270, *WALL ASSOCIATED KINASE2* (*WAK2*), a wall associated kinase involved in uni-dimensional cell growth, cellular osmoregulation and oligosaccharide metabolic processes and responsive to salicylic acid (He et al., 1999; Kohorn et al., 2006). Another candidate gene found in the list of top candidates was also involved in cell growth. This gene At1g20330, *STEROL METHYLTRANSFERASE2* (*SMT2*), is involved in multi-dimensional cell growth and sterol biosynthesis (Carland et al., 1999; Hase et al., 2000; Carland et al., 2002; Hase et al., 2005). The functions of these two genes correspond to the "growth" part of the studied traits. Other top candidates were found that correspond more to the influence of light on the studied traits. Gene At1g15550 *GA REQUIRING4 (GA4)*, a gibberellin 3-beta-dioxygenase involved in gibberellin signaling and biosynthetic process, as well as the response to red/far-red light (Cowling et al., 1998; Hisamatsu et al., 2005; Matsushita et al., 2007; Oh et al., 2007) is one of these genes. Another likely candidate gene, At1g22920, *AJH*, encodes a subunit of the COP9 complex that is involved in protein (cullin) deneddylation and the negative regulation of photomorphogenesis (Karniol and Chamovitz, 2000; Gusmaroli et al., 2007). AJH is also responsive to auxin (Dohmann et al., 2005). Furthermore, a glutathione transferase, At1g17190, *ARABIDOPSIS THALIANA GLUTATHIONE S-TRANSFERASE 26* (*AtGSTU 26*), was found and a GTP binding protein involved in intracellular protein transport At1g16920, *RAB GTPASE HOMOLOG A1B* (*RAB11*). Less was known about four genes that were annotated as being located in the chloroplast, At1g16445, At1g16880, At1g20990 and At1g22170. The last six genes were of unknown function, At1g15760, At1g16840, At1g17330, At1g23050, At1g23070 and At1g23560; these may be possible new players in light affected growth traits.

Transcript regulation networks

We selected the top gene from our combined trait list as a start for a putative network. Possible target genes of *WAK2* were identified by combining co-expressed genes in the RILs, eQTL position and regulation of transcription in LCN1-10. The first level targets and regulators to which *WAK2* transcript abundance was best correlated are visualized in Figure 6.4. Three of the putative first level targets are involved in glucosinolate and auxin synthesis, *CYP83A1* (At4g13770), *CYP79F2* (At1g16400) and a glycosyl transferase (At3g23760). We investigated if more genes in this pathway are affected by the genetic background and the neutral shade treatment. In the second level targets of *WAK2*, more genes involved in the glucosinolate and auxin synthesis pathway were found (Fig. 6.5). These genes, including *CYP83A1*, *CYP79F2*, *CYP79B2*, *CYP79B3*, *CYP83B1*, *UGT74B1*, *ATGSTF11*, *AtSOT17* and *MYB29* are all part of a proposed single enzyme complex (Fig. 6.6) (Grubb and Abel, 2006) synthesizing glucosinolates, auxin and the precursors of phenylpropanoids (lignin) and camalexin. Furthermore, many of the transcript levels of these genes are affected by neutral shade treatment (Fig. 6.6) (Millenaar et al., 2006).

This enzyme complex uses methionine and tryptophan as a substrate. The pathways for the synthesis of these amino acids are also under influence of the genetic background/neutral shade treatment (see Fig. 6.6 for tryptophan). For the synthesis of methionine only two genes had both differently abundant transcripts in the LCN1-10 and an eQTL. A cystathionine beta-lyase (At4g23600) catalyzing the reaction from cystathionine

to L-homocysteine releasing pyruvate and ammonia. One step further downstream a methionine synthase (At3g03780), catalyzing the reaction from L-homocysteine to 2 L-methionine using 5-methyltetrahydropteroyltri-L-glutamate as a co-substrate.

Many genes of the enzymatic route of converting chorismate to tryptophan had also DATs between LCN1-10 and L*er* (Fig. 6.6). Moreover, the genes encoding enzymes for the last two steps (*trp2* and *trp3*) were also affected by light (Fig. 6.6).

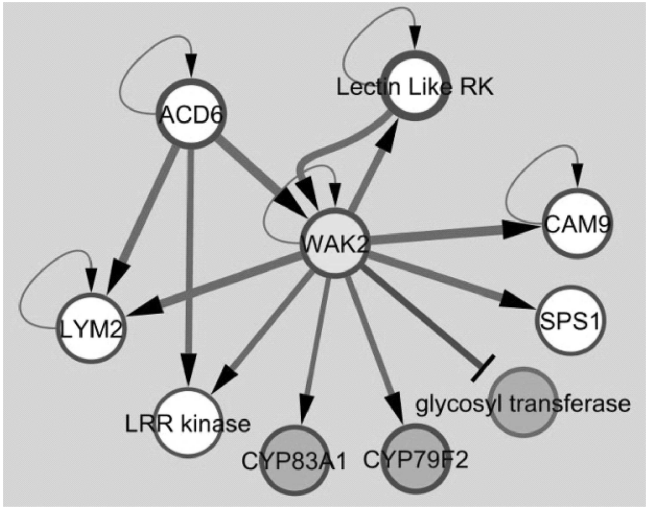


Figure 6.4 Putative (first level) targets and regulators of WAK2 (At1g21270), positive correlation between two genes are indicated by an arrow, negative by a T end. Line width is relative to the absolute correlation. Arrow points towards the target (the regulated) gene. Node border color stands for the relative expression in LCN1-10 compared to L*er* (Dark grey means lower expression in L*er*), thickness is relative to the effect size. Genes in light grey circle are involved in glucosinolate and auxin synthesis. Lectin Like RK (At2g19200); ACD6, ACCELERATED CELL DEATH 6 (At4g14400); WAK2, WALL ASSOCIATED KINASE 2 (At1g21270); CAM9, CALMODULIN 9 (At3g51920); SPS1, SOLANESYL DIPHOSPHATE SYNTHASE 1 (At1g78510); LYM2, LYSM DOMAIN GPI-ANCHORED PROTEIN 2 PRECURSOR (At2g17120); LRR kinase, leucine rich repeat kinase (At1g05700); glycosyl transferase, (At3g23760); CYP83A1, CYTOCHROME P450 MONOOXYGENASE 83A1 (At4g13770); CYP79F2, SUPERSHOOT/BUSHY 1 (At1g16400).

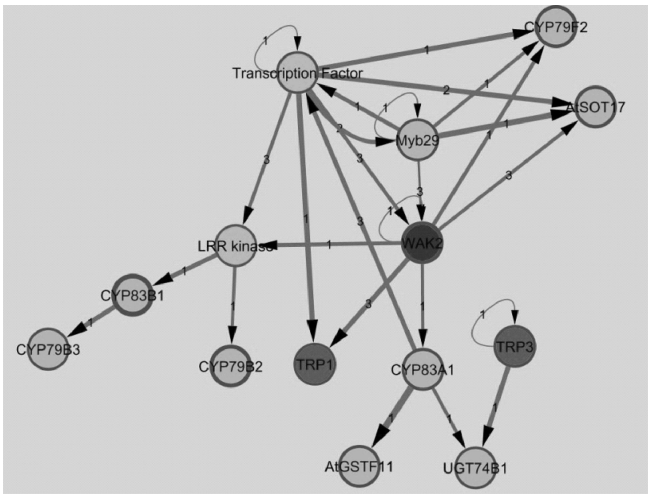


Figure 6.5 Putative regulatory network of glucosinolate/auxin synthesis , initiated by WAK2 (dark grey). Tryptophan synthesis genes in grey. Other putative regulators involved in this network are described as LRR kinase and Transcription Factor.

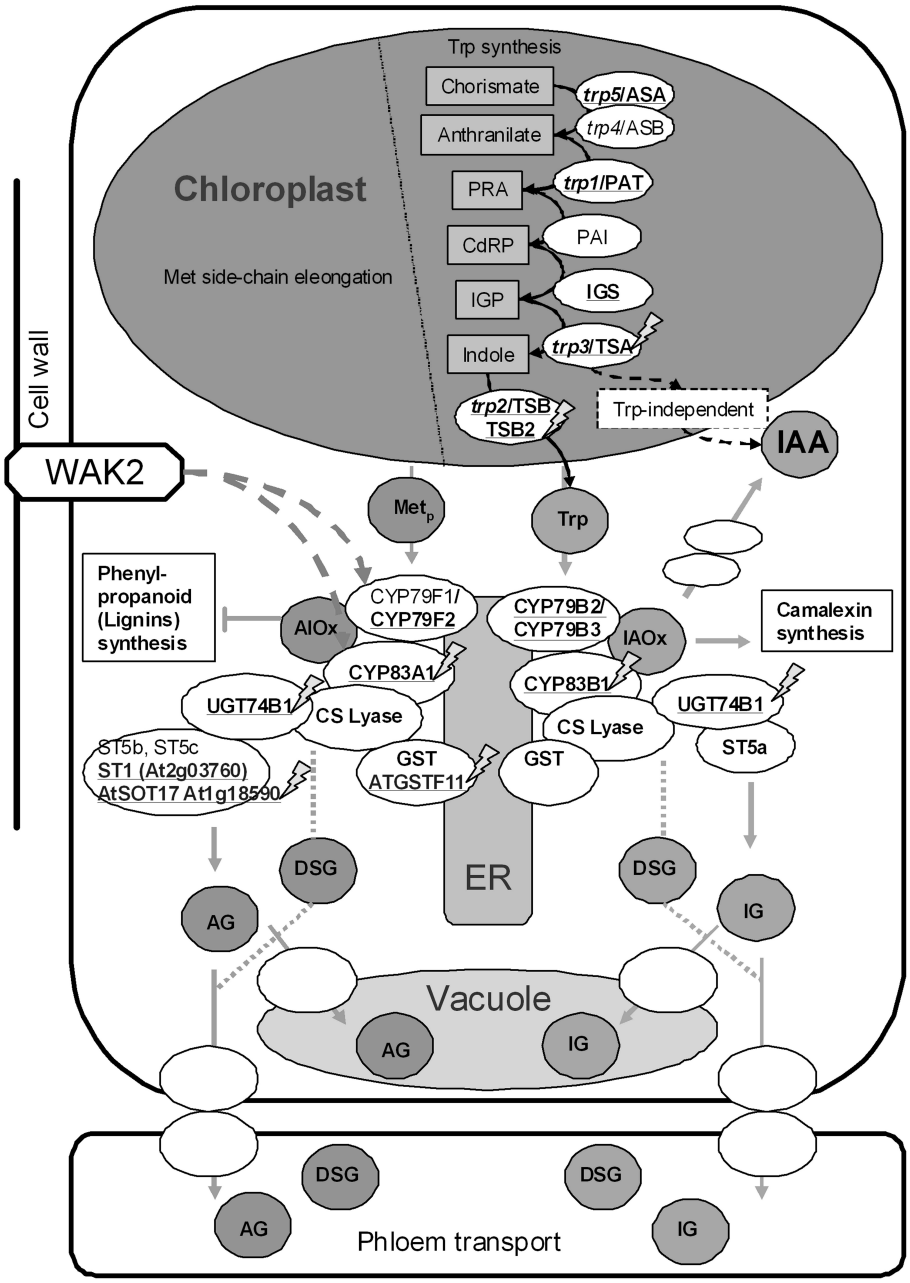


Figure 6.6 Natural variation in transcript levels of genes putatively regulated by WAK2 involved in the proposed subcellular organization of glucosinolate and auxin pathways (modified from Grubb and Abel, 2006) and tryptophan synthetic pathway (modified from Ouyang et al., 2000) preceding the glucosinolate/auxin pathway. Gene symbols underlined in bold have differently abundant transcripts between L*er* and LCN1-10 as well as an

eQTL in the L*er* x Cvi RIL population. Genes affected by neutral shade treatment in the transcriptional profiling experiment by Millenaar et al., 2006 are indicated by the yellow lightning. Both methionine and tryptophan are synthesized in the chloroplast and both function as glucosinolate precursor. Methionine derived metabolites are in blue, tryptophan derived metabolites are in orange.

In the tryptophan synthetic pathway genes are indicated in capitals and their respective mutants are in italics. ASA, anthranilate synthase α subunit; ASB, anthranilate synthase β subunit; CdRP, 1-(O-carboxyphenylamino)-1-deoxyribulose-5-phosphate; IGP, indole-3-glycerol phosphate; IGS, indole-3-glycerol phosphate synthase; PAI, phosphoribosylanthranilate; TSA, tryptophan synthase α subunit; TSB, tryptophan synthase β subunit.

The glucosinolate and auxin pathways also branch to phenylpropanoid/lignin synthesis and camalexin synthesis, indicated in squares. CYP79F1/F2/B2/B3 and CYP83A1/B1 are Cytochrome P450s; C-S Lyase, transaminase *sur1*; UGT71B1, UDP-glycosyl transferase; ST1/5a/5b/5c,At1g18590, Sulfotransferase; GST,ATGSTF11, glutathione-S-transferase; IAOx, indole-3-acetaldoxime; AIOx, aliphatic aldoximes; IAA, indole-3-acetic acid; AG, aliphatic glucosinolate; IG, indolyl glucosinolates; DSG, desulfo-glucosinolate. It is highly unlikely that all these affected genes would have been identified just by chance (example: genes with differently abundant transcripts between L*er* and LCN1-10; glucosinolate/IAA metabolon $p \sim 7.16*10^{-9}$; trp synthesis $p \sim 8.88*10^{-7}$; combined $p \sim 3.09*10^{-10}$).

Discussion

Trait correlation with RIL expression

Transcript variation underlying natural variation in light-influenced traits in Arabidopsis

With the methods employed we could identify many genes to be correlated with one or more traits. These genes are putative candidates for being involved in those traits. Furthermore, the correlated transcripts clustered to the traits mostly according to their phenotypic relatedness. Putative candidates are discussed in supplemental Data 6.1. A good example of identifying functionally different alleles causal for the variation in a trait can be found in the group of genes correlated to compactness (group 4; supplemental Data 6.1). A candidate gene, *PHYTOCHROME B PHYB* (At2g18790), was previously identified to be involved in compactness and was shown to have a functional allelic difference between L*er* and Cvi (Chapter 3, this thesis). Furthermore, a gene downstream of the PHYB initiated signaling cascade was identified: *ARABIDOPSIS PSEUDO RESPONS REGULATOR 9 APRR9* (At2g46790) that is involved in the circadian rhythm and red light response pathway (Ito et al., 2005).

LCN1-10 Phenotype

Expression Phenotype

We found 2146 genes with differentially abundant transcripts (DATs) between LCN1-10 and L*er*. These genes are putatively causal for the observed phenotypic differences. Less transcript level differences were found between another NIL (LCN3-1) compared to L*er* (549 with the least conservative method; Juenger et al., 2006; Keurentjes et al., 2007). This is possibly due to the number of genes on the introgressions of LCN1-10 and LCN3-1, which are ~2000 and 800 respectively. Interestingly, when we compare the number of genes affected, but physically located outside the introgressions the percentage

in both NILs is ~85%, indicating trans regulation is more abundant than cis/local regulation. In chapter 5 we identified 3228 genes with DATs between L*er* and Cvi and a hotspot for transcript regulation (HTR). This is reflected in LCN1-10, which contains roughly 5% of the Cvi genome in a L*er* background but has ~66% of the number of DATs between L*er* and Cvi. Besides this only ~27% of the genes with DATs between L*er* and LCN1-10 had also DATs between L*er* and Cvi (SI Data 6.1). Most DATs between LCN1-10 and L*er* must therefore have been formed from new perturbations of transcription through the Cvi introgression and the rest of the (L*er*) genome. If we combine this with the results of Juenger et al., 2006, this is most probably true for any NIL.

eQTLs on LCN1-10 introgression

An eQTL was found for 74% of the parental DATs of the genes in GG2 (2379 of 3228, chapter 5). In total we found that ~47% of the 2147genes with DATs between LCN1-10 and L*er* had an eQTL (Fig. 6.3) which is more than the genome-wide average of 32% and shows the influence of the natural variation in genetic background on transcript levels.

SNP correlation with Phenotypes

Candidate gene At1g21270 WALL ASSOCIATED KINASE2

WALL ASSOCIATED KINASE2 (*WAK2)* functions in uni-dimensional cell growth (Kohorn et al., 2006). It is expressed in most tissues and GUS-promotor fusions show it to be relatively higher expressed at the junction of organs (Anderson et al., 2001). The extracellular part of the protein suggests that it may be bound to some structural component of the cell wall, probably not pectin. For WAKs in general evidence has been found for their role in a molecular mechanism linking cell-wall sensing (via pectin attachment) to regulation of solute metabolism (Anderson et al., 2001; Wagner and Kohorn, 2001). This in turn is known to be related to turgor maintenance in growing cells. The cytoplasmic kinase domain of WAK2 suggests a signaling function. Some evidence for this was found by expressing a BRI-WAK fusion protein in protoplasts which prevented the protoplasts from shrinking (Kohorn et al., 2006). The BRI-WAK fusion consisted of the WAK kinase domain and the extracellular region of BRI, the BRI region activated the WAK kinase domain. In leaves of *wak-2* mutants a decrease in vacuolar invertases (84 ±9 % of wt) and an increase in cell-wall invertases (123 ±13 % of wt) was found although the p-values of 0.043 and 0.046 were not considered as significant by the authors (Kohorn et al., 2006). Both results obtained from work on protoplasts and mutant leaves hint to putative targets of the WAK2 initiated signaling. They obviously include those components that regulate or influence cellular turgor.

Putative WAK2 initiated transcript network

For several members of the putative transcript network initiated by WAK2 (Fig. 6.4) (part of) the function is known. All but one of the targets of WAK2 is positively correlated with WAK2 and have more abundant transcripts in LCN1-10 compared to L*er*. The only gene negatively correlated with WAK2 and less abundant compared to L*er* is a glycosyl transferase (At3g23760). The other genes were more abundant in LCN1-10 and had diverse functions (listed in supplemental Table 6.1). Two genes, *CYP83A1*

(At4g13770) and *CYP79F2* (At1g16400) are involved in the same pathway(s). The other members of this pathway leading to glucosinolates and auxin synthesis were almost all identified as second level targets (Fig. 6.5).

Natural variation in glucosinolates and auxin synthesis is possibly caused by variation in WAK2 transcript level

A large group of putative first and second level targets of WAK2 are enzymes involved in glucosinolates and auxin synthesis (Fig. 6.5; (Grubb and Abel, 2006). In general, auxin is required for growth and glucosinolates for pathogen defense, their metabolic pathways are largely shared. The transcription of many of these genes is regulated upon shade treatment (Fig. 6.6; Millenaar et al., 2006). All these genes are up regulated in LCN1-10 compared to L*er* and are proposed to function in a metabolon (Fig 6.6; (Grubb and Abel, 2006)). It is highly unlikely that all these affected genes would have been identified just by chance (example: LCN1-10 DATs; glucosinolate/IAA metabolon $p \sim 7.16*10^{-9}$; trp synthesis $p \sim 8.88*10^{-7}$; combined $p \sim 3.09*10^{-10}$). In this metabolon methionine and tryptophan are the basic start components for synthesis of several metabolites. Methionine is mainly converted to aliphatic glucosinolates and phenylpropanoids (a.o. lignins). Tryptophan is converted to indole-3-acetic acid (IAA), camalexins and indolyl glucosinolates.

The flux of metabolites to end-products through this metabolon is regulated by several environmental factors. The synthesis of the three amino acids is limited by nitrogen availability. The branch in the pathway towards either glucosinolates or IAA is regulated by JA, light and sulfur availability (Grubb and Abel, 2006; Hirai et al., 2007). The regulation of one enzyme CYP83B1 acts as a switch between glucosinolate and IAA synthesis (Barlier et al., 2000; Bak et al., 2001; Hoecker et al., 2004; Grubb and Abel, 2006). A more detailed description of the individual genes in this pathway can be found in supplemental Data 6.3.

Based on the regulation of parts of the metabolic routes resulting in glucosinolate and IAA synthesis we were able to construct a putative hierarchy in environmental factors leading to elongation growth when in shade. When nitrogen availability is not limited, biosynthesis of amino acids is ensured. Moreover, when light is not a limiting factor either and thus secures energy production, much of the IAA precursor is gated towards glucosinolate synthesis (Bak et al., 2001; Naur et al., 2003; Hoecker et al., 2004) ensuring a proper defense. This is enforced by jasmonic acid since more glucosinolates are synthesized when under pathogen attack. When plants are in a shaded environment, however, more of the IAA precursor is used for IAA production instead of glucosinolate production, since a lack of energy is less favorable than a preventive defense system, so growth mediated by auxin is needed to escape shade. However, when under shaded conditions pathogen attack occurs jasmonic acid ensures that the glucosinolate synthesis is favored over IAA (Grubb and Abel, 2006; Dombrecht et al., 2007; Hirai et al., 2007; Kusnierczyk et al., 2007).

The fact that we found natural variation in transcript abundance for the genes encoding this metabolon and also for the tryptophan synthesis route is strong indirect evidence that LCN1-10 has altered IAA, glucosinolates and amino acid levels. Further subsequent phenotypes like altered growth, leaf angles and/or pathogen defense are probably also seriously affected by this.

Materials and Methods

Plant growth and treatment

To identify variation in transcript abundance between L*er* and LCN1-10 they were grown all at the same time. Seeds were imbibed on wet filter paper for 3 d after which the germinated seedlings were transferred to soil. These plants were grown under short days (9 h light, 15 h dark), a light intensity of 200 µmol m^{-2} s^{-1} and watered 3 times per week. Further details are described in (Millenaar et al., 2005). All plants were treated with 3 h of neutral shade (ca 10% of standard growth conditions) 21 d after transfer to soil (24 d after germination). The three most responsive leaves (petiole and lamina) per plant were harvested and pooled per genotype for RNA isolation and transcription profiling.

The accessions for local association mapping were grown together in a separate batch under similar growing conditions. Per accession 10-20 individual plants were available for measurements.

Measurement of phenotypes

Leaf angle traits of L*er* and LCN1-10 were previously measured (chapter 2). The rosette describing traits, rosette area, fermax and rosette compactness were measured as described in chapter 3. The accessions for local association mapping were measured 25 d after transfer to soil. The leaf angle traits are described in chapter 2 and the rosette describing traits in chapter 3.

RNA Sample preparation

We prepared three biological replicates of both L*er* and LCN1-10 RNA. All procedures were described in de Jong et al., 2006. In short, RNA was extracted and purified using the RNeasy kit (Qiagen, Valencia, U.S.A). Amplification and labeling was performed with the MessageAmp aRNA kit (Ambion, Austin, U.S.A). Amplified RNA was used to generate labeled cDNA with incorporation of 5-(3-aminoallyl)-dUTP and labeled with either Cy3 or Cy5 mono-reactive dye (Amersham, Piscataway, U.S.A.). All cDNA products were purified using the RNeasy kit (Qiagen, Valencia, U.S.A).

Micro array transcript profiling

The CATMA (Allemeersch et al., 2005) Arabidopsis DNA microarrays were produced from a set of (150-450)-mers representing 20833 unique genes (http://www.catma.org). Hybridizations performed on CATMA arrays and arrays were scanned using ScanArray Express HT (PerkinElmer, Wellesley, U.S.A.) and quantified with Imagene 6.0 (BioDiscovery, El Segundo, U.S.A.).
The Limma package for the statistical work environment R (http://www.R-project.org) was used for normalization (Smyth and Speed, 2003) and for determining differently abundant transcripts (Smyth, 2004). All array data is submitted to ArrayExpress.

Trait transcript correlation

To determine to which trait the transcripts were correlated we calculated the Pearson correlation between the phenotypic values per trait per RIL and the normalized single channel micro-array data per transcript.

Local association mapping

We determined local association per SNP with a trait by a t-test of two groups of phenotypes, a reference SNP group and a non-reference SNP group. The two groups contained single plant values for the trait of interest. Since we specifically were searching for a difference between L*er* and Cvi we limited the tests to those where L*er* and Cvi were in the opposite groups. We further selected the test for group-size to be larger than one. We included 19 of the 20 accessions profiled with the "Perlegen" SNP-arrays available at TAIR ftp://ftp.arabidopsis.org/home/tair/Sequences/Perlegen_Array_Resequencing_Data _2007/). Data from the accessions used are Bay-0, Bor-4, Br-0, Bur-0, C24, Col-0, Cvi-0, Est-1, Fei-0, Got-7, Ler-1, NFA-8, RRS-10, RRS-7, Sha, Tamm-2, Ts-1, Tsu-1 and Van-0. We considered the top 5% associated SNP as useful for candidate gene selection.

Obtaining the most likely candidates from the four criteria

Each of the four criteria: trait/transcript correlation, L*er*/LCN1-10 transcript difference, eQTL mapping to the introgression of LCN1-10 and associated SNPs were rank ordered. All the lists contained 748 members (genes). For the eQTL and SNP lists we determined the score of genes with no eQTL or associated SNP to be 748. The top 15 genes (2%), those with the lowest combined scores by ranked product (Breitling et al., 2004) were selected as the best candidates. In the combined trait criteria we only used the eQTL and L*er* and LCN1-10 transcript difference score once because these are not trait specific. The SNP and trait/transcript correlation were used per trait since they give a trait specific score. Again, the top 15 genes (2%), those with the lowest combined scores were selected as the best candidates.

Transcript regulation network

To construct a putative network for transcript regulation we selected the gene with DATs between L*er* and LCN1-10 as a start group. The transcript correlations between all members were determined by Pearson correlation. Possible interactions were filtered for those that have both members with an eQTL. Furthermore, the peak of the eQTL of the possible target should map to a region within 2 Mbp to either side. After that we used the direction of the correlation and the difference in transcript levels to further decrease the possible interactions. Interaction pairs with a positive correlation should have transcript level differences with similar signs, whereas those with a negative correlation should have transcript level differences with opposite signs. The list of interactions now contains interaction pairs of which the target has a eQTL mapping to the putative regulator and which correlation corresponds to the transcript level difference measured between L*er* and LCN1-10. This list was ordered by absolute correlation of the pairs. Subsequently, for each target the rank of the possible regulator was determined. For visualization only the best correlated regulators were used.

Regulation of the IAA/glucosinolate metabolon

We calculated the significance of the genes belonging to specific parts of the IAA/glucosinolate metabolic pathway by a hyper geometric test. The question asked was "what is the probability of finding x regulated genes when the pathway under study consists of y genes, when 2146 genes are regulated out of 20833 genes in total".

Chapter 7

General discussion

L. Basten Snoek & Anton J. M. Peeters

Mankind has been using the variation in flora and fauna found in nature for ages. Both within and among species variation in traits have determined a particular use, for example as food source, clothing or building material. Traditionally natural variation within a plant species has been used by selecting those individuals with favorable traits and breed accessions for a particular use. Nowadays, with the development of molecular biology and computers this natural variation can be exploited in another way. Not only has the introduction of molecular markers and quantitative trait loci (QTL) mapping aided in selecting those individuals with a favorable trait, it also helped identifying functional different alleles of genes for natural variable traits. Especially in the model plant *Arabidopsis thaliana*, genes have been identified of which the allelic variants lead to variation in phenotype(s) though QTL mapping. These genes, alleles and phenotypes are the products of natural selection and show a "snap shot" of evolution/selection of (part) of the two parental genotypes. In chapter two and three natural variation for traits involving the leaves was found. This natural variation in initial position of the leaves, hyponastic growth and rosette compactness show that these traits are under selection pressure. Moreover, which of the phenotypes (and genotypes) that make up these traits are advantageous depends on the environment they are found in.

Recently, with the establishment of genetical genomics this natural variation can be used in a third way, the uncovering of genetical regulation networks. By combining micro-arrays with a recombinant inbred line (RIL) population, regulatory networks at the transcript level can be constructed in *Arabidopsis thaliana* (Keurentjes et al., 2007; West et al., 2007). This revealed that variation in transcript levels is at least as numerous as the variation in traditional traits identified up to now. This variation in gene expression can be used to map expression QTLs (eQTLs). The eQTLs of a gene identify the loci explaining the variation in transcript abundance for that gene. These loci contain the putative regulators of this gene, either cis or trans, and can be used to construct the network. Not only can the transcriptome be used as such. Recent developments in metabolomics, proteomics and enzyme assays and phenomics show promising results at these levels of regulation too. Furthermore, chapter five shows that a major part of the variation in transcription is the result of an interaction between the genotype and environment. The results of that experiment even indicate that much of the variation is inducible and might only exist temporally.

Important adaptive traits for plants are those that enable them to deal with the variation in growth limiting resources, like light and water. Some plant species have developed mechanisms to alter growth to deal with such variations. A particularly good example is *Rumex palustris* that, when submerged shows a remarkable hyponastic growth

and petiole elongation to escape the suffocating conditions. This reaction is mainly triggered by an increase in ethylene concentration due to physical entrapment. A phenotypic similar response is used by Arabidopsis to escape from plant-induced shade but can also be triggered by ethylene treatment. Most Arabidopsis accessions show hyponastic growth only a few hours after being shaded or treated with ethylene. In chapter 2, natural variation in the hyponastic response to both shade and ethylene was found in two Arabidopsis RIL populations (Col x L*er* and L*er* x Cvi). Moreover, a difference in response of a particular part of the petiole could be observed between the two treatments. This helped to identify two regions of differential growth during hyponastic growth. The basal part of the petiole responds similar to both shading and ethylene treatment, whereas the part where the petiole changes into the lamina responds most to shading. These rapid growth/escape responses are followed by long-term adjustments when the limiting resource persists. Because of the natural variation present in a RIL set, some individual RILs will already be limited by a resource while others do not seem to be bothered.

To study how light sensitivity affects the development and growth of *Arabidopsis thaliana*, we studied the light-affected traits that make up the rosette architecture under normal growth conditions. Natural variation in light sensitivity of hypocotyls was identified in the RIL population L*er* x Cvi (Borevitz et al., 2002). The natural variation in light sensitivity between the RILs of these populations, ensured that both normal and shade avoiding rosette phenotypes were present for our experiment. In chapter 3 the identification of several QTLs is described for the traits rosette area, fermax (a trait most descriptive for leaf length) and rosette compactness, which describes how much of the area of a perfect circle is covered by the leaves. Two genes were found to have functionally different alleles, *PHYTOCHROME B* and *ERECTA*. Furthermore, several other candidates were discovered. We found some indications that a feedback mechanism between the genotype and the phenotype may exist for rosette growth and morphology. The genotype determines both pure genetic-driven growth as well as part of the genotype*environment interaction driven growth. Since the morphology of the rosette creates a micro-environment for the individual leaves and leaves start to shade themselves a possible feedback mechanism is created. We found indications for this by an increase in broad sense heritability of the rosette compactness during time/development. It will be challenging to uncover the mechanisms and components involved in this multi level interaction in which the role of the genotype seems to change over time.

With the establishment of genetical genomics in Arabidopsis, new ways of constructing putative regulation networks are needed. Several statistical methods have been successful by using transcriptome information of the RIL population alone. Usually these methods identify putative regulators which transcripts are also variable within the RIL population. Many regulators, however, will not show a difference on transcript level, but will on another level of regulation. In Chapter 4 we used a combination of approaches to prove the major hotspot for transcript regulation (HTR) in genetical genomics 1 (Keurentjes et al., 2007) is caused by the *erecta* mutation in L*er*. After we showed that this HTR was indeed caused by *erecta* we could use ERECTA as a start to construct a putative regulatory transcriptional network. By combining the differences in the transcriptome of several mutants, co-expression and transcription factor binding site overrepresentation with eQTL position, a set of WRKY transcription factors and genes both downstream of ERECTA

were identified. The additional use of transcript profiles of mutants, even those already publicly available, in combination with genetical genomics shows promising and many possibilities to identify and construct putative transcriptional pathways.

In Chapter 5, a genetical genomics experiment in which gene expression was induced by 3 hours of shading was compared to a previous experiment (Keurentjes et al., 2007) using the same population. Here we uncovered the existence of constitutive and plastic eQTLs. Furthermore, it was shown that the natural variation in the transcriptome can be used to identify processes important at the time the plants were grown/treated/harvested. For the response to shade, we discovered that the whole cellular machinery making the response to change possible is (at least partially) regulated at the transcript level. Annotation groups like “transcription”, “translation”, “protein folding”, even the “organization and biogenesis of ribosomes” with more genes with eQTL than can be expected at random show that all levels of response/regulation are influenced by shade and that natural variation can be observed. It is intriguing to think about how natural variation on different factors (genes) can lead to a similar phenotype.

In chapter 6, the identification of candidate genes and construction of regulatory networks was further extended by a combination of methods. The transcriptome of a near isogenic line (NIL) with a contrasting phenotype was compared to L*er*. This was combined with transcript/phenotype correlation in the RILs, local association mapping and eQTL position to find the genes causal for the difference in phenotype between L*er* and the NIL. We extended this by constructing a putative regulatory network starting with the best candidate gene. By constructing a transcriptional network around this gene natural variation in a possible downstream molecular switch between growth and defense was uncovered. The genes involved in this branching point between a.o. glucosinolates and auxin metabolism all show natural variation in gene expression (in both RILs and NIL) as well as variation in expression through light intensity. These results show how useful genetical genomics can be in combination with a more dedicated experiment in identifying the underling genetic variation that leads to different phenotypes. Moreover, it shows which processes are important.

The five experimental chapters of this thesis show promising results for future unraveling of the interactions within the genotype and between genotype and environment. Especially, the (future) increase in data generated for one RIL population will help to clarify the temporal and developmental influence of genetic variation on phenotypes and adaptation to the environment. More specifically when the transcriptome is investigated along the variation in these three main factors, time/development, environment and genotype the patterns of interacting transcripts will increase the quality of the generated transcriptional networks.

Literature

Ahmad M, Jarillo JA, Smirnova O, Cashmore AR (1998) Cryptochrome blue-light photoreceptors of Arabidopsis implicated in phototropism. Nature **392:** 720-723

Allemeersch J, Durinck S, Vanderhaeghen R, Alard P, Maes R, Seeuws K, Bogaert T, Coddens K, Deschouwer K, Van Hummelen P, Vuylsteke M, Moreau Y, Kwekkeboom J, Wijfjes AH, May S, Beynon J, Hilson P, Kuiper MT (2005) Benchmarking the CATMA microarray. A novel tool for Arabidopsis transcriptome analysis. Plant Physiol **137:** 588-601

Alonso-Blanco C, Bentsink L, Hanhart CJ, Blankestijn-de Vries H, Koornneef M (2003) Analysis of natural allelic variation at seed dormancy loci of *Arabidopsis thaliana*. Genetics **164:** 711-729

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

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

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

Anderson CM, Wagner TA, Perret M, He ZH, He D, Kohorn BD (2001) WAKs: cell wall-associated kinases linking the cytoplasm to the extracellular matrix. Plant Mol Biol **47:** 197-206

Andreasson E, Jenkins T, Brodersen P, Thorgrimsen S, Petersen NH, Zhu S, Qiu JL, Micheelsen P, Rocher A, Petersen M, Newman MA, Bjorn Nielsen H, Hirt H, Somssich I, Mattsson O, Mundy J (2005) The MAP kinase substrate MKS1 is a regulator of plant defense responses. EMBO J **24:** 2579-2589

Armstrong W (1979) Aeration in higher plants. Advances in Botanical Research **7:** 225-332

Armstrong W, Jackson M, Brändle R (1994) Mechanisms of flood tolerance in plants. Acta Botanica Neerlandica **43:** 307-358

Asai T, Tena G, Plotnikova J, Willmann MR, Chiu WL, Gomez-Gomez L, Boller T, Ausubel FM, Sheen J (2002) MAP kinase signalling cascade in Arabidopsis innate immunity. Nature **415:** 977-983

Bailey-Serres J, Voesenek LACJ (2008) Flooding stress: acclimations and genetic diversity. Annu Rev Plant Biol **59:** 313-339

Bak S, Tax FE, Feldmann KA, Galbraith DW, Feyereisen R (2001) CYP83B1, a cytochrome P450 at the metabolic branch point in auxin and indole glucosinolate biosynthesis in Arabidopsis. Plant Cell **13:** 101-111

Ballare CL (1999) Keeping up with the neighbours: phytochrome sensing and other signalling mechanisms. Trends Plant Sci **4:** 97-102

Banga M, Slaa EJ, Blom C, Voesenek LACJ (1996) Ethylene Biosynthesis and Accumulation under Drained and Submerged Conditions (A Comparative Study of Two *Rumex* Species). Plant Physiol **112:** 229-237

Barlier I, Kowalczyk M, Marchant A, Ljung K, Bhalerao R, Bennett M, Sandberg G, Bellini C (2000) The SUR2 gene of *Arabidopsis thaliana* encodes the cytochrome P450 CYP83B1, a modulator of auxin homeostasis. Proc Natl Acad Sci U S A **97:** 14819-14824

Barth C, Conklin PL (2003) The lower cell density of leaf parenchyma in the *Arabidopsis thaliana* mutant lcd1-1 is associated with increased sensitivity to ozone and virulent Pseudomonas syringae. Plant J **35:** 206-218

Benjamini Y, Yekutieli D (2005) Quantitative trait Loci analysis using the false discovery rate. Genetics **171:** 783-790

Benschop JJ, Bou J, Peeters AJ, Wagemaker N, Guhl K, Ward D, Hedden P, Moritz T, Voesenek LA (2006) Long-term submergence-induced elongation in Rumex palustris requires abscisic acid-dependent biosynthesis of gibberellin1. Plant Physiol **141:** 1644-1652

Benschop JJ, Jackson MB, Guhl K, Vreeburg RA, Croker SJ, Peeters AJ, Voesenek LA (2005) Contrasting interactions between ethylene and abscisic acid in Rumex species differing in submergence tolerance. Plant J **44:** 756-768

Benschop JJ, Millenaar FF, Smeets ME, van Zanten M, Voesenek LA, Peeters AJ (2007) Abscisic acid antagonizes ethylene-induced hyponastic growth in Arabidopsis. Plant Physiol **143:** 1013-1023

Bentsink L, Yuan K, Koornneef M, Vreugdenhil D (2003) The genetics of phytate and phosphate accumulation in seeds and leaves of *Arabidopsis thaliana*, using natural variation. Theor Appl Genet **106:** 1234-1243

Bing N, Hoeschele I (2005) Genetical genomics analysis of a yeast segregant population for transcription network inference. Genetics **170:** 533-542

Borevitz JO, Maloof JN, Lutes J, Dabi T, Redfern JL, Trainer GT, Werner JD, Asami T, Berry CC, Weigel D, Chory J (2002) Quantitative trait loci controlling light and hormone response in two accessions of *Arabidopsis thaliana*. Genetics **160:** 683-696

Borevitz JO, Nordborg M (2003) The impact of genomics on the study of natural variation in Arabidopsis. Plant Physiol **132:** 718-725

Botto JF, Coluccio MP (2007) Seasonal and plant-density dependency for quantitative trait loci affecting flowering time in multiple populations of *Arabidopsis thaliana*. Plant Cell Environ **30:** 1465-1479

Breitling R, Amtmann A, Herzyk P (2004) Iterative Group Analysis (iGA): a simple tool to enhance sensitivity and facilitate interpretation of microarray experiments. BMC Bioinformatics **5:** 34

Breitling R, Armengaud P, Amtmann A, Herzyk P (2004) Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. FEBS Lett **573:** 83-92

Brem RB, Kruglyak L (2005) The landscape of genetic complexity across 5,700 gene expression traits in yeast. Proc Natl Acad Sci U S A **102:** 1572-1577

Brem RB, Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. Science **296:** 752-755

Byrne ME, Barley R, Curtis M, Arroyo JM, Dunham M, Hudson A, Martienssen RA (2000) Asymmetric leaves1 mediates leaf patterning and stem cell function in Arabidopsis. Nature **408:** 967-971

Bystrykh L, Weersing E, Dontje B, Sutton S, Pletcher MT, Wiltshire T, Su AI, Vellenga E, Wang J, Manly KF, Lu L, Chesler EJ, Alberts R, Jansen RC, Williams RW, Cooke MP, de Haan G (2005) Uncovering regulatory pathways that affect hematopoietic stem cell function using 'genetical genomics'. Nat Genet **37:** 225-232

Cao S, Ye M, Jiang S (2005) Involvement of GIGANTEA gene in the regulation of the cold stress response in Arabidopsis. Plant Cell Rep **24:** 683-690

Carland FM, Berg BL, FitzGerald JN, Jinamornphongs S, Nelson T, Keith B (1999) Genetic regulation of vascular tissue patterning in Arabidopsis. Plant Cell **11:** 2123-2137

Carland FM, Fujioka S, Takatsuto S, Yoshida S, Nelson T (2002) The identification of CVP1 reveals a role for sterols in vascular patterning. Plant Cell **14:** 2045-2058

Cheng Y, Dai X, Zhao Y (2007) Auxin synthesized by the YUCCA flavin monooxygenases is essential for embryogenesis and leaf formation in Arabidopsis. Plant Cell **19:** 2430-2439

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics **138:** 963-971

Clarke JH, Mithen R, Brown JK, Dean C (1995) QTL analysis of flowering time in *Arabidopsis thaliana*. Mol Gen Genet **248:** 278-286

Clarke JH, Tack D, Findlay K, Van Montagu M, Van Lijsebettens M (1999) The *SERRATE* locus controls the formation of the early juvenile leaves and phase length in Arabidopsis. Plant J **20:** 493-501

Clua A, Bottini R, Brocchi GN, Bogino J, Luna V, Montaldi ER (1996) Growth habit of Lotus tenuis shoots and the influence of photosynthetic photon flux density, sucrose and endogenous levels of gibberellins A(1) and A(3). Physiol Plant **98:** 381-388

Colmer TD (2003) Aerenchyma and an inducible barrier to radial oxygen loss facilitate root aeration in upland, paddy and deep-water rice (*Oryza sativa* L.). Ann Bot (Lond) **91 Spec No:** 301-309

Cowling RJ, Kamiya Y, Seto H, Harberd NP (1998) Gibberellin dose-response regulation of *GA4* gene transcript levels in Arabidopsis. Plant Physiol **117:** 1195-1203

Cox MC, Benschop JJ, Vreeburg RA, Wagemaker CA, Moritz T, Peeters AJ, Voesenek LA (2004) The roles of ethylene, auxin, abscisic acid, and gibberellin in the hyponastic growth of submerged Rumex palustris petioles. Plant Physiol **136:** 2948-2960

Cox MC, Millenaar FF, Van Berkel YE, Peeters AJ, Voesenek LA (2003) Plant movement. Submergence-induced petiole elongation in *Rumex palustris* depends on hyponastic growth. Plant Physiol **132:** 282-291

Cox MC, Peeters AJ, Voesenek LA (2006) The stimulating effects of ethylene and auxin on petiole elongation and on hyponastic curvature are independent processes in submerged Rumex palustris. Plant Cell Environ **29:** 282-290

Crawford R (1992) Oxygen availability as an ecological limit to plant distribution. Advances in Ecological Research **23:** 93-185

Darrah C, Taylor BL, Edwards KD, Brown PE, Hall A, McWatters HG (2006) Analysis of phase of LUCIFERASE expression reveals novel circadian quantitative trait loci in Arabidopsis. Plant Physiol **140:** 1464-1474

David KM, Armbruster U, Tama N, Putterill J (2006) Arabidopsis GIGANTEA protein is post-transcriptionally regulated by light and dark. FEBS Lett **580:** 1193-1197

de Jong M, van Breukelen B, Wittink FR, Menke FL, Weisbeek PJ, Van den Ackerveken G (2006) Membrane-associated transcripts in Arabidopsis; their isolation and characterization by DNA microarray analysis and bioinformatics. Plant J **46:** 708-721

DeCook R, Lall S, Nettleton D, Howell SH (2006) Genetic regulation of gene expression during shoot development in Arabidopsis. Genetics **172:** 1155-1164

Dohmann EM, Kuhnle C, Schwechheimer C (2005) Loss of the CONSTITUTIVE PHOTOMORPHOGENIC9 signalosome subunit 5 is sufficient to cause the cop/det/fus mutant phenotype in Arabidopsis. Plant Cell **17:** 1967-1978

Dombrecht B, Xue GP, Sprague SJ, Kirkegaard JA, Ross JJ, Reid JB, Fitt GP, Sewelam N, Schenk PM, Manners JM, Kazan K (2007) MYC2 differentially modulates diverse jasmonate-dependent functions in Arabidopsis. Plant Cell **19:** 2225-2245

Drew MC, He CJ, Morgan PW (2000) Programmed cell death and aerenchyma formation in roots. Trends Plant Sci **5:** 123-127

Edwards KD, Lynn JR, Gyula P, Nagy F, Millar AJ (2005) Natural allelic variation in the temperature-compensation mechanisms of the *Arabidopsis thaliana* circadian clock. Genetics **170:** 387-400

Ehrenreich IM, Stafford PA, Purugganan MD (2007) The genetic architecture of shoot branching in *Arabidopsis thaliana*: a comparative assessment of candidate gene associations vs. quantitative trait locus mapping. Genetics **176:** 1223-1236

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

El-Lithy ME, Clerkx EJ, Ruys GJ, 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

Eulgem T, Rushton PJ, Robatzek S, Somssich IE (2000) The WRKY superfamily of plant transcription factors. Trends Plant Sci **5:** 199-206

Firn RD, Digby J (1980) The establishment of tropic curvatures in plants. Annu Rev Plant Physiol **31:** 131-148.

Fowler S, Lee K, Onouchi H, Samach A, Richardson K, Morris B, Coupland G, Putterill J (1999) GIGANTEA: a circadian clock-controlled gene that regulates photoperiodic flowering in Arabidopsis and encodes a protein with several possible membrane-spanning domains. EMBO J **18:** 4679-4688

Franke L, van Bakel H, Fokkens L, de Jong ED, Egmont-Petersen M, Wijmenga C (2006) Reconstruction of a functional human gene network, with an application for prioritizing positional candidate genes. Am J Hum Genet **78:** 1011-1025

Franklin KA, Whitelam GC (2005) Phytochromes and shade-avoidance responses in plants. Ann Bot (Lond) **96:** 169-175

Friml J, Wisniewska J, Benkova E, Mendgen K, Palme K (2002) Lateral relocation of auxin efflux regulator PIN3 mediates tropism in Arabidopsis. Nature **415:** 806-809

Fu J, Jansen RC (2006) Optimal design and analysis of genetic studies on gene expression. Genetics **172:** 1993-1999

Gailing O, Macnair MR, Bachmann K (2004) QTL mapping for a trade-off between leaf and bud production in a recombinant inbred population of Microseris douglasii and M. bigelovii (Asteraceae, Lactuceae): a potential preadaptation for the colonization of serpentine soils. Plant Biol **6:** 440-446

Gautier H, VarletGrancher C, Baudry N (1997) Effects of blue light on the vertical colonization of space by white clover and their consequences for dry matter distribution. Annals of Botany **80:** 665-671

Gibbs J, Greenway H (2003) Mechanisms of anoxia tolerance in plants. I. Growth, survival and anaerobic catabolism. Functional Plant Biology **30:** 1-47

Godiard L, Sauviac L, Torii KU, Grenon O, Mangin B, Grimsley NH, Marco Y (2003) ERECTA, an LRR receptor-like kinase protein controlling development pleiotropically affects resistance to bacterial wilt. Plant J **36:** 353-365

Grigg SP, Canales C, Hay A, Tsiantis M (2005) SERRATE coordinates shoot meristem function and leaf axial patterning in Arabidopsis. Nature **437:** 1022-1026

Grubb CD, Abel S (2006) Glucosinolate metabolism and its control. Trends Plant Sci **11:** 89-100

Gusmaroli G, Figueroa P, Serino G, Deng XW (2007) Role of the MPN subunits in COP9 signalosome assembly and activity, and their regulatory interaction with Arabidopsis Cullin3-based E3 ligases. Plant Cell **19:** 564-581

Hangarter RP (1997) Gravity, light and plant form. Plant Cell Environ **20:** 796-800

Harper RM, Stowe-Evans EL, Luesse DR, Muto H, Tatematsu K, Watahiki MK, Yamamoto K, Liscum E (2000) The NPH4 locus encodes the auxin response factor ARF7, a conditional regulator of differential growth in aerial Arabidopsis tissue. Plant Cell **12:** 757-770

Hase Y, Fujioka S, Yoshida S, Sun G, Umeda M, Tanaka A (2005) Ectopic endoreduplication caused by sterol alteration results in serrated petals in Arabidopsis. J Exp Bot **56:** 1263-1268

Hase Y, Tanaka A, Baba T, Watanabe H (2000) FRL1 is required for petal and sepal development in Arabidopsis. Plant J **24:** 21-32

He ZH, Cheeseman I, He D, Kohorn BD (1999) A cluster of five cell wall-associated receptor kinase genes, *Wak1-5*, are expressed in specific organs of Arabidopsis. Plant Mol Biol **39:** 1189-1196

Hegmann JP, Possidente B (1981) Estimating genetic correlations from inbred strains. Behav Genet **11:** 103-114

Hirai MY, Sugiyama K, Sawada Y, Tohge T, Obayashi T, Suzuki A, Araki R, Sakurai N, Suzuki H, Aoki K, Goda H, Nishizawa OI, Shibata D, Saito K (2007) Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis. Proc Natl Acad Sci U S A **104:** 6478-6483

Hisamatsu T, King RW, Helliwell CA, Koshioka M (2005) The involvement of gibberellin 20-oxidase genes in phytochrome-regulated petiole elongation of Arabidopsis. Plant Physiol **138:** 1106-1116

Hoecker U, Toledo-Ortiz G, Bender J, Quail PH (2004) The photomorphogenesis-related mutant red1 is defective in CYP83B1, a red light-induced gene encoding a cytochrome P450 required for normal auxin homeostasis. Planta **219:** 195-200

Hoffmann-Benning S, Kende H (1992) On the Role of Abscisic Acid and Gibberellin in the Regulation of Growth in Rice. Plant Physiol **99:** 1156-1161

Ichimura K, Casais C, Peck SC, Shinozaki K, Shirasu K (2006) MEKK1 is required for MPK4 activation and regulates tissue-specific and temperature-dependent cell death in Arabidopsis. J Biol Chem **281:** 36969-36976

Ingram GC (2005) Plant development: spacing out stomatal pores. Curr Biol **15:** R663-665

Ito S, Nakamichi N, Matsushika A, Fujimori T, Yamashino T, Mizuno T (2005) Molecular dissection of the promoter of the light-induced and circadian-controlled APRR9 gene encoding a clock-associated component of *Arabidopsis thaliana*. Biosci Biotechnol Biochem **69:** 382-390

Jackson MB (1985) Ethylene and the responses of plants to soil waterlogging and submergence. Annual Review of Plant Physiology **36:** 145-147

Jansen RC (1993) Interval mapping of multiple quantitative trait loci. Genetics **135:** 205-211

Jansen RC (2003) Studying complex biological systems using multifactorial perturbation. Nat Rev Genet **4:** 145-151

Jansen RC, Nap JP (2001) Genetical genomics: the added value from segregation. Trends Genet **17:** 388-391

Journot-Catalino N, Somssich IE, Roby D, Kroj T (2006) The transcription factors WRKY11 and WRKY17 act as negative regulators of basal resistance in *Arabidopsis thaliana*. Plant Cell **18:** 3289-3302

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

Juenger TE, Wayne T, Boles S, Symonds VV, McKay J, Coughlan SJ (2006) Natural genetic variation in whole-genome expression in *Arabidopsis thaliana*: the impact of physiological QTL introgression. Mol Ecol **15:** 1351-1365

Kao CH, Zeng ZB, Teasdale RD (1999) Multiple interval mapping for quantitative trait loci. Genetics **152:** 1203-1216

Karniol B, Chamovitz DA (2000) The COP9 signalosome: from light signaling to general developmental regulation and back. Curr Opin Plant Biol **3:** 387-393

Keurentjes JJ, Bentsink L, Alonso-Blanco C, Hanhart CJ, 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 JJ, Fu J, de Vos CH, Lommen A, Hall RD, Bino RJ, van der Plas LH, Jansen RC, Vreugdenhil D, Koornneef M (2006) The genetics of plant metabolism. Nat Genet **38:** 842-849

Keurentjes JJ, Fu J, Terpstra IR, Garcia JM, van den Ackerveken G, Snoek LB, Peeters AJM, Vreugdenhil D, Koornneef M, Jansen RC (2007) Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. Proc Natl Acad Sci U S A **104:** 1708-1713

Keurentjes JJ, Koornneef M, Vreugdenhil D (2008) Quantitative genetics in the age of omics. Curr Opin Plant Biol

Kim GT, Yano S, Kozuka T, Tsukaya H (2005) Photomorphogenesis of leaves: shade-avoidance and differentiation of sun and shade leaves. Photochem Photobiol Sci **4:** 770-774

Kim S, Zhao K, Jiang R, Molitor J, Borevitz JO, Nordborg M, Marjoram P (2006) Association mapping with single-feature polymorphisms. Genetics **173:** 1125-1133

Kliebenstein DJ, Kroymann J, Brown P, Figuth A, Pedersen D, Gershenzon J, Mitchell-Olds T (2001) Genetic control of natural variation in Arabidopsis glucosinolate accumulation. Plant Physiol **126:** 811-825

Kohorn BD, Kobayashi M, Johansen S, Riese J, Huang LF, Koch K, Fu S, Dotson A, Byers N (2006) An Arabidopsis cell wall-associated kinase required for invertase activity and cell growth. Plant J **46:** 307-316

Koornneef M, Alonso-Blanco C, Stam P (2006) Genetic analysis. Methods Mol Biol **323:** 65-77

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

Kozuka T, Horiguchi G, Kim GT, Ohgishi M, Sakai T, Tsukaya H (2005) The different growth responses of the *Arabidopsis thaliana* leaf blade and the petiole during shade avoidance are regulated by photoreceptors and sugar. Plant Cell Physiol **46:** 213-223

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

Kusnierczyk A, Winge P, Midelfart H, Armbruster WS, Rossiter JT, Bones AM (2007) Transcriptional responses of *Arabidopsis thaliana* ecotypes with different glucosinolate profiles after attack by polyphagous Myzus persicae and oligophagous Brevicoryne brassicae. J Exp Bot **58:** 2537-2552

Li H, Lu L, Manly KF, Chesler EJ, Bao L, Wang J, Zhou M, Williams RW, Cui Y (2005) Inferring gene transcriptional modulatory relations: a genetical genomics approach. Hum Mol Genet **14:** 1119-1125

Li Y, Alvarez OA, Gutteling EW, Tijsterman M, Fu J, Riksen JA, Hazendonk E, Prins P, Plasterk RH, Jansen RC, Breitling R, Kammenga JE (2006) Mapping determinants of gene expression plasticity by genetical genomics in C. elegans. PLoS Genet **2:** e222

Lippok B, Birkenbihl RP, Rivory G, Brummer J, Schmelzer E, Logemann E, Somssich IE (2007) Expression of *AtWRKY33* encoding a pathogen- or PAMP-responsive WRKY transcription factor is regulated by a composite DNA motif containing W box elements. Mol Plant Microbe Interact **20:** 420-429

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

Loudet O, Chaillou S, Merigout P, Talbotec J, Daniel-Vedele F (2003) Quantitative trait loci analysis of nitrogen use efficiency in Arabidopsis. Plant Physiol **131:** 345-358

MacAlister CA, Ohashi-Ito K, Bergmann DC (2007) Transcription factor control of asymmetric cell divisions that establish the stomatal lineage. Nature **445:** 537-540

Mallory AC, Reinhart BJ, Jones-Rhoades MW, Tang G, Zamore PD, Barton MK, Bartel DP (2004) MicroRNA control of PHABULOSA in leaf development: importance of pairing to the microRNA 5' region. EMBO J **23:** 3356-3364

Manfield IW, Jen CH, Pinney JW, Michalopoulos I, Bradford JR, Gilmartin PM, Westhead DR (2006) Arabidopsis Co-expression Tool (ACT): web server tools for microarray-based gene expression analysis. Nucleic Acids Res **34:** W504-509

Martin-Tryon EL, Kreps JA, Harmer SL (2007) GIGANTEA acts in blue light signaling and has biochemically separable roles in circadian clock and flowering time regulation. Plant Physiol **143:** 473-486

Matsushita A, Furumoto T, Ishida S, Takahashi Y (2007) AGF1, an AT-hook protein, is necessary for the negative feedback of AtGA3ox1 encoding GA 3-oxidase. Plant Physiol **143:** 1152-1162

McConnell JR, Barton MK (1998) Leaf polarity and meristem formation in Arabidopsis. Development **125:** 2935-2942

Mehrabian M, Allayee H, Stockton J, Lum PY, Drake TA, Castellani LW, Suh M, Armour C, Edwards S, Lamb J, Lusis AJ, Schadt EE (2005) Integrating genotypic and expression data in a segregating mouse population to identify 5-lipoxygenase as a susceptibility gene for obesity and bone traits. Nat Genet **37:** 1224-1233

Meszaros T, Helfer A, Hatzimasoura E, Magyar Z, Serazetdinova L, Rios G, Bardoczy V, Teige M, Koncz C, Peck S, Bogre L (2006) The Arabidopsis MAP kinase kinase MKK1 participates to defence responses to the bacterial elicitor flagellin. Plant J **48:** 485-498

Michael TP, Salome PA, Yu HJ, Spencer TR, Sharp EL, McPeek MA, Alonso JM, Ecker JR, McClung CR (2003) Enhanced fitness conferred by naturally occurring variation in the circadian clock. Science **302:** 1049-1053

Millenaar FF, Cox MC, van Berkel YE, Welschen RA, Pierik R, Voesenek LACJ, Peeters AJM (2005) Ethylene-induced differential growth of petioles in Arabidopsis. Analyzing natural variation, response kinetics, and regulation. Plant Physiol **137:** 998-1008

Millenaar FF, Okyere J, May ST, van Zanten M, Voesenek LA, Peeters AJ (2006) How to decide? Different methods of calculating gene expression from short oligonucleotide array data will give different results. BMC Bioinformatics **7:** 137

Mittler R (2006) Abiotic stress, the field environment and stress combination. Trends Plant Sci **11:** 15-19

Mommer L, Pons TL, Visser EJ (2006) Photosynthetic consequences of phenotypic plasticity in response to submergence: *Rumex palustris* as a case study. J Exp Bot **57:** 283-290

Monks SA, Leonardson A, Zhu H, Cundiff P, Pietrusiak P, Edwards S, Phillips JW, Sachs A, Schadt EE (2004) Genetic inheritance of gene expression in human cell lines. Am J Hum Genet **75:** 1094-1105

Narita NN, Moore S, Horiguchi G, Kubo M, Demura T, Fukuda H, Goodrich J, Tsukaya H (2004) Overexpression of a novel small peptide ROTUNDIFOLIA4 decreases cell proliferation and alters leaf shape in *Arabidopsis thaliana*. Plant J **38:** 699-713

Naur P, Petersen BL, Mikkelsen MD, Bak S, Rasmussen H, Olsen CE, Halkier BA (2003) CYP83A1 and CYP83B1, two nonredundant cytochrome P450 enzymes metabolizing oximes in the biosynthesis of glucosinolates in Arabidopsis. Plant Physiol **133:** 63-72

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

O'Connor TR, Dyreson C, Wyrick JJ (2005) Athena: a resource for rapid visualization and systematic analysis of Arabidopsis promoter sequences. Bioinformatics **21:** 4411-4413

Ogas J, Kaufmann S, Henderson J, Somerville C (1999) PICKLE is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative development in Arabidopsis. Proc Natl Acad Sci U S A **96:** 13839-13844

Oh E, Yamaguchi S, Hu J, Yusuke J, Jung B, Paik I, Lee HS, Sun TP, Kamiya Y, Choi G (2007) PIL5, a phytochrome-interacting bHLH protein, regulates gibberellin responsiveness by binding directly to the GAI and RGA promoters in Arabidopsis seeds. Plant Cell **19:** 1192-1208

Oliverio KA, Crepy M, Martin-Tryon EL, Milich R, Harmer SL, Putterill J, Yanovsky MJ, Casal JJ (2007) GIGANTEA Regulates Phytochrome A-mediated Photomorphogenesis Independently of its Role in the Circadian Clock. Plant Physiol

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

Perata P, Alpi A (1993) Plant responses to anaerobiosis. Plant Science **93:** 1-17

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

Pierik R, Cuppens ML, Voesenek LA, Visser EJ (2004) Interactions between ethylene and gibberellins in phytochrome-mediated shade avoidance responses in tobacco. Plant Physiol **136:** 2928-2936

Pierik R, Millenaar FF, Peeters AJ, Voesenek LA (2005) New perspectives in flooding research: the use of shade avoidance and *Arabidopsis thaliana*. Ann Bot (Lond) **96:** 533-540

Pierik R, Visser EJW, De Kroon H, Voesenek LACJ (2003) Ethylene is required in tobacco to successfully compete with proximate neighbours. Plant Cell and Environment **26:** 1229-1234

Pierik R, Whitelam GC, Voesenek LA, de Kroon H, Visser EJ (2004) Canopy studies on ethylene-insensitive tobacco identify ethylene as a novel element in blue light and plant-plant signalling. Plant J **38:** 310-319

Pillitteri LJ, Sloan DB, Bogenschutz NL, Torii KU (2007) Termination of asymmetric cell division and differentiation of stomata. Nature **445:** 501-505

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

Reed JW, Nagatani A, Elich TD, Fagan M, Chory J (1994) Phytochrome A and Phytochrome B Have Overlapping but Distinct Functions in Arabidopsis Development. Plant Physiol **104:** 1139-1149

Reed JW, Nagpal P, Poole DS, Furuya M, Chory J (1993) Mutations in the gene for the red/far-red light receptor phytochrome B alter cell elongation and physiological responses throughout Arabidopsis development. Plant Cell **5:** 147-157

Rijnders J, Armstrong W, Darwent M, Blom C, Voesenek L (2000) The role of oxygen in submergence-induced petiole elongation in *Rumex palustris*: *in situ* measurements of oxygen in petioles of intact plants using micro-electrodes. New Phytol **147:** 497-504

Ritchie ME, Diyagama D, Neilson J, van Laar R, Dobrovic A, Holloway A, Smyth GK (2006) Empirical array quality weights in the analysis of microarray data. BMC Bioinformatics **7:** 261

Ritchie ME, Silver J, Oshlack A, Holmes M, Diyagama D, Holloway A, Smyth GK (2007) A comparison of background correction methods for two-colour microarrays. Bioinformatics **23:** 2700-2707

Robatzek S, Somssich IE (2001) A new member of the Arabidopsis WRKY transcription factor family, AtWRKY6, is associated with both senescence- and defence-related processes. Plant J **28:** 123-133

Robatzek S, Somssich IE (2002) Targets of AtWRKY6 regulation during plant senescence and pathogen defense. Genes Dev **16:** 1139-1149

Saika H, Okamoto M, Miyoshi K, Kushiro T, Shinoda S, Jikumaru Y, Fujimoto M, Arikawa T, Takahashi H, Ando M, Arimura S, Miyao A, Hirochika H, Kamiya Y, Tsutsumi N, Nambara E, Nakazono M (2007) Ethylene promotes submergence-induced expression of *OsABA8ox1*, a gene that encodes ABA 8'-hydroxylase in rice. Plant Cell Physiol **48:** 287-298

Sauter M (2000) Rice in deep water: "how to take heed against a sea of troubles". Naturwissenschaften **87:** 289-303

Scarcelli N, Cheverud JM, Schaal BA, Kover PX (2007) Antagonistic pleiotropic effects reduce the potential adaptive value of the FRIGIDA locus. Proc Natl Acad Sci U S A **104:** 16986-16991

Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, Linsley PS, Mao M, Stoughton RB, Friend SH (2003) Genetics of gene expression surveyed in maize, mouse and man. Nature **422:** 297-302

Shpak ED, Berthiaume CT, Hill EJ, Torii KU (2004) Synergistic interaction of three ERECTA-family receptor-like kinases controls Arabidopsis organ growth and flower development by promoting cell proliferation. Development **131:** 1491-1501

Shpak ED, McAbee JM, Pillitteri LJ, Torii KU (2005) Stomatal patterning and differentiation by synergistic interactions of receptor kinases. Science **309:** 290-293

Siegfried KR, Eshed Y, Baum SF, Otsuga D, Drews GN, Bowman JL (1999) Members of the YABBY gene family specify abaxial cell fate in Arabidopsis. Development **126:** 4117-4128

Smith H, Whitelam GC (1997) The shade avoidance syndrome: Multiple responses mediated by multiple phytochromes. Plant Cell Environ **20:** 840-844

Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol **3:** Article3

Smyth GK, Michaud J, Scott HS (2005) Use of within-array replicate spots for assessing differential expression in microarray experiments. Bioinformatics **21:** 2067-2075

Smyth GK, Speed T (2003) Normalization of cDNA microarray data. Methods **31:** 265-273

Storey JD (2002) A direct approach to false discovery rates J. R. Statist. Soc. B **64:** 479-498

Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. Proc Natl Acad Sci U S A **100:** 9440-9445

Stowe-Evans EL, Harper RM, Motchoulski AV, Liscum E (1998) NPH4, a conditional modulator of auxin-dependent differential growth responses in Arabidopsis. Plant Physiol **118:** 1265-1275

Suarez-Rodriguez MC, Adams-Phillips L, Liu Y, Wang H, Su SH, Jester PJ, Zhang S, Bent AF, Krysan PJ (2007) MEKK1 is required for flg22-induced MPK4 activation in Arabidopsis plants. Plant Physiol **143:** 661-669

Swarup K, Alonso-Blanco C, Lynn JR, Michaels SD, Amasino RM, Koornneef M, Millar AJ (1999) Natural allelic variation identifies new genes in the Arabidopsis circadian system. Plant J **20:** 67-77

Tang W, Perry SE (2003) Binding site selection for the plant MADS domain protein AGL15: an in vitro and in vivo study. J Biol Chem **278:** 28154-28159

Teige M, Scheikl E, Eulgem T, Doczi R, Ichimura K, Shinozaki K, Dangl JL, Hirt H (2004) The MKK2 pathway mediates cold and salt stress signaling in Arabidopsis. Mol Cell **15:** 141-152

ter Steege MW, den Ouden FM, Lambers H, Stam P, Peeters AJ (2005) Genetic and physiological architecture of early vigor in *Aegilops tauschii*, the D-genome donor of hexaploid wheat. A quantitative trait loci analysis. Plant Physiol **139:** 1078-1094

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

Tsukaya H, Kozuka T, Kim GT (2002) Genetic control of petiole length in *Arabidopsis thaliana*. Plant Cell Physiol **43:** 1221-1228

Turck F, Zhou A, Somssich IE (2004) Stimulus-dependent, promoter-specific binding of transcription factor WRKY1 to Its native promoter and the defense-related gene PcPR1-1 in Parsley. Plant Cell **16:** 2573-2585

Voesenek LA, Colmer TD, Pierik R, Millenaar FF, Peeters AJ (2006) How plants cope with complete submergence. New Phytol **170:** 213-226

Voesenek LA, Harren FJ, Bogemann GM, Blom CW, Reuss J (1990) Ethylene Production and Petiole Growth in Rumex Plants Induced by Soil Waterlogging: The Application of a Continuous Flow System and a Laser Driven Intracavity Photoacoustic Detection System. Plant Physiol **94:** 1071-1077

Voesenek LACJ, Blom CWPM, Pouwels RHW (1989) Root and Shoot Development of Rumex Species under Waterlogged Conditions. Can J Bot **67:** 1865-1869

Vreeburg RA, Benschop JJ, Peeters AJ, Colmer TD, Ammerlaan AH, Staal M, Elzenga TM, Staals RH, Darley CP, McQueen-Mason SJ, Voesenek LA (2005) Ethylene regulates fast apoplastic acidification and expansin A transcription during submergence-induced petiole elongation in Rumex palustris. Plant J **43:** 597-610

Vriezen WH, De Graaf B, Mariani C, Voesenek LA (2000) Submergence induces expansin gene expression in flooding-tolerant Rumex palustris and not in flooding-intolerant R. acetosa. Planta **210:** 956-963

Vuylsteke M, van Eeuwijk F, Van Hummelen P, Kuiper M, Zabeau M (2005) Genetic analysis of variation in gene expression in *Arabidopsis thaliana*. Genetics **171:** 1267-1275

Wagner TA, Kohorn BD (2001) Wall-associated kinases are expressed throughout plant development and are required for cell expansion. Plant Cell **13:** 303-318

Wang S, Basten CJ, Zeng ZB (2007) Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC. (http://statgen.ncsu.edu/qtlcart/WQTLCart.htm)

West MA, Kim K, Kliebenstein DJ, van Leeuwen H, Michelmore RW, Doerge RW, St Clair DA (2007) Global eQTL Mapping Reveals the Complex Genetic Architecture of Transcript-Level Variation in Arabidopsis. Genetics **175:** 1441-1450

Xu L, Xu Y, Dong A, Sun Y, Pi L, Xu Y, Huang H (2003) Novel as1 and as2 defects in leaf adaxial-abaxial polarity reveal the requirement for ASYMMETRIC LEAVES1 and 2 and ERECTA functions in specifying leaf adaxial identity. Development **130:** 4097-4107

Xu X, Chen C, Fan B, Chen Z (2006) Physical and functional interactions between pathogen-induced Arabidopsis WRKY18, WRKY40, and WRKY60 transcription factors. Plant Cell **18:** 1310-1326

Yamasaki K, Kigawa T, Inoue M, Tateno M, Yamasaki T, Yabuki T, Aoki M, Seki E, Matsuda T, Tomo Y, Hayami N, Terada T, Shirouzu M, Tanaka A, Seki M, Shinozaki K, Yokoyama S (2005) Solution structure of an Arabidopsis WRKY DNA binding domain. Plant Cell **17:** 944-956

Yokoyama R, Takahashi T, Kato A, Torii KU, Komeda Y (1998) The Arabidopsis ERECTA gene is expressed in the shoot apical meristem and organ primordia. Plant J **15:** 301-310

Yvert G, Brem RB, Whittle J, Akey JM, Foss E, Smith EN, Mackelprang R, Kruglyak L (2003) Trans-acting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors. Nat Genet **35:** 57-64

Zeng ZB (1994) Precision mapping of quantitative trait loci. Genetics **136:** 1457-1468

Zeng ZB, Kao CH, Basten CJ (1999) Estimating the genetic architecture of quantitative traits. Genet Res **74:** 279-289

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

Zheng Z, Qamar SA, Chen Z, Mengiste T (2006) Arabidopsis WRKY33 transcription factor is required for resistance to necrotrophic fungal pathogens. Plant J **48:** 592-605

Zhu J, Lum PY, Lamb J, GuhaThakurta D, Edwards SW, Thieringer R, Berger JP, Wu MS, Thompson J, Sachs AB, Schadt EE (2004) An integrative genomics approach to the reconstruction of gene networks in segregating populations. Cytogenet Genome Res **105:** 363-374

Supplemental information

Supplemental information and color versions of figures are available on request.
Current e-mail address: basten.snoek@wur.nl

Natuurlijke variatie in schaduw- en ethyleen-geïnduceerde differentiële groei en transcriptie in Arabidopsis.

Planten hebben een vaste standplaats waardoor ze geconfronteerd worden met sterk wisselende leefomstandigheden. Hierdoor is het van essentieel belang voor het succes van een plant dat deze kan reageren op, en zich kan aanpassen aan een veranderende omgeving. Verscheidene mechanismen hebben zich in planten geëvolueerd om zo efficiënt mogelijk om te gaan met de grillige beschikbaarheid van nutriënten en zonlicht. Aanpassingen aan deze vormen van stress zijn erfelijk en betreffen functionele variaties op het gebied van fysiologie, anatomie, morfologie en gedrag of een combinatie hiervan. Variatie in deze eigenschappen is niet alleen terug te vinden tussen verschillende soorten planten maar ook tussen verschillende individuen binnen een soort. Deze verschillen maken het voor een soort niet alleen mogelijk om te gaan met de veranderingen binnen een habitat maar ook te overleven in verschillende habitatten.

Stress als gevolg van overstroming

Overstroming is voor de meeste planten een bron van ernstige stress. Te veel aan water, of dat nu een door water verzadigde bodem is of totale onderdompeling van de gehele plant resulteert vrijwel altijd in een verlies aan productiviteit. Ongeveer 10% van het wereldwijde landbouwareaal krijgt, incidenteel of regelmatig, te maken met een teveel aan water.

Landplanten (de meeste landbouwgewassen) zijn meestal niet goed aangepast aan overstroming en zullen snel dood gaan wanneer dit voorkomt. Dit komt voornamelijk doordat de uitwisseling van gassen met de omgeving als gevolg van het water ernstig is verminderd ten opzichte van lucht. Hierdoor krijgen planten te maken met een tekort aan zuurstof, waardoor er moet worden overgegaan op anaërobe verbranding, wat minder energie oplevert.

In de loop van de evolutie hebben zich twee mechanismen ontwikkeld om met deze situatie om te gaan. Het eerste mechanisme zorgt ervoor dat er energie bespaard wordt ten tijde van een overstroming en deze zodoende langer overleefd kan worden. Het tweede zorgt ervoor dat voldoende zuurstof de overstroomde weefsels bereikt, vanuit de niet overstroomde delen, onder andere door het vormen van luchtkanalen. In het geval van complete overstroming hebben enkele soorten een reactie ontwikkeld waarbij door middel van differentiële groei (hoekverandering) en snelle lengtegroei het blad weer boven het water uitkomt. Deze kunnen dan samen met de de vergrootte luchtkanalen als een soort snorkel gebruikt worden om gasuitwisseling met de lucht weer mogelijk te maken.

Een soort die dit mechanisme uitzonderlijk goed ontwikkeld heeft is moeraszuring, *Rumex palustris*. Deze plant neemt overstroming waar door ophoping van het gasvormig plantenhormoon ethyleen. Ethyleen wordt constant geproduceerd en hoopt zich op

gedurende een overstroming door de beperkte gasuitwisseling met water. Wanneer de ethyleen concentratie is opgelopen reageert *R. palustris* met het omhoog richten van de bladeren door groei aan de onderkant van de bladsteel (hyponastische groei). Hierna gaat de gehele bladsteel groeien zodat de bladpunt boven het wateroppervlak kan komen en de gasuitwisseling met de lucht wordt hersteld. Bij deze differentiële en lengtegroei spelen, naast ethyleen, een aantal andere plantenhormonen zoals abscisine zuur, gibberelline en auxine een belangrijke en specifieke rol. Verscheidene genen die coderen voor enzymen die betrokken zijn bij de productie of afbraak van deze hormonen die de groei mogelijk maken worden gereguleerd op transcriptie niveau.

Arabidopsis thaliana

Ondanks het gegeven dat *Rumex palustris* een ideale soort is om overstromings-geïnduceerde differentiële groei te bestuderen vanuit een fysiologisch oogpunt zijn er ook een aantal nadelen. Vooral genetische en moleculaire is *R. palustris* niet makkelijk te bestuderen. Vergeleken met de modelplant zandraket, *Arabidopis thaliana* heeft *R. palustris* een groot polyploid genoom en is er nog maar weinig moleculaire en genetische informatie beschikbaar. Op genetisch en moleculair niveau is *A. thaliana* echter een van de best bestudeerde planten heden ten dage. Ook al heeft *A. thaliana* niet de volledige groeireactie op overstroming, een deels vergelijkbare groeireactie treedt op wanneer *A. thaliana* wordt geconfronteerd met schaduw, overstroming of een verhoogde ethyleenconcentratie. Dit maakt *A. thaliana* geschikt om de genetische variatie en de achterliggende moleculaire mechanismen van geïnduceerde differentiële groei te bestuderen.

Variatie in verschijningsvormen van individuen van een soort is terug te vinden in vrijwel alle soorten, waaronder *A.thaliana*. Deze variatie wordt voornamelijk veroorzaakt door twee factoren, omgevings- en genetische variatie. Wanneer verschillende ecotypes van *A. thaliana* opgroeien onder gelijke omgevingsomstandigheden, kan de genetische invloed op de verschijningsvorm bestudeerd worden.

Analyse van kwantitatieve eigenschappen

De ontwikkeling van, en het gemak waarmee moleculaire merkers kunnen worden toegepast heeft de analyse van genetische variatie vergemakkelijkt en voor een toename van het gebruik van natuurlijk variatie in de analyse van kwantitatieve eigenschappen gezorgd. Ondanks de complexiteit van de genetische regulatie van zulke eigenschappen is er veel vooruitgang geboekt in het ontrafelen van deze eigenschappen door het combineren van moleculaire merkers en geavanceerde mapping methodes. De ontwikkeling van onsterfelijke mapping populaties heeft de studie naar genetische mechanismen achter natuurlijke variatie sterk geholpen. Vooral de onsterfelijke recombinante inteelt populaties (RIL) die de afgelopen 15 jaar zijn ontwikkeld worden veelvuldig gebruikt.

In tegenstelling tot mutantenanalyses, waarbij de variatie in de verschijningsvorm van één gen bestudeerd wordt om achter de positie en functie te komen, bestaat een RIL

populatie uit variaties van meerdere genen. Daardoor is een RIL populatie geschikt om de variatie in veel verschillende eigenschappen te bestuderen, maar ook van eigenschappen die door meerdere genen beïnvloed worden. Een RIL populatie wordt gemaakt door twee (homozygote) ouders te kruisen. Door recombinatie van de ouderlijke genomen in de nakomelingen, die door zelfbevruchting tot stand komen, ontstaat een mozaïek van nieuwe genotypen. Door herhaaldelijke zelfbestuiving kan een homozygote populatie worden verkregen die keer op keer te gebruiken is om loci te vinden die de variatie in een bepaalde eigenschap bepalen.

Natuurlijke variatie in transcript niveaus

Door de ontwikkeling van 'high-througput' methoden om transcriptniveaus te meten is het mogelijk geworden om dit te meten voor alle genen van een gehele RIL populatie. Deze transcriptniveaus kunnen evenzeer gezien worden als kwantitatieve eigenschappen, waarvan de variatie mogelijk kan worden gerelateerd aan een specifiek locus. Deze loci geven aan waar een mogelijke regulator van het gen ligt. Deze informatie kan vervolgens gebruikt worden om een netwerk te maken van op transcript niveau regulerende en gereguleerde genen. Gedurende dit promotieproject is gebleken dat dit een haalbare en nuttige techniek is om de regulatieprocessen in *A. thaliana* te bestuderen.

Dit proefschrift

Wat zijn de onderliggende regulatie netwerken in door ethyleen- en schaduw-geïnduceerde differentiële groei en hoe worden deze door natuurlijke variatie beïnvloed, zijn de twee belangrijkste vragen van dit promotieproject.

In hoofdstuk twee wordt de opwaartse beweging, hyponastische groei, van het blad als gevolg van een verhoogde ethyleenconcentratie of schaduw in detail beschreven. De natuurlijke variatie hierin wordt bestudeerd met behulp van kwantitatieve eigenschap analyse. Twee, in plaats gescheiden, gebieden van een blad waar differentiële groei plaats vindt werden geïdentificeerd. Eén aan de basis van de bladsteel en één op de grens van de bladsteel met de bladschijf. Voor de variatie in absolute hoeken en reacties van deze twee organen, bladstelen en bladschijven, werden specifieke genoomloci gevonden die een deel van die variatie konden verklaren. Deze twee regio's die de bladhoek verandering veroorzaken hebben enerzijds een specifieke genetische architectuur, maar anderzijds ook genetische componenten gemeenschappelijk. De meeste loci konden bevestigd worden met behulp van introgressie lijnen, planten waarbij een klein deel van het genoom van de ene ouder (de plek van het locus) en de rest uit de andere ouder afkomstig is. Verder werd er een verschil gevonden in de manier waarop de planten reageren op schaduw en verhoogde ethyleenconcentraties. De bladschijven, de tweede regio met differentiële groei, reageerden nauwelijks op een verhoogde ethyleenconcentratie, terwijl schaduw wel een reactie teweeg bracht. Op een locus dat een belangrijk deel van de variatie in reactie verklaarde na beide behandelingen kon het *ERECTA* gen geïdentificeerd worden. Het allelische verschil tussen

de ouders in dit gen veroorzaakt een verschil in reactie op vooral ethyleen, maar ook op schaduw.

Hoofstuk drie beschrijft de door licht beïnvloedde groeieigenschappen, te weten rozet oppervlakte, rozet diameter en de compactheid van de rozet. Ook voor deze eigenschappen is de natuurlijke variatie bestudeerd met behulp van kwantitatieve eigenschap analyse. Verscheidene loci, die elk een deel van de variatie in deze eigenschappen verklaren, werden gevonden. Ook voor de variatie in deze eigenschappen speelde het verschil tussen beide ouders in het *ERECTA* gen een belangrijke rol. Ook de allelische verschillen in een ander gen, *PHYTOCHROME B*, speelden een belangrijke rol in de variatie in deze rozeteigenschappen.

Het *ERECTA* gen wordt in hoofdstuk vier verder bestudeerd. In dit hoofdstuk wordt beschreven hoe en welk transcript-regulatie-netwerk downstream van ERECTA functioneert. Met behulp van genetical genomics en transcriptniveaus van de *erecta* mutant, Landsberg *erecta* (L*er*), en het bijbehorende wildtype, Landsberg (LAN), als startpunt voor het netwerk kon een deel van de signaalroute na ERECTA ontrafeld worden.

Hoofdstuk vijf concentreert zich op een genetical genomics experiment waarin drie weken oude planten zijn blootgesteld aan drie uur schaduw. De resultaten van dit experiment zijn vergeleken met de resultaten van een eerder genetical genomics experiment waarbij gebruik is gemaakt van dezelfde populatie, echter jonger en niet behandeld met schaduw. Hierdoor werden een groot aantal loci ontdekt die de variatie verklaren in transcriptniveaus afhankelijk van de behandeling (of ontwikkeling stadium) van de planten. Ook een aantal nieuwe “hotspots” van transcriptregulatie werden geïdentificeerd.

Deze informatie werd in hoofdstuk zes gebruikt om een regulatienetwerk te construeren door een startpunt te genereren met het transcriptieprofiel van een introgressielijn ten opzichte van een van de ouderplanten, L*er*. Door een kombinatie van genetical genomics, lokale associatie en transcriptie profilering is een gen gevonden dat mogelijk verantwoordelijk is voor de verschillen in verschijningsvorm tussen de introgressie-lijn en L*er*. Verder leverde deze studie inzicht in transcriptregulatie binnen de metabole route die leidt tot auxine productie of glucosinolaat vorming op. Genetische variatie en lichtintensiteit verklaren een groot deel van de variatie in transcript niveau’s binnen deze route.

Dankwoord,

Allereerst wil ik iedereen bedanken die op enigerwijze heeft bij gedragen aan dit proefschrift. Ik heb veel geleerd de afgelopen jaren. Bijvoorbeeld dat wetenschap en onderzoek echt mensenwerk is, met alle facetten die je daarbij tegenkomt. Dat voorstellings- en doorzettingsvermogen erg belangrijke eigenschappen zijn om een proefschrift af te leveren. Maar vooral dat onderzoeken uitdagend en leuk blijft.

Rita bedankt voor alle steun en ons fijne thuis, allebei aan een promotieproject werken is soms best druk maar met jou altijd erg leuk. Niet alleen samen promoveren maar nu ook papa en mama van Jorrit. Jorrit wil ik dan ook bedanken voor zijn vrolijke karakter en voor het slapen op mijn arm, terwijl ik een groot deel van dit proefschrift typte. Het afronden van dit boekje was daardoor een andere ervaring dan ik me een aantal jaren daarvoor had voorgesteld.

Ik wil Rens als promoter en Ton als co-promoter bedanken dat zij me deze baan aanboden. Daaraan voorafgaand, Maarten, bedankt dat je me aangeraden hebt bij Ton. Ik wil Ton bedanken voor de uitstekende begeleiding en de manier waarop ik, mede daardoor, een hoop heb kunnen leren.

Martijn, ik had geen betere kamergenoot kunnen wensen, bedankt voor je gezelligheid, enthousiasme en wetenschappelijke input. Frank en Ronald, ik ben ook nog een tijdje jullie kamergenoot geweest, ook dat was gezellig en bedankt voor het delen van jullie ervaringen als postdocs. Hendrik bedankt voor je humor en het delen van je alternatieve kijk op allerlei zaken. Zohreh and Tanja thanks for your input and insights in other cultures.

Ook met de masterstudenten was het meestal leuk, Fionn, Auke, Feddo, Dirk bedankt voor jullie input. Dirk bedankt voor je geweldige hulp tijdens het eerstejaars practicum. Fionn, Auke en Diederik verdienen ook een extra bedankje voor het meenemen van een afgestudeerde naar een paar studentenfeestjes en door me keer op keer te verslaan met poolen.

Bedanken wil ik uiteraard ook de hele QTL express groep, Joost, Maarten, Sjeng, Sjef, Inez, Guido, Jing Juan & Ritsert, het was spannend en uitdagend om de mogelijkheden van “genetical genomics” in Arabidopsis te verkennen. Inez bedankt voor de voortreffelijke samenwerking en voor de inspiratie om in “R” te leren programmeren. Rene, bedankt voor de uitstekende micro-arrays, de muziek en de verhalen op het lab.

Rob bedankt voor al je hulp, vooral bij het draaiende houden van de kweekkamers was je onmisbaar. Ankie, Yvonne, Judith en Kerstin bedankt voor jullie hulp bij van alles en nog wat, zonder jullie zou het een zooitje zijn geworden. Voor de rest iedereen van de leerstoelgroep “ecofysiologie van planten” bedankt voor de wetenschappelijke hulp en de gezellige input tijdens het werk en de koffie.

Basten.

CURRICULUM VITAE

Education

2003 – 2007	Graduate school for Experimental Plant Sciences PhD education
2001 – 2002	Wageningen University Master of Science; (Plant) Biotechnology Graduation: Plant genetics Graduated: September 2002
1995 – 2000	Hogeschool Enschede Medical and Biological research Biochemistry / Biotechnology Graduated: June 2000
1993 – 1994	HAVO Class 4 and 5
1989 – 1993	VWO Class 1 - 3

Employment

Oct 2007 – present
Employer: Wageningen University, Wageningen; Department of Nematology.
Postdoc; project “Natural variation in complex disease signaling pathways in *C.elegans*”.

Jan 2003 – Summer 2007
Employer: Utrecht University, Utrecht; Department of Plant Ecophysiology.
PhD-student; project “Genetical genomics of plant performance traits”.

Jan 2001 – Dec 2002
Employer: Plant Research International, Wageningen; Business unit Biodiversity.
Technician; project “The influence of the land use management system on genetic diversity of grassland species”.

Aug 2000 – Dec 2000
Employer: Plant Research International, Wageningen; Business unit Biodiversity.
Technician; project “Genetic diversity of threatened animal populations; “korenwolf””.

Oct 1999 – Apr 2000

Employer: Plant Research International, Wageningen; Business unit Cell-cybernetics.
Traineeship, project “Expressing the viciline protein in the potato tuber” and “Shuffling and mutating the DHDPS-gene to obtain a feedback insensitive enzyme”.

May 1999 – Oct 1999

Employer: Plant Research International, Wageningen; Business unit Cell-cybernetics.
Traineeship, “Increasing the concentration of essential amino-acids in the potato tuber”.

General experience

Nov 2005 – Dec 2006

President of the PhD council of the graduate school for Experimental Plant Sciences

Mar 2003 – Nov 2005

PhD member of the educational comity of the graduate school for Experimental Plant Sciences

Publications

2009 van Zanten M, Snoek LB, Proveniers MCG, Peeters AJM (2009) The many functions of ERECTA. Trends in Plant Science.Vol 14, No. 4: 214-218

2007 Keurentjes JJ, Fu J, Terpstra IR, Garcia JM, van den Ackerveken G, Snoek LB, Peeters AJM, Vreugdenhil D, Koornneef M, Jansen RC (2007) Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. Proc Natl Acad Sci U S A 104: 1708-1713

2003 Smulders MJM, Snoek LB, Booy G, Vosman B (2003) Complete loss of MHC genetic diversity in the Common Hamster (Cricetus cricetus) population in The Netherlands. Consequences for conservation strategies. Conservation Genetics 4: 441-451

Conferences

2008, 2009 Netherlands Annual Ecology Meeting (NERN); Lunteren
Poster presentation 2008
Presentation 2008

2008 fourth Annual progress Meeting of the Ecogenomics Consortium; Amsterdam
Presentation

2008 Keystone Conference on Complex Traits: Biologic and Therapeutic Insights; Santa Fe
Poster presentation
Presentation

2005 Gordon Conference on Quantitative genetics; Los Angeles
Poster presentation

2003 – 2007 Annual meeting of Plant Biological Research; Lunteren
Poster presentation 2004 & 2005
Presentation 2006

2004 15th International Conference on Arabidopsis Research; Berlin
Poster presentation