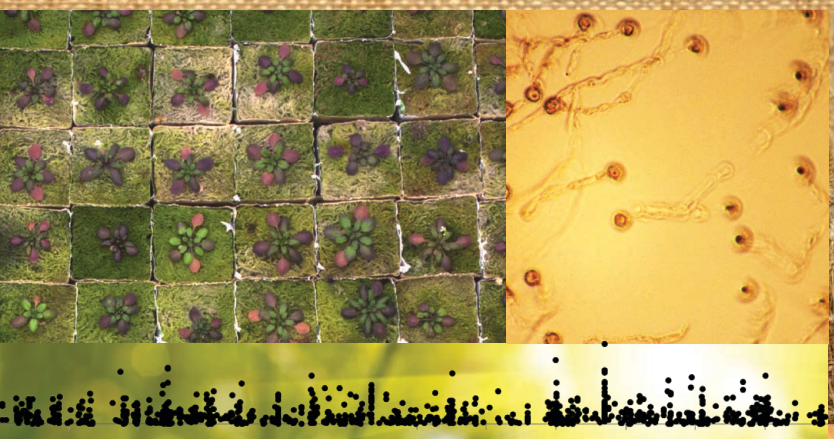
Using natural variation to unravel the dynamic regulation of plant performance in diverse environments



Johanna A. Bac-Molenaar

USING NATURAL VARIATION TO UNRAVEL THE DYNAMIC REGULATION OF PLANT PERFORMANCE IN DIVERSE ENVIRONMENTS

Johanna A. Bac-Molenaar

Thesis committee

Promotors
Prof. Dr H.J. Bouwmeester
Professor of Plant Physiology
Wageningen University

Prof. Dr J.J.B. Keurentjes
Professor of Applied Quantitative Genetics
University of Amsterdam
Associate professor, Laboratory of Genetics
Wageningen University

Co-promotor
Dr D. Vreugdenhil
Associate professor, Laboratory of Plant Physiology
Wageningen University

Other members
Prof. Dr G.C. Angenent, Wageningen University
Dr B. Muller, Institut National de la Recherche Agronomique, Montpellier, France
Dr C.S. Testerink, University of Amsterdam
Dr N. de Vetten, Averis Seeds B.V., Valthermond

This research was conducted under the auspices of
the Graduate School of Experimental Plant Sciences

USING NATURAL VARIATION TO UNRAVEL THE DYNAMIC REGULATION OF PLANT PERFORMANCE IN DIVERSE ENVIRONMENTS

Johanna A. Bac-Molenaar

Thesis
submitted in fulfillment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Wednesday 2 September 2015
at 4 p.m. in the Aula.

Johanna A Bac-Molenaar

Using natural variation to unravel the dynamic regulation of plant performance in diverse environments
186 pages.

PhD thesis, Wageningen University, Wageningen, NL (2015)
With references, with summaries in Dutch and English

ISBN 978-94-6257-344-4

CONTENTS

Chapter 1 — General Introduction

Johanna A. Bac-Molenaar

Goal

Plant performance is determined by an interplay between the environment and the genetic composition of the plant. Plants, as sessile organisms, have the ability to sense changes in the environmental conditions and to respond to them by adapting their growth and metabolism. Genotypic differences are responsible for the diversity in responses observed within and between species. Many experiments are performed in the last years to elucidate the genetic regulation of plant responses to environments. This has increased our knowledge about plant performance significantly, but still not all genes involved in growth and stress responses are identified. And if those genes are identified, their position and role in the underlying regulatory network is in many cases not yet known.

In this thesis I identify and study the role of genes that are underlying natural variation in plant performance in various diverse environments, with a focus on genes involved in the regulation of plant performance under drought and heat stress. To reach this goal a genome wide association (GWA) mapping approach is taken in the model species *Arabidopsis thaliana*.

Natural variation and local adaptation

Central in this thesis are heritable plant traits for which variation is observed in nature. These kind of traits are under selection by humans since some of them settled as farmers. Farmers selected the seeds of the best performing plants to sow them in the next growing season. As a consequence of this domestication process plant performance was increased, but genetic variation was reduced. Later in history, farmers started to breed for new varieties by crossing to combine positive traits of parental lines. Gregor Mendel (1822-1884) discovered that traits are inherited according to rules that are based on logic. Although he could only test simple qualitative traits that are regulated by a limited number of genes, it is nowadays generally accepted that Mendelian genetics also apply for complex quantitative traits that are regulated by large numbers of genes. In the same century, Charles Darwin (1809-1882) and Alfred Wallace (1823-1914) were fascinated by the enormous variety present in nature, both variation between and within species. They suggested that, like humans select plant traits in agriculture, also in nature environmental conditions have selective power on organisms. Their experiments focused on the study of variation in geographically separated groups of the same species and on subsequent speciation, but they extrapolated their findings to hypothesize about the origin of all species. Their ideas inspired many researchers and are still the basis of what nowadays is called the evolution theory. It took until the fifties of the twentieth century before researchers came to the consensus that not proteins, but deoxyribonucleic acids (DNA) are the bearer of genetic information (Avery 1944, Hershey 1952). In the same decade also the double helix structure of DNA was elucidated (Watson 1953). In 1980 the first method to de-

velop molecular markers was published (Botstein 1980). Molecular markers enabled us to genetically characterise individuals and to distinguish them from each other based on differences between their DNA sequences (Semagn 2006). In 2000 the first plant genome was sequenced of the model species Arabidopsis (Arabidopsis Genome Initiative 2000) and thereafter access to genetic information became increasingly easier because of advances in sequencing technologies (Metzker 2010, Wang 2015).

Coupling genotype to phenotype

The enormous amount of genetic information that is available nowadays can only be useful for the understanding of the regulation of heritable traits if a connection can be made between the genotype and the phenotype. Linkage mapping is an effective way to study the regulation of heritable traits (Alonso-Blanco 2000). Linkage mapping is performed in the segregating off-spring of a cross between two accessions of the same species (Alonso-Blanco 1998). Each individual of the segregating population is genotyped with a number of markers that are polymorphic between the parental lines and that are evenly distributed over the whole genome. The relative position of each marker in the genome is determined and through linkage analysis of the markers a genetic map is made. In addition, each individual of the segregating population is phenotypically characterised for the trait of interest. Thereafter each marker is statistically tested to identify linkage between the phenotypic trait and the genotypic marker. Linkage mapping results in the detection of quantitative trait loci (QTLs), genomic regions that are associated with the trait of interest. Statistical methods are optimized to identify accurate QTL positions and to minimize false discovery rates (Ooijen 1992, Jansen 1993). Linkage mapping can be done in all populations of which the pedigree is known, like F2-, backcross-, double haploid (DH) and recombinant inbred (RIL) populations. RIL and DH populations are preferred because they are immortal due to homozygosity and can therefore be replicated. In case of quantitative traits, linkage mapping can result in the detection of several QTLs that all contribute to the explanation of the phenotype. Causal genes underlying QTLs are genes of which the gene sequence (exon, intron, promotor or other cis-regulatory DNA elements) differs between the two parents. Genes that do not differ between the parents will not be detected with this method. Linkage mapping in one experimental population will therefore almost never result in the detection of all important regulatory genes, because within one population only a very limited amount of the total species variation is represented. Furthermore, hundreds of genes are usually located within a QTL support interval. To identify causal genes more recombination events must be introduced within the QTL region by additional rounds of crossing or inbreeding. This process of fine-mapping allows to indicate a small genomic region to be causal. Functional analysis of genes located in this small region will allow to identify the causal gene. Linkage mapping became a popular method in both the model species Arabidopsis and in crops to study the genetic regulation of processes on the whole-plant level, and the metabolic (Keurentjes 2009) and gene expression level (Mackay 2009). It is

also the basis of marker assisted selection which is commonly used nowadays in breeding programs (Collard 2008).

To increase the amount of natural variation included in the population, multi-parental populations have been developed (Balasubramanian 2009, Kover 2009, McMullen 2009, Huang 2011). Multiple rounds of intercrossing are needed to obtain a population with individuals that contain genetic information from all parents. Due to the increased number of meiotic events included in those populations the mapping resolution is increased. Development of these populations takes a long time and therefore only a limited amount of multi-parental populations is developed.

In human genetics, experimental crosses are unethical and therefore linkage mapping could only be applied on data collected in families. To make use of the genetic variation present within human populations and to investigate the genetic regulation of diseases, a mapping method was developed that could be used in populations of individuals with unknown pedigree (natural populations). This method is called genome-wide association (GWA) mapping. Linkage decay is very fast in a population of unrelated individuals because the genetic composition of the members is the consequence of ages of selection on natural variation due to spontaneous mutations and meiotic recombination events. In experimental populations with closely related individuals linkage decay is much less strong, because it only depends on recombination events introduced by crossing. To achieve proper mapping results the presence of at least one marker in linkage disequilibrium with each gene in the genome is required. In experimental populations only hundreds of markers are needed to cover the whole genome, whereas in natural populations more than hundred-thousand markers are needed for full genome coverage (International HapMap Consortium 2003). Although the concept of GWA mapping was already proposed in 1996 (Risch 1996), new genotyping techniques were needed to be able to reach marker densities that allowed full genome coverage in natural populations. In 2006 the first study using GWA mapping in a human population was published (De-Wan 2006, Visscher 2012).

Although it is possible to create experimental mapping populations in plants, GWA mapping in populations of accessions of the same plant species seems to have advantages over linkage mapping. GWA mapping allows to test at once most of the genetic variation present in a species. In addition, the fast linkage decay (Nordborg 2002, Kim 2007) will result in the detection of QTLs with high resolution. Before association mapping was applied to the whole genome, it was used for fine-mapping (Hagenblad 2004) and the determination of causal nucleotides in QTL candidate genes (Olsen 2004).When I started this project in January 2011, only a very limited number of studies performing GWA mapping in Arabidopsis were published (Atwell 2010, Brachi 2010, Chan 2010). In addition a study was published about GWA mapping in barley (Cockram 2010). Thereafter each year more GWA mapping studies in both Arabidopsis and crops were published.

Although GWA mapping is relatively new in the field of plant sciences, in addition to clear advantages also disadvantages were reported (Korte 2013).

Clear advantages of GWA mapping are the possibility to include most of the genetic variation present within the species and the high mapping resolution are. A point of attention is the population structure, because it can be a confounding factor and therefore should be corrected for in the statistical model. This is at the moment included in the standard GWA mapping approaches (Seren 2012), but one should be aware that correction can lead to false negatives (Brachi 2010). Disadvantages are related to three characteristics of natural populations. First, for most causal variants more than two alleles will be present. Most markers used in GWA mapping are bi-allelic (Kim 2007) and therefore the detection of the allelic effects of the causal variants can be hampered. Further, some of the alleles of causal genes can have low frequencies. Low frequency markers are not tested in GWA mapping to avoid false positive detection as a consequence of spurious linkage. Therefore rare causal variants cannot be detected by GWA mapping. In addition, low frequency markers with large effect can mask common causal variants with small effect. Lastly, underlying genetic networks can be different between accessions with the same phenotypic value. For example, within a natural population different mechanisms to deal with drought stress can be present, all having their own genetic regulation. This genetic heterogeneity weakens the detection of causal genes. These three characteristics of natural populations differ from bi-parental populations in which always two alleles segregate in a 50%/50% ratio (as long as no segregating disorders occur due to lethal combinations). Therefore the mapping power to detect causal genes is lower in natural populations than in bi-parental populations. Especially the power to detect rare causal alleles and causal alleles with small effect is substantially lower. So, the success of GWA studies depends on the genetic architecture of the trait, which, for most traits, is *a priori* unknown.

Societal impact of research on stress responses

At the moment two major challenges are greatly influencing the breeding and growing practice of crops. The first challenge is the fast increase of the world population, especially in Asia and Africa, which will lead to an increased demand for food, feed, construction materials, and other plant-derived products. The second challenge is the expected change in the climate that will lead to more extreme and more variable weather conditions. This asks for increased yield on existing farmlands and for expansion of the area used for agriculture by cultivation of crops on less favourable locations (Araus 2008, Tester 2010). In addition, this increase in yield should be secured even when climate conditions change over the years and yearly climatic patterns are distorted (Chaïb 2006) (Mickelbart 2015). These challenges can be dealt with in several ways. A promising approach is the development of new cultivars with improved yield in variable climatic conditions. Knowledge about the natural range of variation for these traits within species and about the underlying genetic networks would be helpful to develop new cultivars.

Short-term heat stress

In the last century, the global temperature has risen with almost one degree Celsius, due to, amongst others, the emission of greenhouse gasses. It is expected that the global temperature will keep rising in the coming years. In addition, more periods of extreme heat are expected, also in temperate climates (Gourdji 2013). Elevated temperatures influence both plant growth and reproduction (Wahid 2007, Bita 2013), but reproductive tissues of plants are more sensitive to short-term heat than vegetative tissues. In tomato reduced fruit set as a consequence of short periods of heat is already a research topic since the beginning of the 20th century (Sugiyama 1965, Smith 2007). Also in several other crops, like cereals (Barnabàs 2008), soybean (Djanaguiraman 2013), and cherry (Hedhly 2004), reduced yield due to short-term periods of heat during the reproductive phase are reported. Heat stress is therefore responsible for great losses in the agro-business sector. Research of these phenomena revealed that during flower development not all processes are equally sensitive to heat. The development of male organs is more heat sensitive than the development of female organs. Upon short-term heat, disruption of the male meiosis and reduction in anther dehiscence (Kim 2001) (Sato 2002), anther length (Sakata 2010), and pollen viability (Endo 2009) are reported. In some crops, such as rice (Ye 2012) and tomato (Grilli 2007), natural variation for the sensitivity to short-term heat was used to perform linkage mapping. Although QTLs were identified, no regulatory genes underlying this natural variation were identified so far. Also in the model species Arabidopsis natural variation for short-term heat stress was described (Warner 2005, Zinn 2010), but the genetic regulation has not been studied yet through gene expression profiling or linkage mapping. A GWA mapping study for this trait, as described in this thesis, could be a great step forward in the understanding of the genetic regulation of the heat response.

Moderate drought

One of the most profound effects expected from the climate change is more irregular rainfall. At the moment a substantial part of the world's food production is done on arable land in temperate climates. It is expected that in the coming years in those regions periods without rainfall will occur more often and will be longer (Fang 2015). During these moderate drought periods crops may survive, but plant growth and yield will be decreased significantly (Skirycz 2011). Breeding has resulted in a limited number of drought tolerant cultivars, but the mechanism behind this tolerance and the causal genes are in most cases unknown (Ashraf 2010, Fang 2015). Research on drought tolerance in the model species Arabidopsis has, for a long time, been focused on the study of responses to severe drought. But both mutant analyses (Skirycz 2011) and gene expression profiling (Harb 2010) revealed that the responses to severe and moderate drought are regulated differently. Therefore in the last years, more attention is given to research of the moderate drought response (Granier 2006, Des Marais 2012, Ma 2014, Clauw 2015). Natural variation for moderate drought responses is reported (Aguirrezabal 2006, Bouchabke 2008)

and linkage mapping for moderate drought related traits was done in two RIL populations (Juenger 2005, Tisné 2010). In addition, gene expression changes upon moderate drought are reported in both vegetative (Harb 2010, Baerenfaller 2012) and reproductive tissue (Ma 2014) for many genes in the genome. These expression changes are partly genotype dependent (Des Marais 2012, Clauw 2015). Moderate drought studies gave important insights into drought tolerance mechanisms, but the knowledge is fragmented. For a refined and comprehensive understanding of the genetic variation in moderate drought responses much more research is required (Fang 2015). Therefore, GWA mapping under moderate drought, as described in this thesis, will add to the reduction of this knowledge gap and has the potential to identify alleles of regulatory genes that increase moderate drought resistance. Orthologues of those genes in crops can possibly be used in breeding programs.

Arabidopsis as model species

In this thesis, I chose to make use of the model species *Arabidopsis thaliana*. Although Arabidopsis does not have any agricultural value, research of this species was and is very valuable for the plant sciences. Knowledge gained in Arabidopsis forms in many cases the basis for experiments performed in other species (Gonzalez 2009). For example, Arabidopsis was the first sequenced plant species (Arabidopsis Genome Initiative 2000) and the functional annotation of its genome is improving each day. When a new species is sequenced nowadays the gene annotation is for a large part based on homology with the Arabidopsis genome. Arabidopsis has outspoken advantages compared to many crop species (Pang 1987). In nature Arabidopsis accessions are found on many locations with large differences in climatic conditions and therefore natural variation within this species is observed for almost all traits (Alonso-Blanco 2000), allowing to perform mapping studies to identify underlying genetic regulation networks. In domesticated plants within species variation is reduced and therefore, at the moment, collections of wild relatives are composed to compensate for this reduction in genetic variation. Because Arabidopsis is a self-pollinator, accessions collected in the wild are almost completely homozygous and as a consequence hardly any segregation is observed in the off-spring. Homozygosity makes it possible to replicate accessions within and between experiments and it allows for exchange of lines by exchange of seeds. The plant is small and has a short lifecycle, enabling the performance of experiments with large sets of plants in relatively short time and small space. Arabidopsis also has a small genome and therefore relatively few genes with redundancy in function and relatively little non-coding DNA. In addition, Arabidopsis is diploid and has only 5 chromosomes whereas many crops have higher ploidy levels and larger numbers of chromosomes. The small genome, the low ploidy level, and the small number of chromosomes are advantageous for re-sequencing, quantitative genetic analysis, and creation of homozygous mutant or overexpression lines. Transformation by floral dipping (Clough 1998) also facilitates the creation of transgenic lines.
In Arabidopsis, the study of plant responses to various stress environments is mostly

done in one natural reference accession: Col-0. Although mutant and overexpression analyses, gene expression studies, and metabolite profiling in this accession resulted in the annotation of many genes, the study of one accession has some limitations. Genes that are not present in this accession or genes with a non-functional Col-0 allele cannot be studied. Furthermore, Col-0 only represents a limited number of response mechanisms to stresses, although within species multiple response mechanisms and large variations in the utilization of these mechanisms are reported. Therefore, exploitation of natural variation present within the species Arabidopsis is the key to the full understanding of the performance of this species in various environments (Alonso-Blanco 2000, Lavagi 2012). The study of experimental or natural populations and subsequent mapping are tools to elucidate the genetic control of many plant traits. Therefore investigation of the natural variation present within the species Arabidopsis and subsequent study of the underlying genetic mechanism by GWA mapping is chosen as the subject of this thesis.

Scope of the thesis

In this thesis I aim at identifying genes that explain natural variation in plant performance under various different environmental conditions. To achieve this, I followed a similar strategy in all chapters: a natural population of 350 accessions of Arabidopsis, genotyped with 215k SNPs, was grown under control and/or stress conditions and the resulting plant performance was evaluated by phenotyping one or several traits. Associations between phenotype and genotype were obtained via GWA mapping. Significant associations were studied in more detail. Based on linkage disequilibrium (LD), functional annotation, expression data, and the literature, genes in the support window were prioritized. Next, the causality of the candidate genes was investigated. Since full validation may be a long process, in most chapters of this thesis only a limited number of steps of the validation process were taken.

Research described in this thesis is part of the STW-funded 'Learning from Nature' research program. This program consists of several research projects that all aim to use natural variation to improve stress resistance of crops. Each research project focusses on specific biotic or abiotic stresses or a combination of both and has specific target crops. Each project consists of two experimental phases. First the same natural population of Arabidopsis is used to investigate variation in tolerance towards the stress of interest. Subsequently, GWA mapping is performed to search for genomic regions associated with stress-tolerance. In the second phase results obtained in Arabidopsis are translated to applications in the target crop. In this second phase the industrial partners play a profound role. The strength of the program is the similar experimental approach of all research projects, which allows easy comparison of results and opens the possibility to distinguish stress-specific from general regulators.

This thesis consists of seven chapters including this introduction (Chapter 1). Chapters 2 to 6 describe the major findings of the research. In Chapter 7 the findings described in previous chapters are compared and evaluated, resulting in general conclusions and remarks.

In **Chapter 2** natural variation in anthocyanin accumulation was studied under osmotic stress conditions. GWA mapping of rosette colour resulted in a very strong association with MYB90, located on chromosome 1. Re-sequence data were used to reveal which allele of MYB90 is associated with higher anthocyanin accumulation and to identify the causal SNP. Subsequently the ecological and evolutionary impact of the sequence variation in MYB90 is discussed.
Chapter 3 describes natural variation in fertility reduction upon heat stress. Measurement of the silique length along the main inflorescence revealed that the heat response is developmental stage specific. GWA mapping was performed to identify QTLs specific for the heat response before and after anthesis. Two QTLs were studied in more detail and knockout lines of several candidate genes were tested to identify the causal genes.

Chapter 4 and 5 describe the study of natural variation in plant growth under control and moderate drought stress conditions. Growth dynamics were followed over time by top-view imaging and growth was modelled, using an exponential function. To identify QTLs involved in growth, GWA mapping was performed on temporal plant size data and on model parameters. Causal genes are suggested based on mutant phenotypes and gene expression data.

Chapter 6 describes multi-environment GWA mapping of plant size data collected in 25 different environments and the identification of more than 100 markers with effects in multiple environments. Each SNP was tested for GxE to distinguish between constitutive and environment specific QTLs, because constitutive QTLs are promising candidates for the translation of findings from Arabidopsis to crops.

Finally, in **Chapter 7** the results in the preceding chapters are discussed in the light of the current status of the research on the genetic regulation of plant performance under different conditions. This results in general conclusions and remarks.

Chapter 2 - GWA mapping of anthocyanin accumulation reveals balancing selection of *MYB90* in *Arabidopsis thaliana*

Johanna A. Bac-Molenaar[1,2], Dick Vreugdenhil[1] and Joost J.B. Keurentjes[2]

[1] Laboratory of Plant Physiology, Wageningen University, The Netherlands
[2] Laboratory of Genetics, Wageningen University, The Netherlands

Abstract

Induction of anthocyanin accumulation by osmotic stress was assessed in 350 accessions of *Arabidopsis thaliana*. A wide range of natural variation, with phenotypes ranging from green to completely red/purple rosettes, was observed. A genome wide association (GWA) mapping approach revealed that sequence diversity in a small 15 kb region on chromosome 1 explained 40% of the variation observed. Sequence and expression analyses of alleles of the candidate gene *MYB90* identified a causal polymorphism at amino acid (AA) position 210 of this transcription factor of the anthocyanin biosynthesis pathway. This amino acid discriminates the two most frequent alleles of *MYB90*. Both alleles are present in a substantial part of the population, suggesting balancing selection between these two alleles. Analysis of the geographical origin of the studied accessions suggests that the macro climate is not the driving force behind positive or negative selection for anthocyanin accumulation. An important role for local climatic conditions is, therefore, suggested. This study emphasizes that GWA mapping is a powerful approach to identify alleles that are under balancing selection pressure in nature.

Submitted to PLoS One

Introduction

Accumulation of anthocyanins in the vacuole of plants is the most clearly visible sign of stress. Anthocyanins have been studied for decades because of their strong color that directly attracts the eye of the observer. They are best known from their appearance during the ripening of fruits (Winefield 2009) and from their accumulation in autumn leaves (Hoch 2001), but they also play a prominent role in many stress responses. Accumulation of anthocyanins upon stress exposure is a conserved and desirable trait. It is found in many plant species upon exposure to a wide variety of biotic and abiotic stresses (Chalker-Scott 1999, Hasdai 2006, Gutha 2010, Park 2012, Sperdouli 2012). Anthocyanins protect plants against stress owing to four properties. First, anthocyanins protect the photosynthetic machinery of plants exposed to an excess of light (especially UV-B) by absorbing blue light and reflecting red light (Zhang 2010). Secondly, anthocyanins possess anti-oxidant properties, which may scavenge reactive compounds known to occur under many stresses, such as reactive oxygen species, and, therefore, increase a plant's tolerance towards stress (Zeng 2010). Thirdly, anthocyanins may bind to various heavy metals, providing a protective role in metal-tolerance by sequestering these toxic elements (Hale 2001). Finally, anthocyanins are assumed to exhibit an osmo-regulatory function, allowing a plant to remain turgid under low water availability (Chalker-Scott 2002, Hannah 2006). Because of these characteristics, induced anthocyanin accumulation is a desirable trait in many environments in which plants are temporarily experiencing stresses.

Anthocyanins are the end products of a specific branch of the flavonoid biosynthesis pathway (Figure S1), (for a review of this pathway see (Winkel-Shirley 2001)). The expression of the structural and regulatory genes of the anthocyanin pathway have been studied in detail in the *Arabidopsis thaliana* accession *Col-0* and for each step in the pathway at least one mutant in the *Col-0* background has been studied (Petroni 2011). These analyses revealed that many structural genes in the pathway are essential for the synthesis of anthocyanins but that functional redundancy exists for transcriptional regulators (Figure S1). Biosynthesis genes early in the pathway are regulated differently than the genes at the end of the pathway (Gonzalez 2008). Early biosynthesis genes, leading to the production of flavonols, are regulated by R2R3-MYB transcription factors (*MYB11*, *MYB12* and *MYB111*). The biosynthesis genes later in the pathway, leading to the synthesis of pro-anthocyanidins and anthocyanins, are regulated by a complex of MYB, bHLH and W40 transcription factor families (Zimmermann 2004). *MYB75* (*PRODUCTION OF ANTHOCYANIN PIGMENT 1*, *PAP1*), *MYB90* (*PRODUCTION OF ANTHOCYANIN PIGMENT 2*, *PAP2*), *MYB113* and *MYB114* display redundant functions and can all be part of this complex (Gonzalez 2008). Overexpression studies of these four MYB transcription factors revealed that the up-regulation of either one of them is sufficient to increase anthocyanin accumulation in young leaves and upon osmotic stress (Borevitz 2000, Gonzalez 2008). In addition to these transcription factors, many other structural and regulatory genes of the pathway are up-regulated in the reference accession Col-0

under various stress conditions (Kimura 2003, Vanderauwera 2005, Solfanelli 2006, Sperdouli 2012). Moreover, a few studies report on variation in constitutive or stress induced anthocyanin accumulation in different accessions of Arabidopsis (Hannah 2006, Hasdai 2006, Atwell 2010, Schulz 2015). For *MYB75* (*PAP1*) and *MYB90 (PAP2),* sequence variation has been reported to be causal for natural variation in anthocyanin accumulation (Teng 2005, Diaz 2006, Ilk 2015). Although these studies identified natural variation in the ability to accumulate anthocyanins and even detected causal genes explaining the observed variation, little insight was gained on their role in natural populations. Genetic analyses of recombinant inbred lines (RILs), for instance, provides no information about the prevalence of allelic differences or on the selective forces that may shape the genetic architecture of populations.

Sequence diversity that causes differences in anthocyanin accumulation will be subject to natural selection. Depending on the environment, low or high constitutive anthocyanins levels and strong or moderate induction of anthocyanin production will be favored. Here we investigated natural variation in constitutive and induced anthocyanin accumulation in 350 natural accessions of Arabidopsis. GWA mapping revealed the transcription factor *PAP2* (*MYB90*) as the major gene responsible for the observed variation in induced anthocyanin accumulation. Two alleles of this gene dominate the world-wide population, suggesting balancing selection for the polymorphism discriminating these two alleles. No link could be found between the origin of collection of the accessions and the amount of constitutive or induced anthocyanin accumulation. These results suggest that not the macro climate, but the local climate conditions might be the driving force behind positive or negative selection for anthocyanin accumulation.

Results

Accumulation of anthocyanins in the leaves can be favorable for a plant because of their light-shielding, metal-binding and antioxidant capacity, and their function in osmotic-regulation. However, this process is also energy-demanding and, therefore, not favorable under all conditions. In this study we investigated the natural variation in constitutive and induced anthocyanin accumulation and used a GWA mapping approach to search for regulatory genes that are underlying the natural variation in anthocyanin accumulation.

Anthocyanin accumulation is genotype dependent

Anthocyanin accumulation for 360 natural accessions of Arabidopsis was evaluated by visual scoring of rosette-stage plants grown under control conditions or osmotic stress conditions. Each plant was assigned to a category ranging from a completely green plant (category 1) to a completely red/purple plant (category 5) (Figure 1A). These categories

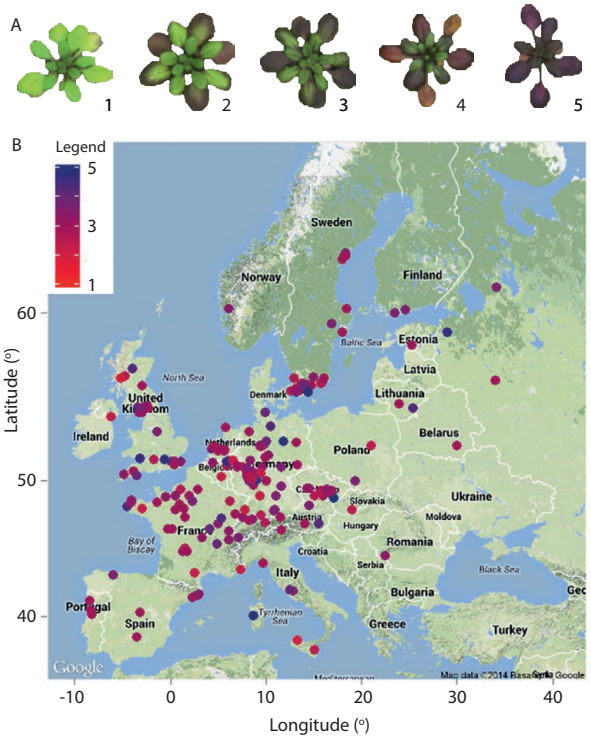


Figure 1 A Representative examples of categories of anthocyanin accumulation. Anthocyanin accumulation was scored visually by assigning each individual plant to an accumulation category. **B** Geographical distribution of the accessions in Europe. The color indicates the anthocyanin accumulation category in stress conditions.

correlated very well with anthocyanin content as determined by spectrophotometric analyses (Figure 2). Most plants (66%) did not produce any visible anthocyanins under control conditions and only very few accumulated substantial levels (Figure 3). PEG treatment, however, induced anthocyanin accumulation in almost all accessions, with most plants being assigned to category 3 and more than 98% to category 2 or higher (Figure 3A). Levels of accumulation were highly reproducible among replicates of accessions within and between experiments. Illustratively, a high correlation was observed between anthocyanin accumulation in the association mapping experiment and the confirmation experiment (Spearman's ρ=0.93, p=0.04*10^{-11}). The broad sense heritability of anthocyanin accumulation under control conditions was 35% and 47% under stress conditions, indicating moderate to strong genetic control. A weak, but significant correlation was detected between anthocyanin accumulation under control and treatment conditions (Spearman's ρ=0.19, p=0.00033), indicating that anthocyanin formation depends partly on the genetic

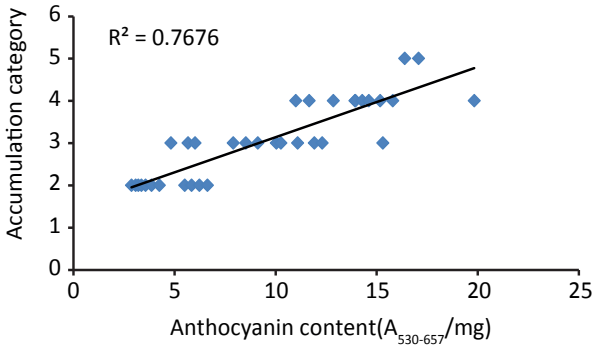


Figure 2 Linear correlation between accumulation category assigned to plants by visual scoring and anthocyanin content quantified spectrophotometrically.

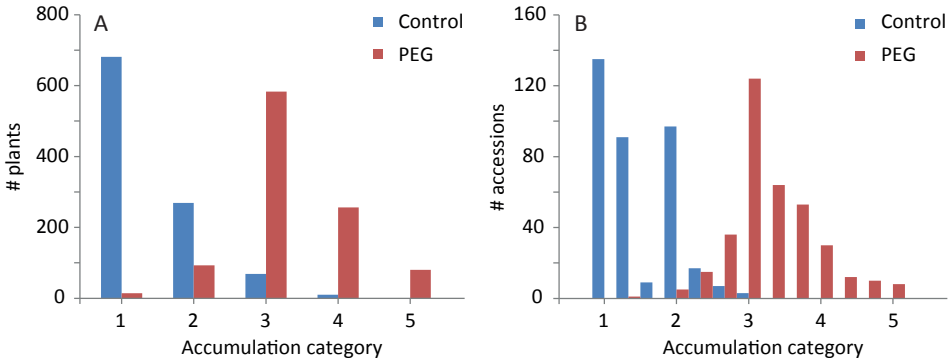


Figure 3 Frequency distribution over the scoring categories. In total 2160 plants were scored, three replicates of each of the 360 accessions in control and in PEG treated conditions. A Frequency distribution based on number of plants. B Frequency distribution based on number of accessions.

background, independent of the environment. One accession (Pa-1, CS76204) responded to the osmotic stress with signs of early senescence. The leaves turned yellow without detectable levels of anthocyanin.

Evolution of anthocyanin accumulation is independent of macro-climatic differences

To identify geographical clines in the ability to accumulate high levels of anthocyanins, the origin of collection of all accessions tested was analysed. Most of the accessions (321 out of 360) originate from Europe. To avoid bias due to low observation numbers in other continents, only the European accessions were used to identify patterns in the distribution of the different accumulation categories. Correlations between anthocyanin accumulation and latitude and longitude were calculated. A weak significant correlation was found between the latitude and the accumulation of anthocyanins under control (ρ=0.15, p=0.006) and stress (ρ=0.13, p=0.023) conditions. This suggests a trend of northern accessions accumulating more anthocyanin than southern accessions (Figure 1B). However, as can be expected from the weak clines, accessions with different accumulation phenotypes often occur in close range of each other. In the United Kingdom, for example, accessions of the whole range of accumulation under both control and stress conditions occur, indicating that macro climate is not the most important factor determining selection of constitutive and stress-induced anthocyanin accumulation.

GWA mapping reveals a strong association on chromosome 1

In order to detect causal loci for the variation in anthocyanin accumulation, GWA mapping was applied. This resulted in the detection of a highly significant association of 12 SNPs with the accumulation of anthocyanins induced by PEG (Figure 4B and Table S2). This locus spans a region of 15 kb on chromosome 1 (Figure 4C). Three genes are located in this region, viz., a calmodulin (AT1G66400) and two transcription factors of the MYB-family, *MYB114* (AT1G66380) and *MYB90* (*PAP2*, AT1G66390). Both

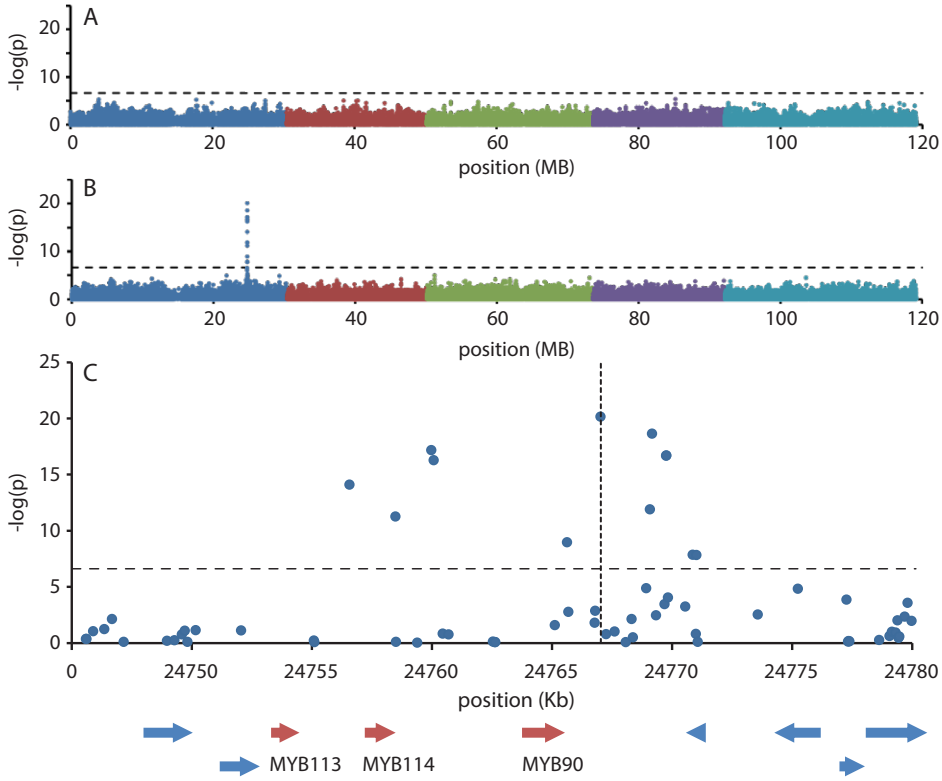


Figure 4 Outcome of GWA for anthocyanin accumulation under control (**A**) and stress conditions (**B**). A Bonferroni-corrected threshold of α = 0.05 is represented by the dashed line. **C** Blow-up of the region around the significantly associated SNPs. All genes positioned in the region are represented by arrows (red color: genes known to play a role in anthocyanin accumulation).

transcription factors have been annotated to play a role in anthocyanin accumulation (Gonzalez 2008). A third transcription factor (*MYB113* (AT1G66370), also annotated to play a role in anthocyanin biosynthesis (Gonzalez 2008), is located just at the border of the associated region. The most significant SNP (–log(p-value)= 20.19) explained 43% of the variance and contributed an effect size of 0.91. The latter indicates that plants of the Columbia haplotype were on average categorized almost one class higher than plants of the non-Columbia haplotype. Genome wide association mapping of anthocyanin accumulation under control conditions did not lead to any association above the Bonferroni corrected significance threshold (Figure 4A), indicating that the chromosome 1 locus does not play a major role in constitutive anthocyanin accumulation. To discover whether the strong association on chromosome 1 masked other weak associations, the most significant SNP (chr1, position 24769177) was used as a cofactor in the mixed model analysis (Segura 2012, Seren 2012). The conditional GWA mapping, however, did not result in additional SNPs above the Bonferroni corrected threshold.

Anthocyanin accumulation is not determined by the level of gene expression of MYBs

Natural variation can act on the level of transcription by modifying promoter regions or on the effectiveness of protein function by modifying coding regions. To assess whether any of the assigned candidate genes displayed expression variation that corresponds to differences in anthocyanin accumulation, the three MYB genes at the associated locus on chromosome 1 were subjected to qPCR analysis on plants grown under control and stress conditions. The analysis included the phylogenetically related *MYB75* transcription factor (*PAP1*, At1G56650)(Stracke 2001), which was previously shown to explain variation in anthocyanin accumulation in a recombinant inbred line (RIL) population (Teng 2005). All genes (*MYB90*, *MYB113*, *MYB114* and *MYB75*) are part of the largest class of MYB-transcription factors, the R2R3-class, and have high similarity in the overall sequence, particularly the DNA binding domains (Stracke 2001). For all accessions tested, all genes were up-regulated upon osmotic stress, except *MYB114,* which was not up-regulated in Ty-0. This up-regulation confirms the role of these genes in stress responses. However, for none of the genes a correlation was found between the observed anthocyanin accumulation and gene expression levels, although natural variation for anthocyanin accumulation and gene expression was detected both under control and stress conditions (Figure S2). Strikingly, *MYB90*, the candidate gene in closest proximity of the associated SNPs, showed the strongest response to the stress treatment in terms of expression, but this did not correlate with the amount of anthocyanins accumulated. These observations suggest that the causal mechanism for variation in anthocyanin accumulation is not transcriptional modification but more likely resides in functional changes.

Analysis of re-sequence data reveals a QTN in MYB90

To gain more insight into the causal sequence variation of the MYB transcription factors, the nucleic acid and translated amino acid sequences of the phenotyped accessions for which re-sequence information was available (157 out of 360) were aligned. Because no correlation was found between the up-regulation of gene expression upon osmotic stress and the observed phenotype, the focus here was on the protein sequences and only information of the exons was taken into account (Figure S3 and Table 1). As expected, the coding sequences of the MYB family members are highly conserved, especially in the R2-R3 DNA binding domain. For each transcription factor, sequence diversity was observed with other MYB family members and also within the same gene between accessions. Because the lowest coding sequence variation was observed for *MYB90*, this gene appears to be the most conserved among the different accessions. Within the coding sequences of *MYB90* only ten mutations occurred of which 4 were non-synonymous (Table 1), whereas between 27 and 41 mutations, of which between 22 and 26 were non-synonymous, were observed for the other three MYB genes. Based on the observed non-synonymous mutations, different alleles were defined for each of the proteins. In line with the observed mutation frequencies, allelic diversity was lowest for *MYB90,* with five different alleles,

Table 1 Statistics of the sequence diversity of MYB75, MYB113 and MYB114, MYB90 and results of tests for deviations from neutrality of selection of Tajima ((Tajima 1989) and Fu and Li ((Fu 1993). SNP=single nucleotide polymorphism; mutation: this number deviates from the number of SNPs if at one position more than two different nucleotides are observed; syn=synonymous, non-syn= non-synonymous. Significance is indicated by * (α=0.05) and & (α=0.10).

		SNPs			Fu and Li		
gene	mutations	syn	non-syn	Tajima's D	D*	F*	# of alleles
MYB75	36	13	21	-1.32	-2.00$^{\&}$	-2.02$^{\&}$	15
MYB113	41	13	25	-2.05*	-3.16*	-3.00*	19
MYB114	27	4	22	-1.18	-2.17$^{\&}$	-2.23$^{\&}$	18
MYB90	10	6	4	-1.14	-6.94*	-5.96*	5

whereas for the other three MYB genes, *MYB75*, *MYB113* and *MYB114*, two to three times more alleles were observed, 15, 19 and 18 alleles respectively (Table 1). For all four genes, the frequency distribution of alleles was skewed, with only a few alleles of each transcription factor dominating the population (Figure S4). Skewed allele distributions were expected because most non-synonymous mutations will have negative effects and therefore will not be maintained in the population. In many cases the most favorable allele will dominate a population but in case of balancing selection or substitution by a newly arisen allele, i.e. a selective sweep, more than one allele can be present in a substantial part of the population. For *MYB75* clearly one allele is dominating with the most frequent allele being present in 77% of the population and the second most frequent allele in only 5% of the population. All of the other 12 alleles have allele frequencies below 5%. For the other three MYBs at least two alleles with allele frequencies above 10% occur, indicating balancing selection or allele substitution. The difference between the two most frequent alleles of the three MYBs is only one amino acid and no signs of genetic hitch-hiking were observed, indicating that the discriminating SNPs have been maintained in the population for a long time. Therefore, balancing selection is expected for *MYB90*, *MYB113* and *MYB114*. However, balancing or positive selection was not supported by Tajima's D or Fu and Li's D* or F* but weak purifying selection was suggested by both tests (Table 1). To test for significant differences in anthocyanin accumulation between allelic haplotypes, analysis of variance (ANOVA) was performed. Alleles with frequencies below 2% were grouped and tested against alleles with higher frequencies. Significant differences were found between alleles of *MYB*90, *MYB113* and *MYB114*, but not of *MYB75*. Subsequent pair-wise comparisons revealed significantly higher anthocyanin accumulation for the second most frequent allelic haplotype of *MYB90* compared to all other alleles (Figure 5B, allele frequency = 17%). In addition, also higher anthocyanin accumulation was observed for the second allele of *MYB114* (Figure S5B, allele frequency = 17%). This difference is significant in pairwise comparisons with most other alleles, including the first, third and fourth most frequent alleles representing 64% of the variation for *MYB114*. However,

A

Position		Alleles						
Nucleo-tide	Amino Acid	1st 122	2nd 26	3rd 5	<2% 1	<2% 1	Codon	Amino Acid
403	135	C	C	A	C	C	Caa/Aaa	Gln/Lys
552	184	C	C	C	G	C	aaC/aaG	Asn/Lys
603	201	T	T	T	T	G	aaT/aaG	Asn/Lys
626	209	A	G	A	A	A	gAg/gGg	Glu/Gly

B

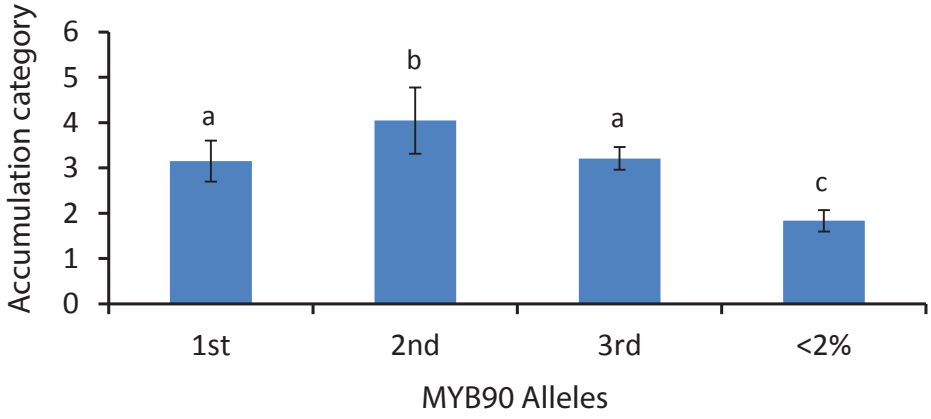


Figure 5 Genotypic and phenotypic differences between alleles of MYB90. **A** Positions of non-synonymous SNPs and corresponding amino acid changes defining the alleles. Number of accessions carrying the allele are indicated above the double line. **B** Anthocyanin accumulation per allele (mean +/- SD). Means per allele are compared by ANOVA following pairwise comparison using Bonferroni corrected significance threshold of α=0.05. Letters above the bars indicate significant differences.

the significance threshold was not reached for three alleles which together account for 11% of the variation for *MYB114*. Not surprisingly, allele MYB90-2 and MYB114-2 are discriminated by the SNPs that were in strong linkage disequilibrium (LD) with the association peak (MYB90-2, LD=0.70; MYB114-2, LD=0.65). Allele MYB90-2 is the only allele that contains a glycine instead of a glutamic acid at AA position 210 (Figure 5A), whereas allele MYB114-2 is the only allele carrying a stop-codon at AA position 141 and the protein is, therefore, truncated (Figure S5A). All other *MYB114* alleles do not contain this pre-mature stop-codon and their encoded proteins have the same size as the other MYB-protein family members. Overexpression of allele MYB114-2, resulting in truncated proteins, did not result in increased anthocyanin accumulation (Gonzalez 2008). The SNP discriminating this allele from the other alleles is, therefore, most probably not causal for the detected association. The two discriminating polymorphisms in *MYB90* and *MYB114* are highly linked (LD=0.93). All accessions, except two, that contain allele

MYB114-2 also contain allele MYB90-2. For *MYB113*, significant differences were observed between the first two alleles, but both are not significantly different from all other alleles. The SNP that is discriminative between those two alleles is also polymorphic in other alleles of *MYB113* and can, therefore, not explain the strong association detected by GWA mapping. Interestingly, significant lower anthocyanin accumulation was observed for the two alleles of *MYB90* with low (<2%) allele frequency. These two alleles were present in two accessions, Kas-2 and Per-1, of which the very low anthocyanin accumulation upon osmotic stress was confirmed in the second experiment used to determine gene expression. This suggests that some rare mutations in *MYB90* or its cis-regulatory regions affect anthocyanin accumulation more severe than the mutation at position 210. Possibly, such severe reduction in induced anthocyanin accumulation is not favorable in any environment and therefore these alleles do not occur at higher allele frequencies and are purged from the population. In conclusion, because differences in gene expression could not explain the observed phenotype and given the allelic differences observed, the polymorphism in *MYB90,* discriminating the second allele from the rest, seems to be causal for the associations detected in GWA mapping. However, the SNP discriminating the first two alleles of *MYB114* was found to be linked to both the polymorphism in *MYB90* and the associated SNP of the GWA mapping. This tight linkage cannot rule out a joint role in the natural variation in anthocyanin accumulation for allele MYB114-2 and allele MYB90-2 completely.

Discussion

Natural variation in MYB90 is causal for anthocyanin accumulation

Variation for constitutive and osmotic stress induced anthocyanin accumulation was observed within natural accessions of Arabidopsis. Osmotic stress induced accumulation resulted in extreme variation, ranging from completely green to completely red/purple plants. Using a GWA mapping approach, a region on chromosome 1 explaining more than 40% of the observed variation was identified. Analysis of the alleles of three MYB-transcription factors in this region, known to be involved in anthocyanin biosynthesis, allowed us to identify two SNPs, one in *MYB90* at AA position 210 and one in *MYB114* at AA position 141, which were discriminative for gene alleles that were associated with higher anthocyanin accumulation. The biosynthesis pathway of anthocyanins reveals that these MYB transcription factors upregulate the genes in the last part of the anthocyanin biosynthesis pathway (Figure S1, (Petroni 2011)). Simultaneous knock-down by RNAi of all four MYB transcription factors involved in the later steps of the pathway results in anthocyanin deficiency under stress (Gonzalez 2008). Overexpression studies of the four MYBs revealed that the up-regulation of either one of them is sufficient to increase anthocyanin accumulation in young leaves under control conditions and upon osmotic stress treatment (Borevitz 2000, Gonzalez 2008). *MYB114* is truncated in Col-0, due to a

mutation at AA position 141, which introduces a pre-mature stopcodon. Overexpression of the truncated protein, corresponding to allele MYB114-2, does not lead to increased anthocyanin accumulation, whereas overexpression of the full protein, corresponding to allele MYB114-1, results in higher levels of anthocyanins (Gonzalez 2008). This observation is in contrast with our results that indicate that the truncated allele MYB114-2 is associated with higher anthocyanin accumulation and allele MYB114-1 with lower accumulation. We, therefore, postulate that the allelic differences in *MYB114* are not causal for the observed variation in anthocyanin accumulation in wild accessions. This leaves the allelic variation in *MYB90* as the cause to explain stress-induced anthocyanin accumulation in natural accessions. Because the causal polymorphism is located at the end of the protein outside the DNA-binding domain we speculate on an enhanced functional activity of the modified protein. However, because of the tight linkage between *MYB114* and *MYB90* we cannot rule out a genetic interaction between the two transcription factors.

GWA mapping in natural populations reveals evolutionary significance of variation

GWA mapping and subsequent analysis of alleles of candidate genes using re-sequence information of a large number of accessions resulted in the identification of a single gene (*MYB90*) to be responsible for most of the variation present in anthocyanin accumulation in natural populations. Classical QTL mapping studies in three different biparental experimental populations also reported natural variation in anthocyanin accumulation induced by stress (Teng 2005, Diaz 2006, Ilk 2015). In one of these studies, anthocyanin accumulation in seedlings of a RIL population derived from a cross between the accessions Landsberg *erecta* (L*er*) and Cape Verde Islands (Cvi) was induced by high sucrose levels (Teng 2005). In another study, RILs derived from a cross between the accessions Bayreuth-0 (Bay-0) and Shahdara (Sha) were grown under low nitrogen conditions and as a result displayed increased levels of anthocyanins (Diaz 2006). Finally, variation for anthocyanin accumulation was observed upon cold acclimation in the L*er* x Eri-1 population (Ilk 2015). These three QTL studies detected a major QTL on chromosome 1 and sequence variation in MYB75 was proven to result in differences in anthocyanin accumulation (Teng 2005, Ilk 2015). It is suggested that besides MYB75 also MYB90 could contribute to this major QTL (Ilk 2015). GWA mapping of our data did not detect the slightest association of *MYB75* with anthocyanin accumulation. The different results of these two approaches (i.e. experimental vs natural population mapping) poses questions about the biological relevance of the identified variants and the impact on selection processes in natural populations.

Available re-sequence data identified the causal GWA SNP at AA position 210 of *MYB90*. This SNP substituted a glycine for glutamic acid in 17% of the accessions analysed. Because this is the only SNP discriminating allele MYB90-2 from all other alleles, it has probably

been present in the population for a long time, suggesting balancing selection between the haplotypes with the glycine and haplotypes with the glumatic acid residue. However, the accessions L*er*, Cvi, Bay-0 and Sha all belong to the glutamic acid haplotype and hence this locus is not segregating in the RIL populations derived from these lines and no QTL could be detected due to the lack of variation in *MYB90*. Re-sequence data of Eri-1 are not available, but it was suggested that this accessions has a different haplotype than L*er* (Ilk 2015).

In contrast, allele specific analysis of *MYB75* revealed that 77% of the natural accessions share the most common allele while the other 23% are represented by 14 different alleles with allele frequencies of 5% or less. This suggests that variation in *MYB75* is the result of random independent mutation events without a clear selection for, or against, a specific allele. Unfortunately, re-sequence data for *MYB75* of Sha and Bay-0 are incomplete with missing data at many positions and Eri-1 data are not available yet. However, L*er* and Cvi each carry a unique allele and accumulate anthocyanins at high and low levels, respectively. Segregation at this locus, therefore, results in the detection of a strong QTL at this position in the *Ler* x Cvi RIL population. GWA mapping failed to detect *MYB75* as a causal locus explaining variation in anthocyanin accumulation because of the high number of rare alleles, of which many may be functionally indifferent, at this locus. The power to detect a rare allele using GWA is rather low especially if the population size is limited. In addition, rare alleles will be highly linked to other rare alleles resulting in high false discovery rates (Korte 2013). As a consequence SNPs with low allele frequency are excluded in most GWA mapping studies. RIL populations are ideal to find such rare causal alleles, but only a very limited amount of the variation that can be found in nature can be tested (Bergelson 2010). In contrast, GWA mapping assesses most variation present in nature and will detect causal genes that are relevant in natural settings because different alleles are present in a substantial part of the population.

MYB75 and MYB90 cannot account for all observed variation

The almost normal distribution of the values of stress-induced anthocyanin accumulation in the natural population indicates that this trait is complex, meaning that the expression and regulation of several genes together result in the observed phenotype. In our GWA study only one region was found to associate with induced anthocyanin accumulation. The causal polymorphism could explain 40% of the phenotypic variation, corresponding to one accumulation category difference (Table S2). However, differences larger than one accumulation category were observed and confirmed between natural accessions. The other 60% of the phenotypic variation may be caused by random variation, but also by genes with small effects or by alleles with allele frequencies lower than 5% that were excluded from the analysis. Clear examples are accessions Kas-2 and Per-1, that have a unique allele of *MYB90* and both show strong deficiency in anthocyanin accumulation upon osmotic stress. That anthocyanin accumulation is not a monogenic trait is also confirmed in QTL mapping studies. Besides the major QTL on chromosome one, also

other small (Teng 2005, Diaz 2008) or large effect (Ilk 2015) QTLs were detected in the bi-parental studies. This suggests that besides the two key-players, *MYB90* and *MYB75,* most probably more regulatory genes are involved, which have a small effect size compared to *MYB90* and *MYB75*.

Allele frequencies of MYB90 are explained by local adaptation

Anthocyanin accumulation can be induced by many different stresses and in many cases anthocyanins have been shown to help plants cope with stress (Hasdai 2006, Jiang 2007, Gutha 2010, Sperdouli 2012). It is reported that anthocyanins protect against high light and ROS (Zeng 2010, Zhang 2010), that they can bind to metals (Hale 2001) and that they function in regulating the osmotic potential within plant tissues (Chalker-Scott 2002). Because of these functions, the production of anthocyanins as a response to stress is a beneficial trait for plants in many climates. However, anthocyanin production is an energy-demanding process and therefore strong natural selection against their accumulation is expected (Mitchell-Olds 2010). The large natural variation observed in this study suggests that anthocyanin production is part of the local adaption process. Allele MYB90-2, which was associated with higher anthocyanin accumulation, is present in almost 17% of the population. The presence of two alleles at high frequency in the world-wide population suggests balancing selection. Allele MYB90-2 might be favored in some environments, whereas the other alleles might be favored in other environments. The large geographical distribution of the different alleles also suggests that the selection pressure for stress-induced anthocyanin accumulation is not specific to one particular environment but widespread.

Interestingly, anthocyanins not only accumulate in response to stress but are also produced in plants grown under control conditions. Even though the accumulation of anthocyanins was much lower under control conditions, variation between accessions was consistent. For almost 100 accessions, 25% of the population, all three replicates were assigned to accumulation category two, whereas for 135 accession, 37% of the population, no anthocyanin accumulation was observed in any replicate. Constitutive production of anthocyanins can be beneficial in local climates where stress occurs on regular basis e.g. daily or annually, and therefore can be anticipated. For example, in plants grown in equatorial regions with high UV-B levels (Hoch 2001) or in winter-annuals that have to cope with low temperatures in autumn and winter (Hannah 2006). However, no association with variation in anthocyanin accumulation under control conditions was detected, suggesting that natural selection for constitutive anthocyanin production does not act on a limited number of genes or has evolved independently in many accessions. A very weak correlation was found between the accumulation of anthocyanins under control and stressed conditions. This further suggests that the selection criteria and targets for constitutive and induced accumulation are different.

For both control and stress conditions, a very weak correlation was found between anthocyanin accumulation in different accessions and the latitude of their origin of collection. Further comparisons of the geographical distribution did not reveal additional geographical trends. This is in contrast with a study of anthocyanin accumulation during autumn leaf senescence in deciduous species, in which a clear latitudinal cline, correlated with UV-B radiation, was observed (Hoch 2001). This emphasizes the fact that anthocyanins have many different functions within and between species, and therefore may be positively selected for in several different environments.

Conclusion

This GWA study reveals that sequence variation in a small 15 kb region on chromosome 1 is the basis of a large part of the observed natural variation in anthocyanin accumulation induced by osmotic stress. Using re-sequence data and expression studies, the causal SNP was identified at amino acid position 210 of *MYB90*. The two most frequent alleles of *MYB90* differ only at this position. These two alleles dominate the population and in less than 5% of the population another allele was observed. This suggests balancing selection between the two alleles. Analysis of the geographical origin of the studied accessions suggests that the macro climate is not the driving force behind positive or negative selection for induced and constitutive anthocyanin accumulation. We assume, therefore, that local climate differences are determining the allelic constitution of natural accessions. These results emphasize that GWA mapping is very powerful in finding sequence variation that plays a major role in natural selection.

Materials and Methods

Plant materials
For association mapping a collection of 360 accessions selected to capture most of the genetic variation present within the species *Arabidopsis thaliana* was used (Baxter 2010, Li 2010, Platt 2010). Each accession is genotyped with approximately 215k SNP markers (Col vs non-Col) (Kim 2007, Atwell 2010). For confirmation of results a selection of 12 accessions from this collection was grown representing the whole range of anthocyanin accumulation: cs76210, cs28786, cs76150, cs28014, cs28279, cs28564, cs22689, cs76113, cs76297, cs28099, cs28420, cs76141.

Experimental design
Association mapping: Plants were grown in six replicate blocks. Each accession was present once in each block. Within each block the plants were grown in nine trays each containing 40 plants. Within a tray the plants had a fixed position. The nine trays were positioned randomly within a block.
In a second experiment the plants were grown in four trays each containing 40 plants. Each tray was treated as a block. In each block all twelve accessions were replicated at least 3 times in a randomized design.

Growth conditions
Seeds were sown in petri-dishes on wet filter paper. After four days of cold treatment, they were placed in the light at room temperature for 1.5 day to germinate. Germinated seeds were placed on rockwool blocks saturated with nutrient solution (Hyponex, 1mM N, 1.1 mM P, 5.9 mM K). Three times a week the perforated trays containing the rockwool blocks were placed in nutrient solution for approximately five minutes. For the osmotic stress treatment, the trays were fed with nutrient solution containing 0.1 g/ml poly ethylene glycol (PEG8000) on day 8, 11, 13 and 15. Control blocks were fed with nutrient solution only. For the association mapping experiment, three blocks received PEG treatment and three blocks received control treatment. In the confirmation experiment, three blocks received PEG treatment and 1 block received control treatment.

Quantification of anthocyanin accumulation and RNA extraction
On day 27 (GWA experiment) or day 21 (confirmation experiment), each plant was scored visually for anthocyanin accumulation. Five categories were considered, ranging from 1 (totally green plant) to 5 (totally red/purple plant) (Figure 1A). Frozen rosette tissue was grinded with a mortar and pestle. RNA was extracted from 10-15 mg tissue (Oñate-Sánchez 2008). Anthocyanins were extracted according to the procedure described by (Neff 1998). In brief, 50-200 mg tissue was incubated overnight in 300 µl of methanol containing 1% fuming HCl at 4°C in the dark. Thereafter 200 µl MilliQ water and 500 µl chloroform were added. After centrifugation (14,000*g*, 5 min) 400 µl of the aqueous top layer was taken and 400 µl 60% methanol containing 1% HCl was added. Anthocyanin content was determined spectrophotometrically as the difference between absorbance at 530 nm and 657 nm.

cDNA synthesis, qPCR and data analysis
For plants of the confirmation experiment, gene expression of *MYB75* (*PAP1*), *MYB90* (*PAP2*), *MYB113* and *MYB114* was determined by qPCR in control and PEG-treated plants. Three biological and 2 technical replicates were analysed. RNA concentration was determined using Nanodrop ND-1000 (Nanodrop Technologies Inc.) cDNA was synthesized from 500 ng RNA using iScript™ cDNA Synthesis Kit (Biorad). cDNA samples were diluted 10× with sterile milliQ water. For each qPCR, 2.5 µl of cDNA, 5 µl of iQ SYBR Green Supermix (Bio-Rad) and 0.25 µl of primer mix (from a 10 µM working solution of forward and reverse primer) was added and supplemented with water to a final volume 10 µl. The RT–qPCRs were run on a CFX96 (Bio-Rad). The qPCR program consisted of a first step at 95°C for 3 min and afterwards 40 cycles alternating between 15 s at 95°C and 30 s at 60°C. Routinely, a melting curve analysis was performed after the qPCR run (between 55 and 95°C with 0.5°C increments for 10 s each). AT1G18610 and AT1G13320 were used as reference genes (Dekkers 2012). See Table S1 for primer sequences. Primer efficiency was determined from the linear part of the amplification curves using the program LinregPCR, (Ruijter 2009). Relative expression values were determined using the program qBase+ (Biogazelle).

Geographical distribution
The latitude and longitude of the place of origin of all accessions was obtained from the ABRC website.

Statistical analysis
GWAS was performed using the EmmaX software package, based on (Kang 2010). A mixed model was used which corrects for population structure based on the identity-by-state matrix (also called kinship matrix) of all SNPs. SNPs with an allele frequency lower than 0.05 were excluded from the analysis. Significance was determined using the Bonferroni correction for multiple testing. Conditional GWAS was performed using the on-line GWApp tool (Seren 2012). The most significant SNP (chr1, position 24769177) was used as co-factor. A significance threshold of α=0.05 was used after Bonferroni-correction for multiple testing. Broad-sense heritabilities were calculated as $H^2= V_g/(V_g+V_e)$ where V_g is the between accessions variation and V_e is the within accessions variation. Spearman's rank correlation coefficient (ρ) (2-tailed, α=0.05) (SPSS, (IBM Corp 2010)) was used to determine correlation between

data series.

Sequence analysis and definition of alleles
The sequences of *MYB75*, *MYB113*, *MYB114* and *MYB90* were extracted from the 1001 genome browser for all accessions for which data were available (157 accessions, for list see S1). For all further analyses only the exon-sequences were used. If an accession contained missing data in an exon it was removed from the analysis of the corresponding protein. For each MYB transcription factor, alleles were defined based on non-synonymous SNP diversity. The most common allele was classified 1st , the second most frequent 2nd etc. Deviations from neutrality were tested by calculating Tajima's D (Tajima 1989) and Fu and Li's D* and F* (Fu 1993) using DNASP software, version 5 (Librado 2009).

Supplementary data

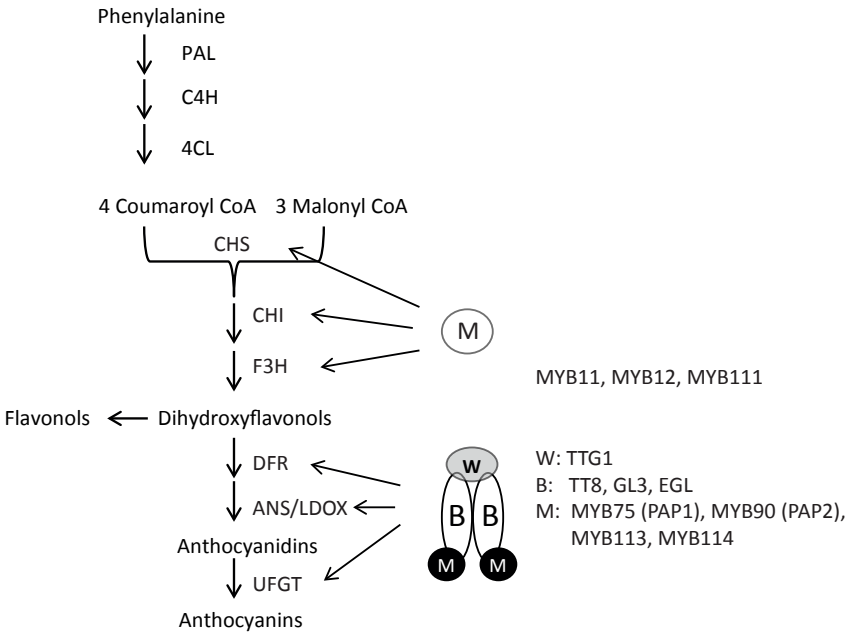


Figure S1 Biosynthesis pathway of anthocyanins and flavonols (simplified from (Petroni 2011)). The early biosynthesis genes are regulated by MYB transcription factors. The late biosynthesis genes are regulated by a complex of two MYB-, two bHLH- and one W40-transcription factor. Enzyme and gene abbreviations are as follows: PAL, phenylalanine ammonia lyase; C4H, cinnamic acid 4-hydroxylase; 4CL, 4 coumarate CoA ligase; CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; DFR, dihydroflavonol reductase; FLS, flavonol synthase; ANS/LDOX, anthocyanidin synthase/leucoanthocyanidin dioxygenase; UFGT, UDP-flavonoid glucosyl transferase; MYB11, At3g62610; MYB12, At2g47460; MYB111, At5g49330; TTG1, At5g24520; TTL8, At4g09820; GL3, At5g41315; EGL, At1g63650; PAP1, MYB75, At1g56650; PAP2, MYB90, At1g66390; MYB113, At1g66370; MYB114, At1g66380

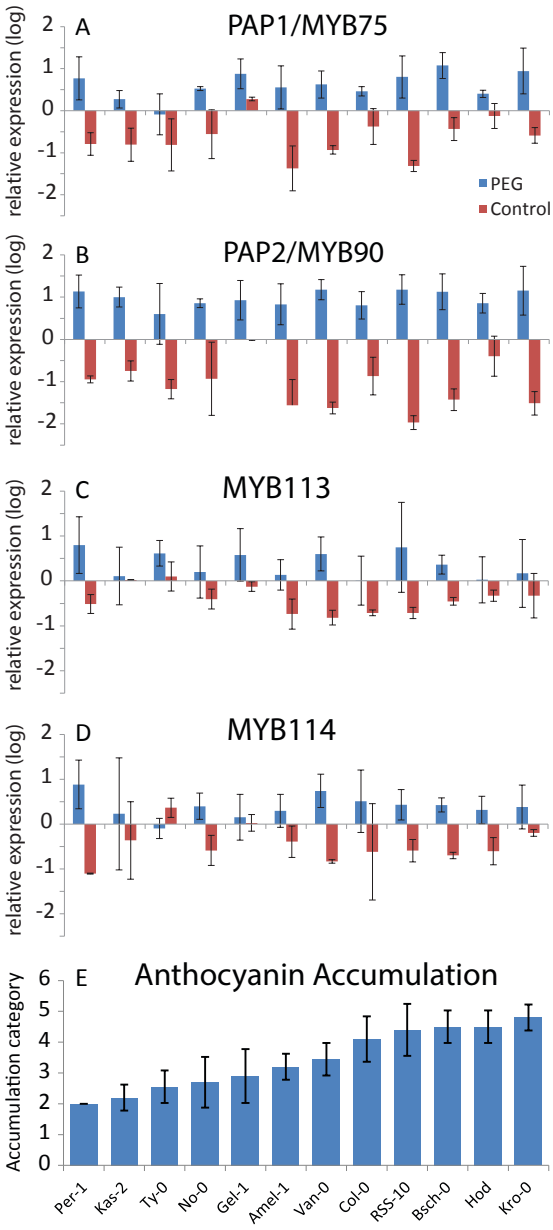


Figure S2 Relative expression of 4 MYB-transcription factors in control and PEG-treated plants. For each gene, the values are relative to the average and are normalized using the expression of two reference genes. Error bars represent standard deviations. Accessions are ordered from low to high anthocyanin accumulation upon osmotic stress.

```
MYB75      1   MEGSSKGLRK GAWTTEEDSL LRQCINKYGE GKWHQVPVRA GLNRCRKSCR LRWLNYLKPS
MYB113     1   MGESPKGLRK GTWTTEEDIL LRQCIDKYGE GKWHRVPLRT GLNRCRKSCR LRWLNYLKPS
MYB114     1   MEGSSKGLRK GAWTAEEDSL LRQCIGKYGE GKWHQVPLRA GLNRCRKSCR LRWLNYLKPS
MYB90      1   MEGSSKGLRK GAWTAEEDSL LRLCIDKYGE GKWHQVPLRA GLNRCRKSCR LRWLNYLKPS
                                                R2

MYB75     61   IKRGKLSSDE VDLLLRLHRL LGNRWSLIAG RLPGRTANDV KNYWNTHLSK KHE~PCCKIK
MYB113    61   IKRGKLCSDE VDLVLRLHKL LGNRWSLIAG RLPGRTANDV KNYWNTHLSK KHDERCCKTK
MYB114    61   IKRGKFSSDE VDLLLRLHKL LGNRWSLIAG RLPGRTANDV KNYWNTHLSK KHE~PCCKTK
MYB90     61   IKRGRLSNDE VDLLLRLHKL LGNRWSLIAG RLPGRTANDV KNYWNTHLSK KHESSCCKSK
                                                R3

MYB75    121 MKKRDITPIP TTPALKNNVY KPRPRSFTVN NDCNHLNAPP KVDVNPPCLG LNINNVCDNS
MYB113   121 MINKNITSHP TSSAQKIDVL KPRPRSFSDK NSCNDVNILP KVDVVPLHLG LNNNYVCESS
MYB114   121 IKRINIITPP NTPAQKVDIF KPRPRFFSIK TGCNHLDGQS EVGVIPPCLG LDNDNVCENS
MYB90    121 MKKKNIISPP TTPVQKIGVF KPRPRSFSVN NGCSHLNGLF EVDLIPSCLG LKKNNVCENS

MYB75    181 IIYNKDKKKD QLVN~NLIDG DNMWLEKFLE ESQEVDILVP EATTTEKGDT LAFDVDQLWS
MYB113   181 ITCNKDEQKD KLINININLLDG DNMWWESLLE ~~~~ADVLGP EATETAKGVT LPLDFEQIWA
MYB114   181 ITCNKDDEKD DFVD~NFMVG DNIWLERLLD ESQEVDVLVT EAAATEKEGT LAFDVEQLWN
MYB90    181 ITCNKDDEKD DFVN~NLMNG DNMWLEMLLE ENQEADAIVP EATTAEHGAT LAFDVEQLWS

MYB75    241 LFDGETVKFDX
MYB113   241 RFDEETLELNX
MYB114   241 LFDGETVIFDX
MYB90    241 LFDGETVELDX
```

Figure S3 Comparison of sequence diversity between and within four MYB family members, MYB75, MYB113, MYB114, MYB90. A Alignment of amino acid sequences of MYB75 (PAP1), MYB113, MYB114, and MYB90 (PAP2). Green and blue high-lights indicate respectively, synonymous and non-synonymous differences in the AA codon of the MYB indicated in the first column between accessions. The most frequent AA codon is represented (NB the most frequent AA codon will in some cases not represent Col-0). Red letter indicates that the amino acid at that position deviates from the common (conserved) amino acid present in the other MYB-proteins. Box around the letter indicates a SNP in the codon that is in strong LD (r^2>0.65) with the highest associated SNP found in the GWAS.

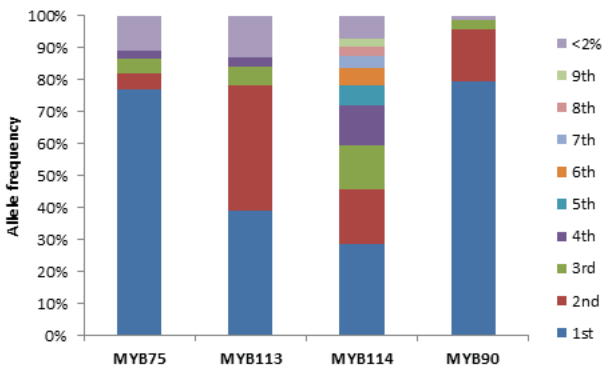


Figure S4 Frequency distribution of alleles of four MYB transcription factors, MYB75, MYB113, MYB114 and MYB90, in the population. The alleles are ordered from most to least frequent and alleles with a frequency lower than 2% are pooled together.

Position		Alleles											
Nucleo-tide	Amino Acid	1st 44	2nd 26	3rd 21	4th 19	5th 10	6th 8	7th 6	8th 4	9th 4	<2% 11	Codon	Amino Acid
41	12	C	C	G	C	C	C	C	C	C	C	aCt/aGt	Thr/Ser
65	20	G	G	G	G	G	G	G	G	G	G A	aGg/aAG	Arg/Lys
103	33	C	C	C	C	C	G	C	C	C	C	Caa/Gaa	Gln/Glu
237	78	G	G	G	G	G	G	G	G	G	G T	aaG/aaT	Lys/Asn
391	129	A	A	A	A	A	A	A	A	A	A G	ACa/GCa/ATa	Thr/Ala/Ile
392		C	C	C	C	C	C	C	C	C	C T		
395	130	C	C	C	T	C	C	T	C	T	C T	cCg/cTg	Pro/Leu
405	134	A	A	A	A	A	A	A	C	A	A	aaA/aaC	Lys/Asn
418	138	A	T	A	A	A	A	A	A	A	A	Tag/Aag	Stop/Lys
442	146	A	A	A	A	C	A	A	A	A	A	Aat/Cat	Asn/His
463	153	C	C	C	C	C	C	T	C	C	C T	Ctc/Ttc	Leu/Phe
481	159	G	G	G	G	G	G	G	G	G	G A	Gtt/Att	Val/Ile
487	161	G	G	G	G	G	G	G	G	G	T G	Gta/Tta	Val/Leu
497	164	C	C	C	C	C	C	C	C	C	T C	cTa/cCa	Leu/Pro
523	173	G	G	G	G	G	G	G	G	G	G A	Gtt/Att	Val/Ile
548	181	A	A	A	A	A	A	A	A	A	A G	aAc/aGc	Asn/Ser
613	203	C	C	C	G	C	C	G	C	C	C G	CCt/GGt/CGt	Pro/Gly/Arg
614		G	G	G	G	G	G	G	G	G	G C		
629	208	G	G	G	G	G	G	G	G	G	G A	gAc/gGc	Asp/Gly
664	220	G	G	G	G	G	G	G	G	G	G A	Gca/Aca	Ala/Thr
670	222	G	G	G	G	A	G	G	G	G	G	Gaa/Aaa	Glu/Lys
678	225	G	G	G	G	G	G	G	G	G	G C	gaG/gaC	Glu/Asp

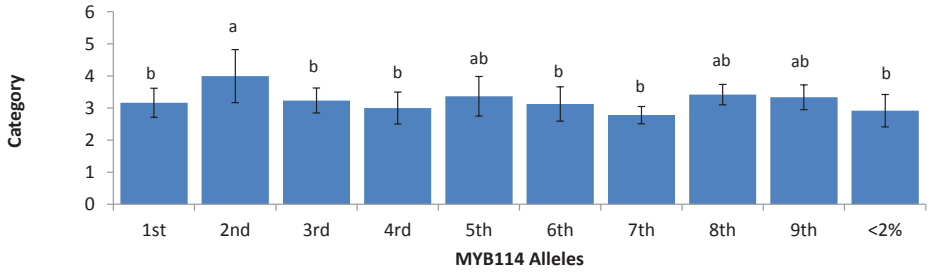


Figure S5 Alleles of MYB114 **A** Positions of non-synonymous SNPs and corresponding amino acid changes defining the alleles. Number of accessions carrying the allele are indicated above the double line. **B** Anthocyanin accumulation per allele (mean +/- SD). Means per allele are compared by ANOVA following pairwise comparison using Bonferroni corrected significance threshold of α=0.05. Letters above the bars indicate significant differences.

Table S1 Primers used for qPCR

gene		primer sequence (5'→ 3')	primer efficiency	remark
AT1G13320	forward	TAACGTGGCCAAAATGATGC	2.13	reference
	reverse	GTTCTCCACAACCGCTTGGT		
AT1G18610	forward	CTCAGGTTGATGAAGCGTCT	2.15	reference
	reverse	AAAACCCACCCTATCTCCAG		
AT1G56650	forward	TCTTCTTCTTCGCCTTCATAG	2.02	*MYB75, PAP1*
	reverse	ATCTTACAACACGGTTCATG		
AT1G66370	forward	TCCGATGAAGTTGATCTTG	1.98	*MYB113*
	reverse	TCATCGTGCTTCTTACTCAA		
AT1G66380	forward	GTCAAGAACTACTGGAACACC	2.04	*MYB114* , not over exon-intron border
	reverse	ATATCGACTTTTTGGGCC		
AT1G66390	forward	CAAGAGAGGAAGACTTAGCAA	2.05	*MYB90, PAP2*
	reverse	GATGGGTGTTCCAGTAATTT		

Table S2 SNPs significantly associated with the accumulation of anthocyanin in stress conditions.

chro-mo-some	position	-log(p)	Minor allele frequency	Effect size	Explained genetic variance	Explained phenotypic variance	gene	gene annotation
1	24756576	13.92	0.15	0.7	27.87	10.02	promoter AT1G66380	MYB domain protein 114 (MYB114)
1	24758488	11.11	0.13	0.67	22.68	8.15	promoter AT1G66380	MYB domain protein 114 (MYB114)
1	24759985	16.98	0.14	0.81	35.06	12.6	promoter AT1G66390	MYB domain protein 90 (MYB90)
1	24760088	16.09	0.13	0.8	33.5	12.04	promoter AT1G66390	MYB domain protein 90 (MYB90)
1	24765632	8.85	0.29	0.42	16.39	5.89	promoter AT1G66400	calmodulin like 23 (CML23)
1	24767033	19.92	0.13	0.91	42.03	15.1	promoter AT1G66400	calmodulin like 23 (CML23)
1	24769084	11.75	0.2	0.57	23.5	8.45	promoter AT1G66400	calmodulin like 23 (CML23)
1	24769177	18.42	0.14	0.82	37.28	13.4	promoter AT1G66400	calmodulin like 23 (CML23)
1	24769757	16.48	0.17	0.73	33.11	11.9	promoter AT1G66400	calmodulin like 23 (CML23)
1	24769770	16.5	0.17	0.73	33.11	11.9	promoter AT1G66400	calmodulin like 23 (CML23)
1	24770871	7.74	0.33	0.38	14.44	5.19	exon AT1G66400	calmodulin like 23 (CML23)
1	24771021	7.72	0.32	0.38	14.43	5.19	exon AT1G66400	calmodulin like 23 (CML23)

Chapter 3 - Genome Wide Association mapping of fertility reduction upon heat stress reveals developmental stage specific QTLs in *Arabidopsis thaliana*

Johanna A. Bac-Molenaar[1,2], Emilie F. Fradin[1,2], Frank F.M. Becker[2], Jurriaan A Rienstra[1], J. van der Schoot[1], Dick Vreugdenhil[1], Joost J.B. Keurentjes[2]

[1] Laboratory of Plant Physiology, Wageningen University, The Netherlands
[2] Laboratory of Genetics, Wageningen University, The Netherlands

Abstract

For crops that are grown for their fruits or seeds, the impact of elevated temperatures on yield is much stronger during flowering and seed or fruit set than during the vegetative stage. Even short-term exposure to heat can have a large impact on yield. In this study we used Arabidopsis to study the impact of short-term heat exposure on flower and seed development. The impact of a single hot day (35°C) was determined in more than 250 natural accessions by measuring the lengths of the siliques along the main inflorescence. Two sensitive developmental stages were identified, one before anthesis, during male and female meiosis, and one after anthesis, during fertilization and early embryo development. For both stages large differences in the heat response were observed between accessions. In addition, a correlation between flowering time and heat tolerance was observed. Genome wide association (GWA) mapping revealed four quantitative trait loci (QTLs) strongly associated with the heat response. These QTLs were developmental stage specific, as different QTLs were detected before and after anthesis. For a number of QTLs T-DNA insertion knockout lines could validate assigned candidate genes. These results emphasize the importance of large phenotypic screens for the detection of quantitative variation in natural populations and the identification of causal genetic factors. Moreover, our findings show that the regulation of complex traits can be highly dependent on the developmental timing, which might be an important consideration for many studies.

Plant Cell, doi: 10.1105/tpc.15.00248

Introduction

One of the major challenges in the coming decades is to secure that the global food production will increase proportionally to the increase of the human world population. High crop yield should therefore be sustained while climate change is expected to result in more extreme weather conditions, including longer drought periods and hotter summers. Plant breeding for resistance towards abiotic stress, like drought or heat, will play a crucial role in facing this challenge. To achieve this goal it is crucial to identify the natural range of variation for these traits within species and to gain a profound understanding of the mechanistic basis of regulatory processes. Both goals can be met by identifying key regulatory genes in large phenotypic screens of genetically diverse mapping populations, either naturally or experimentally derived.

Crops can roughly be divided into two classes: crops grown for total biomass and crops grown for seed or fruit production. For the latter the impact of abiotic and biotic stresses on yield should be investigated during both the vegetative and the generative phases of the plant's life cycle since both are strongly related to reproductive success. Temperature stress is one of the most detrimental abiotic stresses during the generative phase and reduced yields due to high temperatures during flowering have been reported for many species (Bita 2013), (Hedhly 2011). Heat stress during flowering of crops already has a large impact on the global food production and this impact will increase in the coming years because of expected climatic changes. For example, in the year 2000, 15% of the global harvested area of maize was exposed to at least five days of heat stress during the flowering period, and by 2050 this is expected to increase to 44% (Gourdji 2013). Besides maize, many other cereals like wheat, barley, and rice are affected by high temperatures during the reproductive phase (Barnabàs 2008). Negative effects of heat stress have also been reported for other important crops from various plant families, such as tomato (Sato 2002),(Sato 2006) and pepper (Marcelis 2004) from the *Solanaceae* family, canola (Angadi 2000) and mustard (Gan 2004) from the *Brassicaceae* family, soybean (Djanaguiraman 2013) and chickpea (Clarke 2004) from the *Fabiaceae* family, and peach (Hedhly 2005) and cherry (Hedhly 2004) from the *Rosaceae* family.

As mentioned, abiotic and biotic stresses may affect yield during both the vegetative and the generative phase of the plant's life cycle but heat stress is particularly detrimental during the reproductive stage of seed and fruit producing crops (Wahid 2007). However, the sensitivity towards temperature stress is not constant during flower and seed development. Two sensitive periods have been identified: during meiosis and around the fertilization event. Many studies, e.g. in cereals, tomato and Arabidopsis, have shown that the development of male reproductive organs is more sensitive to elevated temperatures than the development of female reproductive organs (Sato 2006), (Kim 2001). These studies showed that pollen release is reduced as a result of the inhibition of anther

dehiscence upon heat stress. Heat stress also leads to shortening of the anthers, due to a reduced amount of auxin (Sakata 2010), which results in a mismatch between the ripening of the stigma and anthers, preventing self-fertilization in bisexual monoecious plants. Moreover, viability of pollen is reduced by disruption of the male meiotic process upon heat stress (Endo 2009). In addition to disturbance of the development of male organs and pollen, changes in the development of the female organs are reported. Besides malformation of the ovule and surrounding tissues (Hedhly 2011), changes in the receptivity of the stigma, resulting in a decrease of the number of pollen tubes reaching the ovary, are described (Saini 1982). The timing and severity of the stress determine to a large extent which physiological processes are disrupted. However, responses are similar in many species, even when they are not closely related (Hedhly 2011). This makes the study of these processes in a model plant like *Arabidopsis thaliana* very relevant. Surprisingly, however, many studies in this species focus on the impact of elevated temperatures on germination (Baskin 1983, Schmuths 2006, He 2014, Silva-Correia 2014), flowering time (Balasubramanian 2006, Li 2013) and plant architecture (Wahid 2007, Patel 2009, Antoun 2013), but only very few address the impact of heat on fertility (Kim 2001), (Warner 2005), (Zinn 2010).

In many species heat tolerant and sensitive cultivars have been identified. Also in Arabidopsis, natural variation for heat response during flowering has been reported (Warner 2005). Unfortunately, much less is known about the genetic factors that cause this variation in heat response. To identify the mechanisms responsible for heat tolerance, heat tolerant and sensitive cultivars or variable natural accessions can be compared at the physiological, molecular and genetic level. In tomato, metabolite and gene expression profiles have been compared in cultivars with different levels of heat tolerance. In anthers of tolerant cultivars higher constitutive expression of heat-shock proteins and heat-shock transcription factors was observed (Bita 2011). Moreover, higher invertase activity was reported upon heat stress in tolerant cultivars (Li 2012, Dorion, 1996 #480), and down-regulation of a proline transporter was observed in sensitive cultivars (Sato 2006). This information may be helpful for understanding tolerance mechanisms, but to be able to breed for temperature-tolerant crops allelic variation in genes promoting heat tolerance needs to be identified. Natural variation in heat-tolerance between cultivars can be used to detect such alleles in segregating mapping populations. Moreover, such studies reveal the actual genetic regulators driving a specific response, in contrast to indicating transcriptional differences upon heat stress in gene expression studies, of which many might be downstream effects of the response. Bi-parental mapping populations have been developed and used for QTL mapping of heat tolerance in various crops, such as rice (Ye 2012), cowpea (Lucas 2013) and tomato (Grilli 2007). These studies identified multiple QTLs (ranging from 2 to 18) with an explained variance between 2 and 20%. These studies indicate that QTL analyses and subsequent fine-mapping and cloning are a promising way of identifying heat tolerance genes. In addition, the number of QTLs

detected in these studies and the relatively low explained variance for most of them indicate that heat-tolerance during flowering is a multi-genic trait. Unfortunately, in the studies referred to, none of the underlying causal genes were identified. Such identification in classical linkage studies is time-consuming because of the wide regions that are associated with the trait of interest. Genome Wide Association mapping (GWA) in natural populations can therefore be of great use because, due to the fast linkage decay, the resolution is much higher and therefore fine-mapping is often not needed to identify candidate genes (Bergelson 2010). Because similar heat responses are observed in many species, heat tolerance genes identified by GWA mapping in a natural population of Arabidopsis will likely have homologous equivalents in crops. Considering the variation in heat stress sensitivity observed for different developmental stages and genotypes in various species it is timely to combine the analysis of these observations in a single study. This offers the opportunity to identify natural variation in the key regulators in the response to heat stress that exert their function in a specific phase of development. These insights might provide a deeper understanding in the regulation of developmental timing and the natural strategies and mechanisms of coping with abiotic stresses.

We aimed to investigate in a systematic way the genetic basis of heat tolerance for various developmental phases in the reproduction of *Arabidopsis thaliana*. Not all stages of flower and seed development were equally sensitive. We were able to distinguish between heat responses before and after anthesis and GWA mapping was performed on data obtained for different regions along the inflorescence, resulting in developmental stage-specific QTLs. Strong associations could be observed for pre- and post-anthesis heat response and several candidate genes were identified that have not previously been associated with heat stress. A number of these candidate genes could be functionally validated by T-DNA insertion knockout analyses and artificially induced mutants, providing confidence in the applied methods.

Results

In order to investigate natural variation in reduction of fertility caused by short-term elevated temperatures a collection of natural accessions of *Arabidopsis thaliana* was subjected to heat stress and subsequent genome wide association (GWA) mapping.

Quantification of heat stress response by determination of silique length

Flowering plants of a wide range of natural accessions of *Arabidopsis thaliana* were exposed to a one-day treatment at 35°C. First, it was tested if the effect of heat stress could be quantified by taking the length of each silique along the main inflorescence as a measure for the number of seeds. Accession Di-1 (origin Dijon, France), showing an intermediate response towards the heat treatment, was selected for detailed analysis of silique length,

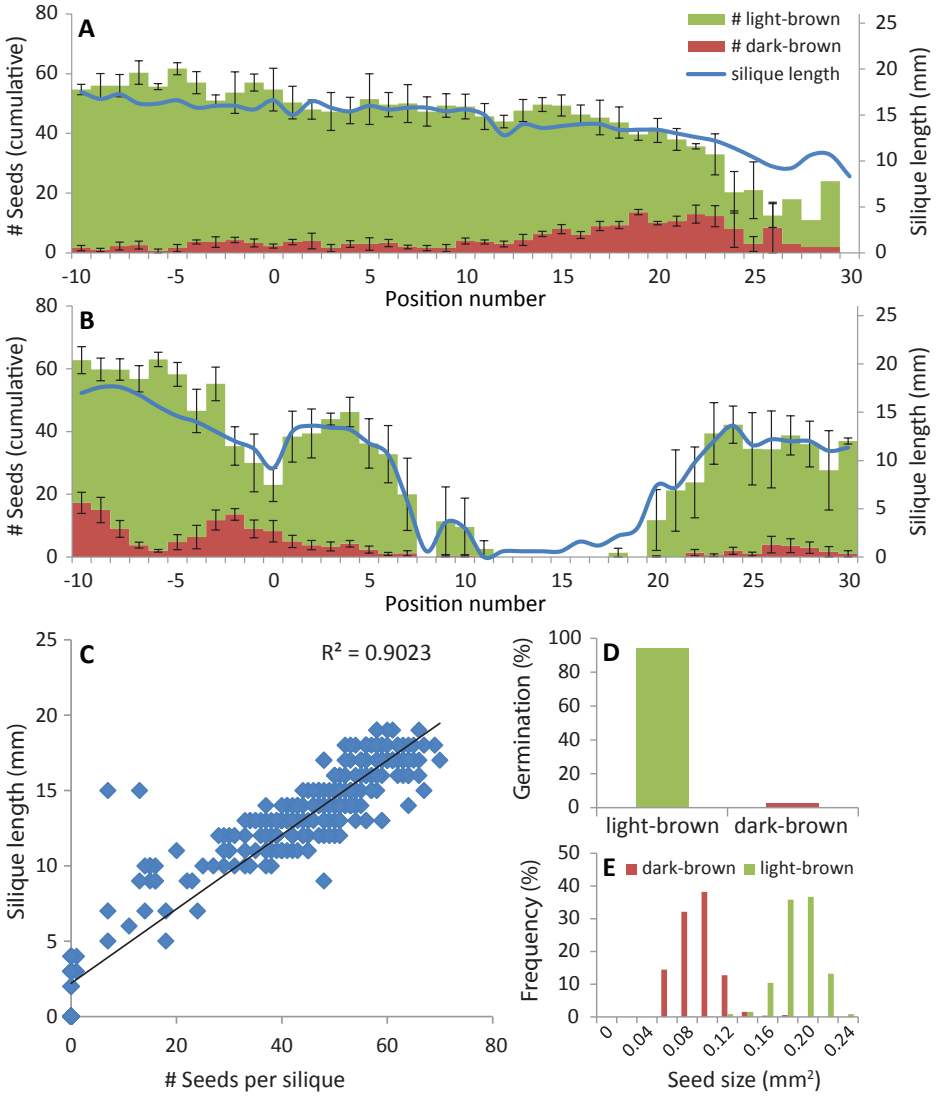


Figure 1 Silique length, seed number, and seed size per silique along the main inflorescence of Di-1 plants and germination rates of the seeds. **A-B** Comparison of silique length and number of seeds (light- and dark-brown) in control (n=3 for seed size, n=5 for silique length) and heat-treated (n=5) conditions (**A** and **B**, respectively). Error bars represent standard deviation. Position numbers are relative to the flower that opened first at the day of the treatment, which received position number 0. **C** Linear correlation between silique length and number of seeds per silique based on data represented in **A-B**. **D** Germination rate of light- (n=213) and dark-brown (n=244) seeds of heat-treated plants. E Size distribution of light- (n=575) and dark brown (n=526) seeds of heat-treated plants.

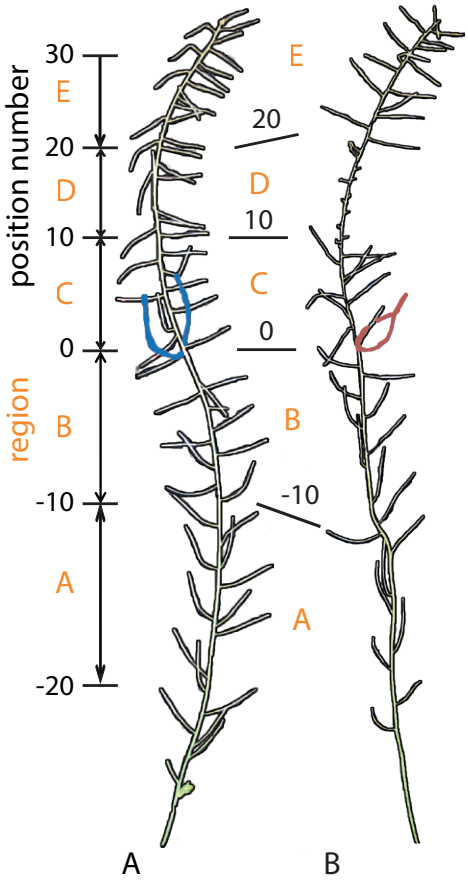


Figure 2 Outline of photographed inflorescences of a representative susceptible accession grown in control (**A**) and heat-treated (**B**) conditions. Definition of position number and developmental regions; the silique bearing the tag originated from the flower that opened first at the day of the treatment. Position numbers are relative to the tag. Five regions (A to E) were defined each containing ten siliques and representing different developmental stages as indicated in table 1.

seed size, and seed number per silique for all siliques along the main inflorescence (Figure 1). A linear correlation between the length of the silique and the number of seeds per silique was observed (Figure 1C, r^2=0.90). Such a high correlation has also been reported previously for accessions Landsberg *erecta* (*Ler*) and Cape Verde Islands (*Cvi*) (Alonso-Blanco 1999). The reduction in silique length is predominantly caused by a lower number of seeds per silique rather than by smaller seeds. Two types of seeds were typically observed: a majority of light-brown seeds and a lower number of dark-brown small seeds (Figure 1E). Germination tests revealed that most small dark-brown seeds were not able to germinate while almost all light-brown seeds did germinate (Figure 1D). The small dark-brown seeds were mainly present in siliques in which embryogenesis had already started at the day of the treatment. Therefore, these small dark-brown seeds are most probably aborted seeds. These putative seed abortions did not lead to shorter siliques, which indicates that the length of the silique is a measure for successful fertilisation, resulting in the initiation of embryogenesis, and might not always reflect the number of viable seeds within the silique. Siliques from control plants also contained a few small dark-brown seeds (Figure 1A), which indicates that seed abortion does not only result from heat stress, but may be caused by other factors as well.

The flower that opened first on the day of the treatment was tagged for each replicate of 285 accessions used in the experiment. This tag enabled us to relate the effect of heat stress to the developmental stage of the flowers and siliques along the inflorescence (Figure 2). In reference to the tag, all siliques received a relative position number and subsequently five developmental regions, A to E, were defined. To link the observed reduction in silique length with the disruption of flower development and embryogenesis, we matched the regions as defined in Figure 2 with data regarding timing and order of events taking place

Table 1 Matching of landmark events in flower and seed development with the developmental regions along the inflorescence as defined in Figure 2.

Stage[a]	Landmark event at beginning of stage[a]	Duration[a] (hr)	Landmark event in reproduction and seed development[b]		Region[c]
1	Flower buttress arises	24			
2	Flower primordium forms	30			
3	Sepal primordia arise	18			
4	Sepals overlie flower meristem	18			
5	Petal and stamen primordia arise	6			
6	Sepals enclose bud	30			
7	Long stamen primordia stalked at base	24			E
8	Locules appear in long stamens	24			
9	Petal primordia stalked at base	60	Male meiosis		
10	Petals level with short stamens	12			D
11	Stigmatic papillae appear	30	Anther elongation	Female meiosis	C
12	Petals level with long stamens	42			
13	Bud opens, petals visible, anthesis	6		Anther dehiscence, pollen tube growth, fertilization	
14	Long anthers extend above stigma	18	Early embryogene-sis until heart stage		B
15	Stigma extends above long anthers	24			
16	Petals and sepals withering	12			
17	All organs fall from green siliques	192		Late embryogenesis from torpedo stage to mature embryo, seed coat formation	A
18	Siliques turn yellow	36			
19	Valves separate from dry silique	24			
20	Seeds fall				

[a] Data are taken from (Smyth et al., 1990)

[b] Data are taken from (Schneitz et al., 1995), (Sanders et al., 1999), (Mansfield and Briarty, 1991)

[c] Regions along the inflorescence as defined in Figure 2. Because of natural variation in number of flowers that open per day, matching with other data in the table relies on estimates.

during these developmental processes in Arabidopsis (Table 1, (Mansfield 1991), (Smyth 1990), (Schneitz 1995), (Sanders 1999)). Region A (silique position numbers -20 to -10) is linked to late embryogenesis. Region B (silique position numbers -10 to 0) is related to processes that take place shortly after anthesis, like fertilization and early embryogenesis. Region C (silique position numbers 0 to 10) is related to processes just before anthesis, like anther elongation and female meiosis and region D (silique position numbers 10 to 20) is

linked to processes that take place some days before anthesis, like male meiosis. Finally, region E (silique position numbers 20 to 30) corresponds to early flower development. Although we defined each region by a fixed number of siliques without differentiating between accessions we note that natural variation was observed for the number of flowers that opened per day. Even though such phase shifts in developmental timing might introduce noise, this did not negatively affect later analyses, probably because the size of the assigned regions is relatively large in comparison with the observed differences between accessions. We, therefore, assume that the developmental processes assigned to the different regions are correct for most accessions and analyses described hereafter.

Natural variation of heat response is developmental stage dependent

Natural variation in seed set upon heat stress was recorded by measuring the length of all siliques along the main inflorescence of each individual plant of all accessions. To determine the general response pattern, the average of the silique length of all plants in the experiment was calculated for all silique positions along the main inflorescence (Figure 3). The effect of the treatment varied along the inflorescence. For instance, region A (late embryogenesis) and region E (early flower development) were least affected by the heat treatment in all accessions (Figure 3A), while the siliques in other regions were at least affected in some accessions. After heat treatment, two minima were observed in the overall average silique length. The first minimum was found for siliques in the phase of early embryogenesis (region B) with the absolute minimum at position minus three. On average silique length at that position decreased from 13.4 to 8.2 mm, corresponding to a decrease of approximately 21 seeds per silique on average (Figure 1C).

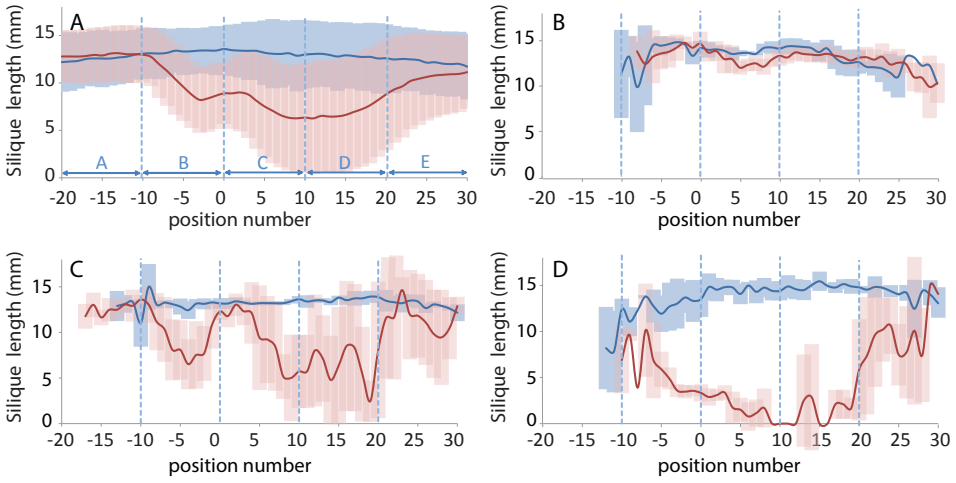


Figure 3 Silique length along the main inflorescences of control (blue) and heat-treated (red) plants (mean +/- SD). **A** Overall average of all plants in the experiment **B** Representative tolerant accession, C24 **C** Representative moderately sensitive accession, Rmx-A180 **D** Representative extremely sensitive accession, Arby-1. **B-D** control: n=3, treatment: n=5. Bars indicate the standard deviation of silique lengths among replicates, solid lines indicate mean values.

Trait	Developmental Region	H^2
Control	B	0.58
Silique	C	0.57
length	D	0.66
Heat	B	0.70
Silique	C	0.84
length	D	0.73
Flowering time		0.96

Table 2 Broad-sense heritability of silique length in control and heat stress conditions per developmental region (B, C, D, as defined in Figure 2) and of flowering time.

The second minimum spanned a large region in which male and female meiosis takes place. The lowest value was reached at position ten with an average decrease in silique length from 13.0 to 6.3 mm, corresponding to a decrease of approximately 27 seeds per silique on average (Figure 1C). Overall, large differences were observed in the response to the treatment, ranging from accessions without a clear response, to complete absence of siliques, and consequently the two minima were not always detected in all accessions. For instance, accession C24 (origin Portugal), displayed no reduction in silique length at any position whereas accession Rmx-A180 (origin Michigan, USA) was clearly affected in both regions. In addition, accession Arby-1 (origin Sweden) was severely affected in silique length along most parts of the inflorescence (Figure 3 B-D). The variation in silique length in control and heat stress conditions was high between accessions, whereas the variation within accessions was low. This resulted in high broad-sense heritabilities (H^2 between 0.58 and 0.73, Table 2), indicating that the observed variation in silique length between accessions is for the major part determined by genetic factors. In addition, substantial variation between the different accessions was observed for the total number of flowers formed along the main inflorescence during the reproductive period (Figure 4). In the majority of accessions, the number of flowers formed in heat-treated plants was higher than in control conditions (Figure 4B). The average number of flowers formed in all heat-treated plants was also significantly larger than the average number in control plants ($p < 0.001$). On average, heat-treated plants developed 12% more flowers than the plants in control condition, suggesting a compensation mechanism for the loss of viable seeds.

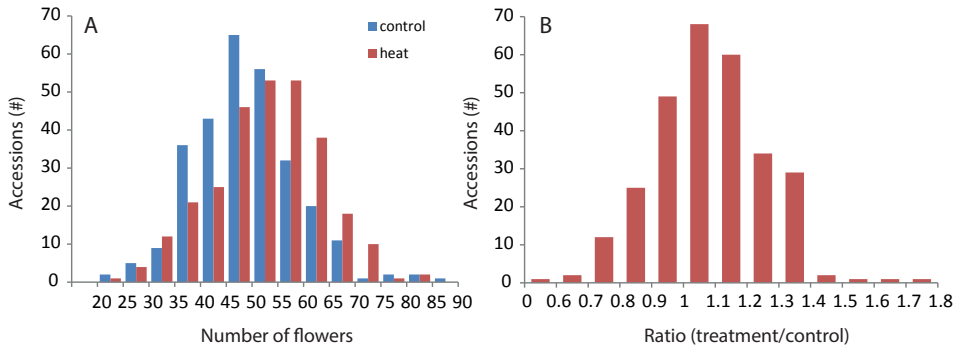


Figure 4 Comparison of total number and ratio of flowers along the main inflorescence in control and heat-treated plants for all accessions. **A** Frequency distribution of the total number of flowers per main inflorescence averaged per accession. **B** Frequency distribution of the ratio between the numbers of flowers along the main inflorescence in heat treated and control conditions per accession.

			Average Silique Length								
			Control			Heat			Residuals		
			B	C	D	B	C	D	B	C	D
Average Silique Length	Control	B	1.00								
		C	0.70	1.00							
		D	0.44	0.82	1.00						
	Heat	B	0.51	0.43	0.37	1.00					
		C	0.22	0.42	0.46	0.55	1.00				
		D	0.13	0.37	0.48	0.40	0.76	1.00			
	Residuals	B	0.02	0.11	0.19	0.83	0.50	0.41	1.00		
		C	-0.04	0.07	0.19	0.42	0.91	0.68	0.53	1.00	
		D	-0.02	0.07	0.11	0.31	0.65	0.91	0.40	0.69	1.00
Flowering Time			0.06	-0.10	-0.25	-0.31	-0.65	-0.58	-0.41	-0.67	-0.53

Table 3 Spearman's correlation between average silique length of control and heat treated plants in three developmental regions (B, C, D, as defined in Figure 2), the corresponding residuals for those regions and flowering time.

To quantify the impact of heat on the different developmental stages, the average silique length per developmental region was calculated for control and heat-treated plants. Regions A and E were not taken into account because for some accessions with a short main inflorescence not all positions in these regions developed a flower. In addition, it was observed, after determination of the developmental sensitivity window, that for most accessions these regions were outside this window. To quantify the impact of the heat independently of the size of the siliques in control conditions, simple residuals of the silique length of control and heat-treated plants were calculated using linear regression. Correlation analysis was performed for all these fertility traits to determine whether they are interconnected (Table 3). Flowering time was also included in this correlation analysis. Significant positive correlations were found between the lengths of the siliques of heat-treated plants in regions B, C and D. This indicates that plants that are sensitive in region B in general are also sensitive in regions C and D and vice versa. Although the heat responses in the different regions can be interdependent, a correlation does not always reflect a direct causal relationship and might result from downstream effects of early events. Furthermore, a strong negative correlation was observed between flowering time and silique length in heat-treated plants (Table 3). To synchronize flowering, the plants were vernalized for ten weeks, but the variation in flowering time still spanned four weeks between the earliest and latest flowering accessions. Therefore, the heat treatment was given in batches to keep the time interval between the start of flowering and the treatment for both early and late flowering accessions similar. The correlation between flowering time and heat response also suggests a role for developmental timing in the sensitivity to heat stress. As for the heat response, also for flowering time high heritability was observed ($H^2 = 0.96$, Table 2).

QTLs associated with stage specific heat responses

In search for the genetic variation that causes the phenotypic variation observed in the fertility traits, GWA mapping was performed on data from control and heat-treated plants and for the residuals. Traits analysed were average silique length in regions B, C and D and flowering time. GWA mapping resulted in the detection of 15 strongly associated SNPs with a –log(p-value) above an arbitrarily set significance threshold of 5.5 (Figure 5). In addition, 297 moderately associated SNPs with a –log(p-value) between 4 and 5.5 were identified. The arbitrary threshold of –log(p-value)=5.5 was set to focus on the SNPs with the largest explained variance and to minimize the chance of detecting false positives. Some of these significant SNPs are in linkage disequilibrium with each other and, therefore, are very likely to represent a single locus. This resulted in seven unique major QTLs, three related to silique length in control conditions and four related to the heat response (Table 4). Variation in silique length as a result of heat stress could partly

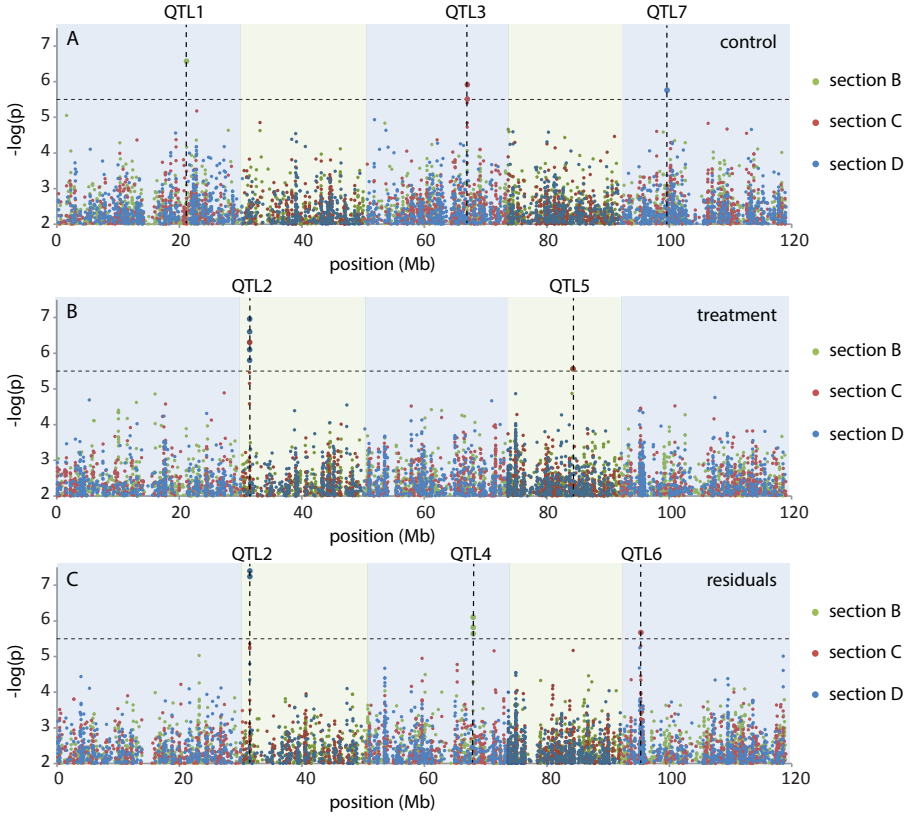


Figure 5 Manhattan plots representing the associations between SNP-markers and silique length in the regions B, C, and D. **A** GWA mapping of silique length in control plants **B** GWA mapping of silique length in heat-treated plants **C** GWA mapping of residuals, representing heat response independent of silique length in control conditions. Dotted line represents the significance threshold of –log(p-value)=5.5.

Table 4 Major QTLs identified by GWA mapping for flowering time and average silique length in developmental regions B, C and D (as defined in Figure 2) in control and heat treated conditions and their corresponding residuals for those regions. Columns represent the exact position of SNPs and their association with above mentioned traits and the effect size (Col-0 allele is positive, non-Col-0 allele is negative) and explained genetic and phenotypic variance for the highest association detected, respectively. Only SNPs which are strongly associated with at least one trait (-log(p-value) > 5.5) are shown. Bold indicate significantly associated SNPs, no value indicates no association (-log(p-value)<2).

QTL	chromosome	position (Mb)	FT	control			heat			residuals			effect size	explained genotypic variance	explained phenotypic variance
				B	C	D	B	C	D	B	C	D			
1	1	21.17		6.6	3.5		3.7						-2.16	9.88	8.43
2	2	1.06					2.2	**6.3**	**6.1**		5.3	4.8	-2.63	9.72	7.93
	2	1.06						**5.5**	**5.8**		3.8	4.3	-3.14	9.45	6.94
	2	1.06						5.1	**7**		5.4	**7.4**	-3.40	25.10	11.54
	2	1.07						4.6	**6.6**		5.2	**7.2**	-3.32	24.41	11.22
3	3	16.89		4.1	**5.9**	3.5		2.1					-1.46	9.99	7.18
	3	16.89		3.5	**5.5**	2.2							-1.92	9.28	6.67
4	3	17.85					2.8			**5.6**			1.53	14.22	8.09
	3	17.85					2.8			**5.6**			1.53	14.22	8.09
	3	17.85					2.7			**6.1**			1.62	15.40	8.76
	3	17.85					2.8			**5.6**			1.53	14.22	8.09
	3	17.85					2.9			**5.8**			1.58	14.73	8.38
5	4	10.72						**5.6**			5.2		-2.09	8.04	6.55
6	5	3.19	4.96					4.5			**5.7**		-2.45	15.66	8.88
7	5	7.48			3.1	**5.8**							1.44	7.02	5.37

be explained by QTL2, QTL4, QTL5 and QTL6. Total explained phenotypic variance of these 4 QTLs reached 35.7%, which is relatively high for GWA studies. Especially, considering that this may be an underestimation due to the possible presence of multiple alleles per locus. However, the large fraction of missing heritability indicates that other genetic factors, possibly with large effects, remained undetected, e.g. due to low allele frequencies or overcorrection for population structure. None of the significant SNPs were identified in data from control conditions, suggesting that the observed variation in silique length explained by these SNPs is solely due to the heat treatment. The QTL with the strongest association and the largest effect size (QTL2), consists of 4 significant SNPs located at the top of chromosome 2 within a region of 0.7 kb (Table 4). All 4 SNPs were strongly associated with regions C and D, but not with region B. This indicates that this QTL is specific for the observed variation for average silique length at the position of the second minimum and can therefore be related to pre-anthesis processes, such as male

and female meiosis (Table 1). QTL 5 and QTL6 were specific for developmental region C. Besides female meiosis also anther elongation takes place in that developmental stage at the time of the heat treatment. QTL4 is the only QTL found to be associated with the first minimum in average silique length and is therefore related to post-anthesis processes like early embryogenesis. Although the main aim of this research was to identify genes involved in heat tolerance, the collected data also give insight into the regulation of silique development under control conditions. Variation observed between the accessions in the silique length in control conditions could be partly explained by three QTLs, viz., QTL1, QTL3, and QTL7. Although each control QTL was identified in a distinct developmental region, moderate associations were identified for at least one other region, indicating that genes that determine silique length in one developmental region may also play a role in another region (Table 4). For QTL1, a weak association was also found for variation in silique length in region B of heat-treated plants in addition to a strong association in control conditions, suggesting that the causal gene has an effect on silique length independent of the environmental conditions. Although a strong negative correlation was observed between flowering time and silique length in heat-treated plants, no significant associations (-log(p-value)>5.5) were detected for flowering time. That said, flowering time was moderately associated with QTL6 (-log(p-value)=4.96). This QTL maps to the position of the well-known flowering time gene *FLC*, which might explain the observed variation in flowering time and suggests a role for the regulation of flowering time in heat stress response. To complete our analyses, data were also collected for silique length averaged over three regions (B, C, and D together) and for the total number of siliques smaller than 5 mm, which probably do not contain viable seeds. These traits are independent of a possible bias introduced by developmental phase shifts between accessions and might, therefore, reveal additional regulatory loci. However, GWA mapping did not reveal significant associations other than the ones described above, further strengthening our assumption that the developmental processes assigned to the different regions are correct for most accessions and that possible deviations do not affect our analyses drastically.

Assignment of candidate genes based on LD, gene annotation and expression

Re-sequence data (1001genomes.org) of 502 accessions were used to assign possible causal candidates to each of the seven observed QTL. For 124 of those 502 accessions the heat response was investigated in this study. Genes in LD with the significant SNPs (threshold LD>0.4) were considered as candidates (Figure S1). In addition, gene annotation and gene expression patterns in different plant tissues were obtained from TAIR and the Arabidopsis eFP browser (Winter 2007), respectively. For a number of candidate genes, selected for likelihood and detection confidence, the heat response of knockout mutants was investigated in two independent experiments (Figure S2).

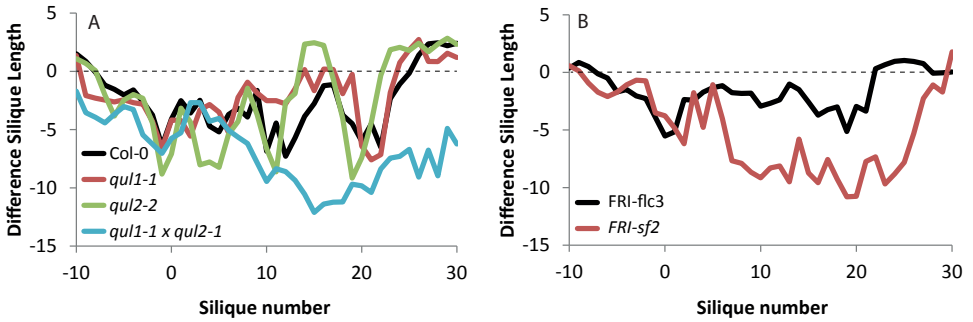


Figure 6 Difference between silique length of heat stressed and control plants (average silique length of heat stress plants minus average silique length of control plants) **A** of Col-0 wildtype and the single and double knock-down mutant lines of QUL1 and QUL2 and **B** of isogenic lines with FRI strong allele and FLC null allele (early flowering, FRI-flc3) or FLC functional allele (late flowering, FRI-sf2). Results of one experiment are shown. Results of a second independent experiment can be found in Figure S2.

3

Pre-anthesis sensitivity to heat stress

The four strongly associated SNPs of QTL2, explaining variation in silique length upon pre-anthesis heat treatment, are in linkage with nine other SNPs, as identified in the re-sequence data. All thirteen SNPs are located in the introns of gene *AT2G03505*, which is, therefore, the only candidate gene in LD with the significant SNPs. *AT2G03505* encodes an unknown protein of the carbohydrate-binding X8 domain superfamily. This gene is relatively high expressed in flowers and especially carpels (Laubinger 2008). As a start of the validation process, a Gabi-kat knock-out line was tested for its heat response in two independent experiments. No differences were observed between the heat response of this line and the Col-0 wildtype (Figure S2). This suggests that this gene might not be the causal gene, although the wildtype Col-0 allele in itself can already be weak or non-functional and additional experiments of knock-out and overexpression lines of non-Col-0 alleles of this gene are needed for final determination. Because causal genes are sometimes also detected outside the LD interval of significantly associated SNPs (Chao 2014), due to, amongst others, low linkage between causal and associated SNPs (Freedman 2011) , other genes in the 20 kb region around the significant SNPs (10kb up- and 10kb downstream) were considered as candidates, in accordance to the average linkage distance in *Arabidopsis thaliana* of 10 kb (Kim 2007). This resulted in the identification of a candidate gene, *QUASIMODO-LIKE 2* (*QUL2*, *AT2G03480*) that is highly expressed during the development of the male gametophyte (Honys 2004). This process takes place in region D, in which the highest association was observed. A triple mutant of *QUL2* and its two paralogs *QUA2* (*AT1G78240*) and *QUL1* (*AT1G13860*) was previously found to have temperature dependent changes in stem diameter and stem lignification (Fuentes 2010). Therefore, the heat response of the T-DNA insertion lines of *QUL2*, its closest paralogue *QUL1* and the corresponding double mutant were investigated (Figure 6, Figure S3). Expression analysis confirmed that the single and double mutants of *QUL1* and *QUL2* were true knock-downs (Figure S3). The double mutant was more heat sensitive in

regions C, D and E (Figure 6) but not in region B, whereas both single mutants resembled the wild type (Figure S2). This indicates that *QUL1* and *QUL2* have probably very similar complementary functions with direct or indirect positive effects on silique formation upon heat. This result also emphasizes that different genes might be responsible for the heat response before and after anthesis.
A second significant association (QTL5*)* assigned to the pre-anthesis stage is located in a chromosomal region with very fast linkage decay (Figure S1). The significant SNP is not in LD with re-sequence SNPs that are located upstream and only with a limited number of SNPs located 4 kb downstream. This leads to the identification of one candidate gene, *AT4G19710*, coding for an aspartate kinase-homoserine dehydrogenase. This gene is expressed in many tissues in control conditions (Schmid 2005) but has not been associated with heat stress before.

Post-anthesis sensitivity to heat stress

QTL4, at the bottom of chromosome three, is specific for the heat response in developmental region B and assumed to be related to fertilization or embryo formation. QTL4 is determined by five tightly linked SNPs, which are in high linkage disequilibrium (LD>0.9, Figure S1) with re-sequence SNPs in a gene encoding a kinetochore protein (*AT3G48210*). This protein is involved in cell division, a process that is crucial for the growing embryo. Functional characterization of this gene has not been performed, but it is expressed in many tissues, including the embryo (Schmid 2005). High linkage (LD>0.7, Figure S1) was also observed between the strongly associated SNPs and re-sequence SNPs located in *AT3G48190* and its promoter. *AT3G48190* is suggested to play a role in DNA repair upon stress (Huefner 2011) and might be important in heat stressed embryos. However, the slow LD decay in this region results in a linkage block of 140 kb (Figure S1) in which 37 genes are located, and further analyses are needed to confirm the causal gene.

A possible genetic link between developmental timing of flowering and response to heat stress

The strongly associated SNP of QTL6, at the top of chromosome five, is in LD with 17 SNPs identified in the re-sequence data (Figure S1). Four of those SNPs were included in the GWA mapping and were moderately associated (-log(p-value)>4) with heat stress response in region C. The strongest associated SNP is responsible for a non-synonymous changes in the protein sequence of AT5G10170, which codes for Myo-Inositol 1-Phosphate Synthase3 (MIPS3). *MIPS3* expression was most prominent in vascular tissue (Donahue 2010) and heat-induced expression was reported during the seedling stage (Khurana 2012). In Arabidopsis, three *MIPS* genes were identified. Analysis of the close relative *MIPS2* showed that the expression of this gene is highly induced upon heat stress treatment in seedlings, flowers and mature siliques (Khurana 2012). Investigation of heat response of the T-DNA knock-out mutant of *MIPS3* did not reveal differences between the heat response of this line and wild type Col-0 (Figure S2), suggesting again that this

gene might not be causal for the observed natural variation. In addition, mutants of two other genes in the LD region of QTL6 were analysed. *AT5G10200* codes for an ARM-repeat/Tetratricopeptide repeat (TPR)-like protein. Like most ARM-repeat proteins in Arabidopsis, the function of AT5G10200 is unknown. In Rice, a genome wide study of ARM-repeat gene expression showed that several genes of this family are regulated by abiotic stresses (Sharma 2014). However, investigation of a T-DNA knock-out mutant did not reveal an immediate role for this gene in the heat response of Arabidopsis (Figure S2). This is in agreement with the observation that another Arabidopsis ARM-repeat protein, *ARO1*, was shown to be required for pollen tube germination and penetration into the stigma (Gebert 2008). These processes take place post-anthesis, whereas QTL6 is associated with pre-anthesis heat responses.

Interestingly, the other gene in LD with QTL6, *AT5G10140*, encodes for FLOWERING LOCUS C (*FLC*), a gene involved in the timing of flowering. Two isogenic lines, containing different *FLC* alleles and a strong functional allele of *FRIGIDA* (FRI; *AT4G00650*) in a Columbia background, were tested for their heat response. A strong *FRI* allele is needed because *FRI* is an up-stream regulator of *FLC* and the presence of a non-functional allele (like in the Columbia wild type accession) will result in early flowering, independent of the *FLC* allele. Line *FRI-sf2*, containing the functional Columbia allele of *FLC*, requires vernalization to flower, whereas line *FRI-flc3*, containing a null-allele of *FLC*, flowers early. Both lines were vernalized for 10 weeks, after which *FRI-sf2* still flowered a week later than *FRI-flc3*. Heat treatment was given approximately one week after the start of flowering. The late flowering line, *FRI-sf2*, was much more sensitive to the heat stress in region C, D and E, than the early flowering line, *FRI-flc3* (Figure 6). Both lines had similar heat response for region B, suggesting that *FLC* is responsible for part of the pre-anthesis heat response. This experiment is in line with the observation in the association panel that heat response and flowering time are negatively correlated and it suggests that the two processes are, directly or indirectly, genetically linked. To investigate whether this result is specific for the Columbia allele, the six most common *FLC* alleles (Li 2014b) were compared for flowering time after 10 weeks of vernalization and for their heat response. This comparison confirmed the negative correlation observed between the two traits (Figure 7). Accessions with the Col-0 allele (haplotype RV2), which flowered early, had significantly larger siliques in region C (Figure 7) and region D (Figure S4) than accessions belonging to haplotypes SV2, SV3, and SV4, which flowered later. Haplotype RV1 was, like RV2, identified because of its strong response to vernalization (Li 2014b), but no significant difference was found for silique length upon heat between accessions of this haplotype and the SV haplotypes with slow response to vernalization, indicating that RV2 represents the *FLC* allele associated with heat resistance. However, large variation in flowering time and heat response was observed between accessions belonging to the same haplotype, which might be a consequence of sequence variation at other flowering time loci (Shindo 2005) or heat response genes.

Natural variation in silique length in control conditions suggests a role for pollen allergens

Genes in linkage with the strongly associated SNPs detected in control conditions were also analysed to identify candidates that could be involved in the determination of silique length and consequently seed yield. For QTL1, the control QTL with the strongest associated SNP and the largest effect size, a second weaker associated SNP (-log(p-value)=4.2) in moderate LD (LD = 0.424) was detected 30kb downstream. The significant SNP is located in an intron of a white rust resistance gene (*AT1G56510*), which is highly expressed in leaves. The weaker associated SNP is located between *AT1G56580*, a gene associated with trichome architecture, and *AT1G56590*, a gene involved in vesicle trafficking between the trans-Golgi network and vacuoles. None of these three genes are obvious candidates. However, the LD decay in this region was very slow, resulting in a linkage block of almost 176 kb, 8kb upstream and 168 kb downstream of the significant SNP (threshold: LD=0.4), in which 127 genes are located (Figure S1). Most of these genes are pre-tRNA's (81 genes) or genes with unknown biological function (22 genes). None of the genes with an annotated function is an obvious candidate and further analyses are needed to identify the causal gene. For QTL3 and QTL7, candidate genes could be identified based on functional annotation. QTL3 is determined by two significant SNPs that are located 0.2kb from each other. Three other moderately associated SNPs were identified in close proximity (within 0.4 kb). These five SNPs are located in the promoter of *AT3G45970*, 1,2kb up-stream of the start-codon, and are in low LD (0.2<LD<0.4) with SNPs in the gene *AT3G45960* (Figure S1). Both genes are annotated as expansin-like A, and are suggested to be involved in cell-wall loosening (Cosgrove 2002). The significant SNP of QTL7 is located in a large linkage block starting 43 kb upstream and ending 76 kb downstream of the significant SNP. Two candidate genes were identified within this linkage block: *AT5G22430*, a gene of a conserved allergen family, which is suggested to play a role in pollen-pistil interactions (Jiménez-López 2011) and *AT5G22640*, a gene essential for proper embryo development (Liang 2010). Remarkably, both identified expansin-like proteins of QTL3 and *AT5G22430* of QTL7 contain a pollen allergen domain. Proteins with such a domain are known to be present in pollen and cause allergic reactions in humans. It is suggested that these proteins loosen maternal cell walls to aid penetration of the stigma by the pollen tube (Cosgrove 2000), (Sharova 2007). Inconsistencies in the interaction between pollen and pistil can lead to reduced fertilization and can therefore result in shorter siliques. In addition, proper embryogenesis is essential for the development of seeds. If this process is disrupted also less seeds will be present per silique, resulting in shorter siliques.

Our results indicate that the observed heat response is dependent on both the genotype and the developmental stage of the flowers and embryo's during the heat treatment. GWA mapping resulted in the identification of developmental stage specific QTLs and covariance between the heat response and flowering time suggests a direct or indirect

link between the two regulatory pathways. In addition, QTLs for silique length in control conditions were identified. Candidate genes were suggested for most QTLs, of which some could be provisionally validated by analysing knockout mutants.

Discussion

Over the past decade much effort has been put in collecting large numbers of natural variants for many species, including the model species *Arabidopsis thaliana*. Technological advancements made it possible to characterise the genetic diversity in such collections in great detail. A number of studies have shown the use of such densely genotyped natural populations in GWA mapping approaches ((Cockram 2010, Chan 2011, Chao 2014, Verslues 2014)). Such studies not only disclose the large amount of natural variation present in species for many traits but also hint at candidate genes causal for the observed variation. The latter is an important argument in favour of GWA mapping over, e.g., gene expression studies because causal genes often represent key regulators of, or essential components in biological processes, whereas differentially expressed genes might merely represent transcriptional program changes in response to genetic or environmental perturbations. As such GWA mapping can add directionality in genetic regulatory networks and identify important anchor points in biological processes. In addition to genetic perturbations, many complex traits might depend on interaction with the environment and the developmental timing of the subject of study, a complication that is often neglected. Including genotypic diversity and developmental dependence in a single study of environmental perturbation can, therefore, provide a deeper understanding of the developmental processes involved in the regulation of abiotic stress responses.

Identification of a developmental sensitivity window

In this study the impact of short-term heat on fertility was measured in a developmental stage specific way. Large variation was observed in the sensitivity of different developmental stages of the flowers and embryo's towards the stress. Early flower development and late embryogenesis, assumed to take place in region A and E, were not sensitive to heat stress in most of the accessions (Figure 3 and Table 1). Two minima in the average silique length, corresponding to reduced seed set, were identified indicating two sensitive periods in development. One of the minima observed in our experiment corresponds to floral stages 9, 10, 11, and 12 (Smyth 1990), representing the developmental stages in which male and female meiosis occur and anther elongation takes place (Table 1). The other minimum corresponds to flowers in floral stages 14 and 15 that were heat-treated after anthesis, during these stages early embryo formation takes place. Male meiosis, which takes place during floral stage 9, is a process that is sensitive to heat in many crops (Hedhly 2011). An earlier study in Arabidopsis investigated the effect of heat on flower development (but not on embryo development) in the accession Columbia by counting

fertile siliques along the main inflorescence. Although the heat treatment in this study was more severe (42°C/4h), a sensitive period at floral stage 9 was observed, similar to our study. In addition, a sensitive period at floral stage 12 was identified (Kim 2001). A more extensive study in Arabidopsis, investigating floral bud abortion upon heat stress in 13 natural accessions, demonstrated that in the Columbia accession floral stage 12, but not floral stage 9, was most sensitive (Warner 2005). That floral stage 9 was not identified could be caused by the longer period of heat that was applied (36°C for 72 h). However, other accessions were shown to be more sensitive than Columbia, and floral bud abortion was also observed in earlier stages than floral stage 12. Floral stage 12 corresponds to anther elongation. Reduction in anther elongation was observed upon heat stress as a result of decreased auxin levels in the developing anthers. Reduced anther elongation upon heat stress has also been found in other species, such as barley (Sakata 2010). The observations of *Kim et al.* and *Warner et al.* are in line with our observation that natural variation exists both in terms of duration of the sensitivity window and the degree of reduction in seed set within the sensitivity window. However, comparison of floral stages between different studies should be done carefully since the matching between the developmental regions and the landmark events in flower and seed development are based on generalization and natural variation was observed for the number of flowers that opened per day.

Also in other *Brassicacea* species, heat response patterns similar to our observations have been reported. Studies of canola and mustard species (*B. napus, and B. juncea*) showed that heat stress during flowering or pod development led to more reduction in yield than heat stress during bud formation (Gan 2004). In addition it was observed that pods that passed a critical developmental threshold were tolerant towards heat stress (Angadi 2000). This threshold corresponds with late embryogenesis, the developmental stage that was also not sensitive to heat in any accession tested in our experiment. Even though the study of *Angadi et al. (2000)* confirms the existence of a developmental sensitivity window in which heat stress affects yield, a comparison between these results and the observations in our experiment also reveals an interesting difference. We observed a higher impact of stress before flowering while in other Brassicacea the impact seems to be more dominant after the start of flowering. This may be related to embryo abortion during early silique development, which was observed in our experiments (Figure 1). Moreover, the findings of *Gan et al. (2004)* are based on total seed yield and should be interpreted carefully. If compensation mechanisms are active (Figure 4 and (Angadi 2003)) total seed yield might not be the best measure to determine a developmental sensitivity window.

In species other than Brassicacea, heat responses similar to our observations have been reported. In green bean (*Phaseolus vulgaris*), the most sensitive period was 10 days before anthesis and could be related to a reduction in pollen viability (Suzuki 2001). This indicates a relationship with male meiosis, which corresponds to the sensitive floral stage 9 in *Arabidopsis thaliana*. In tomato (*Lycopersicon esculentum*), flower buds of 9-5 days before

anthesis and flowers of 1-3 days after anthesis are highly susceptible to high temperatures (Sugiyama 1965). These two periods correspond with the two minima identified in our experiment.

The finding that the present observations on Arabidopsis are in line with data collected for heat responses in various crops, increases the potential of translating the present results about molecular regulation of the heat response to crop species. Interestingly, besides heat stress also drought has an impact on developing flowers and embryos. In cereals two sensitive periods are found upon drought stress, the first one centers around the period of male and female meiosis and the second one during anthesis and initial grain development (Saini 1997). This suggests that similar regulatory mechanisms may be involved, increasing the relevance of the current work.

GWA mapping reveals stage specific QTLs

Although a few studies have reported on the developmental sensitivity window upon heat stress ((Warner 2005), (Kim 2001)), the present study is unique as it investigates this sensitivity window using a large natural population. Therefore, the data could be subjected to GWA mapping, which allowed the identification of stage-specific QTLs. Two choices in our experimental set-up were crucial in identifying stage specific QTLs. First, the 1-day exposure time ensured that developing flowers received stress in only one developmental stage. In experiments in which the stress is applied for longer periods the identification of stage specific QTLs is less accurate, because a single flower may receive stress in multiple developmental stages. Second, the quantification of the impact of the heat stress by measuring the length of individual siliques along the main inflorescence provided high resolution of development-specific heat effects. Studies that quantify the heat stress response by total seed or fruit yield will have less power to identify effects that are only present in a small developmental window. Compensation mechanisms of plants (Figure 4, (Angadi 2003), (Young 2004)) make total seed or fruit yield as a measure even less reliable. Our findings should increase awareness of the fact that the regulation of a trait of interest can change during development. Experimental design is a crucial factor here, which, in most cases, determines whether developmental stage specificity of QTLs or regulatory networks can be detected or not. Accurate measurements or large population sizes will not contribute to a higher power to detect stage specific QTLs when a non-optimal design is chosen.

In our study the main focus was on the identification of stage-specific QTLs. Two minima were identified in the silique length after heat stress corresponding to two distinct processes in flower and seed development (Figure 3 and Table 1). Stage-specific information can be helpful to prioritise candidate genes for future functional analyses. Genes that are not expressed in the developmental stage in which the QTL is identified or that are not expressed in flowers or embryos can be given lower priority. In this study four QTLs

were identified to be highly associated with heat response, one specific for the response after anthesis and three specific for the response before anthesis. Based on LD, annotated gene function, expression profiles and mutant analysis, one or two genes per QTL could be prioritized as candidate genes. The knockout analyses of several candidate genes indicated the involvement of these genes in heat stress responses of reproductive processes. However, for a number of genes their involvement could not be confirmed. This might be due to the genetic background of the mutant line (Col-0), whose wildtype is used for comparison. On the other hand, knockouts that do show a significant effect on the trait under study do not conclusively rule out the involvement of other genes explaining the observed variation in the support window of the QTL. Further evidence should be obtained from functional complementation studies in different genetic backgrounds.

In many studies describing heat stress during flowering a prominent role is given to heat shock proteins (Hsps) and heat shock transcription factors (Hsfs) (Giorno 2013). For instance, constitutive and heat stress induced accumulation of Hsps in reproductive tissue (especially anthers) is associated with heat tolerance (Bita 2011). Hsps stabilize proteins that are unfolding due to heat stress (Al-Whaibi 2011). Hsps, as chaperones, prevent the irreversible aggregation of other proteins and participate in refolding proteins during heat stress conditions (Waters 2012). Surprisingly, in our study no Hsp or Hsf was found in close proximity of a QTL (threshold –log(p)=5.5) and only two Hsps were identified within 20 kb of moderately associated SNPs (threshold –log(p)=4), suggesting that allelic variation in Hsps or Hsfs is not the main cause of natural variation in heat tolerance during flowering. The lack of these obvious candidates in our results shows that an untargeted analysis, like GWA mapping, will, in many cases, result in the identification of novel gene functions while hypothesis driven research, in most cases, results in the proof or dis-proof of an already suggested function.

Genetic co-regulation of heat response and flowering

A high negative correlation and co-incidence of strongly (Table 4) and moderately associated SNPs were observed for heat stress tolerance and flowering time. QTL6 was detected for heat stress response, but was also moderately associated with flowering time. The significant SNP of QTL6 was in linkage with SNPs in *FLC*, a flowering time repressor, and its promoter. The impact of heat stress was evaluated in two isogenic lines, one with the *FLC* functional allele, requiring vernalization to flower, and one containing a *FLC* null-allele resulting in early flowering. Heat stress led to significantly more reduction in silique length in the late flowering line compared to the early flowering line. Analysis of the six most common *FLC* haplotypes further confirmed that heat sensitivity is associated with late flowering and might be based on cross-talk between regulatory networks. Even though variation in heat stress response mapped to *FLC,* we cannot rule out an indirect effect of flowering time differences. Especially because *FLC* expression diminishes during the 10 weeks of vernalisation, severely reducing the variation in FLC expression during

the treatment (Michaels 1999, Shindo 2006). Although other known genetic factors affecting flowering time might also indirectly contribute to the variation in heat stress response, the analysis of this relationship was outside the scope of this study. A positive correlation was also found for chilling tolerance and flowering time in soybean, where high tolerance was associated with late flowering (Kurosaki 2004). Differences in tolerance were shown to be due to natural variation in a well-known maturation locus. While the response to non-optimal temperatures can be linked to natural variation in flowering time or maturation genes, we cannot exclude that the correlation found in our experiment was caused by cold priming. The time between the cold period, needed to synchronize the flowering, and the treatment is not equal for all accessions: accessions that flowered earlier inherently received the treatment sooner after the cold treatment than later flowering ones. It may be that recently cold-primed plants are less sensitive to the heat stress. This priming might be released over time, resulting in increased sensitivity to heat stress over time. However, natural variation was observed for accessions that received the heat treatment on the same day after the cold treatment, indicating that a major part of the natural variation cannot be explained by differences in the period between cold and heat treatment. We, therefore, propose that the variation in heat tolerance, observed between early and late flowering accessions, is a result of local adaptation. Early flowering accessions are summer annuals that germinate in spring and flower in the summer, whereas late flowering accessions are winter annuals that germinate in autumn, survive winter in the vegetative stage and flower in spring. This suggests that early flowering accessions will benefit from a heat tolerance mechanism in their flowers, whereas late flowering accessions will not, because flowers will rarely be exposed to high temperatures during spring.

Homology of heat stress responses between species

As mentioned, similar stage specific heat responses are observed in several crop species, suggesting common regulatory mechanisms of the response of plants towards elevated temperature during flowering, and fruit and seed set. Such a common response mechanism is promising for the translation of findings from model species to crops. Unfortunately comparison of the QTLs observed in this study with earlier identified QTLs in other mapping studies in crops aiming to identify heat tolerance genes during flowering, is difficult because most QTLs have not been cloned yet (Grilli 2007, Ye 2012, Lucas 2013). All earlier studies were done in low resolution bi-parental mapping populations and, therefore, the associated regions still contain many genes. Because none of the tested crops, rice, tomato and cowpea, are closely related to Arabidopsis, comparison of genome regions is difficult due to the lack of synteny. However, because of the high resolution of GWA mapping and the enormous amount of prior information available in Arabidopsis, it is promising to follow up the prioritized candidate genes identified in this study with further functional analyses to identify their functions and subsequently to be able to identify the heat response network. Translation of a response network from one

species to the other can be more fruitful than single gene comparisons (Ferrier 2011). Homologues of proteins in the pathway can be identified, and knowledge about protein function and expression, protein-protein and protein-metabolite interactions can be used to identify a similar response pathway in another species (Assmann 2013).

Although the focus of this paper is on silique length in heat stressed-plants, results obtained from plants grown in control conditions are also relevant for breeding purposes. To improve yield and consequently produce more oil, breeding is intensively done in *Brassica napus* and *B. juncacea* varieties. Total yield is determined by several plant traits, such as total fruits per plant, seed weight, seeds per fruit and also fruit length. The last three are correlated and many coinciding QTLs have been reported (Alonso-Blanco 1999, Dechaine 2014, Li 2014a). Fruit length is considered to be a suitable trait to select for in *Brassica napus* (Samizadeh 2010), because high heritability is observed for this trait which is less sensitive to GxE than, for example, the total number of fruits per plant (Dechaine 2014) (Chay 1989). In the current experiment high heritability was also observed for silique length (Table 3). Because Arabidopsis is a member of the *Brassicacea* family, similar mechanisms are likely to determine silique length and translation of mapping results from this model species to crops should be relatively easy.

Three candidate genes were identified for silique length in control conditions, which all contained a pollen allergen domain. Pollen allergens are known to cause allergic reactions in humans and are therefore studied intensively. Although their biological function in plants is not fully elucidated yet (Jiménez-López 2011), it is suggested that these proteins may weaken the cell wall of the maternal tissue to ease the penetration of the pollen tube into the stigma (Cosgrove 1997),(Mollet 2013). In tomato, competition experiments have been performed between wild type pollen and pollen in which one of these pollen allergens has been mutated. It revealed that pollen with the mutated protein were outcompeted by the wild type pollen because of slower pollen tube growth, indicating an important role of pollen allergens for the access of pollen to the ovule (Valdivia 2007). Silique length is correlated with the number of seeds per silique (Figure 1). Slower penetration of the pollen tube may result in less efficient fertilization and as a consequence less seeds will develop and siliques will be shorter. For clinical immunological reasons variation of these allergens is studied between species and within species (Radauer 2006) (Songnuan 2013), but the impact of natural variation on fertilization success has not been studied yet.

In conclusion, the chosen experimental design was very appropriate to determine developmental stage specific genetic regulators of the heat response of reproductive processes. Provisional validation by knockout analyses was provided for two pre-anthesis QTLs, strengthening the candidacy of two genes in close proximity of the significantly associated SNPs. Furthermore, the high resolution of GWA mapping, the similarities between flower and seed development in different species and the suggested common

regulatory mechanisms of heat response make the translation of findings in Arabidopsis to crops promising. To reach that goal, further functional analyses of candidate genes in Arabidopsis and their homologues in crops are needed.

Materials and Methods

Plant materials
A collection consisting of 285 natural accessions of Arabidopsis thaliana selected to capture most of the genetic variation present within the species and genotyped with approximately 215k SNP markers was obtained from the ABRC stock centre (Li 2010),(Baxter 2010), (Platt 2010),(Kim 2007).
For functional gene analyses, two lines, FRI-sf2 and FRI-flc3, containing two different alleles of FLC in a Columbia background (Michaels 1999), were kindly provided by M. Koornneef (MPI Köln). FRI-sf2 (CS6209) can also be obtained from the ABRC stock centre. In addition, the following SALK T-DNA insertion mutants (Alonso 2003) and GABI-Kat T-DNA insertion mutant (Kleinboelting 2012) were obtained from NASC: qul2-2 (SALK_065604c), AT2G03505 (GABI_745A07), AT2G03490 (SALK_125910), mips3-2 (SALK_120131c, Donahue 2010), AT5G10200 (SALK_040912c). Seeds of the mutants qul1-1 and qul1-1xqul2-1 were kindly provided by L. Østergaard (Fuentes 2010). Plants carrying a homozygote T-DNA insert were genotyped by PCR using the primers listed in Table S1.

Experimental set-up
For each accession, eight plants were grown in controlled conditions. Five replicates received a heat treatment and three replicates served as controls. Seeds were sown in petri-dishes on wet filter paper. After four days of cold treatment, they were placed at room temperature in the light for one and a half day to germinate. Germinated seeds were placed on rockwool blocks saturated with nutrient solution (Hyponex, 1mM N, 1.1 mM P, 5.9 mM K) in a climate room (16h light, 125 umol/s/m^2, 70% humidity, 20/18°C). Three times a week the rockwool blocks were saturated with nutrient solution using an automated flooding system. After three weeks, the plants were vernalized for 10 weeks to synchronize flowering. After vernalization the plants were returned to the climate room (same conditions as above). One to two weeks after the first replicate of an accession started to flower, the heat treatment was applied to all replicates of that accession. A small number of accessions received the treatment outside this window because of large variation in flowering time between the replicates.The first flower that opened on the day of the treatment was tagged with a thread. Three replicates per accession were kept in the climate room as controls. Five replicates per accession were transferred to a climate cabinet where they received heat treatment. At the start of the day, the temperature was raised from 20°C to 35°C in two hours. The temperature was kept at 35°C for 13,5 hours. At the end of the light period, the temperature was decreased again to 20°C in two hours. The day after the treatment the plants were returned to the climate room.
The same procedure was followed to test the heat response in FRI-sf2 and FRI-flc3 For the other T-DNA insertion mutants used in this study, the same procedure was used but without the vernalization period. All lines were tested in two independent experiments (Experiment A and B).

Harvest and analysis of main inflorescence siliques
When the plant stopped producing new flowers or when at least 40 flowers above the thread-tag had developed into siliques, the main inflorescence was cut off and stored in a paper bag for later analysis. Control and treated plants of the same accession were always harvested on the same day. Stored material was used to determine the length of all siliques present at the main inflorescence. The silique tagged with the thread was labelled as position number 0. Siliques below the tag received negative position numbers and siliques above the tag received positive position numbers (Figure 2).

Seed properties
To determine the number of seeds per silique and their seed size, seeds were spread on white paper

and pictures were taken in constant light conditions to enable easy image processing. The pictures were processed with the software package ImageJ (Schneider 2012). Based on the size distribution of dark-brown and light brown seeds a threshold of 0.12 mm^2 was used to define small and normal sized seeds, indicative for non-viable and viable seeds, respectively. To determine germination rates seeds were placed on wet filter paper in petri-dishes sealed with parafilm. The petridishes were placed in the dark at 4°C for 4 days. Thereafter the petri-dishes were placed in the light at room temperature. The numbers of germinated and non-germinated seeds were counted after 48h in the light.

Expression analysis QUL1 and QUL2
Expression of the *QUL1* and *QUL2* genes were measured in the *qul1-1, qul2-2, qul1-1xqul2-1* mutants as well as in Col-0 in three weeks old plant grown at 22°C in16hrs light. RNA was isolated from 20-100mg of plant rosette with the Invitrap Spin Plant RNA Mini Kit (Strateco) according to the manufacturer's instructions. For the cDNA synthesis, 1 µg of RNA was converted to cDNA using the iScript™ cDNA synthesis kit (Bio-Rad) according to the manufacturer's instructions. The cDNA was subsequently diluted to 150 µl end-volume for use in RT-qPCR. Absence of genomic DNA was confirmed by comparing cDNA samples with RNA samples which were not reverse transcribed (minus RT control). The RT-qPCR was performed on the CFX-Connect (Bio-Rad) in 96-well plates (HSP-9645; Bio-Rad) using per well, 10 µl total volume consisting of 2.5 µl of cDNA, 0.5 µl of 10 µM primermix, 5 µl iQ™ SYBR® Green (Bio-Rad) and 2 µl of water. The following program was used for every reaction: 95°C for 3min, followed by 40 cycles of 95°C for 15s and 60°C for 30s, followed by a melt curve analysis from 65 to 95°C. Each of the four biological replicates was measured in duplo. Expression of *QUL1* and *QUL2* was measured using the primers described in Fuentes et al. (2010). For the reference gene, *MONENSIN SENSITIVITY1* (*MON1, AT2G283900*), the primer set described in (Czechowski 2005) was used. Expression data was normalized with the $2^{-\Delta\Delta CT}$ method (Schmittgen 2008).

Statistical analysis
GWA mapping was performed on fertility and architectural traits, and flowering time. GENSTAT (VSN International 2013) was used to calculate the simple residuals, which characterize the impact of the heat, corrected for difference between accessions in the size of the siliques in control conditions. Simple residuals were calculated for the average silique length in the different regions along the inflorescence, representing different developmental stages. To calculate the simple residuals, the mean of data collected under control and heat stress conditions were compared by a linear regression (least sum of squares) including all accessions. GWA mapping was performed with the EmmaX software package (Kang 2010). A mixed model was used correcting for population structure based on an kinship matrix of all SNPs. SNPs with a minor allele frequency below 0.05 were excluded from the analysis. For each SNP, it was calculated how much of the total phenotypic and genotypic variance could be explained by the alleles of that SNP. Spearman's rank correlation coefficient (ρ) (2-tailed, α=0.05) was used to determine correlations between data series using SPSS (IBM Corp 2010). A one-sided, two-sample t-test, assuming unequal variances, was used to compare the mean number of flowers formed on the main inflorescence in control and heat treated plants.

All sequences from the re-sequenced Arabidopsis accessions were obtained from 1001genomes.org. for 525 accessions, 2012 nucleotide variation files compared to Col-0 (TAIR10) were downloaded. Custom Perl scripts were developed to determine positions with an allele frequency >2% (SNPS must be shared by more than 11 accessions). Another perl script parsed these positions per accession and outputs either an 1 or 0 for compliance or no compliance with Col-0. The resulting data is stored as data frames (.csv) on disk. In order to calculate the LD, required data is extracted from the .scv files with the gnu program 'cut' in order to slice out the region of interest. The sliced data frame is read in to R (R Development Core Team 2012) and column wise the LD (r^2 or correlation coefficient) can be determined by invoking the R function 'cor()' followed by a quadratic operation. LD scores are made available to the user via web interface (in house access only). The user can calculate the LD in any region on the genome

Supplementary data

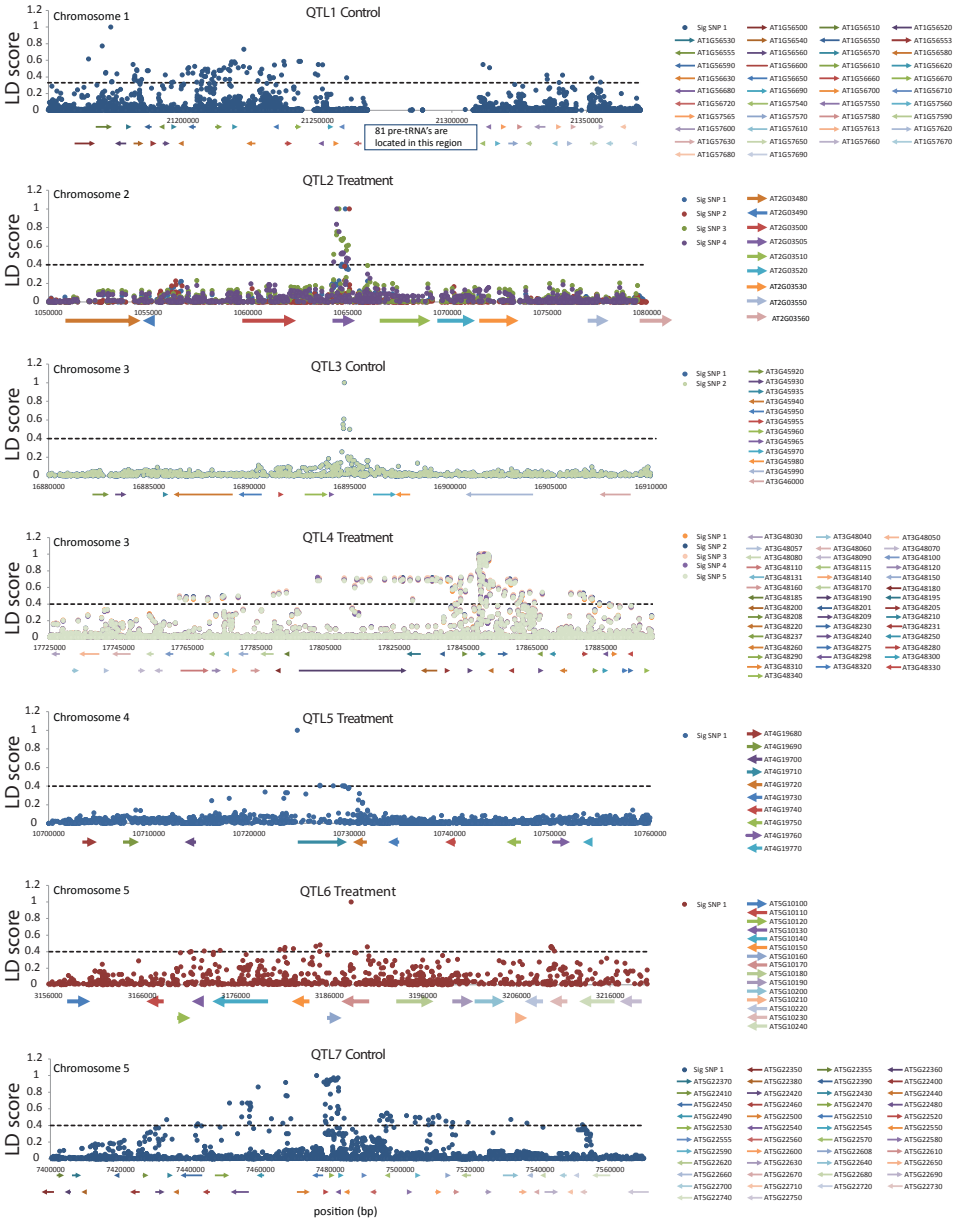


Figure S1 Local linkage decay of each QTL. LD between significant SNPs detected by GWA mapping and surrounding SNPs obtained from re-sequence data.

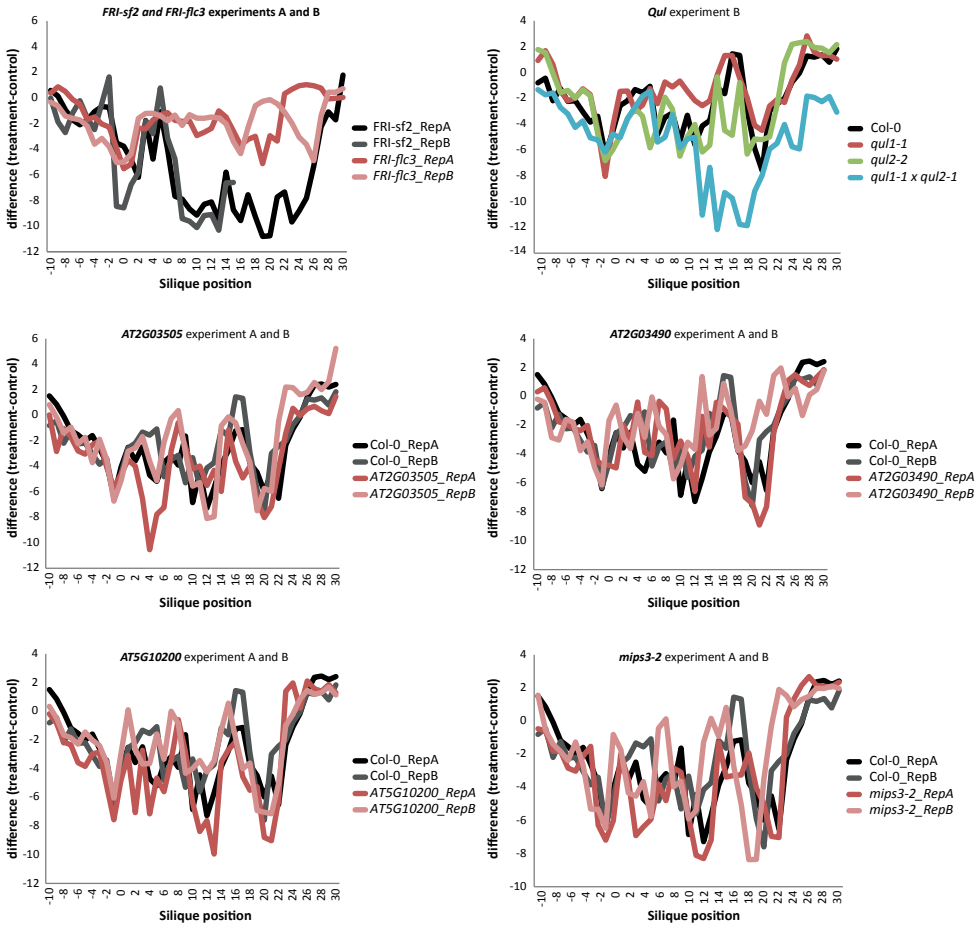


Figure S2 Silique length along the inflorescence of heat-stressed and control plants of several T-DNA insertion lines and two lines with functional (FRI-sf2) and non-functional (FRI-flc3) FLC alleles.

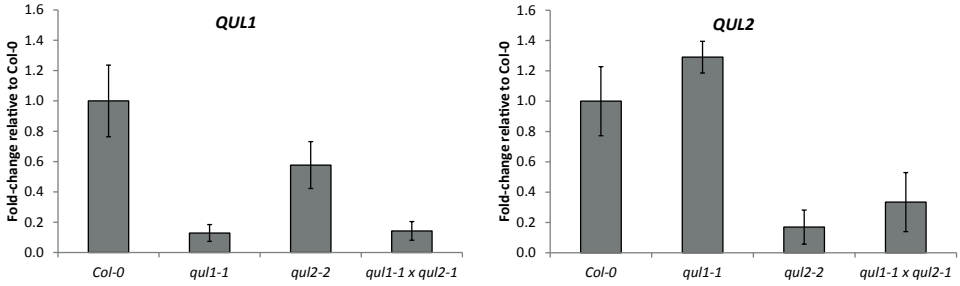


Figure S3 Expression of QUL1 and QUL2 in single and double mutants and Col-0 wildtype.

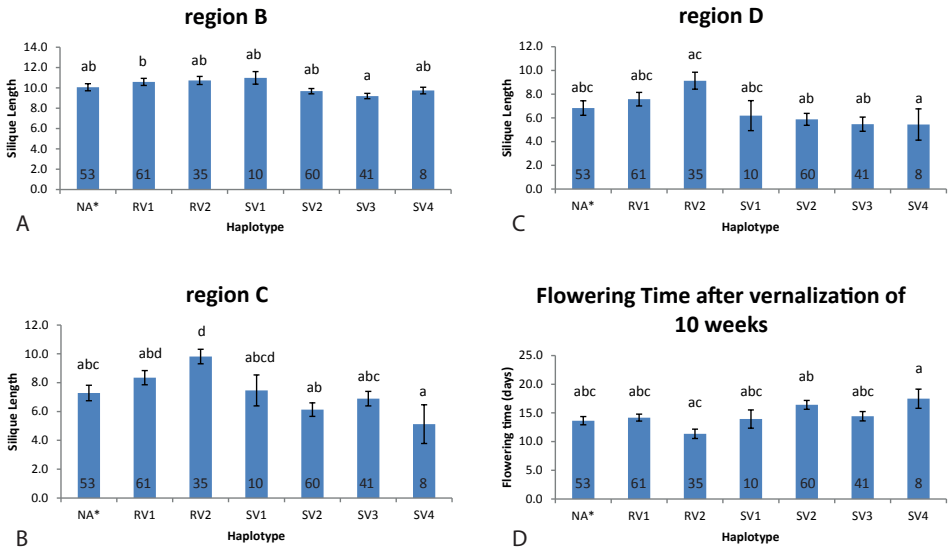


Figure S4 Comparison of silique length and flowering time of different FLC haplotypes.
ABC Length of siliques (average +/- SE) in region B, C and D upon heat stress. **D** flowering time (average +/- SE) after 10 weeks of vernalization for accessions of the six most common FLC haplotypes (Li et al., 2014b). Numbers in bars indicate the number of accessions belonging to the haplotype indicated below the bar, Col-0 belongs to the RV2 haplotype. Significant differences ($p<0.05$), as indicated by letters above the bars, are determined by one-way ANOVA with Bonferroni correction for multiple testing. Li et al., 2014b). *Accessions containg FLC haplotype different from the six most common haplotypes

Table S1 Primers used to determine homozygosity of T-DNA inserts.

Primer name	Sequences (5'-3')	Description
SALK_065604C_LB	TGAATGGTTTTTGACAGACGA	*qul2-2* genotyping
SALK_065604C_RB	CTGGTTGGTCTAGTTCGCAAG	*qul2-2* genotyping
GK-745-A07_LB	TTGGTCTTGTTTTAAAACCGC	*at2g03505* genotyping
GK-745-A07_RB	AACTTGCTTTGGTCCATCATG	*at2g03505* genotyping
SALK_125910_LB	TCAAAATGTCAAATCCTGAGC	*at2g03490* genotyping
SALK_125910_RB	GTGTATATTCGTGGCTGCCTG	*at2g03490* genotyping
SALK_120131_LB	AAGTCATCACCACCGATCAAG	*mips3-2* genotyping
SALK_120131_RB	ATTTCGGGTCATTAACCCAAG	*mips3-2* genotyping
SALK_040912_LB	TTGTCTCTATGCCCAATGAGG	*at5g10200* genotyping
SALK_040912_RB	TACAATCTCGGCCAATCATTG	*at5g10200* genotyping

Acknowledgement

We would like to thank Matthijs van Houwen, Rik Kooke, Laurens Deurhof and Jiaming Yang for their assistance during sowing and harvesting and their efforts to determine silique lengths. In addition, we like to acknowledge Henri van der Geest and Rik Kooke for the development of the LD-SNP tool. Finally, we thank Caroline Dean and Peijin Li for generous sharing of the FLC haplotype information, Maarten Koornneef for the transgenic lines FRI-sf2 and FRI-flc3, and Lars Østergaard for the qul1-1 and qul1-1xqul2-1 seeds.

Chapter 4 - Genome wide association mapping of growth dynamics detects time-specific and general QTLs

Johanna A. Bac-Molenaar[1,3], Christine Granier[2], Dick Vreugdenhil[1], Joost J.B. Keurentjes[3]

[1] Laboratory of Plant Physiology, Wageningen University, The Netherlands
[2] Laboratoire d'Ecophysiologie des Plantes sous Stress Environnementaux, UMR 759, Institut National de la Recherche Agronomique/Ecole Nationale Supérieure d'Agronomie, Place Viala, F–34060 Montpellier, Cedex 1, France
[3] Laboratory of Genetics, Wageningen University, The Netherlands

Abstract

Growth is a complex trait determined by the interplay between many genes, some of which some play a role at a specific moment during development whereas others play a more general role. To identify the genetic basis of growth, natural variation in Arabidopsis rosette growth was followed in 324 accessions by a combination of top-view imaging, high throughput image analysis, modelling of growth dynamics, and end-point fresh weight determination. Genome Wide Association (GWA) mapping of the temporal growth data resulted in the detection of time-specific quantitative trait loci (QTLs) whereas mapping of model parameters resulted in another set of QTLs related to the whole growth curve. The positive correlation between projected leaf area (PLA) at different time-points during the course of the experiment suggested the existence of general growth factors with a function in multiple developmental stages or with prolonged downstream effects. Many QTLs could not be identified when growth was evaluated at a single time-point only. Eleven candidate genes were identified, which were annotated to be involved in the determination of cell number and size, seed germination, embryo development, developmental phase transition or senescence. For eight of these a mutant or overexpression phenotype related to growth has been reported, supporting the identification of true positives. In addition, the detection of QTLs without obvious candidate genes implies the annotation of novel functions for underlying genes.

Journal of Experimental Botany, doi: 10.1093/jxb/erv176

Introduction

Plant growth is a dynamic process that is influenced by many external and internal signals the plant receives. For growth a plant needs light and carbon dioxide to perform photosynthesis to produce sugars, which are the building blocks and energy source for many processes in the plant. Besides that, the plant needs water and nutrients to be able to produce nucleotides, proteins and metabolites. Transport of sugars and other essential molecules from source to sink is important during all stages of growth. Perturbation of these source-sink relationships by changes in the environment or due to the genetic composition of the plant may lead to changes in biomass accumulation. A better knowledge about the genetic factors that are involved in growth regulation would help to understand the mechanisms underlying different growth patterns as observed in nature. Such dynamic patterns are better understood when growth and its regulation are studied over time, instead of at a single time-point (Leister 1999, Granier 2006, Tessmer 2013).

Growth is orchestrated precisely and is controlled by many genes. The functional importance of most growth-related genes is not equal during all developmental stages and in all tissues and many display specific temporal and spatial expression profiles (Schmid 2005). In addition, some genes play an essential role in the overall development of the plant, whereas others are mainly important if the plant has to cope with specific environmental conditions (Geng 2013). These tightly regulated genes form a robust network that enables the plant to complete its life cycle under many different circumstances.

Growth patterns of plants may differ widely between species (Westoby 2002), but also within the same species (Koornneef 2004, Alonso-Blanco 2009, Zhang 2012). Within species, the observed variation can be caused by differences in the local environment or due to natural genetic variation. This genetic variation is a result of random mutation and meiotic recombination and can result in plants that, as a result of centuries of selection, are adapted to the local environment. The growth differences observed between and within species indicate that the regulation of growth is not only robust, but also genetically variable. Natural variation of growth within the same species can be used to search for genes regulating growth (Alonso-Blanco 2009). When growth phenotypes are determined in mapping populations, which are genotyped with many markers, linkage between genotypes and phenotypes can be identified by statistically testing association between molecular markers and observed phenotypes. Many mapping studies have been performed for plant growth and size resulting in the identification of many QTLs (Alonso-Blanco 1999, El-Lithy 2004, Lisec 2008, Chardon 2014). However, in those mapping studies biomass was determined at the end of the experiment to evaluate the differences in growth (end-point measurements). As a result only major players in the regulation of growth, like genes that orchestrate the transition from a vegetative to a generative state (e.g. FLC) (Kowalski 1994) or genes related to dwarf growth (erecta locus or ga20ox1)

have been cloned and confirmed (Komeda 1998, Barboza 2013). However, most of these major players were identified in experimental mapping populations in which only a few QTLs segregate or are artificially introduced (e.g. erecta). Additional players explaining large part of the plant size variation observed in nature seem to be scarce. The mapping studies suggest that growth is a complex trait and that many genes are involved in the regulation of the accumulation of biomass. Genome wide association mapping studies (GWAS) might help to identify these genes because a much larger fraction of a species' diversity is analysed.

Because growth is a dynamic process, the timing of the end-point measurement will highly influence the outcome of the mapping (El-Lithy 2004). High throughput automated phenotyping creates the possibility to follow the growth, or at least the 2D expansion, of plants over time in a non-invasive way (Furbank and Tester, 2011). For plants with a two-dimensional structure, like rosettes of Arabidopsis, the challenge of high throughput imaging is not only the capturing of pictures, but also the automatic image analysis. Different approaches dealing with this issue have been described: A pipeline to automatically determine rosette size (Hartmann 2011), rosette size and compactness (Arvidsson 2011), rosette shape (Camargo 2014) or rosette size accounting for leaf overlap (Tessmer 2013). The next challenge is how to use the additional information present in these time-series data for mapping purposes. Several ways have been described in which data collected over time can be combined with mapping, but no standard approach is agreed upon. The simplest approach is to treat data of different time points as unrelated traits and perform mapping for each time-point separately, here referred to as univariate mapping of trait per time-point (Moore 2013, Wurschum 2014). This approach resulted in the identification of time-specific QTLs for root bending upon a change in the direction of gravity in Arabidopsis (Moore 2013) and for plant height in wheat (Wurschum 2014). Another approach is to perform a two-step procedure. First a growth model is fitted to the growth data, after which the model parameters that describe the characteristics of growth, are used in a standard mapping approach, here referred to as univariate mapping of model parameters (Heuven 2010). This approach resulted in the detection of QTLs for leaf elongation rate in maize (Reymond 2004). A similar approach was used to perform mapping on germination data, resulting in e.g. the detection of QTLs that are related to uniformity of germination (Joosen 2010). Finally, growth data collected over time can be analysed with multivariate mapping approaches (Ma 2002, Malosetti 2006, Yang 2011). The mapping power of multivariate approaches is higher, because they take into account that growth data collected over time and the derived parameters may be correlated, while univariate methods ignore this fact (Wu 2006). Multivariate mapping can be done in a two-step approach in which a growth model is fitted to the growth data, after which the model parameters are used in a multivariate mapping approach (Korte 2012). This can also be applied in a one-step approach that uses one statistical model in which the molecular marker information and the parameters of a growth model are both included (Ma

2002). A multivariate approach was e.g. used to detect QTLs for progression of senescence over time in potato (Malosetti 2006), for leaf age based on the length and number of leaves emerging over time in rice (Yang 2011) and for stem diameter in poplar (Ma 2002). Univariate methods can be performed with standard mapping software packages, while multivariate mapping requires dedicated software (http://statgen.psu.edu/software/funmap.html, (Korte 2012)), which has hampered the adaptation of multivariate mapping by a broad community. Although each of the described methods have their own drawbacks they clearly show that mapping of data collected over time results in the identification of QTLs that would not have been detected if only single-point measurements were used as input for the QTL analyses.

Here, we describe a series of analyses that enable us to observe growth dynamics by automatic imaging. We will show that top-view imaging of Arabidopsis plants in combination with high throughput image analysis can be used to follow rosette growth over time in a large and diverse population of natural accessions. We will further show that comparison of accessions demonstrating large variation in developmental rate and in plant size can be done by modelling of growth. In addition, Genome Wide Association mapping on temporal plant size data was performed using univariate and multivariate mapping approaches. Time specific growth QTLs were detected by performing univariate GWA mapping for each time-point separately, whereas general QTLs related to growth rate during the course of the whole experiment were identified by performing univariate and multivariate GWA mapping on the growth model parameters. Finally, candidate genes involved in the regulation of growth could be indicated.

Results and Discussion

Capturing the dynamics of growth by top-view imaging

A large-scale experiment was performed in the plant phenotyping platform PHENOPSIS (Granier 2006). Three-hundred-twenty-four natural accessions of *Arabidopsis thaliana* were grown and their rosette sizes were monitored over time by capturing top-view pictures daily. The plant architecture of the vegetative stage of Arabidopsis makes this species very suitable for top-view imaging. Because the rosette grows in a horizontal plane, it can be approached as a 2D structure of which the size can be determined accurately from top-view images. Top-view imaging of Arabidopsis rosettes was first reported in the 90s (Leister 1999), but became suitable for large populations only recently because of advances in the automation of image analysis (Berger 2010, Arvidsson 2011, Tessmer 2013). Although at the moment low-cost, high-throughput methods are available to determine the genome of an organism and genetic information is available for many species and for many mutants and natural accessions, the plant science community lags behind in the high throughput measurements of phenotypes (Houle 2010). In this experiment

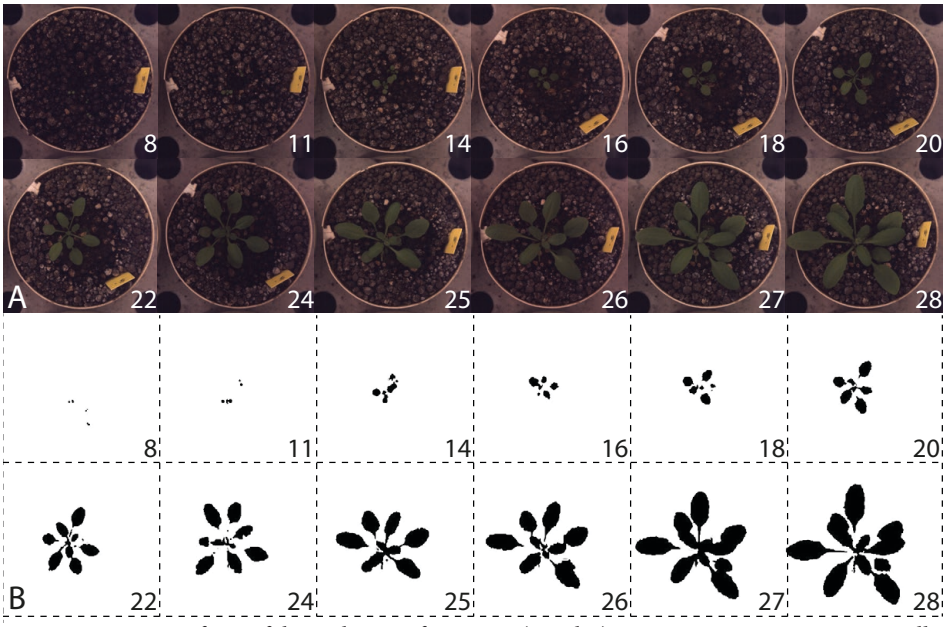


Figure 1 A Images of one of the replicates of CS28014 (Amel-1), a representative accession, on all time-points included in the analyses. **B** Pictures processed by ImageJ to determine the Projected Leaf Area. Pictures were segmented based on colour, saturation and brightness and thereafter made binary. Too small particles (<120 pixels) were excluded from the analysis. In the images of days 8, 11, 14 more than one plant is present, but only the remaining one (days 16 and onwards) is taken into account for PLA determinations.

top-view imaging in combination with high throughput image analysis allowed us to determine the rosette size of plants of 324 accessions in triplicate on 11 time-points during the growth. Projected leaf area (PLA) was determined from day eight onwards and the experiment was ended before too many leaves were overlapping (Figure 1). On day eight all seeds had germinated, the cotyledons were unfolded, but the first true leaves were not yet visible. As the growth rate increased during the course of the experiment, the interval between the time points of PLA determination was decreased, from a three-days interval in the second week to a one-day interval in the fourth week, to ensure that dynamics in growth were accurately captured. Because diurnal leaf movement was observed, PLA was always determined within two hours after the start of the light period. This analysis is one of the first steps in the detailed characterization of the phenomes of these natural accessions (Furbank 2011). Similar approaches can also be used in the future to further characterize the phenomes of these natural accessions by performing similar experiments when plants are grown in different and possibly less favourable conditions, like short-days or under abiotic or biotic stress. For much smaller sets of accessions similar experiments are previously performed, but to be able to use the phenotypes in mapping studies much larger populations need to be screened (El-Lithy 2004, Granier 2006). Growth was not only determined by differences in PLA over time, but also at the end of the experiment by measuring fresh weights of the rosettes (FW).

For PLA and FW, large natural variation was observed of which 28-70% could be explained by genetic differences (Table 1). Broad-sense heritability (H^2) of PLA increased over time (Table 1), most likely because determination of the PLA of small plants was less accurate than that of larger plants. These data demonstrate that top-view imaging of Arabidopsis is a powerful method to compare plant size and growth rate in large panels of plants which do not only differ in size but also in developmental traits like flowering time (Li 2010), number of leaves and leaf emergence rate (Granier 2006, Tisné 2010). FW at the end of the experiment correlated positively with PLA at the end of the experiment (r^2=0.95), as shown earlier (Leister 1999). This high correlation is also reflected in almost equal H^2 of FW and PLA at day 28 (H^2=0.69 and H^2=0.70, respectively). FW also correlated with PLA in week two and three (Figure 2). In the last week of the experiment leaves started to overlap and variation for this trait was observed between accessions. Despite this increase in overlap over time, the correlation between FW at day 28 and PLA at the sequential measuring dates increased over time, reaching the highest correlation at day 27 (r^2=0.96). This correlation suggests the existence of general growth factors of which the effects are visible at the phenotype level during a large part of the plants life cycle. Seedling size at day eight, when the cotyledons are unfolded, but the first true leaves are not yet visible, is for a large part determined by seed size, germination rate and the capacity of the seedling to establish. The correlation of PLA during the experiment also suggests that the effect of genes involved in the regulation of these processes are visible at the phenotype level when seedlings develop into plants with many leaves. The water status of the plant was evaluated by the determination of the Water Content (WC) of the largest leaf on the 24th day. A proper water status is important for the plant to maintain growth. WC was high for all plants (between 0.85 and 0.95) indicating that in our conditions the water status was not limiting for growth. This corresponds with small variation in WC observed in a collection of 20 accessions (El-Lithy 2004). Significant but very weak correlations were observed between WC and A_0, r, FW and PLA at day 8, 27 and 28, whereas the correlation between WC and plant size on other days was not significant. Because of the low vari-

	Days	Avg	Min	Max	Stdev	H^2
FW (g)	28	0.26	0.01	0.74	0.12	0.69
PLA(mm^2)	8	6	1	42	3	0.28
	11	15	3	48	7	0.52
	14	36	4	153	16	0.51
	16	61	6	185	29	0.52
	18	110	8	362	54	0.55
	20	188	21	590	91	0.55
	22	314	14	986	153	0.59
	24	495	32	1438	235	0.62
	25	520	27	1627	257	0.61
	26	640	40	2041	312	0.62
	27	769	43	2450	362	0.65
	28	911	52	2832	417	0.70
Expo2: A_0		5.33	0.21	47.22	4.49	0.63
Expo2: r		0.19	0.10	0.31	0.02	0.28

Table 1 Natural variation and broad-sense heritabilities for growth traits and growth model parameters. FW: Fresh Weight of Rosette in gram. PLA: Projected Leaf Area of Rosette in mm2. Expo2: Growth model using an exponential function with two parameters, r (growth rate) and A0 (initial size and magnification factor). Days: days after transfer from cold to climate room. Avg, Min, Max, Stdev: Average value, minimum value, maximum value and standard deviation observed for the indicated trait on the indicated date, individual plants are used instead of averages per accession. H2: Broad-sense heritability

		FW	Projected Leaf Area												Expo2		WC
		28	8	11	14	16	18	20	22	24	25	26	27	28	A0	r	24
FW	28		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.006	.010
PLA	8	0.50		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.031
	11	0.70	0.73		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.107
	14	0.71	0.67	0.89		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.321
	16	0.78	0.67	0.89	0.92		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.514
	18	0.84	0.65	0.86	0.90	0.97		.000	.000	.000	.000	.000	.000	.000	.000	.000	.849
	20	0.87	0.63	0.85	0.88	0.95	0.98		.000	.000	.000	.000	.000	.000	.000	.000	.921
	22	0.90	0.59	0.82	0.85	0.92	0.96	0.98		.000	.000	.000	.000	.000	.000	.027	.531
	24	0.92	0.57	0.79	0.83	0.89	0.94	0.96	0.98		.000	.000	.000	.000	.000	.387	.210
	25	0.94	0.55	0.77	0.81	0.88	0.93	0.95	0.97	0.98		.000	.000	.000	.000	.899	.081
	26	0.95	0.55	0.76	0.79	0.86	0.91	0.94	0.96	0.98	0.99		.000	.000	.000	.256	.053
	27	0.96	0.53	0.74	0.76	0.84	0.89	0.92	0.95	0.97	0.98	0.99		.000	.000	.025	.045
	28	0.95	0.52	0.71	0.74	0.81	0.87	0.89	0.93	0.95	0.96	0.97	0.98		.000	.001	.033
Expo2	A0	0.28	0.44	0.55	0.59	0.60	0.59	0.59	0.54	0.48	0.43	0.38	0.34	0.29		.000	.006
	r	0.15	-0.26	-0.32	-0.32	-0.30	-0.24	-0.21	-0.12	-0.05	0.01	0.06	0.12	0.19	-0.71		.000
WC	24	0.14	-0.12	-0.09	-0.05	-0.04	-0.01	0.01	0.03	0.07	0.10	0.11	0.11	0.12	-0.15	0.27	

Figure 2 Pearson correlations between Fresh Weight of Rosette (FW) at the end of the experiment (day 28), Projected Leaf Area (PLA) over time (day 8 till 28) and parameters 'r' and 'A_0'of growth model Expo2. r^2-values are given in the left-down part of the figure, whereas corresponding p-values are given in the right-up part of the figure. Blue and red indicate positive and negative correlations, respectively. The stronger the intensity of the colour, the stronger the correlation.

ation observed WC did not play a prominent role in determining growth differences in this experiment. Because PLA of the rosette was on average doubled during the last four days of the experiment we decided not to correct for the absence of the largest leaf. In the growth curve of some accessions between day 24 and 25 a dip is observed, however for many accessions this dip was hardly visible, suggesting a huge compensation investment in the growth of the remaining leaves. Without correction for the absence of the leaf, the growth modelling resulted in very reliable curve fits for the Expo2 and Gom, indicating that growth rate was hardly influenced by the removal of the largest leaf.

Natural variation in bolting time was also observed among the accessions analysed. Plants that started bolting before the end of the experiment were significantly larger than vegetative plants. A similar pattern was observed when plants were classified as winter or summer annuals of which the first require vernalisation to flower. Summer annuals, of which many flowered at day 28, were significantly larger than winter annuals, of which none were flowering in our experiment. Winter annuals germinate in autumn and survive winter as small plants, for which fast growth is not a priority (Gazzani 2003, Grennan 2006). Summer annuals, on the other hand, germinate in spring and have to finish their life cycle before the dry and hot summer period and fast growth might, therefore, be an advantage. When grown in optimal growing conditions these properties may result in the observed differences in size.

Comparison of models to describe early vegetative growth

To be able to quantify the dynamics in rosette growth over time, growth was modelled using different mathematical functions (Table 2). Determinate growth, i.e. growth that terminates before the end of the life cycle of an organism, can in many cases be described

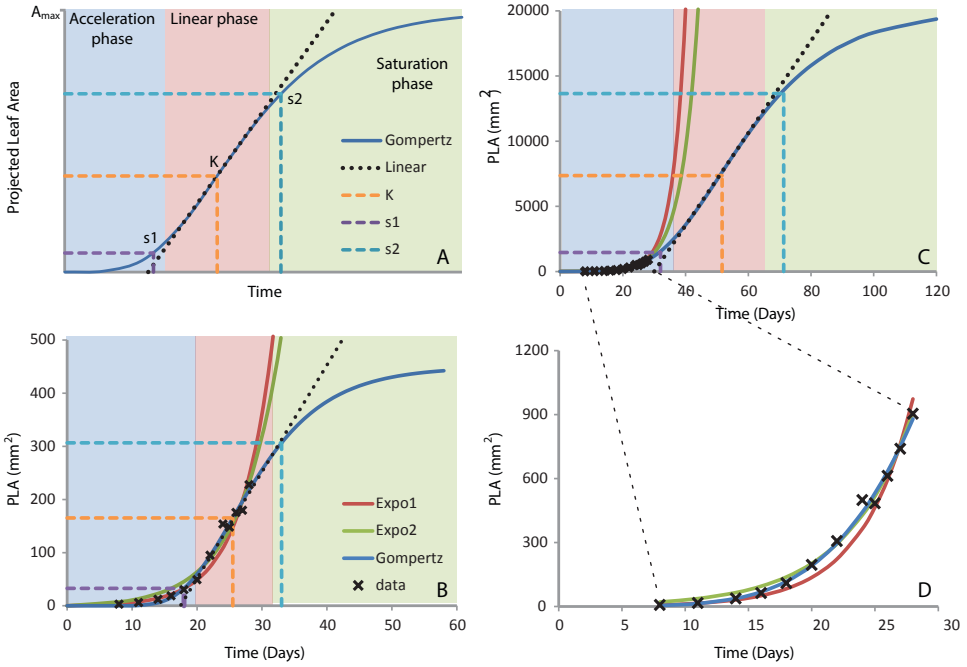


Figure 3 Modelling of growth using three mathematical functions, Expo1, Expo2 and Gom (see Table 2 for details). **A** S-curve as model for determinate growth consisting of three phases: Acceleration phase, Linear phase and Saturation phase. The S-curve has three characteristic points: The inflection point 'K' where the growth changes from increasing to decreasing, the point of maximal acceleration 's1' and the point of maximal deceleration 's2'. **B** Data and curve fits for line CS76226 (Se-0), representative for the 11% of the plants for which the growth curve reached K. The data were fitted to the three models. It shows that even if the inflection point 'K' is reached, model Expo1 and Expo2, which both assume indeterminate growth, resulted in good fits. **C.** Data and curve fits for line CS76308 (ZdrI2-25), representative for 56% of the plants for which the growth curve did not reach 's1'. The data were fitted to the three models. **D** Zoom of C allowing better comparison of the various models for the time-window of the experiment.

by a sigmoid function (S-curve). Rosette growth of *Arabidopsis* is known to be determinate, following such a S-curve (Boyes 2001). S-curves are characterized by an accelerating phase, a linear phase, and a saturation phase (Figure 3A). Within the linear phase, which is not really linear, but can be approached by a linear function, the inflection point 'K' is located. At 'K' the curve changes from increasing growth to decreasing growth. Near the end of the acceleration phase, which can be approached by an exponential function, the point of maximal acceleration 's1' is located. Near the beginning of the saturation phase the point of maximal deceleration 's2' is reached and, thereafter, the growth gradually stops and the final rosette size 'A_{max}' is reached. Determinate growth was modelled using the Gompertz function (Gom), which results in an S-curve (Gompertz 1825, Winsor 1932), (Table 2). This function is a slightly modified form of the basic logistic function, which was first described by Pierre Verhulst in 1838 (Verhulst 1838). The modifications

Table 2 Mathematical functions used to model growth and their properties

Model	Formula	Description parameters	Remarks
Expo1	$\mathrm{PLA} = e^{r*t}$	r: Growth rate	A_0=1
Expo2	$\mathrm{PLA} = A_0 * e^{r*t}$	A_0: Initial size r: Growth rate	A_0 is also a magnification factor
Gompertz	$\mathrm{PLA} = \mathrm{A_{max}} * e^{-b*e^{-r*t}}$	A_{max}: Final rosette size b: Position along time-axis r: Growth rate at inflection point K	Limits used for fitting of data: A_{max} = 20000, b=60

of the basic logistic function change this basic symmetric growth curve into an asymmetric one. The Gompertz function used here contains three parameters: 'A_{max}', the final rosette size; 'b' that determines the position of the curve along the time-axis; and 'r' that determines the growth rate at the inflection point 'K' (Table 2). The combination of these three parameters determines on which day 's1', 's2' and 'K' are reached. As we investigated the growth curves fitted with Gom, none of the plants in this experiment reached 'A_{max}' and only 4% reached 's2' within the window of the experiment. Even 'K' and 's1' were not reached by the majority of the plants: for 89% of the plants 'K' was not reached before day 28 and for 57% of the plants 's1' was not reached before day 28. This means that for most plants the collected data points are located in the accelerating phase and the beginning of the linear phase of the growth curve, implying that the plants did not yet enter the saturation phase. This was expected for the plants that did not bolt yet, but for the 30% of the plants that were bolting on the last day it was expected that they would have reached at least the saturation phase, because earlier studies reported that Arabidopsis rosettes reach the final size when they start to flower (Boyes 2001). Because most plants were in the acceleration phase even on the last day of the experiment, the growth was not only modelled with the Gompertz curve that describes determinate growth, but also with models that describe indeterminate growth, e.g. exponential growth. The simplest indeterminate growth model used was based on an exponential function (Expo1) with one parameter 'r', which represents the growth rate (Table 2). This model assumes that the growth rate is equal during the whole growth period and that the initial size (A_0) is 1 (Table 2). Exponential growth was also modelled with a function (Expo2) with two parameters, growth rate 'r' and the initial size (A_0) (Table 2). A_0 does not only represent the starting value it is also a magnification factor. This means that for two plants with equal 'r' and a factor two difference in A_0, plant size is also a factor two different during the whole experiment. To illustrate the use of the three models, data from two plants that showed determinate and indeterminate growth were used for curve-fitting (Figure 3 BCD). The plant with determinate growth is representative for 11% of the plants in which growth reached 'K' within the course of the experiment (Figure 3B). The plant with indeterminate growth is representative for the 56% of the plants for which growth did not reach 's1' within the course of the experiment (Figure 3CD).

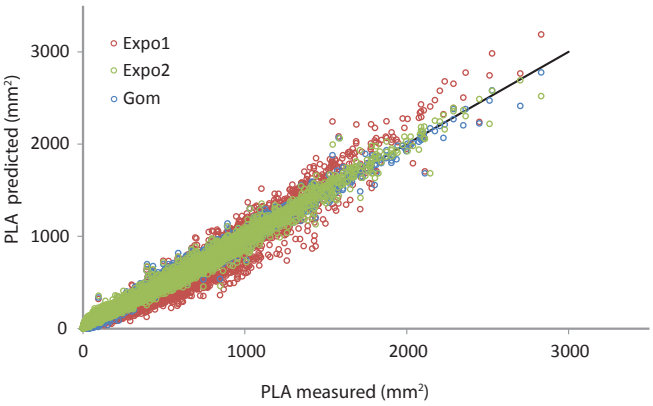


Figure 4 Comparison of the goodness-of-fit for the three growth models used: Exponential function with one (Expo1) or two (Expo2) parameters and Gompertz function (Gom). Plot of the measured PLA on days 8, 11, 14, 16, 18, 20, 22, 24, 25, 26, 27, and 28 against the predicted PLA on the same days. Black line represents y=x (PLA measured = PLA predicted).

For each model the goodness-of-fit was evaluated (Figure 4). As expected, r^2 was on average higher when more parameters were introduced into the model . Expo1 predictions were in general too low at small PLA and too high at large PLA (Figure 4), which indicates that this model is too simplistic. Interestingly, the differences in goodness-of-fit between Expo2 and Gom were not large, emphasizing that most plants in this experiment do not reach K and that growth rate thus does not decrease significantly during the duration of the experiment. So, for most plants in this experiment, determinate growth cannot be concluded from the PLA data collected. This was supported by the smaller confidence intervals for the parameters of Expo2 compared to the parameters of Gom. Especially for A_{max} very large confidence intervals were observed. Based on Figure 4, the confidence intervals and the principle of parsimony, stating that the simplest of two competing models is to be preferred, Expo2 was chosen to be used in the GWA analyses. This model is counter-intuitive because it describes indeterminate growth, while it is known that the Arabidopsis rosette follows determinate growth (Leister et al., 1999). In this case a model describing determinate growth, like the Gompertz function, results in parameters that are more informative (or speculative) for growth outside than inside the experimental window. Growth models that describe an S-curve always contain a parameter representing the final rosette size and the other parameters that are estimated are dependent on that parameter. In our case, Gom, which describes an S-curve, would have functioned as a (not very reliable) predictive model instead of a descriptive model as was aimed for. If curve fitting using the growth model results in reliable fits, as it did for most plants in this experiment, it allows for comparison of plants which differ in developmental timing, growth rate and plant size. However, this comparison only leads to valuable insight if the right model is chosen. Conclusions based on a non-optimal model, should be interpreted carefully as they easily can become very speculative (Tessmer et al., 2013).

Quantification of growth dynamics by exponential growth model parameters

Moderate to high heritabilities was observed for the growth parameters estimated by Expo2 (Table 1), indicating that these growth characteristics are determined partly by

the genotype. The parameter 'r' of Expo2 is weakly correlated with plant size in week 2 and week 3, but no correlation was observed with plant size at the end of the experiment (Figure 2). This indicates that the growth characteristic represented by this parameter goes beyond simple information about plant size. The model assumes a certain growth pattern and, given the parameters, the model describes how this growth exactly takes place. So the mathematical function used in the model expresses the overall shape of the growth curve shared by all the data. The details of the growth curve, determined for each individual plant, are described by a specific set of model parameters. Parameter 'A_0', which represents the initial size of the plant is positively correlated with plant size in week 2, 3 and 4. This is in accordance with the observation that the size of the plant at different time-points is highly correlated throughout the whole experiment. Natural variation is observed for seed size and seed germination (Schmuths 2006, Vallejo 2010, Herridge 2011) and these traits have also been determined for the accessions used in this experiment (Joosen 2013). Correlation between A_0 and PLA determined in this experiment and the seed traits reported by of Joosen *et al.* (2013) is limited: only weak (r^2 between 0.1 and 0.3) but significant correlations were found for seed size (dry and imbibed), but not for germination traits. The environment in which the parental plant matures has a large impact on seed weight and germination rate (Elwell 2011), and therefore differences between seed batches are expected. Thus, germination rate could have influenced plant size in our experiment although no correlation was found between seed germination traits measured in Joosen *et al.* (2013) and PLA and A_0 determined in this experiment. A strong negative correlation was found between the two model parameters 'A_0' and 'r' (r^2=-0.71), which can partly be explained by the fact that a relatively high value for A_0 (A_0>10) was never found in combination with a relatively high r (r>0.20) (Figure S1). This has to do with the boundaries of the natural variation. Probably, for Arabidopsis, too fast growing is not favourable in nature and therefore gene combinations that would lead to both a large A_0 and a large value for r, and hence would result in enormous plants, are not selected for. In addition, enormous plants are probably also physically not possible. Taking all these correlations into account it can be concluded that determination of growth over time and subsequent modelling of growth results in the quantification of growth dynamics that gives insight into the growth patterns that could not have been obtained from single-time-point measurements.

Added value of dynamic growth modelling in GWA mapping

To identify the genetic basis of growth GWA mapping was performed on PLA data (12 different dates), FW data (end point) and on the parameters derived from the selected growth model Expo2 (Figure 5). Parameters of models with fits of r^2<0.9 (11 out of 965) were excluded to avoid bias in detected associations due to outliers created by poor fits. PLA, FW and model parameters were mapped as independent traits, even though they display co-variance. The two parameters of Expo2, 'r' and 'A_0' were also mapped simultaneously using a Multi Trait Mixed Model (MTMM) approach, which takes cova-

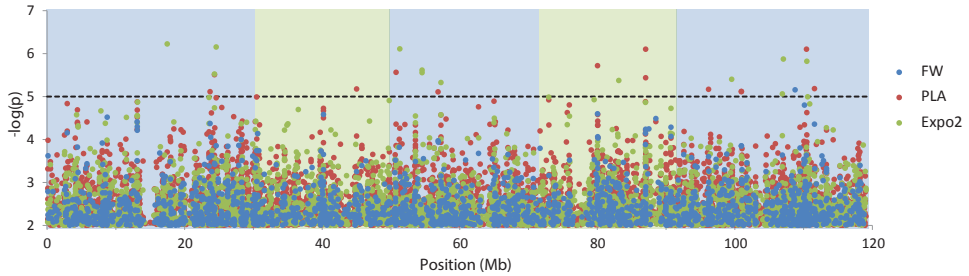


Figure 5 Genome wide association mapping of FW, PLA and parameters of growth model Expo2. Univariate GWA analyses were performed for all traits, in addition the model parameters 'r' and 'A_0' were also analysed together in an MTMM-GWA approach. Manhattan plot of the –log(p) marker-trait association for FW, PLA and model parameters of Expo2 is shown. PLA on the different days are represented by one value, for each SNP only the –log(p) value of the day with the highest association is plotted. Univariate analyses of 'r' and 'A_0' and the MTMM analyses of 'A_0' and 'r' jointly are also represented by one value, for each SNP only the –log(p) value of the analysis with the highest association is plotted. The total number of tested SNP markers was 214K, but only the ca. 10K SNPs with –log10(P)>2 are plotted. Dotted line indicates the arbitrary threshold of –log(p)=5.

riance of parameters into account (Korte 2012). In total 26 SNPs were highly associated (–log(p)>5) with one or more of the traits. One of these SNPs was associated with FW, 13 SNPs were associated with PLA and 12 SNPs were associated with the model parameters. For each of these 26 strongly associated SNPs an association profile was made to identify whether associations were specific for a trait or time-point or whether they were more general (Figure 6). SNPs displaying a profile with strong associations for FW and PLA over time, were not or moderately associated with the Expo2 parameters (Figure 6B). For example, the association profile of two SNPs at chromosome five at 8.8 Mb, was moderate to high for PLA at week three and four (-log(p) between 3.88 and 5.11), moderate for FW (-log(p)=3.85), and low for model parameters (-log(p)<2). This trend was also observed conversely, although quite some SNPs that were highly associated with model parameters were also found to be moderately to highly associated with PLA at some time points (Figure 6). For example, the association profile of the SNP at chromosome three at 1.2 Mb that was high for parameter 'A_0' (-log(p)= 6.15) and for the multi-trait analysis of 'r' and 'A_0' (-log(p)=5.33) was also high for PLA in the third week (-log(p) between 4.01 and 4.97). Remarkably, SNPs that were highly associated with model parameters were never associated with FW at day 28. This emphasizes that the model parameters reveal characteristics of growth that would not have been detected if only final plant size data were considered. Growth modelling, therefore, resulted in the detection of QTLs that would not have been found otherwise. Nonetheless, GWA mapping of model parameters cannot replace GWA mapping of plant size data, because both methods resulted in the detection of unique, highly associated, SNPs. SNPs that were selected because of strong association with PLA at a specific time point had an association profile for PLA that was relatively high (-log(p)>2) during the whole course of the experiment. This observation is in accordance with the significant positive correlation between PLA at different time points

chromo-some	position (Mb)	FW	PLA (mm^2)												Expo2		
		28	8	11	14	16	18	20	22	24	25	26	27	28	A_0	r	full
1	24.6	0.67	3.25	3.42	3.70	3.12	2.90	2.76	2.73	2.22	2.02	1.60	1.31	1.12	6.22	1.81	5.96
3	0.6	1.04	1.75	2.38	3.39	3.40	3.14	2.69	2.16	1.81	1.40	1.14	1.04	0.82	5.55	4.91	4.84
3	1.2	1.60	1.48	3.50	4.85	4.97	4.35	4.01	3.08	2.93	2.58	2.02	1.83	1.98	6.15	2.49	5.33
3	4.4	0.75	0.07	0.26	0.19	0.45	0.18	0.22	0.10	0.10	0.32	0.45	0.64	0.94	3.53	5.61	4.09
3	4.4	0.65	0.08	0.29	0.19	0.54	0.27	0.31	0.17	0.03	0.22	0.34	0.51	0.78	3.64	5.52	4.01
3	7.2	2.16	0.39	0.62	0.73	0.97	1.30	1.44	1.61	1.73	2.06	2.45	2.68	2.96	0.03	3.24	5.37
4	9.6	0.71	1.61	2.21	2.51	2.09	2.03	1.97	2.12	1.84	1.55	1.20	0.97	0.97	5.40	1.16	5.31
5	14.9	2.02	0.55	2.02	2.87	2.75	3.34	3.24	3.56	3.11	3.00	1.98	2.05	1.66	5.87	1.55	5.46
5	18.3	1.96	2.31	4.26	3.22	3.90	4.14	4.05	3.97	3.49	3.49	2.68	2.48	1.80	6.11	3.31	5.23
5	18.3	1.61	2.49	3.20	4.03	3.29	3.64	4.10	3.96	2.91	3.00	2.45	2.44	2.10	5.82	1.79	5.52
5	18.5	0.88	2.60	2.37	2.09	3.13	2.66	2.56	1.93	1.67	1.32	1.00	0.77	0.81	5.06	3.59	4.38
5	19.4	0.08	2.89	3.12	4.06	2.20	1.94	1.64	1.40	1.08	0.88	0.72	0.53	0.40	5.32	3.23	4.55
1	17.5	3.63	6.10	4.65	4.51	4.20	3.78	3.99	4.25	3.49	3.72	3.36	3.18	2.78	2.64	0.40	3.82
1	23.7	4.36	3.28	3.83	3.91	4.49	4.69	4.59	4.64	4.98	4.34	4.67	4.57	5.18	1.30	0.35	3.15
1	24.4	2.22	5.56	4.70	3.80	3.77	4.06	3.34	3.22	3.11	2.95	2.87	2.66	2.75	2.09	0.44	2.37
1	24.4	2.06	5.50	4.76	3.74	3.64	3.59	3.02	2.94	2.73	2.71	2.62	2.41	2.65	1.73	0.28	2.19
2	14.6	2.48	3.05	5.17	4.69	4.45	4.24	3.92	3.61	3.04	2.96	3.19	3.21	3.04	1.69	0.17	2.16
3	6.8	4.37	3.59	2.80	3.35	4.24	5.11	4.49	3.63	3.79	3.87	4.18	3.89	4.46	1.27	0.04	1.49
4	6.5	4.60	3.45	3.12	4.81	4.23	4.32	4.59	5.55	4.93	4.75	5.38	5.16	5.72	1.33	0.17	2.41
4	13.5	4.23	4.54	4.69	4.45	5.42	5.43	5.21	5.01	4.84	4.28	4.05	4.35	3.69	3.54	0.87	3.31
4	13.5	2.72	6.10	5.07	5.11	6.00	5.60	4.96	4.73	4.72	4.07	3.55	3.54	3.69	2.93	0.67	2.79
5	4.0	3.97	3.60	5.11	3.48	4.55	4.12	5.04	4.52	4.01	3.98	4.10	3.98	2.90	2.45	0.71	2.58
5	8.8	3.85	2.02	2.29	2.16	3.88	4.23	3.90	4.27	4.39	4.83	4.63	4.41	5.11	1.12	0.18	1.80
5	8.8	3.85	2.02	2.29	2.16	3.88	4.23	3.90	4.27	4.39	4.83	4.63	4.41	5.11	1.12	0.18	1.80
5	16.4	5.15	2.90	2.32	2.10	2.06	2.58	2.73	3.07	3.49	3.08	3.56	3.52	3.63	0.99	0.64	2.34
5	16.6	3.60	2.61	4.12	5.16	3.63	3.68	3.51	3.56	3.75	3.84	3.51	3.26	3.02	2.60	0.77	2.33

Figure 6 Association profiles of SNPs that were identified by GWA mapping to be highly associated with the traits FW, PLA over time and Expo2 model parameters. Each number in the columns with heading 'FW' or 'PLA (mm^2)' represents the association found by univariate GWA mapping of growth trait by time-point as indicated at the top of the column (FW at day 28 or PLA on one of the indicated days) and the SNP at position indicated in the first two columns. In the last three columns, with heading 'Expo2', the numbers indicate the association found between SNPs and parameters of model Expo2. Columns with heading 'Expo2:A_0' and 'Expo2:r' refer to univariate GWA mapping of model parameters 'A_0' and 'r' respectively, and the column with heading 'Expo2:full' refers to multivariate GWA mapping of both growth model parameters. All SNPs with –log(p)>5 for at least one trair are shown. The intensity of the grey colour corresponds to strength of the association. Accolades indicate that associated SNPs are located within 10 kb and are considered as one QTL.

throughout the experiment (Figure 2). These data indicate that the growth phenotype of a plant is the result of the interplay of many different genes and that the composition or contribution of this set of growth factors will change during the development of the plant. Some genes only play a role at a specific time point, whereas other more general regulators may have a function in growth for a longer period. Many transcription factor are, for example, known to be expressed both developmental-time-specific and tissue-specific (Turnbull 2011), whereas their influence on plant development is visible during several developmental stages and in other tissues, due to the expression of down-stream targets. Similarly, levels of phytohormones are tightly regulated over time, whereas prolonged down-stream effects are often observed (Schachtman 2008). The relative effect size of these regulators might change over time as a result of the dynamic balance between

different regulatory components during development. The effect of these general growth factors will, therefore, only be large enough on specific time periods to be detectable with GWA mapping. SNPs that were selected because of strong association with PLA at a specific time point may, therefore, point to genes that play a role at a very specific period of development, but they may also point to more general regulators. If plant size had only been measured at one time-point, many of these time-specific associations would not have passed the threshold, and thus would have been missed. Most striking is the observation that only one SNP was strongly associated (-log(p)>5) with FW at day 28, so if growth was only evaluated by end-point FW determination, all except one of the associations would have been missed. Thus, our analyses, therefore, show that evaluation of growth over time is more powerful to identify the underlying genetic factors than the evaluation of growth by end-point measurements. This is especially true when many small effect genes, whose relative contribution may change over time, are underlying the trait of interest.

Novel candidate genes for growth dynamics

Because GWA mapping was done in a natural population of accessions with strong LD decay, causal genes are expected to be located in close proximity of the associated SNPs. Because the LD decay in Arabidopsis is on average 10kb (Kim 2007) significant SNPs that were located within this distance from each other were considered to be associated to the same causal gene. This approach resulted in 11 QTLs for the model parameters and another 11 QTLs for FW and PLA over time (Figure 6). Genes that were located within a support interval of 10kb up-stream and down-stream from these QTLs were selected for further analyses (Table 3). We are aware of the fact that LD decay is not equal along the whole genome and that the causal gene can, therefore, be located outside the 10 kb window. However, even in studies of linkage mapping in RIL populations the confirmed causal genes were in most cases located very close to the associated marker even when the support interval was large (Price 2006). For QTLs represented by multiple SNPs in close proximity, the support window was taken 10kb upstream of the first SNP until 10kb downstream of the last SNP, therefore candidate genes in such QTLs could be located more than 20 kb apart. For example, the two associated SNPs on chromosome four at 13.5Mb, are located 6,3kb from each other, so candidate genes for this QTL can be located 26,3 kb from each other. Large differences were observed in gene density in the support windows, ranging from 2 up to 12 genes (Table 3). As expected the number of genes was in general higher in the support window of QTLs that represent multiple associated SNPs. In total 97 genes were selected, 41 for QTLs associated with model parameters and 56 for QTLs associated with plant size.

The annotation of the 97 genes within the support windows was analysed and the 17 genes with GO terms 'developmental processes' and the 13 genes with GO term 'transcription' (TAIR10) were studied in more detail. Both GO terms were not significantly

Table 3 Information about the support window around the 26 SNPs that are highly associated with the growth traits (-log(p)>5). The order of SNPs is corresponding with Figure 6 to enable easy comparison of the data presented. Accolades indicate that associated SNPs are located within 10 kb and can be considered as one QTL.

		max -log(p)		# genes		Candidate genes	
chromosome	position (Mb)	FW + PLA	Expo2	within 10 kb window	with unknown biological function	Code	Name + process involved
1	24.6	3.70	6.22	5	3		
3	0.6	3.40	5.55	5	4		
3	1.2	4.97	6.15	4	2	AT3G04460, PEX12	role in peroxisome formation, mutant embryo-lethal, RNAi knock down results in slower development (Fan 2005).
3	4.4	0.94	5.61	5	0	AT3G13540, MYB5	role in seed coat formation and trichome morphology, mutation of gene has no effect on plant growth (Li 2009)
3	4.4	0.78	5.52			AT3G13550, CIN4	mutant embryo-lethal, weak alleles result in stunted growth (Vogel 1998)
3	7.2	2.96	5.37	4	2	AT3G20550, DAWDLE	role in seed germination, mutation of gene results in slower growth and later flowering (Morris 2006)
4	9.6	2.51	5.40	4	1		
5	14.9	3.56	5.87	3	1	AT5G37600, GSR1	role in leaf senescence, knock-down of isoform (AT1G66200) results in reduced biomass accumulation (Lothier 2011)
5	18.3	4.26	6.11	2	0		
5	18.3	4.10	5.82	2	1		
5	18.5	3.13	5.06	5	3		
5	19.4	4.06	5.32	2	0		
1	17.5	6.10	3.82	2	0		
1	23.7	5.18	3.15	4	0		
1	24.4	5.56	2.37	8	7		
1	24.4	5.50	2.19				
2	14.6	5.17	2.16	3	1	AT2G34650, PINOID	mutant shows abnormal flower development, longer and wider leaves, delayed senescence (Bennett 1995)
3	6.8	5.11	1.49	6	1		
4	6.5	5.72	2.41	4	2		
4	13.5	5.43	3.54	12	2	AT4G26740, ATS1	role in oil-body breakdown, mutant shows delay in development directly after germination, but no effect on biomass was observed later (Poxleitner 2006)
4	13.5	6.10	2.93			AT4G26760, MAP65-2	post-embryonic root development and axial cell growth, double mutant with map65-1 (AT5G55230) results in retarded growth (Lucas 2011, Lucas 2012)
5	4.0	5.11	2.58	6	3	AT5G12840, NF-YA1	overexpression leads to abnormal embryogenesis and seed development, as a consequences plants have retarded growth and delayed flowering time (Mu 2013)
5	8.8	5.11	1.80	3	0	AT5G25380, CYCA2-1	core cell cycle gene, coordinate cell proliferation and endoduplication, mutant has no morphologicalo visual phenotype (Yoshizumi 2006)
5	8.8	5.11	1.80			AT5G25390, SHN3	over-expression leads to smaller plants with higher drought tolerance (Aharoni 2004)
5	16.4	5.15	2.34	3	1		
5	16.6	5.16	2.60	5	2		

overrepresented within the candidate genes (Plant GSEA, (Yi 2013)). Overrepresentation is not expected for common GO terms, because only one or a few causal genes at most are expected per QTL and therefore only a limited number of the 97 candidates genes will be causal. All other genes are presumed to be randomly distributed over GO categories. The QTL on chromosome four at 13.5Mb illustrates the significance of our findings. This QTL contains two SNPs that are highly associated with plant size in week two and three and moderately associated with plant size in week four (Figure 6). A weak association with these SNPs was also found in the univariate GWA mapping of parameter 'A0' and the multivariate analyses of 'A0' and 'r'. Within the support interval of this QTL 12 genes are located of which only two are annotated to be involved in plant development and none are transcription factors. The first gene, AT4G26760 (MAP65-2) is involved in post-embryonic root development (Lucas 2012) and axial cell growth in hypocotyls (Lucas 2011). This gene is most closely located to one of the associated SNPs, at only 2.1kb. Double mutants of this gene and the closest homologue AT5G55230 (MAP65-1) show retarded growth and therefore AT4G26760 is a strong candidate for the causal gene underlying this QTL. The second gene, AT4G26740, is only expressed in the embryo and plays a role in the breakdown of oil-bodies. Directly after germination mutants show delay in growth, but this does not affect biomass accumulation at later stage and, therefore, this gene is less likely to be the causal gene. However, a role in natural variation of plant size cannot be excluded for this gene, because no accessions other than Columbia have been analysed and redundancy in function may be present (Briggs 2006). Two other genes in the region encode unknown proteins and the remaining genes play roles in processes that are not directly linked to biomass accumulation. To confirm that AT4G26760 is responsible for the observed natural variation in growth, it is needed to perform experiments in which more accessions than only Columbia are investigated. Exchange of alleles between natural accessions can be a powerful tool to identify allele specific growth phenotypes. The QTL on chromosome one at 24.6Mb demonstrates that follow-up research of GWA mapping is not straight forward. This QTL contains the strongest associated SNP identified in this experiment (-log(p)=6.22). Strong association was found for univariate GWA mapping of parameter 'A0' and the multivariate analyses of 'A0' and 'r' and weak associations were found with PLA in week two and three. Three genes with unknown function and two genes related to defence are located in the support interval. Additional information about gene-expression or the phenotypes of mutant or over-expression lines is needed to be able to prioritize these genes and finally confirm the causal gene. This is a laborious and time-consuming process, which is probably the reason why not so many non-obvious candidates have been confirmed in GWA mapping yet.

Because we noticed a weak relationship between growth rate and flowering time we screened our candidate gene list also for genes involved in the regulation of flowering. Only one of the 97 candidate genes was related to flowering time (Table 3). Mutants of this gene (viz. DAWDLE (AT3G20550))) show delayed flowering and slower growth

(Morris 2006). This corresponds with our observation that plants that bolted within the experimental time were larger than the ones that did not yet bolt. DAWDLE stabilises the hairpin formation of miRNAs (Yu 2008). Three miRNAs influence the expression of FT and SOC1 (Yamaguchi 2012), two key-players in the flowering time regulatory network downstream of FLC. The SNP associated with DAWDLE (chr 3, 7.2 Mb) was only identified in the MTMM analyses of A_0 and r simultaneously, which is probably the reason that this gene was not identified previously in any RIL or GWA mapping study regarding flowering time or biomass. (Atwell 2010) found strong association between flowering and FLC using a natural population of 95 accessions, of which 4 are overlapping with our set, confirming allelic differences for FLC between natural accessions (Gazzani 2003) (Guo 2012). In our study a weak association (-log(p)=3.19) was found between plant size on day 27 and a SNP (chr5 pos 95343751) in LD with SNPs in FLC. However, the population is very diverse and consequently sequence variation for many other flowering time genes is expected. This might be the reason why the association between FLC and plant size is not stronger.

For one-third of the genes within the support windows no biological function has been was annotated (34 genes,TAIR10). GWA mapping is an approach that is not hypothesis but data driven. It aims to find genes that contribute to the explanation of variation observed for a trait, without the need to know the pathway or mechanism by which the phenotype and the genotype are correlated. GWA mapping is a powerful method to find novel functions for genes, or to identify functions of unknown genes. Unfortunately, such new functions are hardly reported yet, because most studies that report associations identified by GWA mapping are not coupled with studies to confirm candidate genes. Therefore, the attention in GWA mapping studies is biased towards genes for which the functions are already known and the genes with unknown biological functions are given less attention. However, in the field of plant sciences, a report of the confirmation of an unknown gene identified by GWA mapping was recently published (Meijon 2014) and hopefully many publications will follow soon.

In summary, considering all 97 candidate genes, 11 are annotated to play a role in the determination of cell number, cell size, seed germination, embryo development, transition from vegetative to generative stage or senescence. For eight of these genes a mutant or over-expression phenotype related to growth has been reported. These eight genes are located in the support window of eight QTLs of which four were associated with model parameters and four were associated with plant size (details in Table 3). This emphasizes that mapping of growth model parameters is complementary to the mapping of plant size data at several time-points separately. For none of the eight candidate growth dynamics are reported and it is therefore not known whether allelic variants affect growth from the start, only in a specific developmental stage, or from a specific developmental stage onwards. Therefore additional temporal growth and gene expression data need to be

collected to determine whether the candidate genes play a time-specific or a general role in plant growth regulation. For none of the genes in the support window of the other thirteen QTLs a mutant or overexpression phenotype related to biomass accumulation has been reported yet. These findings indicate that the observed associations are likely to be true positives and that many more genes are involved in growth regulation than is currently known.

Conclusions

Here, we describe a series of analyses that started with the observation of growth dynamics by automatic imaging and that, by subsequent image analysis, growth modelling and GWA mapping, resulted in the indication of candidate genes involved in growth regulation. Top-view imaging of Arabidopsis plants in combination with high throughput image analysis allowed us to follow rosette growth over time in a large and diverse population of natural accessions. During the experiment most rosettes were in the acceleration and linear phase of growth, which could be modelled best by an exponential function (Expo2) describing indeterminate growth. Modelling ensured proper comparison of the diverse panel of accessions demonstrating large variety in the rate of development and in plant size. To identify the genetic basis of growth, GWA mapping was performed on PLA data (12 different dates) and FW data (end point) and on the parameters derived from the growth model Expo2. This resulted in the detection of 22 growth QTLs which were highly associated (-log(p)>5) with the growth traits. Many of these QTLs would not have been identified if growth had only been evaluated at a single time-point. Eight candidate genes were identified for which a mutant or overexpression phenotype related to growth has previously been reported, suggesting that the identified QTLs are true positives. For some QTLs no obvious candidates were found, opening up the way to identify new functions for underlying genes or to annotate unknown underlying genes by performing follow-up experiments.

Materials and Methods

Plant materials
A collection consisting of 324 natural accessions of *Arabidopsis thaliana* was used to investigate the growth of rosettes over time. These accessions were selected to capture most of the genetic variation present within the species. (Li 2010),(Baxter 2010), (Platt 2010). Each accession was genotyped with approximately 215k SNP markers (Col vs non-Col)(Kim 2007).

Experimental set-up
The PHENOPSIS phenotyping platform was used to perform the experiments (Granier 2006). The climate conditions within the growth chambers of PHENOPSIS are precisely regulated, preventing differences in growth because of position in the chamber. The plants were grown in four rounds of experiments each containing 84 accessions. Each experiment was designed as a randomized complete block (3 blocks), with four reference accessions grown in each experiment: Col-0 (CS76113), KBS-

Mac-8 (CS76151), Lillo-1 (CS76167) and Wc-2 (CS28814). Note that the reference accessions (checks) were used to connect the different experiments as the accessions do not overlap across experiments. Cylindrical pots (9 cm high, 4.5 cm diameter) filled with a mixture (1 : 1, v : v) of a loamy soil and organic compost were used and the seeds (at least two per pot) were sown directly on the soil. Cold treatment (4°C) was given to the seeds and pots directly after sowing. To enable harvesting of the rosettes within a time-frame of 1.5 hours, the three blocks were transferred from the cold to the growth chamber (PHENOPSIS, 16h light, 125 umol $s^{-1}m^{-2}$, 70% humidity, 20/18°C) on sequential days, four, five, or six days after sowing. The day the plants were transferred to PHENOPSIS was marked as day one. The water content of the soil was kept at 0.35 to 0.40 g $H_2O \cdot g^{-1}$ dry soil by robotic weighing and watering the pots twice a day. After two weeks the plants were thinned to one plant per pot.
An separate experiment was performed to determine whether the accessions were summer or winter annuals. All accessions (three replicates) were grown on rockwool blocks in the greenhouse and were watered regularly. Flowering time of the first replicate of each accessions was recorded. Accessions that flowered within 75 days were called summer-annuals, accessions that did not flower within this period were called winter-annuals.

Determination of rosette growth traits
All plants were daily inspected for visible signs of bolting and bolting dates were recorded. At day 24 the largest leaf of each plant was harvested. The fresh weights (FW) and dry weights (DW) of the leaves were determined to calculate the water contents (WC) by WC = (FW-DW)/FW. At day 28 the rosettes were harvested and the fresh weights were determined. Rosette growth was monitored by taking pictures from above twice a day. These pictures were processed in ImageJ using the macros developed for PHENOPSIS. All pictures and the ImageJ macros are publically available on PHENOPSISDB ((Fabre 2011), http://bioweb.supagro.inra.fr/phenopsis. The projected leaf area of each plant (PLA) was determined semi-automatically on the following days: 8, 11, 14, 16, 18, 20, 22, 24, 25, 26, 27 and 28. When more than one plant was present in a pot before thinning, the largest one close to the middle of the pot was taken for analysis. For each individual plant the growth, based on PLA, was modelled using three functions; two functions describing indeterminate growth by an exponential curve, Expo1 and Expo2, and one function, describing determinate growth by a S-curve, Gom (Table 2). The optimal parameter values were estimated using the Growth Fitting Toolbox™ of MATLAB with the following settings; Expo1: optimization algorithm: 'Trust-Region', fitting method: Non-linear least square, bounds: r [0, Inf]; Expo2: optimization algorithm: 'Trust-region', fitting method: Non-linear least square, bounds: A_0[0,Inf], r[0,Inf]; Gompertz: optimization algorithm: 'Levenberg-Marquardt', fitting method: robust linear least square, using bisquare weights, bounds: A_{max} [0,20000], b[0,60], r[0,Inf]. Goodness of fit indicators (SSE, r^2, and RMSE) and 95% confidence intervals of the parameters of all three models were calculated in MATLAB. Based on these data and the principle of parsimony, Expo2 was chosen as best model and therefore used for further analysis; fits with r^2<0.9 (11 out of 965) were excluded from further analysis. The removal of the largest leaf from day 24 onwards was not corrected for when the fresh weight of the rosette and the model parameters were determined because plants rapidly compensated for this loss.

Statistical analysis
GWA mapping was performed on the FW at day 28, PLA at days 8, 11, 14, 16, 18, 20, 22, 24, 25, 26, 27 and 28,and the estimated model parameters of Expo2. For all traits, adjusted means were obtained with GenStat by fitting the following mixed model: . The terms Checks and Accessions within Round were assumed fixed to obtain Best Linear Unbiased Estimates (BLUE), and all the remaining terms were considered random effects as they are all essentially different sources of experimental error due to Round, Blocks within Rounds (and the interaction with check genotypes) and residual variation. The term Check refers to a factor in which reference accessions Col-0 (CS76113), KBS-Mac-8 (CS76151), Lillo-1 (CS76167) and, Wc-2 (CS28814). GWA mapping was performed on the predicted means using the EmmaX software package for R, which is based on (Kang 2010). A mixed model was used that corrects for population structure, based on the kinship matrix of all SNPs. SNPs with a minor allele frequency < 0.05 were excluded from the analysis. The parameters 'A_0' and 'r' of model Expo2 were

also mapped together using a Multi Trait Mixed Model (MTMM) approach (Korte 2012) (El-Soda 2014). Pearson correlations were used to determine correlations between data series. To calculate the correlation between traits, the corrected means were used. Broad-sense heritabilities were obtained with GenStat by fitting the same model as above. This times the terms Check were assumed fixed and all the remaining terms were considered random effect. Broad sense heritability at the mean level was calculated as: $H^2 = V_g/(V_g + V_e/r)$, where V_g is the genetic variance and V_e is the error variance.

Differences in FW between the rosettes of plants that were bolting at day 28 and plants that were still vegetative, and between summer and winter annuals were determined using a t-test assuming non equal variances and $\alpha=0.05$.

Additional analyses
For each of the candidate genes the annotations and GO terms were retrieved from TAIR10 (arabidopsis.org).

Acknowledgement
We would like to thank Myriam Dauzat for operating and maintaining PHENOPSIS, and Gaelle Rolland and Crispulo Balsera for their assistance during sowing and harvesting.

Supplementary data

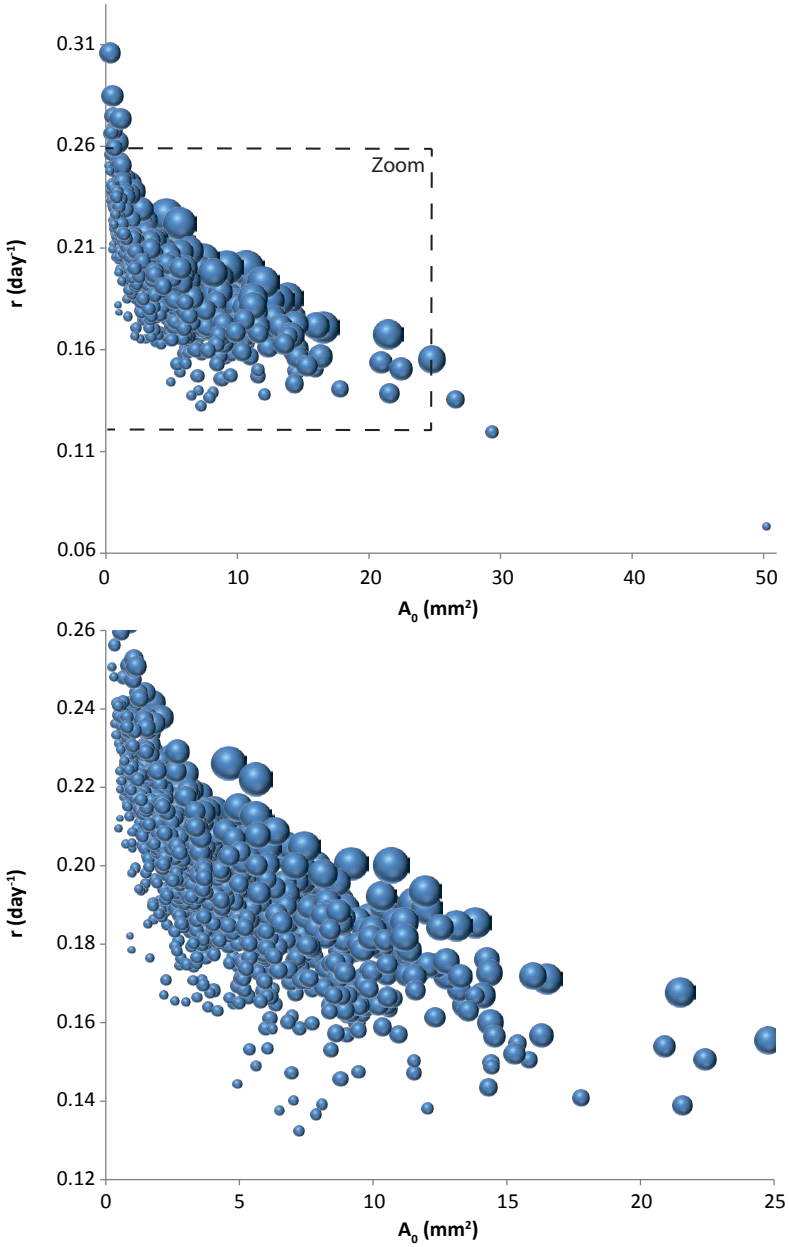


Figure S1: Scatter plot of the Expo2 model parameters 'A_0' and 'r'. Area of the bubbles corresponds with the FW of the rosette at day 28. **A** Scatter plot containing data of all individual plants in the experiment. **B** Zoom as indicated in A.

Chapter 5 - Genome wide association mapping of time-dependent growth responses to moderate drought stress in Arabidopsis

Johanna A. Bac-Molenaar[1,3], Christine Granier[2], Joost J.B. Keurentjes[3], Dick Vreugdenhil[1]

[1] Laboratory of Plant Physiology, Wageningen University, The Netherlands
[2] Laboratoire d'Ecophysiologie des Plantes sous Stress Environnementaux, UMR 759, Institut National de la Recherche Agronomique/Ecole Nationale Supérieure d'Agronomie, Place Viala, F–34060 Montpellier, Cedex 1, France
[3] Laboratory of Genetics, Wageningen University, The Netherlands

Abstract

Large areas of arable land are often confronted with irregular rainfall resulting in limited water availability for part(s) of the growing seasons, which demands research for drought tolerance of plants. Natural variation was observed for biomass accumulation upon controlled moderate drought stress in 324 natural accessions of A*rabidopsis.* Improved performance under drought stress was correlated with late flowering and vernalization requirement, indicating overlap in the regulatory networks of flowering time and drought response or correlated responses of these traits to natural selection. In addition, plant size was negatively correlated with relative water content (RWC) independent of the absolute water content (WC) indicating a prominent role for soluble compounds. Growth in control and drought conditions was determined over time, and modelled by an exponential function. GWA mapping of temporal plant size data and of model parameters resulted in the detection of six, time-dependent, QTLs strongly associated with drought. Most QTLs would not have been identified if plant size was determined at a single time-point. Analysis of earlier reported gene expression changes upon drought enabled us to identify for each QTL the most likely candidates.

Plant, Cell & Environment, doi:10.1111/pce.12595

Introduction

Drought is one of the major causes of yield loss in agriculture. Large areas of arable land are often confronted with irregular rainfall resulting in limited water availability for part(s) of the growing season. Although in temperate climates drought is often only moderate, it results in limited crop growth and reduced yields (Skirycz 2011b). In view of the climate change it is expected that these moderate drought periods will occur more often resulting in an increasing demand for crops that are tolerant to moderate drought.

For decades breeders have put effort into the development of drought tolerant crops (Pennisi 2008). This resulted in the development of a limited number of drought tolerant cultivars, but the mechanism behind this tolerance and the causal genes are in most cases unknown (Ashraf 2010). In addition, many studies aiming to elucidate the molecular mechanism test for tolerance towards severe drought, in which often plant or crop survival is scored rather than growth or yield. Expression studies (Harb 2010) and the analysis of mutants (Skirycz 2011b) both reveal that the responses of *Arabidopsis thaliana* to moderate and to severe drought are to a large extent different. When drought occurs, plants sense it and respond to it in an acute way. Often the growth is highly retarded, stomata close and the plant is preparing for more severe conditions. Thereafter the plant is adapting its metabolism to the new situation and as soon as a new steady state is established, growth will be (partially) regained (Harb 2010, Skirycz 2011a). In a moderate drought scenario this new steady state will be obtained fast, whereas upon severe drought this steady state will be reached much later or not at all.

Arabidopsis thaliana is used as model species to study natural variation in stress responses, because its geographical distribution in combination with local adaptation resulted in large variation in stress responses among natural accessions (Aguirrezabal 2006, Bouchabke 2008). In *Arabidopsis*, accessions may have adapted to drought in two ways: either the plant accelerates its life cycle to ensure reproduction before the drought becomes lethal (drought escape), or the plant optimizes the balance between water uptake and water-loss (drought avoidance) (Berger 2010, Verslues 2011). The latter strategy may include expansion of the root system, adjustment of the osmotic potential, or strategic opening and closing of the stomata (Farooq 2012) (Sanders 2012). Drought avoidance aims to prevent wilting that may damage the membranes of cells and organelles (Hincha 1987, Chołuj 2014). The result of both adaptation strategies, is similar, viz., reduced above-ground vegetative growth, due to increased investment in generative tissues or roots and due to reduced photosynthesis as a consequence of stomatal closure. Similar strategies are observed in many crops (Farooq 2012), making the translation of findings from *Arabidopsis* to crops promising. The occurrence of multiple adaptation mechanisms within one species, each determined by many genes, makes breeding for drought tolerance a challenge, but if breeders succeed to combine favourable alleles of several

drought tolerance genes in elite lines, stability of the food production will be increased.

An effective way to elucidate genes underlying complex traits is the use of natural variation through mapping of these traits in bi-parental, multi-parental or natural populations (Koornneef 2004, Kover 2009, Korte 2013). However, the use of these resources in the elucidation of genes involved in drought tolerance is hampered by large environmental variation due to experimental difficulties in establishing consistent drought conditions, both in pot and in field experiments. Phenotyping platforms like PHENOPSIS (Granier 2006) and WIWAM (Skirycz 2011b) diminish the unequal drying of the soil by carefully measuring and adjusting soil water content. These platforms also allow following the shoot growth response to drought over time by automated imaging of plants, opening up the possibility to determine the dynamics of the drought response and the unravelling of the underlying gene network.

Many studies reported natural variation in *Arabidopsis* upon various degrees of drought stress (McKay 2003, Aguirrezabal 2006, Bouchabke 2008, Vile 2012, El-Soda 2014), but only a limited number of accessions has been tested for moderate drought stress (Granier 2006, Clauw 2015). In *Arabidopsis* the molecular mechanism of moderate drought adaptation has been studied using bi-parental populations and, therefore, only a limited part of the total genetic variation present in this species has yet been investigated (Juenger 2005, Tisné 2010, Schmalenbach 2014). Genome Wide Association (GWA) mapping can be used to explore much more of the variation in moderate drought adaptation present in this species and is expected to result in high resolution QTLs that do not require much further fine-mapping to identify the underpinning genes.

In the current study, the response of 324 natural accessions to moderate drought was investigated in the PHENOPSIS phenotyping platform. Homogenous and controlled soil water content was ensured by twice a day weighing and watering of all pots, and top-view imaging enabled us to follow rosette expansion over time. GWA mapping of rosette growth-related parameters resulted in the detection of time-dependent QTLs. Likely candidates could be identified for most QTLs based on previously reported expression studies.

Results

Reduction of growth upon moderate drought is genotype dependent

To investigate natural variation in rosette growth response to drought, 324 accessions of *Arabidopsis thaliana* were grown under control (0.40 g H_2O g^{-1} soil) and moderate drought (0.22 g H_2O g^{-1} soil) conditions in the PHENOPSIS phenotyping platform (Granier 2006). Under these moderate drought conditions, water is still available for the plants,

Trait	Time (day)	Control			Drought		
		Mean	SD	H^2	Mean	SD	H^2
Projected Leaf Area (mm2)	8	6.5	2.79	0.28	6.5	2.60	0.49
	11	15.3	6.55	0.53	15.3	6.45	0.42
	14	36.2	16.15	0.52	35.4	14.99	0.54
	16	61.2	28.63	0.53	56.6	26.47	0.56
	18	109.7	53.70	0.56	92.8	47.47	0.53
	20	188.4	90.86	0.56	148.4	74.30	0.60
	22	314.3	152.82	0.60	235.1	113.84	0.58
	24	495.4	235.49	0.63	365.6	176.48	0.60
	25	519.9	256.64	0.62	379.1	182.68	0.64
	26	639.6	312.07	0.63	464.5	223.06	0.64
	27	768.7	362.10	0.66	551.7	258.67	0.64
	28	911.2	417.11	0.71	652.0	298.66	0.67
Exponential function A_0		4.69	3.84	0.63	4.25	3.12	0.52
Exponential function r		0.20	0.02	0.61	0.19	0.02	0.51
FW rosette (g)	28	0.27	0.15	0.70	0.17	0.09	0.68
RWC (g/g)	23	0.32	0.07	0.39	0.77	0.06	0.30

Table 1 Averages, SD, and broad-sense heritabilities (H^2) for growth traits (Projected leaf Area (PLA), Fresh Weight (FW), Relative Water Content (RWC)) of plants on indicated days and growth model parameters (r (growth rate) and A_0 (initial size)). Time: days after transfer of stratified seeds from cold to climate room. Drought stress was initiated at day 14. Mean, SD: Overall mean and standard deviation observed for the indicated trait on the indicated date. Trait values of individual plants, not corrected for block and round effects, were used to calculate Mean and SD. Drought conditions: n=965 and control conditions: n=961.

but the amount is limited and identical among pots independently of rosette size and transpiration. Rosette expansion was monitored over time by top view imaging. Reduced growth under moderate drought stress resulted in an increasing difference in plant size between control and drought treated plants during the time-course of the experiment. Already two days after the start of the drought treatment (day 16), reduction in growth due to drought resulted in significantly smaller plants compared to control (Table 1). Finally, biomass was, on average, reduced by 35% upon moderate drought stress (Table 1). However, large variation in plant size was observed among the accessions in control and drought conditions (Figure 1A). These differences were, for a large part, determined by the genotype, since moderate to high heritabilities were found for Projected Leaf Area (PLA) and Fresh Weight of the rosette (FW) in both conditions (Table 1). A strong positive correlation between control and drought conditions was observed for PLA and FW (Table 2) and this correlation was strongest at the end of the experiment (r^2= 0.79) and weakest at day 8 (r^2=0.58) (Table 2), indicating more variation due to non-genetic factors at younger age. This was also reflected in the broad sense heritabilities of PLA that increased over time (Table 1). Variation between accessions was not only observed for plant size in both conditions, but also for the drought response, ranging from accessions that had similar growth in control and drought conditions (Figure 2A) to accessions in which the final rosette size was reduced by more than 50% (Figure 2B). Also variation

		PLA (control)										FW			RWC			r		A_0	
		8	11	14	16	18	20	22	24	25	26	27	28	C	D	res	C	D	res	C	D	res	C	D	res
PLA (drought)	8	0.58	0.61	0.58	0.55	0.57	0.55	0.52	0.50	0.50	0.49	0.50	0.47	0.46	0.57	0.34	-0.28	-0.32	-0.21	-0.17	-0.36	-0.32	0.48	0.64	0.47
	11	0.60	0.67	0.66	0.66	0.67	0.65	0.61	0.60	0.59	0.58	0.57	0.54	0.52	0.71	0.48	-0.27	-0.31	-0.21	-0.22	-0.39	-0.32	0.59	0.77	0.58
	14	0.58	0.68	0.69	0.70	0.71	0.69	0.67	0.65	0.65	0.64	0.63	0.60	0.59	0.79	0.53	-0.28	-0.39	-0.30	-0.16	-0.39	-0.36	0.59	0.82	0.65
	16	0.55	0.66	0.69	0.72	0.74	0.72	0.71	0.69	0.69	0.68	0.67	0.64	0.62	0.82	0.52	-0.32	-0.44	-0.33	-0.16	-0.40	-0.37	0.61	0.85	0.66
	18	0.54	0.66	0.70	0.75	0.78	0.77	0.76	0.75	0.76	0.75	0.74	0.71	0.69	0.87	0.52	-0.36	-0.46	-0.35	-0.13	-0.35	-0.34	0.63	0.86	0.66
	20	0.52	0.64	0.68	0.73	0.77	0.77	0.76	0.75	0.76	0.75	0.75	0.72	0.69	0.89	0.53	-0.37	-0.48	-0.36	-0.10	-0.32	-0.32	0.60	0.85	0.66
	22	0.51	0.64	0.68	0.74	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.75	0.72	0.92	0.54	-0.34	-0.43	-0.31	-0.05	-0.22	-0.23	0.60	0.81	0.60
	24	0.49	0.63	0.67	0.73	0.78	0.78	0.78	0.78	0.79	0.79	0.78	0.76	0.74	0.93	0.53	-0.37	-0.45	-0.33	-0.03	-0.19	-0.21	0.58	0.79	0.59
	25	0.47	0.62	0.66	0.71	0.77	0.77	0.77	0.77	0.78	0.78	0.78	0.76	0.73	0.94	0.56	-0.36	-0.44	-0.31	0.00	-0.14	-0.17	0.55	0.76	0.58
	26	0.46	0.60	0.64	0.69	0.75	0.76	0.76	0.77	0.78	0.78	0.78	0.77	0.75	0.95	0.56	-0.36	-0.43	-0.31	0.05	-0.09	-0.14	0.52	0.72	0.55
	27	0.45	0.59	0.62	0.68	0.74	0.75	0.76	0.77	0.78	0.79	0.79	0.78	0.76	0.96	0.55	-0.37	-0.43	-0.30	0.10	-0.02	-0.09	0.49	0.68	0.53
	28	0.40	0.57	0.60	0.66	0.72	0.73	0.75	0.77	0.78	0.79	0.80	0.79	0.77	0.96	0.54	-0.37	-0.44	-0.31	0.16	0.06	-0.04	0.43	0.61	0.49
FW	c	0.54	0.70	0.73	0.79	0.85	0.88	0.91	0.93	0.93	0.95	0.95	0.95	1.00	0.80	0.03	-0.44	-0.40	-0.22	.170	.034	-.082	0.52	0.47	0.33
	d	0.45	0.58	0.60	0.65	0.71	0.72	0.73	0.75	0.77	0.78	0.79	0.78	0.80	1.00	0.56	-0.37	-0.43	-0.30	.169	.002	-.111	0.42	0.64	0.53
	res	0.03	0.03	0.04	0.04	0.07	0.05	0.03	0.03	0.06	0.05	0.05	0.05	0.03	0.56	1.00	-0.05	-0.25	-0.26	.032	-.068	-.111	0.03	0.44	0.48
RWC	C	-0.27	-0.30	-0.31	-0.35	-0.39	-0.42	-0.42	-0.43	-0.43	-0.44	-0.45	-0.43	-0.44	-0.37	-0.05	1.00	0.45	0.00	.016	.112	.127	-0.26	-0.22	-0.11
	D	-0.24	-0.29	-0.33	-0.37	-0.39	-0.41	-0.41	-0.40	-0.41	-0.41	-0.42	-0.40	-0.40	-0.43	-0.25	0.45	1.00	0.89	-.031	-.005	.019	-0.28	-0.36	-0.27
	res	-0.13	-0.17	-0.21	-0.24	-0.24	-0.24	-0.24	-0.23	-0.24	-0.23	-0.24	-0.23	-0.22	-0.30	-0.26	0.00	0.89	1.00	.034	.127	.131	-0.18	-0.30	-0.24
r	C	-.267	-.314	-.312	-.291	-.234	-.200	-.123	-.043	.013	.067	.126	.200	.170	.169	.032	.016	-.031	.034	1	.576	-.019	-.692	-.374	.050
	D	-.310	-.293	-.301	-.296	-.245	-.210	-.160	-.104	-.079	-.045	-.014	.035	.034	.002	-.068	.112	-.005	.127	.576	1	.799	-.486	-.712	-.573
	res	-.180	-.122	-.133	-.145	-.124	-.110	-.106	-.099	-.104	-.105	-.109	-.108	-.082	-.111	-.111	.127	.019	.131	-.019	.799	1	-.083	-.604	-.696
A_0	C	0.62	0.78	0.80	0.85	0.83	0.81	0.76	0.71	0.66	0.63	0.58	0.51	0.52	0.42	0.03	-0.26	-0.28	-0.18	-.692	-.486	-.083	1.00	0.69	0.26
	D	0.53	0.61	0.64	0.68	0.68	0.66	0.62	0.58	0.57	0.55	0.53	0.49	0.47	0.64	0.44	-0.22	-0.36	-0.30	-.374	-.712	-.604	0.69	1.00	0.78
	res	0.29	0.31	0.33	0.38	0.38	0.35	0.34	0.34	0.35	0.36	0.35	0.34	0.33	0.53	0.48	-0.11	-0.27	-0.24	.050	-.573	-.696	0.26	0.78	1.00

Table 2 Pearson correlation between Projected Leaf Area (PLA), Fresh Weight (FW), Relative Water Content (RWC), and Exponential Model parameters, 'A_0' and 'r' of plants grown in control (C) and moderate drought (D) conditions and their corresponding residuals (res), representing the drought response independent of the size of the plant in control conditions. For each accession, the estimated average (corrected for block- and round-effect) is used to determine the correlations. Data from plants grown in C and D conditions were log-transformed to approach normal distribution of variables as assumed by the Pearson correlation. Residuals did not need transformation because they were normally distributed. Intensity of the colour corresponds with strength of correlation, blue indicates positive correlations and red indicates negative correlations.

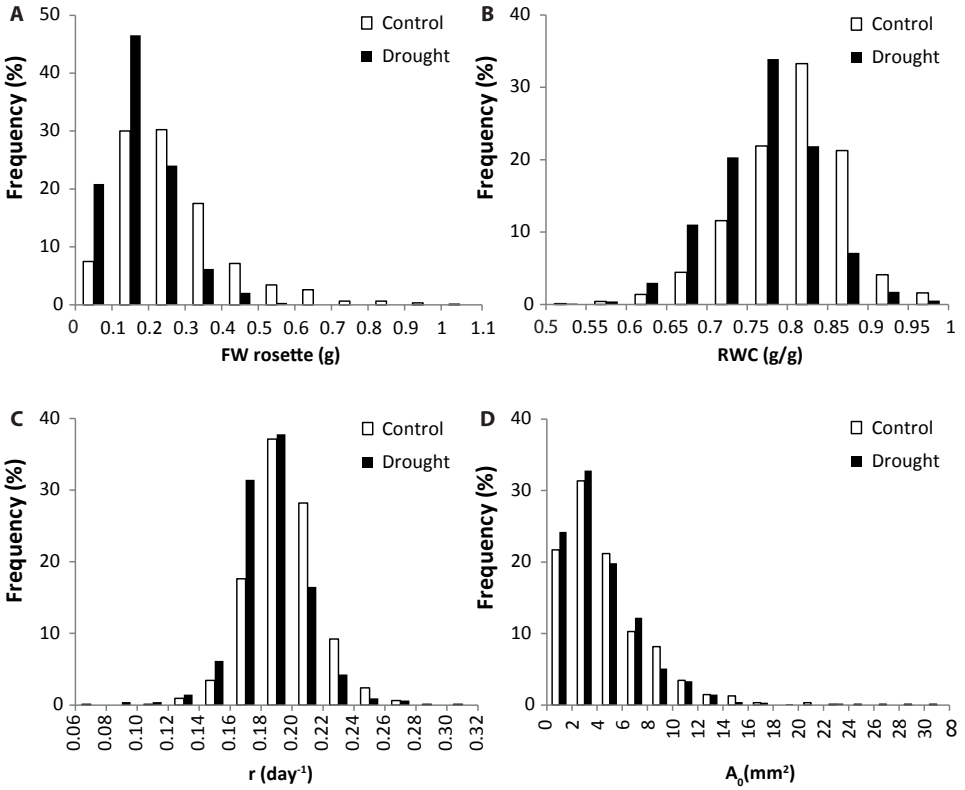


Figure 1 Frequency distribution of rosette Fresh Weight (FW) (**A**), Relative Water Content (RWC) (**B**) and exponential model parameters: growth factor (r) (**C**) and initial size (A_0) (**D**), in control and drought stressed plants (324 accessions with 3 replicates in control and in drought conditions).

was observed for the timing of the start of growth reduction upon drought. To compare the growth response to drought of accessions with such large differences in plant size, we calculated residuals of a linear regression between plant size in control and drought conditions. Residuals represent the drought response independent of plant size in control conditions. Positive residuals indicate a smaller growth reduction than average, whereas negative residuals indicate a larger growth reduction than average. No correlation was observed between plant size in control conditions and the residuals confirming independency of the two measures.

To quantify the dynamics of rosette expansion over time, growth was modelled using an exponential function: $PLA = A_0 * e^{r*t}$ (Bac-Molenaar *et al.*, 2015). The exponential function assumes that plant size (PLA) increases with the same factor (r) during the whole experiment and in addition, it assumes that plant size during the whole experiment is determined by an initial value (A_0) which acts as a magnification factor in the function. An exponential growth model was chosen because at the end of the experiment growth

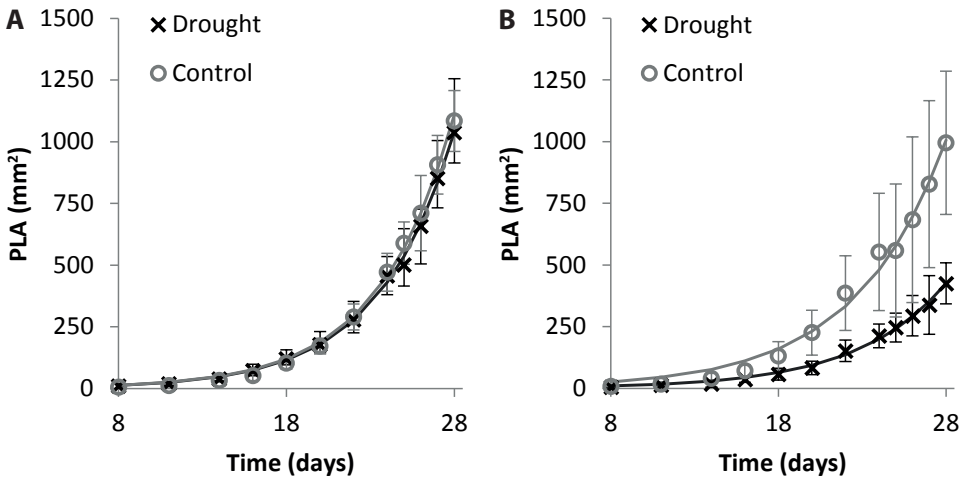


Figure 2 Growth curves of plants under control and drought stress conditions for two accessions with opposing growth response to drought. Error bars represent standard deviations of three replicated measurements. **A** Chat-1, CS28135, an accession that maintained its rosette expansion under drought. **B** Lisse, CS76171, an accession that reduced its rosette expansion under drought.

was still in the exponential or linear phase for most accessions and none of them had reached the final rosette size. The model was very accurate, in 98.5% of the cases a good fit was obtained ($r^2>0.9$), both under control and drought conditions. As for plant size, also for these parameters variation between accessions was determined for a large part by the genotype, corresponding to moderate to high heritabilities ($0.51<H^2<0.63$, Table 1) (Figure 1CD). Despite this variation in parameters, certain combinations of A_0 and r were not observed. For example, relatively large values of A_0 ($A_0>11$) were never observed in combination with relatively large values of r ($r>0.2$). If existing, this combination would have resulted in extremely fast growing plants. Possibly such growth behaviour is biologically impossible, or is not favourable in nature and therefore was not found in our natural population. Comparing the expansion dynamics of control and drought treated plants, positive correlations were observed for A_0 and r between control and drought conditions, which is in accordance with the positive correlation observed between plant size values in control and drought conditions (Table 2). In summary, growth and growth response to drought are dynamic traits that are determined for a large part by the genotype.

Changes in Relative Water Content are independent of changes in Water Content

To investigate whether natural variation in plant performance under drought conditions was related to plant water status, the Relative Water Content (RWC) and Water Content (WC) of the largest leaf of each plant were determined at day 24. Both parameters represent different aspects of the water status. WC signifies the actual percentage of water in the leaf, low WC indicating that leaves start to wilt and metabolic processes stop, thereby blocking further growth. RWC, on the other hand, compares the actual

water content with that of a fully rehydrated leaf. Low RWC is obtained for leaves which contain less water or more soluble compounds or a combination of both. One of the strategies of a plant to maintain high WC during limited water availability is to increase the amount of soluble compounds inside the leaf to improve the flux of water to the leaf and to limit evaporation. No wilting was observed under the applied moderate drought conditions, which corresponds with a WC of 75% or higher observed for all the plants in drought conditions (Figure 1B). Lack of wilting confirmed that the aim to impose moderate instead of severe drought was reached. Variation for WC was limited, 93% of all plants (including both control and moderate drought treated plants) had levels between 88% and 93%, indicating that the differences in RWC are for a major part due to variation in the internal osmotic potential. Variation in RWC was observed between accessions in both control and drought conditions (Figure 1B). This variation is partly genotype-dependent, resulting in heritabilities of 0.39 and 0.30 for control and drought conditions, respectively (Table 1). Interestingly, negative correlations were observed between the size of the plant and RWC in both conditions ($-0.27<r^2<-0.48$) (Table 2). The observation that large plants have lower RWC than small plants, but similar WC, indicates that full-grown leaves of large plants have lower internal osmotic potential at the end of the day, when the measurements were taken, suggesting a role for soluble metabolites. To determine the response in RWC upon drought independent of the RWC in control conditions, the residuals of RWC and plant size were calculated. Residuals of RWC and plant size were weakly negatively correlated, indicating that changes in RWC can be for a small part responsible for changes in drought response. The negative correlation indicates that decrease of the internal osmotic potential by accumulation of soluble metabolites might lead to better plant performance upon drought. Preventing wilting by the accumulation of soluble metabolites is called osmotic adjustment and is observed in many species (Hummel 2010, Farooq 2012, Sanders 2012). In conclusion, differences in soluble metabolites levels between accessions may partly explain the negative correlation between RWC and plant size, and RWC and drought response.

Early flowering accessions are larger and more drought tolerant

Life history, the timing of developmental events within the life cycle of an organism, has an influence on many morphological and stress-related traits (Mitchell-Olds 2013, Huang 2014). Sensitivity towards stresses may differ between developmental phases and therefore timing of the transition from one developmental phase to another is crucial to ensure reproductive success. Large differences in flowering time were observed between the accessions of the natural population used in this experiment, indicating different life histories (Li 2010). Thirty percent of the accessions are winter annuals that require vernalization to flower. To investigate whether the change from vegetative to reproductive phase is linked to variation in drought response, the date on which the first visible sign of this transition, bolting of the inflorescence, was observed was recorded for each plant. A significant difference was observed in plant size between early bolting and late bolting

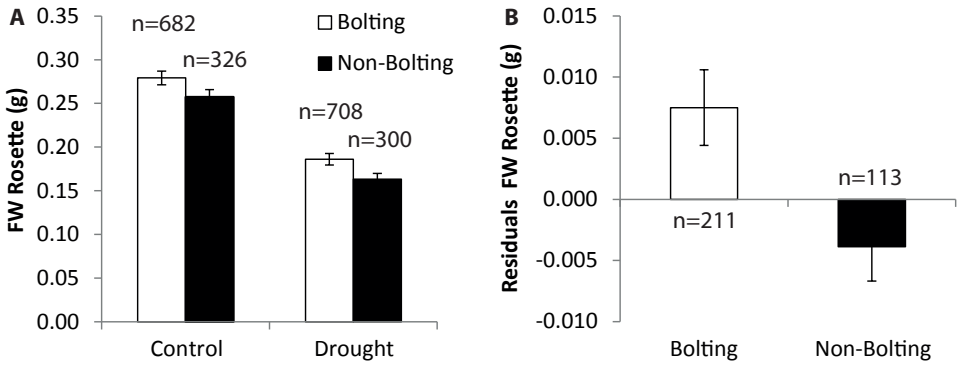


Figure 3 Comparison between plants that were and that were not bolting at day 28. **A** Comparison based on fresh weight (FW) of the rosette **B** Comparison based on the residuals. In A, individual plants were compared whereas in B the residuals were calculated based on the adjusted means per accession. Residuals are taken from a linear regression between the predicted means in control and drought conditions.

individuals long before the first accessions started bolting (from day 11 onwards). Bolting or flowering plants had larger rosettes than vegetative plants after 28 days in both control and drought conditions (Figure 3A). Moreover, the drought response, represented by the residuals of FW, was opposite in the two bolting time classes (Figure 4B). Early bolting accessions performed better than average upon drought, whereas late bolting accessions performed worse than average. A similar trend was observed when the same comparisons were done between winter and summer annuals. Summer annuals that flowered in the greenhouse within 75 days, were larger at any time point from day 11 onwards and more drought tolerant. Our observations indicate that drought response depends both on life history and on plant size. Because also a link was observed between life history and plant size, it is difficult to relate a specific drought response to either of the two plant traits, because it is hard to distinguish cause from consequence.

GWA mapping of drought response reveals strongly associated QTLs

Because the variation in FW, PLA over time, RWC and model parameters is dependent on the genotype, GWA mapping was used to identify genomic regions associated with these traits. Twenty-one strongly associated SNPs (-log(p)>5.5) were detected (Table 3). SNPs in linkage disequilibrium (LD, r^2>0.4) were considered to associate with the same causal genes and were therefore assigned to the same QTL, resulting in fifteen unique QTLs. Six QTLs were associated with model parameters in control conditions (Figure 4A) and three QTLs with FW and PLA in control conditions (Figure 4D). An extensive discussion of these control-QTLs can be found in Bac-Molenaar *et al,* 2015. Three QTLs were associated with FW and PLA in drought conditions (Figure 4E) and another three with the corresponding residuals (Figure 4F). No strong associations were identified for the model parameters in drought conditions and the corresponding residuals (Figure 4B and 4C), for RWC in control or drought conditions, or the corresponding residuals (Figure 4G, 3H and 3I). Between 6% and 9% of the total phenotypic variance could be ex-

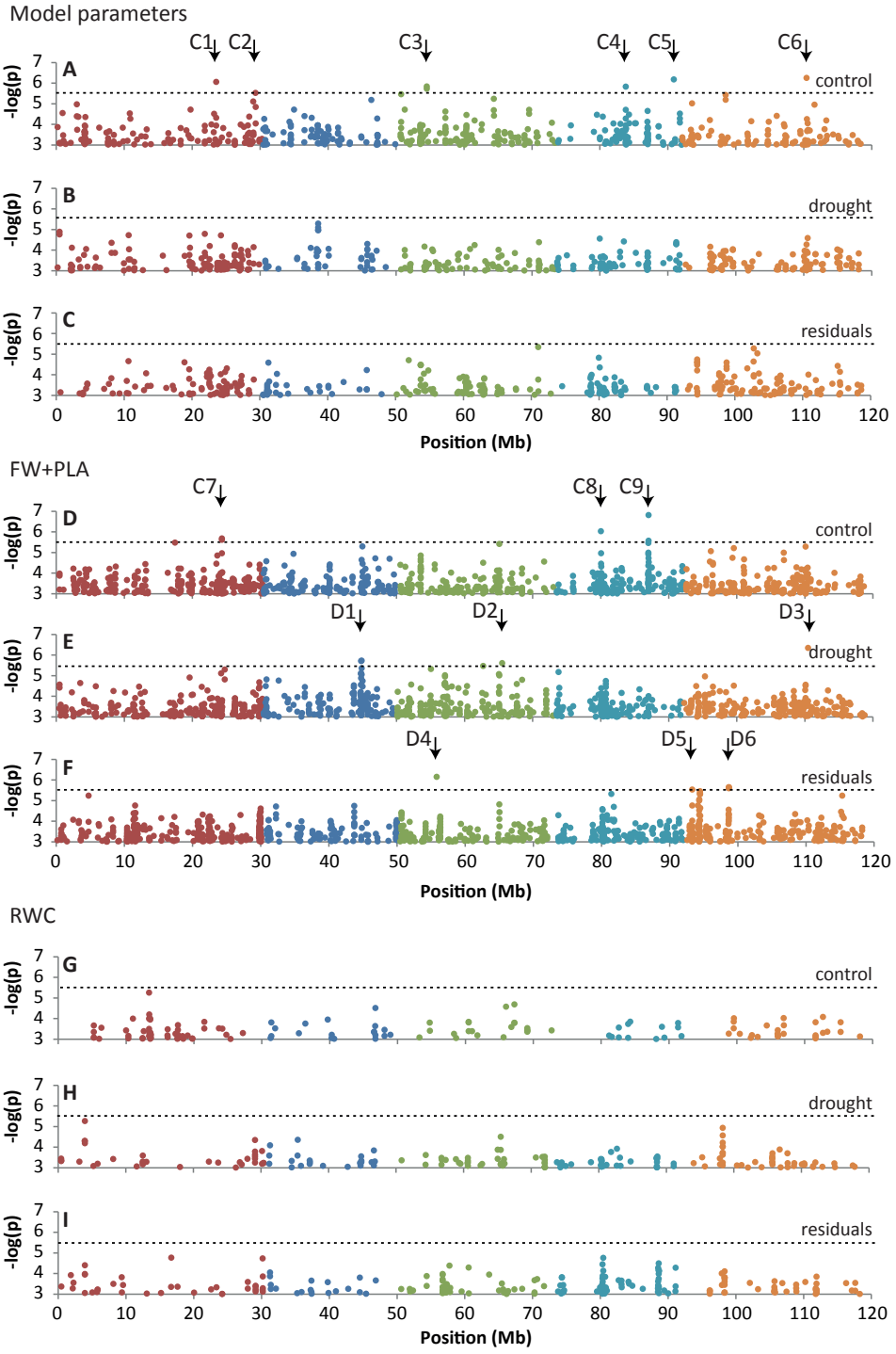

Model parameters
C1
C2
C3
C4
C5
C6
A
control
B
drought
C
residuals
-log(p)
Position (Mb)
FW+PLA
C7
C8
C9
D
control
D1
D2
D3
E
drought
D4
D5
D6
F
residuals
RWC
G
control
H
drought
I
residuals

Figure 4 Manhattan plots representing the associations between SNP-markers and several traits in control and drought conditions and the corresponding residuals, representing the drought response independent of the trait value in control conditions. For clarity, traits were grouped: model parameters 'A_0' and 'r' were taken together, and FW on day 28 is grouped with PLA on day 8,11, 14, 16, 18, 20, 22, 24, 25, 26, 27 and 28. In each group, for each SNP, only the –log(p)-value with the highest association is plotted. Each chromosome is represented by a different colour. **ABC** GWA mapping of exponential model parameters control (A) and drought (B) conditions and the corresponding residuals (C) **DEF** GWA mapping of Fresh Weight (FW) and Projected Leaf Area (PLA) in control (D) and drought (E) conditions and the corresponding residuals (F) **GHI** GWA mapping of Relative Water Content (RWC) in control (G) and drought (H) treated plants and the corresponding residuals (I). Horizontal dotted line represents the significance threshold of –log(p-value)=5.5. QTLs are indicated with arrows. The total number of tested SNP markers was 214K, but only SNPs with –log10(P)>3 were plotted.

plained by each SNP (Table 3). This percentage was in the same range for SNPs associated with traits in control and drought conditions. These percentages emphasize that biomass accumulation is a complex trait, which is regulated by many small effect genes both in control and drought conditions.

To visualise the association pattern of the identified SNPs over time a heat map was constructed (Figure 5). SNPs identified to be associated with PLA on a specific day were, in all cases, also moderately (-log(p-value)>2) associated with PLA on other days. This suggests that the underlying causal genes play a role in determining plant size during the whole experiment, but that the effects of these genes on the phenotype are changing over time. Time-dependent QTLs were identified for control and drought conditions and the corresponding residuals. QTLs that were strongly associated with plant size in control conditions were moderately associated with plant size in drought conditions and *vice versa*, indicating overlap in the network of genes that regulate plant size in both conditions. Hardly any coincidence was observed between the QTLs detected for the residuals of plant size and the QTLs detected for plant size in control or drought conditions, emphasizing that the residuals represent the response of the plant to drought, independent of plant size in control conditions. The residuals are therefore a good measure to identify genes specifically affecting the drought response. In control conditions, strong associations were observed for the two model parameters 'A_0' and 'r'. QTL C6, which was strongly associated with A_0, was also moderately associated with PLA over time. These associations follow the observed correlation between those traits (Table 2). QTLs C1 to C5 were strongly associated with parameter 'r', but not or only weakly associated with plant size traits, indicating that growth modelling is able to identify growth-related genes that would not have been identified if only plant size data were used for mapping.

Selection of candidate genes for drought-QTLs

In our search for genes that could be causal for the observed QTLs, the support window of each QTL was determined (Table 3). For each of the strongly associated SNPs detected by GWA, the SNPs in LD (r^2>0.4) were determined. The support window was defined as the region between the first and last SNP in LD. Large differences were observed in the

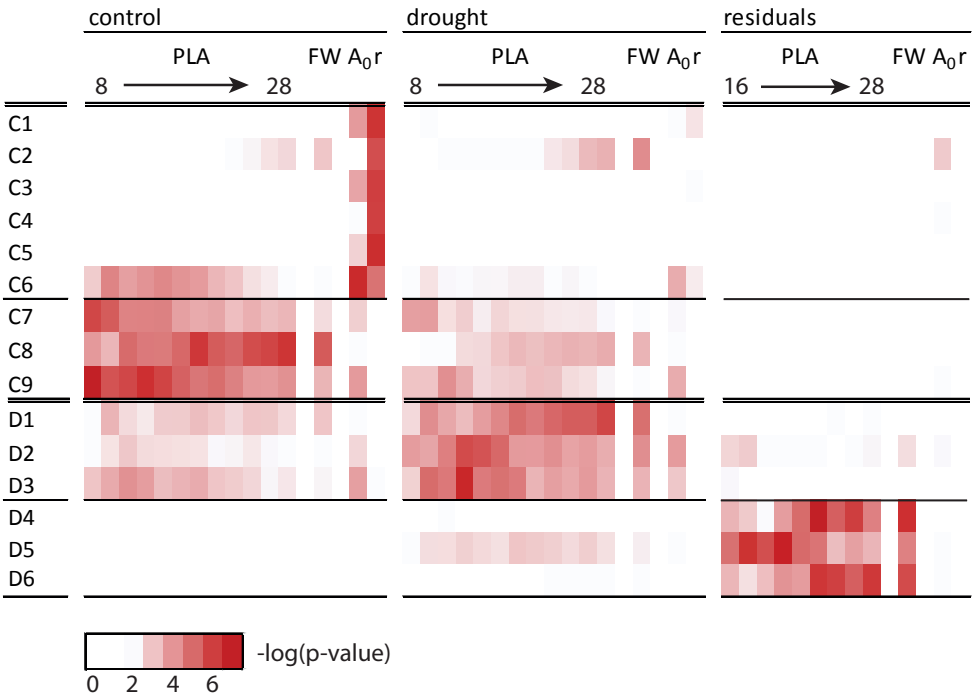


Figure 5 Heat map of the QTLs detected by GWA mapping of PLA over time and FW in control and drought conditions and the corresponding residuals. QTLs labels (Control: C1- C9 and drought: D1-D6) refer to the QTL positions as depicted in Figure 4. Colour indicates association between the QTL indicated in the first column and the trait indicated above the column. Intensity of the colour corresponds with strength of association as indicated by the colour scale.

number of linked SNPs varying from two to 458. This variation indicates that the LD decay is varying between genomic regions and that, therefore, the resolution of the GWA mapping is different for various genomic regions. The number of candidate genes located in the support window was also counted. In general, the number of candidate genes was correlated with the size of the support window, but because the gene density is not equal along the chromosomes, support windows with similar size sometimes had more than two-fold difference in number of candidate genes (compare QTLs C1 and C6 or QTLs C9 and D2, Table 3). For nine of the fifteen QTLs, very high resolution of mapping was reached, with ten or less genes located in the support windows. Such a low number of candidate genes ensures that fine-mapping is not needed and functional analyses of genes can be the immediate next step.

For the six drought-QTLs, the resolution of the GWA mapping was high for QTL D2, D3 and D5 and lower for QTL D1, D4 and D6. Depending on the mapping resolution of the drought-QTLs, all or a limited number of candidate genes were studied in detail. For the QTLs with lower resolution, genes were prioritized based on reported expression studies (Table S1 (Huang 2008, Harb 2010, Baerenfaller 2012, Des Marais 2012).

QTL	Chro-mo-some	SNP position (Mb)	Model parameters C	D	res	FW + PLA C	D	res	Minor allele frequency	Phenotypic variance explained	# SNPs in LD	LD up-stream (kb)	LD down-stream (kb)	# genes in support window
C1	1	23.56	**6.1**	2.4					0.05	7.53	22	-5	+7	7
C2	1	29.35	**5.5**		2.8	2.9	4.1		0.05	6.89	149	-46	+83	32
C3	3	4.42	**5.8**						0.18	7.49	5	-0	+12	5
C3	3	4.42	**5.7**						0.18	7.46	7	-0	+12	
C4	4	10.24	**5.8**						0.05	6.99	134	-86	+99	55
C5	4	17.33	**6.2**						0.06	7.80	5	-6	+4	4
C6	5	18.29	**6.2**	3.4		4.3	2.4		0.09	7.87	6	-1	+11	2
C7	1	24.35	2.7	2.0		**5.7**	3.7		0.19	6.96	44	-15	+11	9
C7	1	24.36	2.4			**5.6**	3.0		0.20	6.91	41	-8	+7	
C8	4	6.48				**6.0**	3.4		0.46	6.84	8	-0	+1	2
C9	4	13.48	4.6	3.7		**5.6**	3.1		0.20	7.65	47	-57	+30	32
C9	4	13.49	3.8	3.4		**6.8**	4.0		0.30	8.58	12	-24	+11	
D1	2	14.38				3.3	**5.7**		0.13	7.44	243	-92	+102	135
D1	2	14.46	2.1	2.1		3.5	**5.7**	2.1	0.06	8.48	343	-278	+106	
D2	3	15.43	2.6	3.7	2.0	2.9	**5.6**	2.8	0.38	7.05	290	-67	+23	10
D3	5	18.35	3.7	2.9		4.0	**6.3**	2.1	0.42	7.95	3	-0	+0	1
D4	3	5.72						**6.1**	0.05	7.35	98	-77	+98	53
D5	5	1.18						**5.5**	0.46	6.53	2	-0	+0	1
D6	5	6.55						**5.6**	0.08	6.06	458	-175	+673	237
D6	5	6.55						**5.6**	0.08	6.06	406	-139	+651	
D6	5	6.56						**5.6**	0.08	6.01	400	-183	+670	

Table 3 Significant associations and corresponding QTLs detected by GWA mapping of PLA, FW and model parameters of control and drought treated plants. Only SNPs that were strongly associated (-log (p-value)>5.5) with at least one trait are displayed. Bold indicates significant SNPs (-log(p-value>5.5), no value indicates no association (-log(p-value)<2). QTLs labels (Control: C1- C9 and drought: D1-D6) refer to the QTL positions as depicted in Figure 4.

For two high-resolution drought-QTLs, D3 and D5, only one candidate gene was identified. For drought-QTL D3, the associated SNP detected in the GWA analysis and the three SNPs in LD were all located in an intron of *PAE11* (AT5G45280), a pectin acetyl esterase (Table S1). For this gene, a functional analysis has not been performed yet, but mutants of two functional homologs, *PAE8* (AT4G19420) and *PAE9* (AT5G23870), were studied in more detail. Mutants have higher acetate content of the cell walls and reduced inflorescence height, indicating that removal of acetate from pectin compounds in the cell wall is needed for normal growth (de Souza 2014). Interestingly, this gene is found to be down-regulated upon moderate drought in two expression studies (Table S1) (Harb 2010, Des Marais 2012). The gene is also differentially expressed upon severe drought, but two studies reveal opposing directions of change (Table S1) (Huang 2008, Harb 2010). For the other single candidate drought-QTL, D5, the associated SNP and one of the two linked SNPs were located in the 5'UTR of AT5G04250. The other linked SNP was located in the promoter region of the same gene. AT5G04250 is a member of the cysteine proteinases superfamily, but no detailed functional analysis has been performed on this gene. It was found to be up-regulated upon moderate drought in summer annuals (Des Marais 2012) and in the Col accession upon severe drought (Harb 2010) (Table S1). For nine out of the ten genes located in the support window of the other high resolution drought-QTL, D2,

it is unknown in which biological process they are involved. The tenth gene, AT3G43550, is annotated to be involved in lipid metabolism, but detailed functional analysis of this gene is also not reported yet. None of these ten have been reported to be differentially expressed upon moderate drought, but two genes with unknown function, AT3G43520 and AT3G43540, were differentially expressed upon severe drought (Harb 2010).

Drought-QTLs D1, D4 and D6 were located in genomic regions with slower linkage decay and therefore larger support windows were observed. Genes within these regions were considered more likely candidates when they were differentially expressed in at least three of five expression studies (Table S1) (Huang 2008, Harb 2010, Baerenfaller 2012, Des Marais 2012, Clauw 2015). For D1, three of the 135 genes in the support window were considered more likely to be causal. AT2G33590 was up-regulated upon severe drought (Østergaard 2001, Huang 2008, Harb 2010) and ABA treatment (Nemhauser 2006) (Table S1). However, this gene was significantly down-regulated in winter annuals upon moderate drought stress (Des Marais 2012) (Table S1). Similar expression patterns were observed for AT2G33770, also located in the support window of drought-QTL D1, which is also up-regulated upon severe drought and down-regulated upon moderate drought in summer annuals (Table S1). These data suggest that both genes may play a different role in moderate and severe drought responses. Expression levels of the third candidate gene of drought-QTL D1 were inconsistent between expression studies. Upon moderate drought both up- and down-regulation is observed, whereas severe drought and ABA treatment consistently results in down-regulation (Table S1). For the other two drought-QTLs with large support windows, four out of 53 genes were likely candidates for D4 and nine out of 237 genes were likely candidates for D6. All but one of those prioritized genes were consistently up- or down regulated in the response to severe drought, indicating that the drought treatments given by Harb *et al.* (2010) and Huang *et al.* (2008) resulted in similar drought responses. However, the three moderate drought expression studies (Harb 2010, Baerenfaller 2012, Des Marais 2012) show inconsistency in the response of the more likely candidates to moderate drought stress. Differences in up- and down-regulation between the different studies indicate that moderate drought may lead to diverse stress responses. However, when a gene is differentially expressed upon drought in several experiments, whether up or down, it suggests that it is part of the network that is responding to drought. It might be that the balance between the genes in the network is changing when drought level is increasing or the drought period is prolonged. Whether the observed expression changes upon drought lead to drought adaptation or resistance is not known yet for any of the likely candidates. In general, for the candidate genes for which expression is induced or reduced by ABA treatment, expression change with the same sign is observed in the severe drought treatment, indicating a profound role of ABA in severe drought responses, as was suggested earlier (Gosti 1995, Nakashima 2014).

In conclusion, GWA mapping is a powerful tool to identify genomic regions associated with drought response and plant size upon drought treatment. The investigation of drought response over time was crucial for the successful identification of most of the observed QTLs, because they would not have been identified if plant size was only determined on a single time point. Size of the support window was varying, resulting in one to 237 candidate genes per QTL. Analysis of gene annotation and previously reported expression data allowed us to select the more likely candidates for each QTL. Differences were observed in the direction of the expression change of likely candidates between moderate and severe drought treatments and even between independent moderate drought treatments indicating that drought responses are complex and may vary upon changes in severity and length of the drought period.

Discussion

In this study we determined the response to stable moderate drought stress in a large number of natural accessions of Arabidopsis. Rosette expansion was determined over time enabling us to add a time-dimension to the GWA mapping. Six time-dependent drought-QTLs were identified and likely candidate genes, explaining part of the observed variation, were identified based on previously reported expression changes upon drought. Interesting correlations were observed between plant size and RWC, and plant size and flowering time. Both correlations suggest overlap in regulatory networks of both traits or correlated selection of two traits in the same environments.

Mapping of drought adaptation over time leads to time-dependent QTLs

Already two days after the start of the drought treatment (day 16) significant differences were observed in plant size between control and drought treated plants. This indicates that drought is already sensed when the level of stress is still very mild and water is still available. Adjustment of growth upon the sensing of the stress was a gradual process since no temporary strong inhibition with subsequent regain of growth was observed for any accession. Temporary strong growth inhibition was reported in many species (Skirycz 2010), but regain of growth can be very fast (Veselov 2002) and therefore will only be visible in experiments with higher time resolution. The gradual reduction in plant size upon moderate drought allowed us to model the growth in control and drought stress conditions with the same exponential equation, allowing comparison of the parameters between the two conditions.

GWA mapping of model parameters resulted in strong associations for the control conditions, but no strong associations were detected for the drought conditions or the corresponding residuals. In both conditions, more than 40% of the variation in the parameters could be explained by the genotype (H^2>0.5). Detection of strong associations can be hampered by, amongst other factors, epistatic interactions and the absence of

causal genes with moderate to large allelic effects. However, many moderate associations explaining less than 8% of the phenotypic variation were observed for the model parameters in drought conditions and the residuals, indicating that lowering of the threshold might lead to the identification of genes with small allelic effects.

GWA mapping of temporal plant size data resulted in the detection of six time-dependent drought-QTLs. Time-dependency is probably a consequence of genotypic differences for both the timing of the start of growth inhibition and the level of growth inhibition. For example, for some accessions the plant size reduction at the end of the experiments was very similar, whereas the timing of the start of the growth retardation was not equal. The time-dependent QTLs were strongly associated with plant size on one specific day, whereas on the other days only moderate or weak associations were observed. This suggests that the underlying genes contribute to the phenotype on multiple days, but that the contribution is highest on the day for which the strong association was observed. Whether the causal gene is expressed on multiple days and whether the causal proteins are active on multiple days cannot be concluded from the data. We suggest four scenario's. First, the causal gene is highest expressed on the day that the strongest association is observed but is not zero on the other days. Second, the causal gene is constantly expressed over a longer period, but protein activity is changing over time due to difference in protein degradation and post-translational modifications. Third, the causal gene is expressed during a specific time-period, but protein degradation is slow and therefore the effect of the gene on the phenotype is observed also when the causal gene is no longer expressed. Fourth, the causal gene is expressed during a specific time-period, but downstream effects will contribute to the phenotype even when the causal gene is not expressed anymore. Only the first two scenarios can explain the association pattern of QTLs with strongest associations at the middle or end of the experiment and weaker associations on earlier days, like drought-QTL D1. In all four scenarios it is expected that overexpression or mutation of the causal gene has influence on the drought response on multiple days.

Time-dependency of the QTLs emphasizes that measuring growth over time is to be preferred over single time point measurements, because most QTLs would have been missed if plant size had only been determined on one specific day.

Most candidates were not annotated to be involved in drought response regulation

For each drought-QTL it was determined whether genes in the support window were reported to be differentially expressed upon moderate and/or severe drought (Huang 2008, Harb 2010, Baerenfaller 2012, Des Marais 2012, Clauw 2015). If differential expression was observed in three or more independent experiments a gene was labelled 'more likely' (Table S1). This approach resulted in the selection of genes with several functions. Some functions could be easily connected with drought. In the support window of drought-QTL D6 four more likely candidates were involved in sugar metabolism. These genes

were also reported to be differentially expressed upon other abiotic stresses (Dejardin 1999, Jayashree 2008, Li 2009, Maruyama 2009). Increase of soluble sugars is associated with drought tolerance in several species (Mohammadkhani 2008, Loukehaich 2012, Redillas 2012). Several of the more likely candidates were not only differentially expressed upon drought, but also upon ABA treatment (Nemhauser 2006). Drought response is tightly connected with ABA concentration, since this hormone regulates the opening and closing of the stomata (Schroeder 2001). However, drought response is also partly regulated in an ABA-independent manner (Gosti 1995). An ethylene response factor (ERF72) was identified for drought-QTL D4. Several other ERFs were reported to be involved in drought tolerance in *Arabidopsis* (Cheng 2013, Licausi 2013) and other species (Joo 2013, Rong 2014). Some of these ERFs are called drought responsive element binding (DREB) factors and DREBs are central in the regulation of the ABA independent drought response (Nakashima 2014). In addition to genes with known function, also genes of which the biological function is not known yet were identified as likely candidates based on expression patterns. For example, the function of one of the three likely candidates in the support window of drought-QTL D1, AT2G34510 is not known yet, but it was reported that this gene is differentially expressed upon ABA and drought treatment. This shows that GWA mapping does not only confirm the presence of allelic differences in genes known to be involved in drought tolerance, but that this method also identifies QTLs of which the causal gene is not annotated yet to be involved in drought tolerance. For drought-QTL D2 none of the ten genes in the support window could be selected as more likely candidate. Because the environmental conditions and the timing of harvest were different between the expression experiments, genes that respond to drought for a longer time or in several developmental stages have a higher chance to be labelled 'more likely' candidates. In addition, causal genes of which the non-Col allele is drought responsive, but the Col-0 allele is not, could only be observed as differentially expressed in the studies of Des Marais *et al.* (2012) and of Clauw *et al.* (2015), since the other studies investigated only the Col-0 accession. These can be reasons why for drought–QTL D2 no likely candidate could be selected.

Recently, a study was published that describes GWA mapping of severe drought responses in the same population (El-Soda 2014). In this study not any strong association was detected and none of the moderate associations reported coincided with the strong associations observed in the present study. This suggests that severe and moderate drought are regulated differently as also concluded from the comparison of mutant behaviour (Skirycz 2011b) and gene expression (Harb 2010). Both breeders and researchers should be aware of these differences, especially in the translation of findings from one experiment to another.

Carbon availability, RWC and plant size are linked

In both control and drought conditions variation was observed between accessions for rosette FW and RWC (Figure 2). On average smaller plants with lower RWC were observed upon moderate drought. When trait values observed in control and drought stress conditions were compared, a positive correlation was observed for both traits. This correlation was very strong (r^2=0.8) for rosette FW, but moderate (r^2=0.4) for RWC (Table 2). In a study of 24 accessions in control and mild drought stress conditions, also positive correlations were observed between trait values obtained in control and drought conditions for plant size and RWC, and also in this study the correlation for plant size was much higher than for RWC (Bouchabke 2008). However, when FW and RWC were correlated with each other in either condition, we observed a moderate negative correlation whereas Bouchabke *et al.* (2008) did not observe any correlation. Absence of correlation can be due to small sample size or due to differences in growing conditions that influence growth severely, like length of light-period or light intensity. In our experiment, the variation in RWC was independent of variation in WC and therefore mainly caused by differences in metabolite content. In a study of osmotic adjustment upon salt stress and drought in cereals, also low relative water content was observed for turgid leaves, but only for stressed and not for control leaves (Boyer 2008). Because we observe a negative correlation between RWC and FW, independent of the growth conditions, osmotic adjustment is probably not the main cause. We propose that differences in carbon availability are underlying this negative correlation, because in a study of 95 accessions grown in non-stress conditions, a negative correlation was observed between biomass and starch content at the end of the day (Sulpice 2009). Less starch, but more growth, means probably that a smaller fraction of the glucose produced through photosynthesis is converted to starch and therefore more glucose and other simple carbohydrates are available as energy source in glycolysis during the day. Simple carbohydrates influence the osmotic potential much more than starch, therefore it is expected that plants with higher carbon availability will have a lower RWC independent of the conditions in which they were grown. These observations do not allow conclusions on the causality of processes; it is open whether plant size influences RWC or *vice versa*.

A weak negative correlation was observed between the residuals of FW and the residuals of RWC indicating that plants that perform well under drought have relatively larger reduction in RWC upon drought (Table 2). Similarly, a weak negative correlation was observed when the residuals are calculated using the plant size and RWC data of 24 accessions in control and mild drought stress conditions reported by Bouchabke *et al.* (2008). This suggests that drought tolerance can be partly explained by changes in the content of soluble metabolites. Soluble sugars and potassium content are reported to increase upon drought. Higher soluble sugar content is a consequence of reduced growth, but maintained photosynthetic activity (Hummel 2010). Changes in soluble metabolite content can only explain a part of the variation in drought response, because the observed

correlation between the residuals of FW and RWC was weak.

Flowering time, plant size and drought response are linked

In both control and drought stress conditions it was observed that early flowering plants were larger than late flowering plants form day 11 onwards. In addition early flowering plants performed on average better upon drought stress than late flowering plants (Figure 3). This suggests a genetic relationship between plant size, drought response and flowering time, but we did not find any flowering time gene in the support window of any of the control- or drought-QTLs identified by GWA mapping. However, a SNP in LD with the well-known flowering time gene FLC was moderately associated with plant size in drought conditions (-log(p)=4.96) possibly indicating a weak role for this gene in the regulation of growth upon drought. Such a genetic relationship is supported by findings in several studies of drought responses in different set of genetic plant material. In two RIL populations (L*er* x An-1, (Tisné 2010) and Sha x Col (El-Soda 2014)) colocation of flowering time and drought-QTLs was observed. Further research of near isogenic lines of the three QTLs detected in the L*er* x An-1 population confirmed that late flowering plants were more drought tolerant, because of better adaptation to the drought over time (Schmalenbach 2014). In two other RIL populations (L*er* x Cvi (Juenger 2005) and Kas-1 x Tsu-1(McKay 2008)) and also in 39 natural accessions (McKay 2003), late flowering was connected to high water use efficiency (WUE) and therefore to good performance upon drought. In drought conditions, plants with higher WUE are expected to perform better than plants with low WUE, but in control conditions the opposite is true. In a set of 35 genetically modified *Arabidopsis thaliana* lines impaired in
different functional pathways including carbohydrate metabolism, circadian clock, signalling or hormonal status, individual leaf area of early flowering genotypes was more affected by drought than late flowering ones (Massonnet 2015). In our experiment late flowering accessions were smaller than early flowering ones, independent of the conditions. In addition, considering the drought response independent of the plant size in control conditions using the residuals, late flowering was connected with poor performance upon drought which is the opposite of other findings in the same species. This discrepancy could be due to the different experimental procedures used in the different studies. In the present study, we compared leaf growth and its response to a moderate and stable soil water deficit imposed soon after germination until day 28. This means that drought response was compared at same date for all accessions. In other studies, drought treatment was imposed until flowering for all genotypes and revealed compensation in the leaf growth response to drought with a lengthening of the duration of the vegetative phase in late flowering genotypes (Schmalenbach *et al.* 2014; Massonnet *et al.* 2015). However, both our mapping results (moderate association) and the mapping results of McKay *et al* (2008) suggest FLC as one of the regulators of differences in drought response. Associations between flowering time and plant performance upon drought have also been reported in various crops (Sadras 2009, Tuberosa 2012). All mentioned

observations can be explained as pleiotropic effects of flowering time genes, but they can also be the consequence of correlated responses of several traits to selection. In nature in many locations early flowering accessions germinate in spring and flower in summer. For plants with such a short lifecycle fast growth is favourable. Drought is mostly experienced in summer and therefore local adaptation would lead to correlated responses of early flowering, fast growth and drought tolerance to selection. Differences in flowering time can, however, only explain a part of the differences observed in plant size and drought response, indicating that many other processes besides flowering time have impact on these traits.

Conclusion

Large natural variation was observed for biomass accumulation upon moderate drought. GWA mapping resulted in the detection of six strong drought-QTLs. These QTLs were time-dependent due to variation between accessions for the timing of the start of the drought response and the magnitude of the response. For five of the six drought-QTLs most likely candidates could be selected based on previously reported expression changes. For many candidate genes their function in drought response was described, but for some their suggested role in drought response is unknown. GWA mapping is therefore a method that not only detects allelic differences in known drought response genes, but is also able to identify allelic differences in unknown drought-related genes, opening up new avenues for research on drought tolerance. Plant size was correlated to flowering time and RWC, indicating overlap in regulatory networks or correlated responses of these traits to selection.

Materials and Methods

Plant materials
A collection consisting of 324 natural accessions of *Arabidopsis thaliana* was used to investigate growth of the rosette over time in control and moderate drought stress conditions.These accessions were selected to capture most of the genetic variation present within the species. (Baxter 2010, Li 2010, Platt 2010). Each accession has been genotyped with approximately 215k SNP markers (Col vs non-Col)(Kim 2007, Atwell 2010).

Experimental set-up
The PHENOPSIS phenotyping platform was used to perform the experiments (Granier 2006). The climate conditions within the growth chambers of PHENOPSIS are precisely regulated, preventing difference in growth because of position in the chamber. The plants were grown in four rounds of experiments each containing 84 accessions. Four of these accessions were reference accessions that were grown in each round (Col-0, CS76113; KBS-Mac-8, CS76151; Lillo-1, CS76167; Wc-2, CS28814). Note that the reference accessions (checks) were used to connect the different experiments as all of the other accessions were only grown in one experiment. Each experiment contained 3 completely randomised blocks and all 84 accessions were present in each block in two replicates, one replicate was

given sufficient water and one replicate was given a moderate drought treatment. Cylindrical pots (9 cm high, 4.5 cm diameter) filled with a mixture (1 : 1, v : v) of a loamy soil and organic compost were used and the seeds were sown directly on the soil. These pots contained holes to guarantee equal soil drying throughout the whole pot (Granier 2006). At the start of the experiment the perforated pots were place in cover pots to minimize fluctuation in water content. Cold treatment (4°C) was given to the pots directly after sowing. To avoid too much difference in harvesting time of the rosette, the three blocks were transferred from the cold to the climate room (PHENOPSIS, 16h light, 125 µmol m^{-2} s^{-1}, 70% air humidity, 20/18°C air temperature during day and night respectively) on sequential days, four, five or six days after sowing. The day the plants were transferred to PHENOPSIS was marked as day zero. The first four days the pots were sprayed with water. Thereafter each pot was weighed and adjusted to the desired soil water content twice a day. The first 2 weeks all pots were adjusted to 0.35-0.40 g H_2O g^{-1} soil and on day 15 the plants were thinned to one plant per pot. Thereafter the pots that received control treatment were adjusted to 0.40 g H_20 g^{-1} soil until harvest, but watering was stopped for the pots that received the drought treatment and the cover pots were removed. If the soil water content of the drought treated pots reached 0.22 g H_2O g^{-1} soil the cover pot was placed back and the pot was kept at this new soil water content until the end of the experiment.
A separate experiment was performed to determine whether the accessions were summer or winter annuals. All accessions (three replicates) were grown on rockwool blocks in the greenhouse (long summer days and 70% air humidity) and were watered regularly. Flowering time of the first replicate of each accession was recorded. Accessions that flowered within 75 days were classified as summer-annuals, whereas accessions that did not flower within this period were classified as winter-annuals.

Determination of growth and water status
All plants were daily inspected for visible signs of bolting and bolting dates were recorded. At day 24, the largest leaf of each plant was harvested. Directly after harvest the leaf was weighed to determine the fresh weight (FW). Thereafter the leaf was put with its petiole in water and after 24 h at 4°C the leaf was weighed to determine the turgid weight (TW). Subsequently, the leaf was dried at 65°C for at least 48 hrs and weighed to determine the dry weight (DW). Water content (WC) was calculated by WC = (FW-DW)/FW and the relative water content (RWC) by RWC= (FW-DW)/(TW-DW). At day 28, the rosette of each plant was harvested and the rosette FW was determined. Growth was monitored by taking pictures twice a day. These pictures were processed in ImageJ using the macros developed for PHENOPSIS. All pictures and the ImageJ macros are publically available on PHENOPSISDB ((Fabre 2011), http://bioweb.supagro.inra.fr/phenopsis). The projected leaf area of each plant (PLA) was determined semi-automatically on the following days: 8, 11, 14, 16, 18, 20, 22, 24, 25, 26, 27 and 28. When more than one plant was present in a pot before thinning, the largest one close to the middle of the pot was taken for analysis. For each individual plant the growth dynamics, based on PLA, was modelled with an exponential function ($PLA= A_0 * e^{r*t}$) in which PLA represents the Projected Leaf Area, A_0 represent the initial size and this parameter is also a magnification factor, r represents the growth rate and t represents the time after transfer to the PHENOPSIS climate chamber. The optimal parameter values were estimated using the Growth Fitting Toolbox™ of MATLAB with the following settings; optimization algorithm: 'Trust-region', fitting method: Non-linear least square, bounds: A_0[0,Inf], r[0,Inf]. The removal of the largest leaf from day 24 onwards was not corrected for when the fresh weight of the rosette and the model parameters were determined because plants rapidly compensated for this loss.

Statistical analysis
GWA mapping was performed on the rosette FW and RWC at day 28, PLA at days 8, 11, 14, 16, 18, 20, 22, 24, 25, 26, 27 and 28, the estimated model parameters and the corresponding residuals of a linear regression between trait values in control and drought conditions. Residuals represent the drought response independent of plant size in control conditions. Model parameters from curve fits with $r^2<0.9$ were not taken into account. For all traits, adjusted means were obtained with GenStat by fitting the following mixed model: Trait (FW,PLA or model parameters) = µ + Treatment + Accession(Round x Treatment) + Round + Block(Round) + Round × Check + Block(Round) × Accession + error. The terms Treatment and Accessions within Round and Treatment were assumed fixed to obtain Best Linear Unbiased Estimates (BLUE), and all the remaining terms were considered random effects as they are all

essentially different sources of experimental error due to Round, Blocks within Rounds (and the interaction with check genotypes) and residual variation. The term Check refers to a factor in which reference accessions were distinguished from the rest. All other terms are used as defined in the experimental set-up. GWA mapping was performed on the adjusted means (log-transformed) and the corresponding residuals using the EmmaX software package for R, which is based on (Kang 2010). A mixed model was used that corrects for population structure, based on the kinship matrix of all SNPs. SNPs with a minor allele frequency < 0.05 were excluded from the analysis. Pearson correlations were used to determine correlations between data series. To calculate the correlation between traits, the adjusted means (derived from log-transformed data) were used. Broad-sense heritabilities were obtained with GenStat by fitting the following mixed model separately to the data of control and drought treated plants: μ + Accession(Round) + Round + Block(Round) + Round × Check + error. All terms were considered random effects. Broad sense heritability at the mean level was calculated as: $H^2 = V_g/(V_g+V_e/r)$, where V_g is the genetic variance and V_e is the residual variance.Differences in FW between the rosettes of plants that were bolting at day 28 and plants that were still vegetative, and between summer and winter annuals were determined using a t-test assuming non equal variances and α=0.05.

Acknowledgement
We thank Sabine Schnabel and Marcos Malossetti for their mathematical and statistical support, Myriam Dauzat for operating and maintaining PHENOPSIS, and Gaelle Rolland and Crispulo Balsera for their assistance during sowing and harvesting.

5

Supplementary data

Table S2 Likely candidate genes per drought-QTL. We indicate whether the candidate gene was found to be up- or down-regulated in the mentioned expression studies or upon ABA treatment. mDr = moderate drought, sevDr = severe drought, d1 = 1 day treatment, d10 = 10 days treatment, Summer = summer annuals, Winter = winter annuals.

QTL	Candidate gene	Name + function	Baeren-faller[a]	Des Marais[b]		Clauw[c]	Harb[d]			Huang[e]	Nem-hauser[f]
				Summer	Winter			d10			
			mDr	mDr	mDr	mDr	d1 mDr	mDr	sevDr	sevDr	ABA
D1	AT2G33590	CRL1, generation of vascular tissue, upregulated during drought (Ostergaard 2001)			Down				Up	Up	Up
	AT2G33770	PHO2, regulation of phosphate homeostasis (Liu 2012)		Down					Up	Up	
	AT2G34510	biological function unknown		Up	Up		Down	Down	Down	Down	Down
D2		No likely candidate									
D3	AT5G45280	PAE11, pectin esterase		Down	Down			Down	Up	Down	
D4	AT3G16670	Pollen Ole e 1 allergen and extensin family protein	Up	Down	Down			Down	Down		
	AT3G16770	ETHYLENE RESPONSE FACTOR 72, transcription, connects GA and ethylene signaling pathways (la Rosa 2014)			Down		Up		Up	Down	
	AT3G16990	Haem oxygenase-like, multi-helical			Down		Up		Up	Up	Up
	AT3G17000	UBC32, brassinosteroid-mediated salt stress tolerance (Cui 2012)		Down	Down		Up		Up	Up	Up
D5	AT5G04250	Cysteine proteinases superfamily protein		Up					Up		Up
D6	AT5G19120	Eukaryotic aspartyl protease family protein				Down		Down	Down	Down	Down
	AT5G20040	isopentenyltransferase 9, cZ-type cytokinin production		Up	Up	Down			Down		
	AT5G20160	involved in ribosome biogenesis			Up				Down	Down	Down
	AT5G20230	SENESCENCE ASSOCIATED GENE 14	Up				Down	Down		Down	Up
	AT5G20250	RAFFINOSE SYNTHASE 6, response to cold and dehydration (Maruyama 2009)	Up			Down	Down	Down	Down	Down	Down
	AT5G20280	SUCROSE PHOSPHATE SYNTHASE 1F, overexpression leads to large plants with higher sucrose content and elongated fibers (Park 2008)			Up				Up	Up	
	AT5G20630	GERMIN 3, encodes a germin-like protein	Up						Down	Down	
	AT5G20720	CHLOROPLAST CHAPERONIN 10, negatively regulates ABA signaling (Wu 2009)		Up					Down	Down	
	AT5G20830	SUCROSE SYNTHASE 1, expression induced by cold and drought (Dejardin 1999)			Down				Up	Up	Up
	AT5G21170	AKINbeta1, involved in regulation of nitrogen and sugar metabolism (Li 2009)			Down				Down	Down	
	AT5G21090	Leucine-rich repeat (LRR) family protein	Up		Down		Up				

[a] (Baerenfaller 2012), [b](Des Marais 2012), [c](Clauw 2015), [d](Harb 2010), [e](Huang 2008), [f](Nemhauser 2006)

Chapter 6 - Detecting constitutive QTLs for plant biomass by Multi-Environment Genome Wide Association mapping

Johanna A. Bac-Molenaar[1,2], M. Joao Caldas Paulo[3], Rik Kooke[1,2], Willem T. Kruijer[3], Nelson H. Davila Olivas[4], Nihal Oztolan-Erol[2], Charles N. Moreira[2], Mohamed El-Soda[2], Mark G.M. Aarts[2], Dick Vreugdenhil[1], Joost J.B. Keurentjes[2], Fred A. van Eeuwijk[3], Marcos Malosetti[3]

[1] Laboratory of Plant Physiology, Wageningen University
[2] Laboratory of Genetics, Wageningen University
[3] Biometris, Wageningen University and Research centre
[4] Laboratory of Entomology, Wageningen University

Abstract

To study the regulation of biomass accumulation across a range of different environments, rosette fresh weight (FW) data of 350 natural accessions of *Arabidopsis thaliana* grown in 25 different environments were analysed jointly. Large variation in biomass accumulation was observed between accessions and environments. Only weak and moderate correlations were observed between FW data in the different environments suggesting an important role for Genotype-by-Environment (GxE) interactions. Multi-environment Genome Wide Association (GWA) mapping resulted in the detection of 64 independent loci, with significant effects in 7 to 16 environments. These strong QTLs were predominantly main effect QTLs but also significant but weaker QTL-by-environment (QxE) effects were identified. Twelve candidate genes involved in the regulation of flowering time, starch metabolism, and leaf size were located in close proximity of the strong QTLs. In addition, three candidate genes reported to be involved in fitness trade-offs, were identified in proximity of QTLs with significant QxE effect. Whole genome prediction was performed with the strong QTLs as well as with the combination of strong and weak QTLs. The first prediction could explain less of the total genetic variance per environment than the latter (21%-57% and 26%-96%, respectively). Correlations between the environments based on the first prediction were moderate to high, whereas the correlations based on the second prediction were predominantly weak and reflected the correlations based on the phenotypic data. From the comparison between observations on the phenotypic level with both predictions we conclude that GxE plays an important role in biomass accumulation, but only a small part can be explained by the QxE effects of the strong QTLs. Rather, GxE is mainly based on small QxE effects of weak QTLs.

Introduction

Phenotypic plasticity, the ability to sense changes in the environment and respond to them, is observed in all plant species. Because plants are sessile, phenotypic plasticity is crucial to survive harsh conditions. In wild plant species plasticity is larger than in crop species as a consequence of the domestication process (Bohnert 1995, Des Marais 2013). In *Arabidopsis thaliana,* for example, large differences are observed in the phenotypic plasticity of natural accessions for biomass accumulation under optimal and biotic or abiotic stress conditions (Bouchabke 2008, Lisec 2008, Chardon 2010, Jha 2010, Todesco 2010). These differences, which might be the result of local adaptation, are the basis of QTL and Genome Wide Association (GWA) mapping studies searching for biomass accumulation and stress tolerance genes.

When plants sense changes in their environment, usually their growth rate changes. Biomass accumulation is therefore a trait that is less stable over environments than other traits, such as the number of leaves or flowering time (Koornneef 2004). Biomass accumulation is a complex trait determined by many genes with often small allelic effects that can work synergistically or antagonistically. It is of great interest to know whether the allelic effects of these genes are similar in all environments or that a specific effect is observed in one environment only or in a number of related environments. Alleles with a biomass enhancing effect in multiple environments are of special interest for translational research. In crops, QTLs (and underlying genes) for which QTL by environment interactions (QxE) play a minor role in the determination of biomass accumulation, are of interest in breeding programs aiming at the development of cultivars with stable biomass production in multiple environments (Chaïb 2006, Collins 2008). To identify these 'stable' or 'constitutive' Quantitative Trait Loci (QTLs) (Lempe 2013), a good estimation of the QxE effect is crucial (Des Marais 2013, Malosetti 2013, El-Soda 2014b).

To effectively use natural variation in the search for genes that determine plant biomass in different environments, large investments in the phenotyping of populations in different locations and climates are a prerequisite. To detect the relevant QTLs, acquired phenotyping data can be analysed with several statistical methods, which can be grouped in three classes. The most basic type of analysis is to perform univariate mapping for each environment separately (Hittalmani 2003, Li 2007, El-Soda 2014b, Norton 2014). This approach does not allow to test for the effect of the same marker in all environments but it allows to separate between environment specific and constitutive QTLs. Another type of analysis is univariate mapping of parameters that summarise genotypic trait responses across environments, such as principal components or regression parameters (Malosetti 2013). The most advanced analysis is multi-environment QTL or GWA mapping, which jointly analyses the data collected in all environments (Korte 2012, Segura 2012, Alimi 2013, El-Soda 2014a). This method has the power to detect QTLs by a general model

with separate effects per environment, and also permits testing for main effects and QxE effects. In addition, the information about the marker effect in each environment and the test for QxE give valuable information about the selection pressure in nature leading to local adaptation and adaptive plasticity of genotypes (Via 1995).

In the search for genes that determine vegetative plant biomass in a wide variety of environments, natural variation in rosette biomass was determined in 25 different environments. Differential adaptation of accessions to the environments (GxE) was observed. Joint analysis of the data by multi-environment GWA mapping resulted in the detection of many constitutive QTLs with significant effect in several environments. For most QTLs the same allele was associated with better performance in all environments. Whole genome prediction (WGP) revealed that differential adaptation of accessions can be explained by QxE interactions. Limited QxE was observed for strong QTLs. Most of the GxE could be explained by the moderate and weak QTLs with small QxE effects.

Results

Variation observed between accessions within and between environments

Rosette biomass was determined under 25 different conditions (environments) defined by differences in growth location (climate room or greenhouse), substrate, daylight duration, intensity of light, relative humidity, and day and night temperature. In addition, in some environments the plants were exposed to abiotic stresses, such as drought, N or P deficiency and osmotic stress (PEG8000 or NaCl) (details in Table S1). Large differences were observed between mean rosette fresh weight (FW) in the different environments (Table 1). As expected, plants grown in environments with no stress (identified as 'control') were, on average, larger than plants that were exposed to abiotic stresses (Table 1). Environment 20_Control was the non-stress environment with lowest mean FW, which can partially be explained by the short day length (8 hours)(Sulpice 2014). Plants were relatively large in four stress environments, 4_Pdef, 5_Salt, 11_Ndef and 12_Drought, pointing to relatively favourable conditions despite the stress. The large range in mean FW indicates that the environmental conditions in which plants were grown had a large impact on plant performance. For most environments, heritabilities (H^2) ranged from 0.40 to 0.85 indicating that a substantial part of the observed phenotypic variation in rosette biomass within an environment could be explained by genotypic differences (Table 1). The lowest heritability (H^2= 0.24) was observed for environment 19_Pdef, resulting from both lower genetic variance and high residual variance. No relationship was found between mean FW and H^2, indicating that the amount of variation between genotypes is not dependent on the overall performance in an environment.

Table 1. Estimates of mean FW (log(g/rosette)), standard deviation (log(g/rosette)), number of replicates, genetic variance (V_g, (log(g/rosette))^2), residual variance (V_e, (log(g/rosette))^2) and heritability (H^2) of rosette FW in 25 environments. In addition, genetic variance explained by whole genome predictions (WGP) with only strong or with strong and weak associations. Environment labels reflect the ranking (based on mean FW) by the prefix number, plus a short reference to the stress/treatment in that environment (details in Table S1).

							Explained genetic variance by WGP	
Environment	mean FW (log(g))	SD (log(g))	Number of replicates	V_g	V_e	H^2	Strong	Weak + Strong
1_Control	0.04	0.15	3	0.019	0.010	0.85	0.57	0.60
2_Control	-0.11	0.16	3	0.013	0.024	0.62	0.47	0.78
3_Control	-0.16	0.20	1	*	*	*	0.24	0.26
4_Pdef	-0.20	0.24	1	*	*	*	0.25	0.28
5_Control	-0.20	0.14	3	0.009	0.033	0.44	0.30	0.88
6_Control	-0.21	0.13	2	0.010	0.014	0.59	0.47	0.78
7_Control	-0.32	0.14	4	0.008	0.042	0.45	0.35	0.55
8_Control	-0.39	0.12	3	0.012	0.009	0.80	0.55	0.61
9_Control	-0.44	0.15	3	0.014	0.022	0.65	0.38	0.81
10_Control	-0.45	0.16	1[a]	*	*	*	0.36	0.86
11_Ndef	-0.49	0.13	3	0.012	0.015	0.70	0.51	0.71
12_Drought	-0.50	0.22	3	0.025	0.060	0.56	0.30	0.74
13_Salt	-0.52	0.18	3	0.014	0.050	0.45	0.33	0.39
14_Control	-0.59	0.21	2[a]	0.031	0.027	0.70	0.42	0.90
15_Salt	-0.61	0.24	3	0.021	0.076	0.46	0.29	0.58
16_Control	-0.65	0.20	3	0.030	0.026	0.78	0.52	0.52
17_Control	-0.65	0.25	3	0.026	0.072	0.52	0.21	0.92
18_Ndef	-0.66	0.10	3	0.008	0.008	0.74	0.54	0.49
19_Pdef	-0.74	0.17	4	0.006	0.082	0.24	0.27	0.93
20_Control	-0.76	0.14	6	0.101	0.164	0.79	0.33	0.95
21_Drought	-0.83	0.20	3	0.028	0.026	0.76	0.48	0.55
22_Pdef	-0.87	0.13	3	0.009	0.021	0.55	0.36	0.96
23_Drought	-0.99	0.21	6	0.035	0.051	0.80	0.26	0.72
24_Pdef	-1.24	0.18	2	0.014	0.035	0.44	0.28	0.73
25_PEG	-1.29	0.15	1[a]	*	*	*	0.29	0.52

[a] pool of 3 plants

To explore differential genotypic responses across environments, correlations between environments for rosette FW were calculated (Figure 1A). Most environments showed weak to moderate positive and significant correlations, but also weak, non-significant, correlations between environments were observed. The lack of strong correlations indicates differential adaptation of accessions to the environments, suggesting that genotype by environment (GxE) interactions play an important role in the determination of FW.

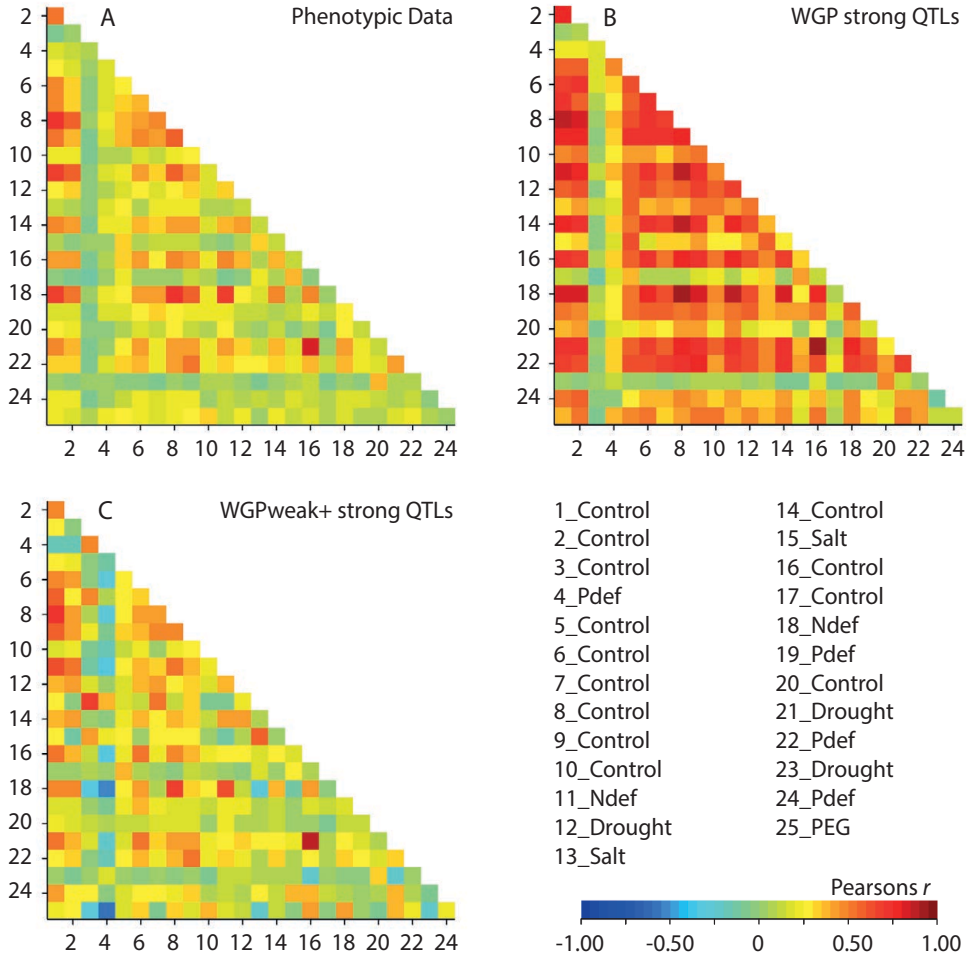


Figure 1 Pearson correlations for the 25 environments based on **A** phenotypic observations (BLUES of FW), **B** whole genome prediction (WPG) of the strongest 64 QTLs **C**. whole genome prediction with weak and strong QTLs. Environments are ordered on basis of the mean FW (Table 1). Colour corresponds with correlation according to the colour legend.

Observed correlations between environments did not necessarily agree with a priori expectations, for example control environments or environments with a similar stress like drought and osmotic stress were not always significantly correlated. On the other hand, significant correlations were found between environments that were expected to be different due to differences in growth conditions. For example, environment 2_Control and 8_Ndef are strongly correlated although the growth location (greenhouse vs climate room), the growth medium (sand vs rockwool), and some other growth condition parameters were different. This indicates the presence of an environmental factor crucial for biomass accumulation that was not monitored, but that was equal in the correlated environments. This is supported by the observation that, in most cases, experiments that

were grown in the same room and at the same time correlated positively, even when both experiments were grown under contrasting conditions (e.g. control versus stress); examples are 8_Control and 18_Ndef, 9_Control and 22_Pdef , and 16_Control and 21_Drought. FW data collected in the 25 different environments were also compared using an Additive Main Effects and Multiplicative Interaction (AMMI) model (Table S2) to estimate how much of the total variance could by explained by the genotypes (G), the environments (E) and the interaction between the two (GxE). The GxE interaction could explain more of the total variance than the genotypes (G) emphasizing the important role of GxE in FW determination. The AMMI model was also used to investigate the structure of the GxE interaction by determining interaction principal components (IPCA). Each of the first five IPCAs explained between 7% and 15% of the GxE (Table S2), suggesting that the GxE component has a complex structure that probably is mediated by many QTLs with QxE effects.

Multi-environment GWAS detects constitutive QTLs

In the search for genes that determine plant size in multiple environments, rosette biomass data of all the accessions grown in the 25 environments were used to perform multi-environment GWA mapping. The multi-environment GWA analysis resulted in the detection of 106 strongly associated SNPs (threshold –log(p) > 10, Figure 2A) located on different chromosomes. The arbitrary threshold of –log(p)=10 was set to focus on the SNPs with the largest explained variance and corresponding causal genes with large effects. The strongly associated SNPs were sometimes located in close proximity suggesting they are linked to the same causal SNP. Therefore 64 independent QTLs were selected corresponding to 64 SNPs that were located at least 10 kb from each other. Because QxE interactions were expected based on the analysis of the phenotypic data, the model used to perform the multi-environment GWA mapping allowed the marker effect to be different for each SNP in each environment. For all strongly associated SNPs large differences in marker effect were observed between environments (Figure 2B). The associated SNPs had a significant effect in on average 11 environments, ranging from 7 to 16 environments (Figure 2C), and therefore were called constitutive QTLs. Opposing marker effects were rare; for only five SNPs the direction of the effect was, in one or two environments, opposite from the direction in the majority of environments. For all other SNPs the same allele was associated with better performance in all environments. For each of the strongly associated SNPs the marker effect was split in a main and a QxE effect. For all these SNPs the main effect was larger than the QxE effect although significant QxE effects (-log(p) > 3) were observed for 20 of the 106 SNPs, representing 15 of the 64 independent QTLs.

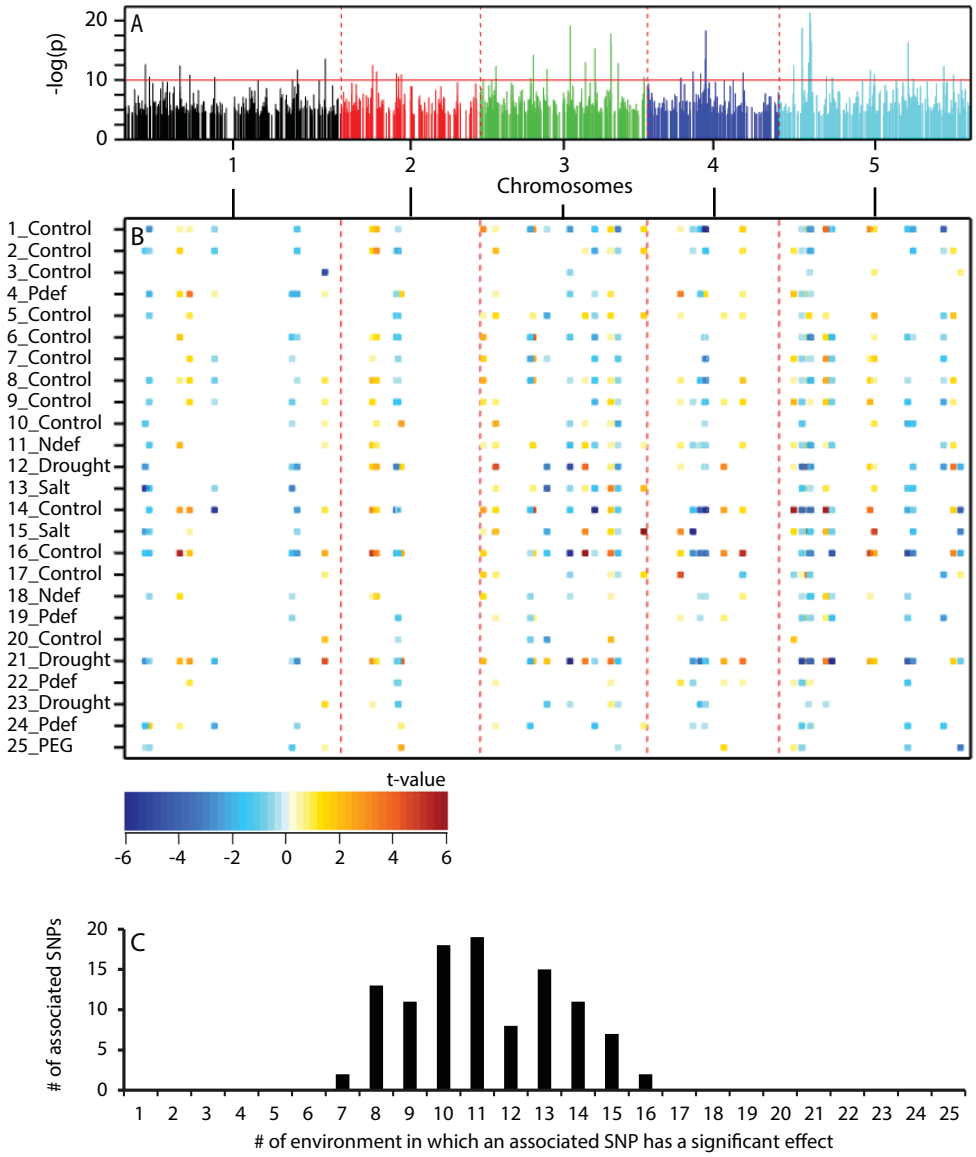


Figure 2 Multi-environment GWA analysis. **A** Manhattan plot of the -log(p) marker-multi-environment associations. Only the SNPs with -log(p)>5 are plotted. Each chromosome is represented by a different colour. Threshold of -log(p)=10 is indicated with a horizontal red line. **B** Heatmap of the effect (t-value = marker effect/standard error) of all 106 significant SNPs (-log(p)>10) in the 25 environments. Colors correspond with t-value according to the color legend. Environments are ordered on basis of the mean FW (Table 1). C Frequency distribution of the number of environments in which an associated SNP had a significant marker effect($|t| > 2$).

Whole genome predictions reveals importance of small effect QTLs in explanation of GxE interactions

Whole-genome prediction (WGP, (de los Campos 2013)) was performed to evaluate how much of the genetic variation estimated from the observation at the phenotypic level could be explained by the effects of the strong QTLs. WGP results in a predicted value for each accessions in each environment. WGP with the strong QTLs could explain between 21% and 57% of the total genetic variance per environment (Table 1). Moderate to very strong correlations were detected between predicted phenotypes in almost all environments (Figure 1B). These correlations correspond with the observation that the main effect was prominent for all strong QTLs. Only the environments 3_Control, 4_Pdef, 17_Control and 23_Drought were not strongly correlated to any other environment. In those environments the genetic variance that could be explained by the strong QTLs was also low (Table 1, below 30%) suggesting that different QTLs not detected in our analysis determine large part of the biomass accumulation in those environments. Comparison of the correlations between environments based on the predicted values obtained by WGP with strong QTLs (Figure 1B) and based on the phenotypic data (Figure 1A) reveals that most of the GxE observed at the phenotypic level cannot be explained by QxE effects of the strong associations. WGP was also performed with all SNPs (weak and strong) that had a significant effect in at least 1 environment. WGP with weak and strong QTLs could explain between 26% and 96% of the genetic variance per environment (Table 1). The high explained genetic variance observed for most environments shows that the multi-environment GWAS approach is able to detect most of the G and GxE effects present in this data set. This corresponds with the similarities that were observed between the correlation pattern of phenotypes predicted with weak and strong QTLs (Figure 1C) and the correlation pattern based on the phenotypic data (Figure 1A). The low explained genetic variance observed for some environments is due to non-additive effects that were not considered in this approach, like epistasis. Also low frequency markers (<5%) were not included in this approach.

Results were obtained on three levels: the phenotypic level, the level of the WGP with strong QTLs and the level of the WGP with the combination of weak and strong QTLs. Comparison of the three levels results in interesting conclusions about the importance of GxE in biomass accumulation. From the observations at the phenotypic level is was concluded that GxE interactions are important (Figure 1A) and that the GxE structure is complex (Table S2). Surprisingly, many strong QTLs detected by multi-environment GWA mapping did not have significant QxE effects. WGP confirmed that most of GxE interactions could not be explained by the QxE effects of the strong QTLs and that the QxE effects of the weak QTLs contributed more to the explanation of the GxE interactions. Combining the observations at the three levels, we conclude that large effect QTLs with large QxE effects were not observed. The observed GxE is, therefore, probably, a consequence of many moderate and weak effect QTLs with small QxE effects.

Table 2 Candidate genes located in close proximity (<32 kb) of SNPs detected by multi-environment GWAS. Candidate genes are observed in four categories: Flowering time related, starch metabolism, leaf size control and fitness trade-off.

Gene	Abbreviation	Name	
Flowering time			
AT3G02380	COL2	CONSTANS-like 2	(Ledger 2001)
AT3G25730	EDF3	ETHYLENE RESPONSE DNA BINDING FACTOR 3	(Matías-Hernández 2014)
AT3G44460	DPBF2	Basic-leucine zipper (bZIP) transcription factor family protein	(Bensmihen 2005)
AT3G62090	PIL2;PIF6	PHYTOCHROME-INTERACTING FACTOR 6,	(Yamashino 2003)
AT4G14110	COP9	CONSTITUTIVE PHOTOMORPHOGENIC 9	(Wang 2003)
AT5G10140	FLC	FLOWERING LOCUS C	(Michaels 1999)
AT5G13480	FY	mRNA processing factor	(Simpson 2003)
AT5G13790	AGL15	AGAMOUS-like 15	(Fernandez 2000)
AT5G46210	CUL4	cullin4	(Chen 2006)
Leaf size control			
AT5G13060	ABAP1	ARMADILLO BTB PROTEIN 1	Masuda 2008
Starch metabolism			
AT1G10760	SEX1	STARCH EXCESS 1	(Yu 2001)
AT3G52180	SEX4	STARCH EXCESS 4	(Kötting 2009)
Fitness trade-off			
AT3G07040	RPM1	RESISTANCE TO P. SYRINGAE PV MACULICOLA 1	(Tian 2003)
AT4G14400	ACD6	ACCUMULATED CELL DEATH 6	(Todesco 2010)
AT5G19473	RIN family	RPM1-interacting protein 4 family protein	(Afzal 2011)

Candidate genes related to flowering time, starch metabolism, leaf size and fitness trade-off

Growth is a complex trait that results from several biological processes, and regulation may occur at different levels. Therefore genes causal for the observed QTLs will probably be part of various regulatory networks. Life history, the timing of the developmental processes within the life cycle, partially determines the size of a plant (Mitchell-Olds 1996, Melzer 2008). Flowering time, for example, is one of the most prominent differences in the life history of Arabidopsis accessions with a large effect on biomass. When flowering starts, the investment in the growth of the rosette will decrease and resources will be allocated to the inflorescence, the flowers, and the seeds. Therefore, many studies observed that changes in flowering time result in changes in biomass of the rosette (Alonso-Blanco 1999, Salehi 2005, Wu 2006). Nine genes, reported to be involved in flowering time regulation (Chen 2012), were observed in close proximity of the 106 SNPs significantly associated with growth (Table 2). Among these genes is *FLOWERING LOCUS C* (*FLC*,

AT5G10140) (Michaels 1999) and one of its repressors (*FY*, AT5G13480)(Simpson 2003). Also starch has a central position in the regulation of plant growth (Sulpice 2009). Developmental progress is often changed in mutants impaired in starch metabolism, suggesting that natural variation in those genes could also result in changes in plant size (Streb 2012). Two genes involved in starch metabolism (*STARCH EXCESS 1, SEX1*, AT1G10760, (Yu 2001), *STARCH EXCESS 4, SEX4*, AT3G52180, (Kötting 2009)) were observed in close proximity of the associated SNPs (Table 2). Both genes are involved in starch degradation and mutation results in increased starch accumulation (Yu 2001, Kötting 2009). Finally, changes in leaf size, as consequence of changes in cell number and/or cell size, will in many cases result in a change of total vegetative biomass (Gonzalez 2012). In search for biomass-enhancing genes, (Gonzalez 2009) listed all leaf size control genes that are reported to lead to vegetative biomass, when over-expressed or mutated. One of these biomass enhancing genes, *ARMADILLO BTB PROTEIN 1* (*ABAP1*, AT5G13060) was located in close proximity (<20Kb) of one of the associated SNPs (Table 2). *ABAP1* participates in a negative feedback loop regulating mitotic DNA replication during leaf development and mutation of *ABAP1* results in increased cell division rates (Masuda 2008).

Genes representing biological trade-offs, with some alleles being favourable in some environments but unfavourable in other environments, are prime candidates to underlie QTLs with significant QxE effects. The SNP with the largest QxE effect (-log(p) = 5.5) was found within 20Kb from *ACCELERATED CELL DEATH 6* (*ACD6*, AT4G14400). Large differences in marker effect between environments were observed for this SNP. A very strong negative effect of the Col-0 allele was observed in environment 1_Control (t-value = -6.16), whereas a non-significant effect (t-value = 0.99) was observed for 17_Control. Some natural alleles of this *ACD6* are known to cause early senescence resulting in a reduction in biomass. The same alleles improve the resistance to *Pseudomonas syringae* by higher accumulation of SA (salicylic acid) suggesting a fitness trade-off between the environments (Todesco 2010). Another gene in close proximity of a SNPs with significant QxE effect (-log(p)= 3.3) is *disease resistance protein RPM1* (AT3G07040). Natural variation in resistance against *P. syringae* could be explained by the absence or presence of expression of this gene and was independent of expression changes in the salicylic acid pathway (Tian 2003). Plants that expressed *RPM1* had a lower biomass, again suggesting a fitness trade-off. In addition, a protein reported to interact with *RPM1* (*RIN4* family, AT5G19473 , (Afzal 2011)) was observed within 20Kb of another SNP with significant QxE effect (-log(p) = 3.2).These results suggest that alleles of *ACD6* and *RPM1* influence growth positively or negatively depending on the abiotic growth conditions.

The candidate genes discussed above correspond to 15 of the 64 independent QTLs. The fact that only a fourth of the QTLs associated with growth could be associated with genes with a known function related to growth, is an indication that growth and development

of the rosette of *Arabidopsis* is not yet fully understood at the molecular level. Further exploration of the QTLs without obvious candidates would be of great interest, because that could result in the identification of new genes regulating plant growth without current functional annotation or could identify new functions (in growth) for genes currently known to be involved in different biological processes.

Discussion

This study shows that, although biomass is influenced by almost all environmental conditions, we were able to identify QTLs that determine vegetative biomass in a range of different environments. Interestingly, for most QTLs the same allele was associated with better performance in all environments and only a limited amount of SNPs with significant QxE were identified. Multi-environment GWA mapping in combinations with WGP revealed that the strongest associated SNPs are mostly main effect QTLs, while the weaker associations explain most of the GxE.

Comparison with earlier biomass accumulation data

In addition to data included in the present analysis (El-Soda 2014a, Bac-Molenaar 2015), natural variation for biomass accumulation under non-stress conditions was reported earlier for 94 Arabidopsis accessions (Sulpice 2009) of which 40 were also included in the present analysis. As far as we know, only information about small numbers of accessions is publicly available for biomass accumulation upon several abiotic stresses. For example, the impact of salt stress on biomass accumulation was reported for 4 accessions (Jha 2010) and the impact of nitrogen deficiency on biomass accumulation was reported for 19 accessions (Chardon 2010). Furthermore, biomass accumulation upon mild drought was studied in respectively 24 and 17 accessions (Bouchabke 2008, Des Marais 2012). Many accessions used in these studies were also included in our analysis. Sulpice *et al*. (2009) reported rosette FW data ranging from 0.1 to 0.4 g per rosette after 5 weeks growth in short-days corresponding with plant sizes in the environments with lower average FW. Plant sizes upon stress differed two- to threefold between accessions within the experiments, but up to 10 fold between the same accessions grown in independent experiments (Des Marais *et al.* (2012); Jha *et al.* (2010)). These differences can only partially be explained by age differences at harvest. Plant sizes in the environments with lowest average FW were similar to those reported by Des Marais *et al.* (2012), whereas plant size data in the environments with high average FW were similar to data described by Jha *et al.* (2010) (Table 1). This indicates that the wide range of plant sizes observed in our experiments is reflected in other independent experiments, emphasizing that our results can be used for comparisons with past and future studies about natural variation.

Confirmation of the interplay between flowering time and biomass accumulation

Genetic architecture of biomass accumulation has been studied in several RIL populations: Columbia-4 (Col-4) x Landsberg *erecta* (L*er* (Quesada 2002), Cape Verde Islands (Cvi) x L*er* (Alonso-Blanco 1999, Harada 2006), Columbia-0 (Col-0) x Coimbra-24 (C24) (Lisec 2008), L*er* x Antwerpen-1 (An-1) (Tisné 2010, Schmalenbach 2014), and Shahdara (Sha) x Bayreuth-0 (Bay-0) (Loudet 2003, El-Lithy 2004). Depending on the parents of the population, different QTLs were identified, which indicates that only a limited amount of the natural variation present in Arabidopsis is tested in each RIL population. Therefore multiple RIL populations should be used to assess a substantial part of the genetic variation present in the species. In our GWA analysis a large part of the natural variation present in the species Arabidopsis was tested at once, but the limitation of this analysis is that rare alleles will be missed. The *ERECTA* locus (AT2G26330) is a good illustration of this limitation. Many QTLs at the *ERECTA* locus were found for plant size in RIL populations in which L*er* was one of the parents, and evidence was found for a role of the *ERECTA* locus in leaf shape (Rédei 1962) and biomass accumulation (Koornneef 2004). Multi-environment GWA mapping did not identify any associated SNPs in close proximity of *ERECTA*, which was expected in view of the fact that this mutation is artificially introduced in the accessions Landsberg and therefore solely present in this accession (Rédei 1962). For some RIL populations co-location was observed of QTLs for flowering time related traits and biomass accumulation, giving strong evidence for a role of *FLC* (AT5G10140) and other flowering time genes in the regulation of biomass accumulation (Demura 2010, Tisné 2010, Schmalenbach 2014). In the multi-environment GWA analysis, nine flowering time related genes were observed in close proximity of an associated SNP. Four of them are located on chromosome 5, the chromosome on which most biomass QTLs were identified in the RIL populations. One of these four genes is *FLC* of which the natural variation in sequence has been described in detail (Guo 2011, Li 2014) and of which the ectopic expression in tobacco leads to plants with higher biomass (Salehi 2005).

Evidence for a role of *ABAP1*, *SEX1* and *SEX4* in biomass accumulation

Leaf size control and starch metabolism are two processes related to biomass accumulation. Leaf size is determined by the number of cells per leaf and the size of those cells (Gonzalez 2012). Increase of cell number without decrease of cell size and *vice versa* results in larger leaves. Plants with larger leaves do not always have higher biomass, because also the leaf initiation rate is of importance. A mutation in *ABAP1*, a gene involved in the regulation of cell division rate, resulted in larger plants (Gonzalez 2009). *ABAP1* was located in close proximity of one of the associated SNPs (Table 2). This gene is mentioned as one of the promising genes to be used to develop genetically modified (GM) crops with improved biomass production (Rojas, 2010). Its constitutive effect detected in the present study makes it even more attractive. Orthologs in crops with similar function have not been reported yet, but if they will be identified and natural variation in the sequence

influences the growth phenotype, this gene could be used in classical breeding programs. In a comparative study of biomass accumulation and the content of several metabolites in 94 Arabidopsis accessions, starch was identified as a major integrator of growth regulation (Sulpice 2009). A strong negative correlation was observed between biomass and starch content of the rosette at the end of the day under different day-night regimes, indicating a tight connection between growth and carbon supply. Mutant studies revealed that plants with low starch content, or plants that were unable to degrade their starch during the night, showed retarded growth and later flowering (Streb 2012), indicating that both starch synthesis and degradation are crucial for proper starch turnover and plant growth. How biomass accumulation is exactly coupled to starch availability is still largely unknown. In our study two genes involved in starch degradation were located in close proximity of the strong QTLS. A knockout mutant of candidate gene *SEX1*, a water dikinase, in Arabidopsis showed reduced growth because of improper starch degradation at night, whereas seed-specific down-regulation of the orthologous gene in wheat led to an increase in vegetative biomass and grain yield (Ral 2012). This indicates that the balance between starch synthesis and degradation is different in source and sink tissues, but in both cases related to biomass accumulation. Absence of the second candidate gene involved in starch degradation, *SEX4*, in Arabidopsis led to dramatic accumulation of starch at the end of the night and a decrease in growth (Zeeman 1999). At*SEX4* and its orthologs in rice and barley were stably expressed during several stresses and their diurnal expression profiles were similar during plant development (Ma 2014). Partial knock-down of *SEX4* expression in Arabidopsis and maize did not lead to hampered plant growth, but led to improved sugar release from plant biomass, an interesting trait for the bio-energy industry (Weise 2012, Stamatiou 2013).

QxE effects are more prominent for weak QTLs than for strong QTLs

Large differences in the average performance of the accessions was observed between environments, even when the conditions were supposed to be optimal (classified as control) (Table 1). In addition, in some cases no correlation was observed between the FW data collected in different environments (Figure 1A), indicating differential adaptation. These observations in combination with the ANOVA results (Table S2) suggest an important role for GxE interactions in the determination of biomass accumulation. An earlier study compared the growth of three natural accessions in ten different laboratories (Massonnet 2010) and indicated that, although the growth conditions were standardized among the labs and the same seed batch was used, large differences were observed between laboratories in both the overall and the genotype-specific performance of the three accessions. E.g, for accession Col a six fold difference in Projected Rosette Area was observed between the two most different laboratories. In addition, in some laboratories Col, L*er,* and Ws had similar sizes, whereas in another laboratory 5 fold size differences were observed between the accessions. This experiment shows that it is necessary to grow the same genotypes in multiple locations in order to be able to determine which part of the observed variation

is explained by the genotype (G), the environment (E), or the interaction between the two (GxE). The quality of light was identified as an important parameter affecting growth (Massonnet 2010). In the 25 environments tested in our experiment the light intensity was 125 or 200 μmol/m²/s in all climate rooms, but different types of lamps were used in different rooms, which can probably explain the large differences observed between the performance in environments with nearly identical conditions. This may also explain why some experiments that were grown in the same climate room at the same time, but with contrasting treatments, were strongly correlated. An important role for GxE in biomass accumulation was also suggested in a study in which a DH population descending from two rice varieties that were grown in nine locations throughout Asia (Hittalmani 2003). Univariate mapping identified several QTLs for plant biomass, but none of them was detected at more than two locations, indicating a large impact of the environment. In that study a multi-environment mapping approach was not used, but would be the proper method to identify constitutive QTLs with minor influence of GxE in multiple environments. Although GxE interaction seems to play an important role in biomass determination in the 25 environments tested in the present study, this was not reflected in the QxE effects determined for the 64 strongest associated regions. Multi-environment GWA mapping of flowering time and germination traits identified hardly any or no SNPs with QxE effect, whereas GxE effects were identified (Korte 2012, Morrison 2014). It was, therefore, suggested that multi-environment mapping lacks power to detect QxE effect (Korte 2012, Morrison 2014). In our experiment for 24% of the strongly associated regions, significant QxE was detected indicating that multi-environment GWA mapping has the power to identify at least a part of the QxE effects. WPG based on strong and weak QTLs even suggests that the QxE was identified properly, but that most of the GxE observed is the consequence of many moderate or small effect SNPs that were not detected because of our strict threshold. For those weaker QTLs QxE is probably in many cases more prominent than for the 64 strong QTLs detected above our threshold. In a multi-environment GWA analysis of 7 agricultural traits of barley determined in 17 to 28 environments the main effects explained more of the genetic variance than the QxE effects for all investigated traits (Zhao 2012). Possibly, some of the GxE interactions in our study can also be explained by markers below the threshold.

Constitutive QTLs promising for utilization in crops

In our study, multi-environment GWA mapping resulted in the detection of strong constitutive QTLs with significant effect in at least 6 independent environments. Those strong QTLs were independent of seed batch, and of place and time of data acquisition. Constitutive QTLs with no or small GxE effects in many environments are promising candidates for the translation of findings from lab to field, because they can be utilized in a broad range of conditions, whereas environment specific QTLs can only be used in the tested environment (Zhao 2012). It is reported that multi-environment-multi-trait mapping approaches are even more effective to identify small effect constitutive QTLs

than multi-environment mapping of one trait (Alimi 2013). Biomass is a complex trait that is flexible and adaptive upon perturbations in the environment. Therefore one of the promising ways to detect more small effect constitutive QTLs, is to perform multi-environment-multi-trait GWA on traits which contribute to vegetative aboveground biomass accumulation, like leaf initiation rate, number of cells per leaves, size of leaf cells, number of leafs, length of vegetative growth period, and leaf thickness.

Conclusion

Our study shows that multi-environment GWA mapping is an effective method to identify constitutive QTLs with effects in diverse environments. Several candidate genes related to flowering time, starch metabolism, and leaf size control were located in close proximity of constitutive QTLs. In addition, *ACD6*, *RPM1*, and an *RPM1*-interaction family gene were located in proximity of strong QTLs with significant QxE effects. Some alleles of *ACD6* and *RPM1* have been reported to cause resistance to bacterial infection as well as reduction in biomass, suggesting a fitness trade-off between alleles in some environments. Last, most of the GxE observed at the phenotypic level is not the consequence of strong QTLs with strong QxE effects, but of moderate and weak QTLs with weak QxE effects.

Materials and Methods

Plant materials and experimental design
A collection of 350 natural accessions of Arabidopsis thaliana was used. These accessions were selected to capture most of the genetic variation present within the species (Baxter 2010, Li 2010, Platt 2010). Each accession was genotyped with approximately 215k SNP markers (Col vs non-Col) (Kim 2007, Atwell 2010). This collection was grown in 25 different conditions (environments) defined by differences in location (climate room or greenhouse), substrate, daylight duration, intensity of light, relative humidity, and day and night temperature. In addition, in some environments the plants were exposed to abiotic stresses, such as drought, N or P deficiency and osmotic stress (PEG8000 or NaCl) (details in Table S1). Plants were harvested just before flowering (28 to 37 days after sowing) and rosette fresh weight was determined (g plant^{-1}). Some of the data analysed were published before (El-Soda 2014a, Bac-Molenaar 2015).

Statistical analysis
Firstly, the data collected in each of the 25 environments were analysed individually to estimate genetic variance components and heritabilities, In addition, genotypic means per environment were obtained. The resulting genotypic means were compiled in a genotype by environment table that was used in the further analyses: multi-environment GWA mapping and whole-genome prediction (WGP).

$$y_{ik} = \mu + rep_k + acc_i + e_{ik} \qquad (1)$$

The model for the individual trial analysis was:
where replicates (rep_k) and accessions (acc_i) were taken as random effects. The response variable fresh weight (y_{ik}) was transformed to a log10 scale after initial inspection of the residuals to stabilize the variances. The genetic and error variance componentsper environment were used to estimate broad

sense heritability on a mean line basis, $H^2 = \frac{\sigma^2_{acc}}{\sigma^2_{acc} + \frac{\sigma^2}{r}}$ where r is the number of replications per accession in the particular environment (depending on the experiment r ranged from 2 to 6). Some of the experiments did not have replications, in which case, H^2 was not estimated. To obtain genotypic means, the basic model was modified to include accessions as fixed effects (instead of random terms) to obtain best linear unbiased estimates (BLUE) per accession, $\hat{\mu}_i$. By applying model (1) to each of the 25 environments (j = 1 ... 25), a genotype by environment table of means $\hat{\mu}_{ij}$ (BLUE) was compiled and used as the starting point for the rest of the analyses. The BLUES for 350 genotypes in five (5) environments were subjected to Additive Main Effects and Multiplicative Interaction (AMMI) analysis of variance and subsequently interaction principal components (IPCA) were determined (Gauch Jr 1992). In addition correlations between BLUES were calculated using Pearson's correlations.

For the GWA the following multi-environment mixed model was used:
where is the mean of accession i in environment j, μ is a general intercept, envj is the fixed environmental

$$\hat{\mu}_{ij} = \mu + env_j + x_i\alpha_j + g_{ij} + e_{ij} \qquad (2)$$

main effect, xi is an indicator variable with the number of non-Columbia alleles at a particular SNP, xi = {0,1,2}, αj are the environment-specific marker allele substitution effects (i.e. the effect of replacing each copy of the Columbia allele at the particular SNP), gij is the residual polygenic effect per environment that follows a multivariate normal distribution MVN ~(0,Σ), with Σ a variance-covariance matrix that combines kinship relationships between the accessions (estimated from SNP markers) and the genetic correlations between the environments, and eij a residual consisting of non-additive genetic effects and experimental error. The model used here is similar to the one described in (El-Soda 2014a), but the dimension of Σ was much larger here because of the large number of genotypes and environments involved. Therefore, to alleviate the computational burden, the covariance matrix Σ was approximated by: 1) the grouping of the accessions and estimating an unstructured covariance matrix among the groups instead of among all accession, so reducing the dimensions from 350x350 to 10x10 (since ten groups were formed), and 2) by estimating the genetic covariances between environments with the more parsimonious factor analytic model instead of the unstructured model. Significant marker-trait associations were declared whenever the null hypothesis H0: αj = 0 was rejected (i.e. a SNP has a significant effect in at least one of the 25 environments). For hypothesis testing a stringent threshold of –log(p) >10 was used to allow focusing on only major SNP effects. To avoid overestimation of explained variance as a consequence of multiple SNPs which are linked to the same causal gene, significant SNPs located within 10 Kb from each other were thinned selecting only the strongest associated SNP in the given region. For each individual SNP, an explicit test for marker by environment interaction was performed, by partitioning the environment-specific SNP effect αj into α + α*j, where α and α*j are the SNP main effect and the interaction respectively. Testing for interaction implies testing the null hypothesis H0 : α*j = 0. Significance of the SNP effects per environment were individually assessed by a t-test within the environment, that is, $t = \frac{\alpha_j}{se_{\alpha_j}}$, with $|t| > 2$ pointing to a significant effect. After the GWA

detection step, the effects of the selected significant SNPs were used to obtain predicted values per environment, $\hat{u}_{ij}^{SNP}$. The explained variance per environment was assessed by the square correlation coefficient between observed means per environments ($\hat{u}_{ij}$) and the predicted value from the selected SNPs ($\hat{u}_{ij}^{SNP}$). Because the GWA model will only contain strong effect SNPs, a GWP model that allows minor effect markers was used (for a review of this type of models see (de los Campos 2013). The model is similar to the GWA model except that not one but all markers are simultaneously included in the model. So, for each environment the following model was used:

$$\hat{\mu}_i = \mu + \sum_M x_{im}\,\alpha_m + e_i, \qquad (3)$$

where the term $\Sigma_M x_{im}\alpha_m$ stands for the contribution of all M SNP markers (note that the polygenic effect

is not in the model anymore since all the additive genetic signal is expected to be absorbed by the marker effects). Predicted values from the model were obtained $\hat{u}^{WGR}$ and were correlated with the means in the particular environment ($\hat{u}_{ij'}$). The square of the correlation coefficient gives an upper bound of the explained variance by the SNP in a particular environment (i.e. collecting all, the strong and weak SNP effects). Unexplained variance, contained in the residual term e_i, are caused by genetic effects not captured by the additive model (non-additive effects such as complex interactions) and experimental error. Environments were correlated based on the phenotypes predicted by WGP with strong effect SNPs as well as by WGP with the combination of weak and strong effect SNPs using Pearson's correlations.using Pearson's correlations.

Table S1 Growth conditions and the experimental design of the 25 different environments in which the same natural population (HapMap population) was grown. Environments are ranked based on the mean FW. Environment labels reflect the ranking by the prefix number, plus a short reference to the stress/treatment in that environment.

Environment	Stress applied	Room	Substrate	# hours light per 24h	Light intensity (umol/m^2/s)	Relative humidity (%)	Temperature (C°) (Days/Night)	Harvest (days)	# replica's	# Accessions	Design
1_Control	-	Climate room	rockwool	10	200	60	20/18	35	3	360	Complete random split block
2_Control	-	Greenhouse	silver sand + toplayer of soil	10	Daylight + lamps (Oct-Nov)	69%/74%	21.3/17.5	34	3	360	Complete random split block
3_Control	-	Climate room	rockwool	12	200	60	20/19	28	1	360	Split block
4_Pdef	P_deficiency (0.1 mM)	Climate room	rockwool	12	200	60	20/19	28	1	360	Split block
5_Control	-	Climate room	rockwool	16	125	70	20/18	28	3	360	Complete random split block
6_Control	-	Climate room	rockwool	10	200	60	20/19	32	2	360	Complete random split block
7_Control	-	Climate room	rockwool	10	200	60	20/19	32	4	360	Complete random block
8_Control	-	Climate room	rockwool	10	200	60	22/20	28	3	360	Complete random split block
9_Control	-	Climate room	rockwool	10	200	60	20/19	32	3	360	Complete random split block
10_Control	-	Climate room	rockwool	16	125	70	20/18	28	1[a]	360	Complete random split block
11_Ndef	N_deficiency (1 mM nitraat/ammonium (9:1))	Climate room	rockwool	10	200	60	20/18	35	3	360	Complete random split block
12_Drought	Drought (day 12 until 34 no watering)	Greenhouse	silver sand + toplayer of soil	10	Daylight + lamps (Oct-Nov)	69%/74%	21.3/17.5	34	3	360	Complete random split block
13_Salt	Salt (25mM NaCl)	Climate room	rockwool	16	125	70	20/18	28	3	360	Complete random split block
14_Control	-	Climate room	rockwool	10	125	70	20/18	37	2[a]	360	Complete random
15_Salt	Salt (25mM NaCl)	Climate room	rockwool	16	125	70	20/18	28	3	360	Complete random split block
16_Control	-	Climate room	soil	16	125	70	20/18	28	3	324	4 rounds of 80 accession, 4 accessions as reference, within each round 3 complete blocks
17_Control	-	Climate room	rockwool	16	125	70	20/18	28	3	360	Complete random split block
18_Ndef	N_deficiency (1 mM nitraat/ammonium (9:1))	Climate room	rockwool	10	200	60	22/20	28	3	360	Complete random split block
19_Pdef	P_deficiency (0.1 mM)	Climate room	rockwool	10	200	60	20/19	32	4	360	Complete random block
20_Control	-	Climate room	soil:sand 1:1	8	200	70	21	33	6	360	all treatments at the same time, rounds of 37 accessions, 3 accessions as reference
21_Drought	Drought (PHENOPSIS, start drought at day 8, water content 0.22 g H2O/g soil)	Climate room	soil	16	125	70	20/18	28	3	324	4 rounds of 80 accession, 4 accessions as reference, within each round 3 complete blocks
22_Pdef	P_deficiency (0.1 mM)	Climate room	rockwool	10	200	60	20/19	32	3	360	Complete random split block
23_Drought	Drought (no watering day 21-27)	Climate room	soil:sand 1:1	8	200	70	21	33	6	360	all treatments at the same time, rounds of 37 accessions, 3 accessions as reference
24_Pdef	P_deficiency (0.1 mM)	Climate room	rockwool	10	200	60	20/19	32	2	360	Complete random split block
25_PEG	Osmotic stress (10% of PEG8000 from day 8	Climate room	rockwool	16	125	70	20/18	28	1[a]	360	Complete random split block

Table S2 ANOVA Table for AMMI Analysis and determination of the interaction principal components (IPCA)

AMMI Analysis

ANOVA table for AMMI model

Source	d.f.	s.s.		m.s.	v.r.	F pr
Genotypes	349	55.7		0.159	6.27	<0.001
Environments	24	935.1		38.964	1531.29	<0.001
Interactions	8376	213.1		0.025		
IPCA 1	372	31.6	(15%)	0.085	4.97	<0.001
IPCA 2	370	23.1	(11%)	0.062	3.65	<0.001
IPCA 3	368	18.0	(8%)	0.049	2.86	<0.001
IPCA 4	366	14.6	(7%)	0.040	2.34	<0.001
IPCA 5	364	14.2	(7%)	0.039	2.29	<0.001
Residuals	6536	111.6	(52%)	0.017		

Chapter 7 - General Discussion

Johanna A. Bac-Molenaar

Introduction

Wherever on earth a plant is growing, its environmental conditions are not constant. Conditions change diurnally and seasonally. And, although it follows diurnal and seasonal patterns, the weather on two sequential days or years is never exactly the same. Plants, as sessile organisms, have the ability to sense the unstable growing conditions and respond to them by adapting their growth, development and metabolism. This process of adaptation is largely genetically regulated. Because not all individuals within a species are genetically identical also their ability and strategy to respond to the environment differs. Natural variation in plant performance is used in mapping approaches that couple genetic variation to phenotypic variation and hence tries to elucidate the genes that are involved in the regulation of plant performance. This information is valuable for breeders who aim to develop new cultivars that have stable yields also in years with unpredictable periods of drought, heat or other stresses. The development of new abiotic stress tolerant and stable yielding cultivars is of vital importance to deal with the expected human population growth and climate changes. In this thesis I studied the genetic regulation of plant performance in several environments. I focussed on two abiotic stresses, drought and heat. A genome wide association (GWA) mapping approach was chosen that uses the variation present in nature with the aim to identify and study genes that explain natural variation in plant performance under various diverse environmental conditions.

When I started the project in 2011, GWA mapping was already common practice in the field of human genetics, but was relatively new in plant science and only a very limited number of studies had been published at that time (Atwell 2010, Brachi 2010, Chan 2010, Cockram 2010). It was expected that GWA mapping would greatly enhance the knowledge of the biological function of genes. This thesis shows that GWA mapping can indeed be very successful in the detection of associations between different plant performance traits and genomic regions in which expected and unexpected regulatory genes are located.

In my General Discussion I will discuss the main findings of this thesis and their commonalities and will compare my results with the research of others in the field of quantitative genetics and plant phenotyping. First I will discuss the many strong associations detected by GWA mapping. Thereafter I will discuss the choices that I made related to plant phenotyping and experimental design that increased the impact of my findings. Further, I will discuss the candidate genes, their validation and the information they give about stress response mechanisms and about their ecological role. Finally, I will discuss the GWA mapping method itself and the impact of this research outside the Arabidopsis research field.

Discussion

GWA mapping detects many strong plant performance QTLs

Genes involved in the regulation of plant performance can be active in one specific environment or in various related or unrelated environments. Studies of co-expression and co-regulation of genes are performed to detect common and environment-specific regulators (Rizhsky 2004, Nakashima 2014, Polanski 2014). Especially the role of transcription factors in various stress responses has been studied in detail (Chen 2012, Mizoi 2012, Nakashima 2012). Most of these studies in Arabidopsis exclusively use the reference accession Col-0. But differences are observed at the sequence, gene expression, metabolite and phenotype level between Col-0 and other natural accessions of Arabidopsis. In this thesis the variation between various natural accessions has a central place. This natural variation is used to identify sequence variation[1] underlying variation in plant performance. In Chapters 2 to 5 of this thesis the identification of environment-specific plant performance QTLs was the subject whereas in Chapter 6 the focus was on the identification of growth regulators that are important over a range of environmental conditions. Strong associations were observed for both environment-specific as well as common plant performance regulators. GWA mapping of anthocyanin accumulation upon osmotic stress resulted in the detection of one very strong association (-log(p)>20) which could explain more than 15% of the phenotypic variance (Chapter 2). In addition, GWA mapping revealed seven QTLs strongly associated with silique length under control conditions and upon short-term heat stress (Chapter 3) each explaining between 5 and 12% of the phenotypic variation. Furthermore, I found 22 QTLs for growth related traits under control conditions (Chapter 4) and six QTLs for growth under moderate drought (Chapter 5). Each of these 28 QTLs is explaining more than 5% of the phenotypic variance of the corresponding trait. No overlap in QTLs was observed between the different environments except for one association with the region in which the flowering gene *FLC* (*FLOWERING LOCUS C*) is located. In Chapter 6, multi-environment GWA mapping resulted in the detection of more than 100 SNPs associated with growth in 7 to 16 environments. For most QTLs the sign of the effect was the same in all 25 environments included in the study, but for a limited number of QTLs opposing effects were detected. Interestingly also in this multi-environment study *FLC* was detected as common growth regulator. Therefore the role of this gene will be discussed in detail below. The identification of so many strong associations, each explaining more than 5% of the phenotypic variance, is remarkable. GWA mapping studies of other complex traits in Arabidopsis, like flowering time under field conditions (Brachi 2010), photosynthetic adaptation to high light (van Rooijen 2015), plant performance upon severe drought (El-Soda 2014) or low water potential-induced proline accumulation (Verslues 2014), resulted in the detection

1 Each time I speak about sequence variation in a gene that results in changes at the phenotype level, I mean sequence variation in the coding sequence, introns, promotors or other cis-regulatory elements that result in changes at the phenotype level.

of no or a very limited number of strong associations. Atwell *et al* (2010) reported GWA mapping of 107 traits and the detection of strong associations for several of these traits. But the traits with strong associations were in most cases expected to be regulated by a limited number of genes (i.e. qualitative traits), such as defence against bacterial infection, which is expected to be activated through the recognition of avirulence proteins by one or a few plant disease resistance proteins. This thesis shows that also for traits that are expected to be regulated by many genes (i.e. quantitative traits), GWA mapping can result in the detection of several strong associations that together explain a substantial part of the total variance. Successful identification of these strong associations was partly due to choices made in the experimental design and plant phenotyping as I will discuss below.

Important choices in experimental design and plant phenotyping

Plants continuously sense their environment and respond to it. To study the natural variation in plant performance in a particular environment, the trait under consideration needs to be quantified in a large number of genotypes, preferably with replicates. In earlier days, genotyping was the limiting factor but with the recent advances in sequencing technologies the identification of SNP markers at high density became easy, accurate and cheap (Jiménez-Gómez 2011). Consequently, plant phenotyping became the bottleneck in quantitative genetics studies (Houle 2010, Furbank 2011). In my work I made certain choices regarding phenotyping and experimental design that influenced the outcomes significantly.

The first issue regards the fact that plant growth is a dynamic trait, which is constantly influenced by changes in the environment. I, therefore, decided to investigate the natural variation in the drought response by measuring plant size on multiple days along the growth period by top-view imaging. The value of top-view imaging in the study of natural variation in plant responses to environments was already shown before (Aguirrezabal 2006, Zhang 2012). Phenotyping over time allows to compare the response of accessions to the environment during different stages of the development and allows to model growth. Parameters estimated by the growth model can be used to compare the growth of accessions based on the whole growth period instead of a single timepoint. Statistical reports indicated that the acquisition of plant performance data over time is valuable for the detection of QTLs (Ma 2004, Malosetti 2006), but only a limited number of publications describe research in which the investigation of dynamic traits over time is combined with mapping (Busemeyer 2013, Moore 2013, Wurschum 2014). GWA mapping of temporal plant size data and growth model parameters resulted in the detection of many QTLs which would not have been identified if plant size had only been determined at a single-time point (Chapters 4 and 5).

The second issue regards a related subject, namely, that the impact of stress on plant development is not equal in all developmental stages. Short-term heat stress has hardly

any influence on vegetative growth, but can impact the development of flowers and seeds significantly (Bita 2013). In addition, not all floral and embryo developmental stages are equally sensitive to short-term heat treatments (Hedhly 2011). To include the differences between the heat responses of the various developmental stages in our analysis, I choose to determine the length of the siliques along the whole inflorescence as a proxy of seed set. In this way I could distinguish the response of flowers that opened before, during and after the treatment. Two minima in silique length were observed, indicating the two most sensitive stages. The first minimum was observed for siliques that resulted from flowers that opened just before the treatment and, therefore, received the treatment after (post) anthesis, during early embryogenesis. The second minimum was observed for siliques that resulted from the flowers that opened some days after the heat stress and hence, the treatment was given during the developmental stage in which male and female meiosis takes place (pre-anthesis treatment). GWA mapping of the heat responses of the various developmental stages resulted in the detection of different QTLs for the pre- and post-anthesis heat responses, suggesting differences in the genetic regulation of the response in both phases. Focus on total yield would have been more logical from a breeders perspective because this is the final trait of interest. However, this focus would ignore differences between developmental time windows and, therefore, would never have led to these in-depth insights in the developmental stage specific regulation of the heat response. My approach, which takes developmental differences in sensitivity into account, showed to be of vital importance to reach my aim, which was, to elucidate and study genes involved in regulation of plant performance upon heat stress and to improve our understanding of the mechanisms involved.

The third issue that improved our results was the simplification of the quantification of the phenotype. The large number of individuals in a mapping population makes detailed studies of plant performance often unfeasible in practice because of time and manpower constraints. In Chapter 2 to 5, simplification of quantification was applied successfully. The most simplified method is described in Chapter 2 where anthocyanin accumulation in the leaves upon stress treatment was quantified by visual scoring of the rosette colour in 5 classes. This method turned out to be very effective because the classes strongly correlated with the concentration of anthocyanins in the leaves. GWA mapping of this simplified trait resulted in the detection of one very strong QTL, associated with a well-known transcription factor, involved in the regulation of anthocyanin accumulation, for which sequence variation underlying natural variation in anthocyanin accumulation was not described yet. In Chapter 3, the influence of heat stress on fertility was quantified by the length of the siliques along the inflorescence. This was a good measure to quantify seed set, because silique length strongly correlated with the number of seeds in the siliques. An additional advantage was that silique length could be determined also after drying of the inflorescence and, therefore, could be done on any moment after the harvest of the inflorescences. Unfortunately, this method did not allow to detect aborted seeds

but analysis of the number of seeds per silique would have been a precarious task because soon after seed ripening siliques will open and seeds will shatter. This fast phenotyping method allowed me to distinguish pre- and post-anthesis heat responses. In Chapters 4 and 5 we chose to quantify biomass accumulation in a non-destructive way using top-view imaging to determine the projected leaf area (PLA). PLA is just an estimate of biomass since leaves might overlap and leaf thickness and water content will influence the correlation between this trait and aboveground biomass. Nonetheless, strong correlations between PLA and total biomass were observed. Time-series measurements and subsequent addition of the time-dimension to GWA mapping, however, was crucial for the detection of time-dependent QTLs.

In this era in which phenotyping is the bottleneck for successful quantitative genetics, issues as addressed above should be considered before the experiment starts. Many plant biologist are interested in traits that are changing over time. In addition, not only the impact of heat is varying between developmental stages but probably the impact of all environmental stresses depends on the developmental stage (Chan 2011). In many cases high throughput phenotyping platforms that use imaging techniques to estimate plant performance parameters can be of great help because in these platforms simplified phenotypes can be acquired automatically and frequently for many plants. High throughput determination of aboveground (Granier 2006, Zhang 2012, Tisné 2013) and belowground (Normanly 2012, Moore 2013, Richard 2015) plant biomass and plant architecture estimates have been reported. In addition, high throughput analysis of photosynthesis, based on chlorophyll fluorescence, has been described (Harbinson 2012). In the future we can expect high throughput imaging with infrared cameras to determine leaf water content, estimated from leaf temperatures, and high throughput imaging of plant stress status based on hyperspectral camera's (Fahlgren 2015). All these imaging techniques will allow the acquisition of phenotyping data of many plants in a non-destructive way and therefore can be very useful for GWA mapping of dynamic traits.

Candidate genes suggest regulatory pathways and evolutionary trajectories

The ultimate goal of the project was not to identify QTLs but to identify genes involved in the regulation of plant performance. Therefore, genes located in close proximity of the strongly associated SNPs were listed as candidate genes. The size of the support region in which the causal gene was expected was determined based on the local LD. The candidates within these regions were prioritized based on prior knowledge, such as reported mutant phenotypes (Chapters 2 and 4), TAIR annotation (all Chapters), gene expression (Chapters 2, 3 and 5), and information about pathways reported to play a role in the regulation of the trait of interest (Chapters 4 and 6). In some cases additional experiments were performed to confirm the suggested role of the candidate or to determine the causal nucleotide. In each of the chapters of this thesis our search for the causal genes resulted in valuable information about the regulatory pathways involved in the stress responses and

about the ecological role of the causal genes, as I will discuss below.

GWA mapping of anthocyanin accumulation upon osmotic stress (Chapter 2) showed that this mapping method enables to hypothesize about the connection between natural selection on the phenotype level and sequence variation at the genotype level. GWA mapping of anthocyanin accumulation upon osmotic stress resulted in the detection of a strong association on chromosome 1 which was in linkage disequilibrium with three genes, located in tandem, that all were close relatives and all were annotated to play a role in anthocyanin accumulation. Expression analysis did not reveal a correlation between the expression of one of these genes and anthocyanin accumulation. Therefore variation in the coding sequences was investigated in detail. It was observed that the second most common allele of one of the three genes, *MYB90*, was associated with higher anthocyanin accumulation and therefore the nucleotide that distinguishes this allele from the other alleles of *MYB90* was identified as the quantitative trait nucleotide (QTN). The two most common alleles were both present in the population at relatively high frequency (>15%) suggesting balancing selection, assuming that, depending on the environmental conditions, the first or the second allele of *MYB90* provides fitness benefits. Interestingly, sequence variation in *MYB75*, another relative of the three genes in the QTL region and located elsewhere in the genome, was previously reported to be causal for variation in anthocyanin accumulation upon osmotic stress (Teng 2005), Ilk et al. 2015). In detail sequence analysis of this gene revealed that one allele is dominating the population and, therefore, is confounding the detection by GWA mapping.

GWA mapping of silique length upon short-term heat stress resulted in the detection of 4 QTLs (Chapter 3). For 2 QTLs strong evidence was gained that we identified the causal genes. The strong association on chromosome 5 was caused by sequence variation in *FLC*. Two isogenic lines containing either a functional or non-functional *FLC* allele showed a large difference in the heat response. The line with the functional *FLC* was very sensitive to heat, whereas the line with the non-functional allele was hardly affected. *FLC* is well-known for its role in flowering time determination (Michaels 1999), but was not associated earlier with heat stress. This study could therefore add a new function to this well-known gene. In addition, mutant analyses of several genes located in close proximity with the strongly associated SNPs on chromosome 2 identified *QUL2* as causal for the association. Hardly anything is reported about the biological or chemical function of this gene. Identification of *QUL2* shows that GWA mapping can lead to new discoveries.

From our search for genes involved in the regulation of plant growth under control conditions (Chapter 4) it can be concluded that many (independent) pathways influence growth within the natural population and possibly also within a single accession. GWA mapping of temporal data and growth curve parameters resulted in the detection of many QTLs. Within the support window of these QTLs eight genes were found for

which an overexpression or mutant phenotype related to growth was reported earlier. These eight genes were annotated to play a role in very different biological processes, such as germination, embryo development, senescence or flowering, suggesting that natural variation in growth is caused by sequence variation in genes which are part of various gene networks.

The candidate genes located in the support window of QTLs detected by GWA mapping of traits related to growth under moderate drought were prioritized based on previously reported expression data (Chapter 5) (Huang 2008, Harb 2010, Baerenfaller 2012, Des Marais 2012, Clauw 2015). Genes in the support window of a QTL were selected when differential expression was observed in at least three independent expression studies. For five, out of six, QTLs this approach resulted in the selection of probable causal genes that drive the stress response. When plants are exposed to drought many genes are differentially expressed (sometimes up to half of the total number of expressed genes). GWA mapping was helpful to distinguish between differentially expressed genes that are regulators that drive the stress response and genes that just change expression as a downstream effect of the action of the regulators.

The multi-environment mapping of fresh weight data recorded in 25 diverse environments suggested a remarkable number of flowering genes to be involved in the regulation of biomass accumulation. Among them were well characterised genes like *FLC, FY (*mRNA processing factor) and *COL2* (*CONSTANCE-LIKE 2*). This suggests a mechanistic link between flowering time and plant size (see next paragraph). In addition, for many QTLs none of the candidates genes was an obvious candidate based on prior knowledge suggesting that GWA mapping can lead to new discoveries. However, two obvious candidates were located in close proximity of QTLs with high QxE, *ACD6* (*ACCELERATED CELL DEATH 6*) and *RPM1* (*RESISTANCE TO P. SYRINGAE PV MACULICOLA 1*). Some alleles of both genes improve biomass accumulation in case of bacterial infections but reduce biomass accumulation in absence of bacteria (Tian 2003, Todesco 2010). This suggests a biological trade-off, because it depends on the environmental conditions which alleles are favourable. The possibility to estimate the size of the interaction between the QTL and the environment is one of the advantages of multi-environment GWAS over univariate GWAS.

In summary, GWA mapping allows to connect natural selection on the phenotype level to sequence variation at the genotype level. Detailed analysis of the sequence variation will help to hypothesize about the evolutionary trajectories (Bergelson 2010). GWA mapping is applied without an *a priori* hypothesis and therefore the ideal method to discover novel and unexpected connections between plant performance traits and genes. In addition, univariate GWA mapping will mostly discover environmental specific associations, whereas multi-environment GWA mapping is powerful in the detection of genes involved

in plant performance regulation in many diverse environments.

Validation and suggested follow-up work

In the previous section I focused on the conclusions that can be drawn from the putative and confirmed causal genes. In this section I will pay attention to the methods that can be used to validate the causality of the underlying genes. In addition I would like to suggest which focus follow-up work should have.

For three selected candidates, *MYB90,* a regulator of anthocyanin accumulation upon osmotic stress (Chapter 2), and *FLC* and *QUL2*, both regulators of heat responses (Chapter 3), I was able to perform additional validation experiments. In Chapter 2, first the expression of the three candidate genes upon osmotic stress was determined but no correlation between anthocyanin accumulation and gene expression was observed. Subsequently, re-sequence data were used to determine the alleles of the three candidate genes and a clear association was observed between higher anthocyanin accumulation and the second most frequent allele of *MYB90*. This allowed us to conclude that amino acid 209 of this gene is most probably causal for the association. Fortunately, this validation process was relatively easy due to the small support interval, the large allelic effect size, the simple treatment and the large amount of appropriate prior knowledge. Such an easy confirmation process is only possible if the QTL points to a well-described gene in one of the pathways known to be involved in the trait of interest.

In Chapter 3 the validation process was started by the analysis of eight T-DNA or mutant lines of genes in the support windows of two of the four heat-stress QTLs. Among all the mutants analysed, only one mutant phenotype was observed. Mutation of *FLC*, a flowering time gene, resulted in earlier flowering and improved heat tolerance. To be able to conclude that sequence variation in *FLC* is causal for the association, allele complementation should be done. For this, knock-out lines of different alleles should be constructed in their own background and subsequently each of the mutants is complemented with other alleles. If a knock-out line of one allele cannot be complemented by another allele, proof is obtained that the sequence difference between the alleles is responsible for the phenotypic difference. In most cases alleles of accessions with opposing phenotypes are tested and weak and strong alleles are compared. For *FLC* six major alleles are known (Li 2014) with different function during vernalization. Comparing them all pairwise would give a complete view of the functions of the alleles in heat responses. However, construction of complemented knockout lines takes several generations. If, a priori, a mutant line of the gene of interest is available in any background another method based on crossing can evaluate complementation capacities of alleles as well (Mackay 2004, Weigel 2012), reducing the number of generations needed. Complementation tests, through transformation or crossing, are needed to prove causality of the genes but additional functional analyses of the causal genes are as important because insights in the biological and mo-

lecular function of the gene will help to elucidate the molecular mechanisms underlying the trait of interest. In case of *FLC*, additional functional analysis would allow us to reveal whether *FLC* interacts directly with genes important in the heat response, or whether this connection is mediated via several intermediate downstream steps.

For one of the tested candidate genes, *QUL2*, no mutant phenotype was observed but redundancy in function was suggested for two family members because a temperature dependent phenotype was reported for a triple mutant of *QUL2* and two homologs (Fuentes 2010). I tested a double mutant of *QUL2* and its closest homolog *QUL1* and a decrease in heat tolerance was observed compared to both single mutants and the wild type, revealing redundancy between the function of *QUL1* and *QUL2*. This suggests that *QUL2* could be causal for the observed association but that sequence variation only leads to variation detectable at the phenotype level in accessions with weak or non-functional *QUL1* alleles. Not any functional analysis has yet been done for *QUL2* and follow-up experiments would, therefore, probably lead to new discoveries. Such analyses would be helpful in the identification of the pathway of which *QUL2* is part and to understand why sequence variation in *QUL2* results in changes in the heat response.

In summary, sequence and mutant analyses of three candidates resulted in strong suggestions of causality. Although complementation tests are needed to validate the causality, I think that follow-up research should be focused on the functional analysis of the candidate genes. This will result in improved understanding of the complex and dynamic regulation of plant performance and the interaction between genotype and environment.

Is plant performance regulated by *FLC*?

In the natural population used for all our experiments, large variation in flowering time was observed. Based on flowering time, two basic plant types can be distinguished within the population: summer annuals that flower without vernalization (i.e. without the requirement of a cold period) and winter annuals that require vernalization to flower. Summer annuals typically grow closer to the equator, whereas winter annuals typically grow in northern regions although many exceptions to this general rule are known (Stinchcombe 2004). In addition, summer annuals germinate in spring and flower in summer, whereas winter annuals germinate in autumn, overwinter in the rosette stage and flower in spring (Grennan 2006). These two types of life cycles have a large impact on the environmental conditions that plants experience during their life cycle and, therefore, also on the selection of traits.

Variation in vernalization requirement could, for a large part, be explained by sequence variation in two genes, *FRI* (*FRIGIDA*) and *FLC* (*FLOWERING LOCUS C*) (Michaels 1999, Michaels 2001, Li 2014). Accessions with strong alleles of both genes are winter annuals. Non-functional *FRI* results in loss of the vernalization requirement for flowering,

independent of the *FLC* allele. The vernalization requirement is also lost when the *FLC* allele is weak or non-functional, independent of the *FRI* allele. (Note: Although vernalization is not any more required for flowering, application is still advancing flowering.) It is suggested that the vernalization requirement has an evolutionary advantage in temperate climates since it prevents plants to flower in winter and it promotes rapid flowering in spring. Closer to the equator, where winters are less cold, the vernalization requirement is a disadvantage since it will prevent flowering in spring after a relatively warm winter. In Chapter 3 we report that a QTL was observed at chromosome 5 at the position of *FLC* for both flowering time (after vernalization) and pre-anthesis heat response. An association with a lower LOD score was observed at the same position for growth upon drought (Chapter 5) and a constitutive QTL was found at that position with effects on biomass accumulation in a range of different environments (Chapter 6). Furthermore, 8 additional genes involved in flowering time regulation are present in close proximity of the other SNPs associated with biomass accumulation (Chapter 6). In addition, in the Chapters 4 and 5 we report that, on average, summer annuals have larger rosettes than winter annuals. These observations suggest a genetic link between flowering time and plant performance. Colocation of flowering time QTLs with other QTLs has been reported before in several RIL populations, for example for drought response (Juenger 2005, McKay 2008, Tisné 2010, El-Soda 2014), leaf rhythmic movement (Swarup 1999), temperature dependent seed germination (Chiang 2009), and leaf senescence (Wingler 2010). In most cases the colocation was found with the *FLC* locus. All these observations suggest pleiotropic effects of *FLC*, which means that one gene, *FLC*, influences multiple seemingly unrelated traits. This influence is probably indirect via diversification of the downstream effects.

Besides pleiotropy, the link between flowering time and several plant performance traits can also be explained by correlated responses of the traits to selection. For example, loss of vernalization requirement is advantageous in warm climates. In these climates it is also advantageous to be drought and heat tolerant. Therefore, in warm climates the optimal organism has lost the vernalization requirement and, in addition, is drought and heat tolerant. Co-occurrence of these traits in the same accessions does therefore not per se require overlap in the regulatory network but might be a consequence of population structure. However, we observed a significant difference in heat response between two lines that contained either a Col-0 *FLC* allele or a non-functional *FLC* allele in the same background, which supports the suggestion of overlap in the regulatory networks and therefore a pleiotropic role for *FLC* (Chapter 3). This pleiotropic role is also supported by the analysis of NILs of three flowering time QTLs, of which one covers the genomic area around *FLC*. Significant differences were observed between early and late flowering NILs for drought response (Schmalenbach 2014). In addition, analysis of a NIL covering the genomic region around *FLC* and an *FLC* overexpression line revealed that high *FLC* expression leads to high germination rates at low temperatures (Chiang 2009). Additional

evidence for overlap in regulatory networks was gained from analyses of mutants of other genes in the regulatory pathway downstream of *FLC*. These mutants also showed changes in the germination rate at low temperatures. However, these results could not be fully reproduced by others (Chen 2014).

To be conclusive about the role of *FLC* in the determination of plant performance traits additional experiments are needed. To begin with, experiments to determine whether a general relation exists between timing of phase transition and plant performance (like drought and heat response) independent of the genotype. In this way we may answer questions like: Do external factors that influence flowering (such as photoperiod, temperature, externally applied GA) also influence plant performance? Do all mutants that change flowering time also change plant performance or is this specific for *FLC* (and its downstream targets)?
Secondly, experiments are needed to confirm overlap in regulatory networks by answering the following questions: Will mutation and overexpression of genes up- and downstream of *FLC* lead to changes in plant performance? Does *FLC* physically interact with genes known to be involved in plant performance [protein-protein interactions, protein-DNA interactions (cis-regulatory elements)]?

In summary, my work, and supporting literature, suggests that *FLC* has direct or indirect influence on different traits, but the molecular mechanisms and evolutionary background are still unclear. More detailed studies of *FLC* and its different roles are needed to clarify these points.

GWA mapping and the future

In literature, several advantages and shortcomings of GWA mapping have been described (Visscher 2012, Korte 2013). Considering these remarks and the results obtained in this thesis the following question arises: Is GWA mapping a valuable approach to use in future studies that aim to use natural variation to elucidate the genetic regulation of plant traits?

The strongest advantage of GWA mapping over classical QTL mapping is the detection of QTLs with small support windows because the linkage decay in natural populations is much faster than in experimental populations. My results confirmed this small support interval size. The second advantage of GWA mapping is the possibility to simultaneously test most of the natural variation present in the species. This thesis reports the detection of many associations that were not found earlier in similar experiments in RIL populations. The large variation included in the population is responsible for these new discoveries. Described disadvantages of GWA mapping are: spurious associations as a consequence of population structure, rare causal alleles that cannot be detected but hamper the detection of common causal alleles and lack of power to detect small effect causal alleles (Korte 2013). At the moment most statistical models used to perform GWA mapping allow for

correction for population structure to avoid false positive discovery as a consequence of population structure (Astle 2009, Seren 2012). Also in this thesis a model was used that corrects for population structure based on kinship. I think that presence of population structure should not stop us from using GWA mapping, because models that correct and that do not correct for population structure can always be used in parallel. Although using those models in parallel does not solve the problem of false positives and false negatives, it increases ones understanding of the problem. The two other disadvantages are in my opinion more serious obstacles. Analysis of sequence variation in *MYB90*, the gene causal for natural variation in anthocyanin accumulation upon osmotic stress (Chapter 2), revealed several rare alleles for this gene and for its close relative *MYB75*. For some rare alleles of *MYB75* it is known that they can have a large effect on the phenotype (Teng 2005). The QTN causal for the reported association could only explain 15% of the total phenotypic variance. No additional QTLs were observed. Therefore I suggest that the unexplained phenotypic variance is partially due to the presence of rare causal variants and due to small effect QTLs of which the detection of the effect was obscured by the rare causal alleles. When another natural population is used (more natural accessions or different natural accessions) allele distributions will be shifted and may lead to the detection of alleles that were rare in the natural population used in this thesis. But for optimal identification of rare causal variants and small effect QTLs bi-parental populations, in which all alleles have a 50%/50% distribution, are more suitable.

This thesis reports the detection of several strong associations for various plant performance traits (Chapters 3, 4 and 5). The strong associations that are related to the same trait together, in many cases, explained more than 25% of the total phenotypic variance and even more of the genotypic variance. For example, the three QTLs associated with pre-anthesis heat response together explained 27% of the total phenotypic variation and almost 50% of the total genotypic variation. This suggests that GWA mapping is able to detect the most important regulators of plant performance effectively. The minor regulators maybe resulted in moderate or weak associations and are therefore not considered in this thesis. Lowering the threshold would probably detect many of them, but the chance of false discoveries will likely also increase.

Interestingly, multi-environment GWA mapping resulted in the detection of large numbers of strong associations (more than 100 association above the threshold of -log(p)=10). For a substantial part of these associations the marker effect is small in all tested environments, suggesting that this method is more powerful to detect small effect QTLs.

Considering all these remarks and observations I would encourage scientists not to hesitate to use GWA mapping because, first of all, an inventory of the natural variation present for a trait of interest gives a lot of valuable information. It tells you whether Col-0 is representative for many or only for a few accessions. If several plant performance traits

are investigated in the same population it is possible to search for correlations between them. Correlations between plant traits and ecological conditions can then also be discovered. Investigation of natural variation can, therefore, lead to the formulation of hypotheses about favourable trait combinations in certain environments. Secondly, the amount of genetic variation that influences a trait of interest is for most traits unknown beforehand. GWA mapping will give an indication of the genetic architecture of a trait. But because for many complex traits several mechanisms are described that result in the same phenotype I recommend to focus not only on the overall phenotype but also to pay attention to sub-traits contributing to the overall phenotype. For example, drought resistance can be acquired, amongst others, by expansion of the roots, smart opening and closure of the stomata, and osmotic adjustment. The regulatory networks of these drought response mechanisms are most probably not, or only partially, overlapping. Therefore, it is expected that genetic variation in the corresponding genes will have a large impact on the sub-trait but only a small effect on the overall phenotype. It would be interesting, for example, to determine the root-shoot ratio, the stomatal conductance or the amount of osmotically active compounds in the leaves. The same can be considered for many other complex traits. It is expected that GWA mapping of less complex sub-traits will increase the power to detect regulatory genes (Kloth 2012). Similar arguments are used to perform QTL analysis on genome wide gene expression data (Keurentjes 2007) and metabolite profiles (Keurentjes 2009). Multi-trait GWA mapping of related traits allows the detection of regulatory genes with effects on multiple sub-traits.

In conclusion, GWA mapping is an attractive method to search for genes regulating complex traits, because of the small support intervals and the large natural variation that can be assessed. However, the genetic architecture of some complex traits may hamper GWA mapping. In such cases it may be helpful to also measure sub-traits for which the genetic architecture fits GWA mapping better. I expect that multi-trait GWA mapping would support the detection of regulatory genes influencing several sub-traits.

Impact results on research outside Arabidopsis field

Functional characterisation of candidate genes should result in fundamental insights in the regulation of plant performance. These insights likely provide information that is not only valid for Arabidopsis but also for other plant species. One of the desired follow-ups of my work is to use the findings to develop crops that are more stress tolerant and that have a stable performance in a wider range of environments. Translation of our findings to crops may be done in several ways. It is possible to search for orthologs of candidate genes in crops and additionally exploit the natural variation of those orthologs within a crop species. To confirm similarity in function, these crop orthologs could be expressed in Arabidopsis mutant lines to assess if any of the natural alleles can complement the mutant phenotype. It is also possible to express the Arabidopsis gene in a crop species. If this results in better plant performance, this insight could be directly used to develop

genetically modified crops but it might also be a research tool to understand which genes in crops are involved in obtaining better performance. This may be done via studies of crop proteins that interact with the Arabidopsis proteins or through studies of crop genes that are differentially expressed as a consequence of the presence of the Arabidopsis gene. If homology with the candidate is weak or when the observed orthologs seem to have a different function, one should focus on understanding the underlying mechanism, since translation of networks from one species to another might be more fruitful than single gene comparisons (Ferrier 2011).

Not all crops are equally suitable for translation of my findings. Availability of sequence information (genome or transcriptome sequence) will speed up the search for orthologs and their natural alleles. Without sequence information orthologs could be found using primers designed to amplify conserved regions of the genes and subsequent amplification of the adjacent regions by (RACE-)PCR, but this is more time- and labour-intensive. Fortunately, with the recent advances in sequencing techniques the sequencing of cDNA libraries is relatively straight forward and, therefore, in many cases suitable and preferred. In addition, availability of a transformation protocol is needed for expression of Arabidopsis genes in the crop. Sequence information of wild relatives is also valuable, because it allows to search for sequence variation of the candidate genes. If sequence variation can be coupled to changes in the phenotype, marker assisted breeding programs could be started.

GWA mapping of agronomic traits is also performed in some crops (Cockram 2010, Famoso 2011, Norton 2014, Thirunavukkarasu 2014). Because of the limited genetic variation present within breeders' germplasm, the support windows of the QTLs reported are usually larger than for Arabidopsis and, in addition, it is expected that for some important regulatory genes no sequence variation is present anymore within the population due to selection against it. Addition of wild relatives of crops to populations of cultivars would enhance the genetic variation tested and therefore improve the resolution of GWA mapping in crops. The positive site of reduced genetic variation is a reduction in number of genetic pathways that result in the same phenotype and, therefore, it is expected that the genetic network behind the trait of interest is less complex. Due to the disadvantages of GWA mapping in crops, GWA mapping and subsequent functional analysis of genes in Arabidopsis is still a valuable tool in understanding plant responses to varying environments (Assmann 2013).

Conclusion

My work has resulted in the identification of a substantial number of genomic regions involved in the regulation of plant performance under heat, drought or other abiotic stress conditions. For several genes in these genomic regions strong evidence was obtained that sequence variation was causal for natural variation at the phenotype level. This thesis shows that GWA mapping in Arabidopsis is a useful tool to identify new functions for known and unknown genes. Such insights will be useful in the interpretation of future mapping results in Arabidopsis and in crops. Follow-up translational experiments are challenging, but successful translation will have great economic and societal impact, because abiotic stresses cause large worldwide yield losses.

References

Afzal A.J., da Cunha L. & Mackey D. (2011) Separable Fragments and Membrane Tethering of Arabidopsis RIN4 Regulate Its Suppression of PAMP-Triggered Immunity. *The Plant Cell*, **23**, 3798-3811.

Aguirrezabal L., Bouchier-Combaud S., Radziejwoski A., Dauzat M., Cookson S.J. & Granier C. (2006) Plasticity to soil water deficit in Arabidopsis thaliana: dissection of leaf development into underlying growth dynamic and cellular variables reveals invisible phenotypes. *Plant, Cell & Environment*, **29**, 2216-2227.

Aharoni A., Dixit S., Jetter R., Thoenes E., van Arkel G. & Pereira A. (2004) The SHINE Clade of AP2 Domain Transcription Factors Activates Wax Biosynthesis, Alters Cuticle Properties, and Confers Drought Tolerance when Overexpressed in Arabidopsis. *The Plant Cell Online*, **16**, 2463-2480.

Al-Whaibi M.H. (2011) Plant heat-shock proteins: A mini review. *Journal of King Saud University - Science*, **23**, 139-150.

Alimi N.A., Bink M.C.A.M., Dieleman J.A., Magán J.J., Wubs A.M., Palloix A. & van Eeuwijk F.A. (2013) Multi-trait and multi-environment QTL analyses of yield and a set of physiological traits in pepper. *TAG Theoretical and Applied Genetics*, **126**, 2597-2625.

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

Alonso-Blanco C., Blankestijn-de Vries H., Hanhart C.J. & Koornneef M. (1999) Natural allelic variation at seed size loci in relation to other life history traits of Arabidopsis thaliana. *Proceedings of the National Academy of Sciences*, **96**, 4710-4717.

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

Alonso-Blanco C., Peeters A.J.M., Koornneef M., Lister C., Dean C., van den Bosch N.(...)Kuiper M.T.R. (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. *The Plant Journal*, **14**, 259-271.

Alonso J.M. & Stepanova A.N. (2003) T-DNA Mutagenesis in Arabidopsis. In: *Plant Functional Genomics*, pp. 177-187. Humana Press.

Angadi S.V., Cutforth H.W., McConkey B.G. & Gan Y. (2003) Yield Adjustment by Canola Grown at Different Plant Populations under Semiarid Conditions. *Crop Science*, **43**, 1358-1366.

Angadi S.V., Cutforth H.W., Miller P.R., McConkey B.G., Entz M.H., Brandt S.A. & Volkmar K.M. (2000) Response of three Brassica species to high temperature stress during reproductive growth. *Canadian Journal of Plant Science*, **80**, 693-701.

Antoun M. & Ouellet F. (2013) Growth temperature affects inflorescence architecture in Arabidopsis thaliana. *Botany*, **91**, 642-651.

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

Araus J.L., Slafer G.A., Royo C. & Serret M.D. (2008) Breeding for yield potential and stress adaptation in cereals. *Critical Reviews in Plant Sciences*, **27**, 377-412.

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

Ashraf M. (2010) Inducing drought tolerance in plants: Recent advances. *Biotechnology Advances*, **28**, 169-183.

Assmann S.M. (2013) Natural Variation in Abiotic Stress and Climate Change Responses in Arabidopsis: Implications for Twenty-First-Century Agriculture. *International Journal of Plant Sciences*, **174**, 3-26.

Astle W. & Balding D.J. (2009) Population Structure and Cryptic Relatedness in Genetic Association Studies. *Statist. Sci.*, **24**, 451-471.

Atwell S., Huang Y.S., Vilhjalmsson B.J., Willems G., Horton M., Li Y.(...)Nordborg M. (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature*, **465**, 627-631.

Avery O.T., MacLeod C.M. & McCarty M. (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types: Induction of transformation by a desoxyribonucleid acid fraction isolated from pneumococcus type III. *The Journal of Experimental Medicine*, **79**, 137-158.

Bac-Molenaar J.A., Vreugdenhil D., Granier C. & Keurentjes J.J.B. (2015) Genome-wide association mapping of growth dynamics detects time-specific and general quantitative trait loci. *Journal of Experimental Botany*.

Baerenfaller K., Massonnet C., Walsh S., Baginsky S., Buhlmann P., Hennig L.(...)Gruissem W. (2012) Systems-based analysis of Arabidopsis leaf growth reveals adaptation to water deficit. *Molecular systems biology*, **8**, 606.

Balasubramanian S., Schwartz C., Singh A., Warthmann N., Kim M.C., Maloof J.N.(...)Weigel D. (2009) QTL Mapping in New *Arabidopsis thaliana* Advanced Intercross-Recombinant Inbred Lines. *PLoS ONE*, **4**, e4318.

Balasubramanian S., Sureshkumar S., Lempe J. & Weigel D. (2006) Potent Induction of *Arabidopsis thaliana* Flowering by Elevated Growth Temperature. *PLoS Genet*, **2**, e106.

Barboza L., Effgen S., Alonso-Blanco C., Kooke R., Keurentjes J.J.B., Koornneef M. & Alcázar R. (2013) Arabidopsis semidwarfs evolved from independent mutations in GA20ox1, ortholog to green revolution dwarf alleles in rice and barley. *Proceedings of the National Academy of Sciences*, **110**, 15818-15823.

Barnabàs B., Jäger K. & Fehér A. (2008) The effect of drought and heat stress on reproductive processes in cereals. *Plant, Cell & Environment*, **31**, 11-38.

Baskin J.M. & Baskin C.C. (1983) Seasonal changes in the germination responses of buried seeds of Arabidopsis thaliana and ecological interpretation. *Botanical Gazette*, 540-543.

Baxter I., Brazelton J.N., Yu D., Huang Y.S., Lahner B., Yakubova E.(...)Salt D.E. (2010) A coastal cline in sodium accumulation in Arabidopsis thaliana is driven by natural variation of the sodium transporter AtHKT1;1. *PLoS genetics*, **6**, e1001193.

Bennett S.R.M., Alvarez J., Bossinger G. & Smyth D.R. (1995) Morphogenesis in pinoid mutants of Arabidopsis thaliana. *The Plant Journal*, **8**, 505-520.

Bensmihen S., Giraudat J. & Parcy F. (2005) Characterization of three homologous basic leucine zipper transcription factors (bZIP) of the ABI5 family during Arabidopsis thaliana embryo maturation. *Journal of Experimental Botany*, **56**, 597-603.

Bergelson J. & Roux F. (2010) Towards identifying genes underlying ecologically relevant traits in Arabidopsis thaliana. *Nature Reviews Genetics*, **11**, 867-879.

Berger B., Parent B. & Tester M. (2010) High-throughput shoot imaging to study drought responses. *Journal of Experimental Botany*, **61**, 3519-3528.

Bita C. & Gerats T. (2013) Plant tolerance to high temperature in a changing environment: scientific fundamentals and production of heat stress tolerant crops. *Frontiers in Plant Science*, **4**.

Bita C.E., Zenoni S., Vriezen W.H., Mariani C., Pezzotti M. & Gerats T. (2011) Temperature stress differentially modulates transcription in meiotic anthers of heat-tolerant and heat-sensitive tomato plants. *Bmc Genomics*, **12**.

Bohnert H.J., Nelson D.E. & Jensen R.G. (1995) Adaptations to environmental stresses. *The Plant Cell*, **7**, 1099.

Borevitz J.O., Xia Y., Blount J., Dixon R.A. & Lamb C. (2000) Activation Tagging Identifies a Conserved MYB Regulator of Phenylpropanoid Biosynthesis. *The Plant Cell Online*, **12**, 2383-2393.

Botstein D., White R.L., Skolnick M. & Davis R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, **32**, 314-331.

Bouchabke O., Chang F., Simon M., Voisin R., Pelletier G. & Durand-Tardif M. (2008) Natural Variation in *Arabidopsis thaliana* as a Tool for Highlighting Differential Drought Responses. *PLoS ONE*, **3**, e1705.

Boyer J.S., James R.A., Munns R., Condon T.G. & Passioura J.B. (2008) Osmotic adjustment leads to anomalously low estimates of relative water content in wheat and barley. *Functional Plant Biology*, **35**, 1172-1182.

Boyes D.C., Zayed A.M., Ascenzi R., McCaskill A.J., Hoffman N.E., Davis K.R. & Gorlach J. (2001) Growth Stage-Based Phenotypic Analysis of Arabidopsis: A Model for High Throughput Functional Genomics in Plants. *Plant Cell*, **13**, 1499-1510.

Brachi B., Faure N., Horton M., Flahauw E., Vazquez A., Nordborg M.(...)Roux F. (2010) Linkage and Association Mapping of *Arabidopsis thaliana* Flowering Time in Nature. *PLoS Genet*, **6**, e1000940.

Briggs G.C., Osmont K.S., Shindo C., Sibout R. & Hardtke C.S. (2006) Unequal genetic redundancies in Arabidopsis – a neglected phenomenon? *Trends in Plant Science*, **11**, 492-498.

Busemeyer L., Ruckelshausen A., Moller K., Melchinger A.E., Alheit K.V., Maurer H.P.(...)Wurschum T. (2013) Precision phenotyping of biomass accumulation in triticale reveals temporal genetic patterns of regulation. *Scientific reports*, **3**.

Camargo A., Papadopoulou D., Spyropoulou Z., Vlachonasios K., Doonan J.H. & Gay A.P. (2014) Objective Definition of Rosette Shape Variation Using a Combined Computer Vision and Data Mining Approach. *PLoS ONE*, **9**, e96889.

Chaïb J., Lecomte L., Buret M. & Causse M. (2006) Stability over genetic backgrounds, generations and years of quantitative trait locus (QTLs) for organoleptic quality in tomato. *TAG Theoretical and Applied Genetics*, **112**, 934-944.

Chalker-Scott L. (1999) Environmental Significance of Anthocyanins in Plant Stress Responses. *Photochemistry and Photobiology*, **70**, 1-9.

Chalker-Scott L. (2002) Do anthocyanins function as osmoregulators in leaf tissues? In: *Advances in Botanical Research*, pp. 103-106. Academic Press.

Chan E.K.F., Rowe H.C., Corwin J.A., Joseph B. & Kliebenstein D.J. (2011) Combining Genome-Wide Association Mapping and Transcriptional Networks to Identify Novel Genes Controlling Glucosinolates in *Arabidopsis thaliana*. *PLoS Biology*, **9**, e1001125.

Chan E.K.F., Rowe H.C. & Kliebenstein D.J. (2010) Understanding the Evolution of Defense Metabolites in Arabidopsis thaliana Using Genome-wide Association Mapping. *Genetics*, **185**, 991-1007.

Chao D.-Y., Chen Y., Chen J., Shi S., Chen Z., Wang C.(...)Salt D.E. (2014) Genome-wide Association Mapping Identifies a New Arsenate Reductase Enzyme Critical for Limiting Arsenic Accumulation in Plants. *PLoS Biol*, **12**, e1002009.

Chardon F., Barthélémy J., Daniel-Vedele F. & Masclaux-Daubresse C. (2010) Natural variation of nitrate uptake and nitrogen use efficiency in Arabidopsis thaliana cultivated with limiting and ample nitrogen supply. *Journal of Experimental Botany*, **61**, 2293-2302.

Chardon F., Jasinski S., Durandet M., Lécureuil A., Soulay F., Bedu M.(...)Masclaux-Daubresse C. (2014) QTL meta-analysis in Arabidopsis reveals an interaction between leaf senescence and resource allocation to seeds. *Journal of Experimental Botany*, **65**, 3949-3962.

Chay P. & Thurling N. (1989) Identification of Genes Controlling Pod Length in Spring Rapeseed, Brassica napus L., and their Utilization for Yield Improvement. *Plant Breeding*, **103**, 54-62.

Chen C., DeClerck G., Tian F., Spooner W., McCouch S. & Buckler E. (2012a) *PICARA*, an Analytical Pipeline Providing Probabilistic Inference about *A Priori* Candidates Genes Underlying Genome-Wide Association QTL in Plants. *PLoS ONE*, **7**, e46596.

Chen H., Shen Y., Tang X., Yu L., Wang J., Guo L.(...)Deng X.W. (2006) Arabidopsis CULLIN4 Forms an E3 Ubiquitin Ligase with RBX1 and the CDD Complex in Mediating Light Control of Development. *The Plant Cell Online*, **18**, 1991-2004.

Chen L., Song Y., Li S., Zhang L., Zou C. & Yu D. (2012b) The role of WRKY transcription factors in plant abiotic stresses. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, **1819**, 120-128.

Chen M., MacGregor D.R., Dave A., Florance H., Moore K., Paszkiewicz K.(...)Penfield S. (2014) Maternal temperature history activates Flowering Locus T in fruits to control progeny dormancy according to time of year. *Proceedings of the National Academy of Sciences*, **111**, 18787-18792.

Cheng M.-C., Liao P.-M., Kuo W.-W. & Lin T.-P. (2013) The Arabidopsis ETHYLENE RESPONSE FACTOR1 Regulates Abiotic Stress-Responsive Gene Expression by Binding to Different cis-Acting Elements in Response to Different Stress Signals. *Plant Physiology*, **162**, 1566-1582.

Chiang G.C.K., Barua D., Kramer E.M., Amasino R.M. & Donohue K. (2009) Major flowering time gene, FLOWERING LOCUS C, regulates seed germination in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences*, **106**, 11661-11666.

Chołuj D., Wiśniewska A., Szafrański K.M., Cebula J., Gozdowski D. & Podlaski S. (2014) Assessment of the physiological responses to drought in different sugar beet genotypes in connection with their genetic distance. *Journal of Plant Physiology*, **171**, 1221-1230.

Clarke H.J. & Siddique K.H.M. (2004) Response of chickpea genotypes to low temperature stress during reproductive development. *Field Crops Research*, **90**, 323-334.

Clauw P., Coppens F., De Beuf K., Dhondt S., Van Daele T., Maleux K.(...)Inzé D. (2015) Leaf Responses to Mild Drought Stress in Natural Variants of Arabidopsis. *Plant Physiology*, **167**, 800-816.

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

Cockram J., White J., Zuluaga D.L., Smith D., Comadran J., Macaulay M.(...)O'Sullivan D.M. (2010) Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. *Proceedings of the National Academy of Sciences*, **107**, 21611-21616.

Collard B.C.Y. & Mackill D.J. (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **363**, 557-572.

Collins N.C., Tardieu F. & Tuberosa R. (2008) Quantitative Trait Loci and Crop Performance under Abiotic Stress: Where Do We Stand? *Plant Physiology*, **147**, 469-486.

Cosgrove D.J. (2000) Loosening of plant cell walls by expansins. *Nature*, **407**, 321-326.

Cosgrove D.J., Bedinger P. & Durachko D.M. (1997) Group I allergens of grass pollen as cell wall-loosening agents. *Proceedings of the National Academy of Sciences*, **94**, 6559-6564.

Cosgrove D.J., Li L.C., Cho H.-T., Hoffmann-Benning S., Moore R.C. & Blecker D. (2002) The Growing World of Expansins. *Plant and Cell Physiology*, **43**, 1436-1444.

Czechowski T., Stitt M., Altmann T., Udvardi M.K. & Scheible W.-R. (2005) Genome-Wide Identification and Testing of Superior Reference Genes for Transcript Normalization in Arabidopsis. *Plant Physiology*, **139**, 5-17.

de los Campos G., Hickey J.M., Pong-Wong R., Daetwyler H.D. & Calus M.P.L. (2013) Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. *Genetics*, **193**, 327-345.

de Souza A., Hull P., Gille S. & Pauly M. (2014) Identification and functional characterization of the distinct plant pectin esterases PAE8 and PAE9 and their deletion mutants. *Planta*, **240**, 1123-1138.

Dechaine J., Brock M. & Weinig C. (2014) QTL architecture of reproductive fitness characters in Brassica rapa. *BMC Plant Biology*, **14**, 66.

Dejardin A., Sokolov L. & Kleczkowski L. (1999) Sugar/osmoticum levels modulate differential abscisic acid-independent expression of two stress-responsive sucrose synthase genes in Arabidopsis. *Biochemical Journal*, **344**, 503-509.

Dekkers B.J.W., Willems L., Bassel G.W., van Bolderen-Veldkamp R.P., Ligterink W., Hilhorst H.W.M. & Bentsink L. (2012) Identification of Reference Genes for RT–qPCR Expression Analysis in Arabidopsis and Tomato Seeds. *Plant and Cell Physiology*, **53**, 28-37.

Demura T. & Ye Z.-H. (2010) Regulation of plant biomass production. *Current Opinion in Plant Biology*, **13**, 298-303.

Des Marais D.L., Hernandez K.M. & Juenger T.E. (2013) Genotype-by-Environment Interaction and Plasticity: Exploring Genomic Responses of Plants to the Abiotic Environment. *Annual Review of Ecology, Evolution, and Systematics*, **44**, 5-29.

Des Marais D.L., McKay J.K., Richards J.H., Sen S., Wayne T. & Juenger T.E. (2012) Physiological Genomics of Response to Soil Drying in Diverse Arabidopsis Accessions. *The Plant Cell Online*, **24**, 893-914.

DeWan A., Liu M., Hartman S., Zhang S.S.M., Liu D.T.L., Zhao C.(...)Hoh J. (2006) HTRA1 promoter polymorphism in wet age-related macular degeneration. *Science*, **314**, 989-992.

Diaz C., Lemaître T., Christ A., Azzopardi M., Kato Y., Sato F.(...)Masclaux-Daubresse C. (2008) Nitrogen Recycling and Remobilization Are Differentially Controlled by Leaf Senescence and Development Stage in Arabidopsis under Low Nitrogen Nutrition. *Plant Physiology*, **147**, 1437-1449.

Diaz C., Saliba-Colombani V., Loudet O., Belluomo P., Moreau L., Daniel-Vedele F.(...)Masclaux-Daubresse C. (2006) Leaf Yellowing and Anthocyanin Accumulation are Two Genetically Independent Strategies in Response to Nitrogen Limitation in Arabidopsis thaliana. *Plant and Cell Physiology*, **47**, 74-83.

Djanaguiraman M., Prasad P.V.V., Boyle D.L. & Schapaugh W.T. (2013) Soybean Pollen Anatomy, Viability and Pod Set under High Temperature Stress. *Journal of Agronomy and Crop Science*, **199**, 171-177.

Donahue J.L., Alford S.R., Torabinejad J., Kerwin R.E., Nourbakhsh A., Ray W.K.(...)Gillaspy G.E. (2010) The Arabidopsis thaliana Myo-Inositol 1-Phosphate Synthase1 Gene Is Required for Myo-inositol Synthesis and Suppression of Cell Death. *The Plant Cell Online*, **22**, 888-903.

El-Lithy M.E., Clerkx E.J.M., Ruys G.J., Koornneef M. & Vreugdenhil D. (2004) Quantitative Trait Locus Analysis of Growth-Related Traits in a New Arabidopsis Recombinant Inbred Population. *Plant Physiology*, **135**, 444-458.

El-Soda M., Kruijer W., Malosetti M., Koornneef M. & Aarts M.G.M. (2014a) Quantitative trait loci and candidate genes underlying genotype by environment interaction in the response of Arabidopsis thaliana to drought. *Plant, Cell & Environment*, **38**, 585-599.

El-Soda M., Malosetti M., Zwaan B.J., Koornneef M. & Aarts M.G.M. (2014b) Genotype × environment interaction QTL mapping in plants: lessons from Arabidopsis. *Trends in Plant Science*, **19**, 390-398.

Elwell A.L., Gronwall D.S., Miller N.D., Spalding E.P. & Durham Brooks T.L. (2011) Separating parental environment from seed size effects on next generation growth and development in Arabidopsis. *Plant, Cell & Environment*, **34**, 291-301.

Endo M., Tsuchiya T., Hamada K., Kawamura S., Yano K., Ohshima M.(...)Kawagishi-Kobayashi M. (2009) High Temperatures Cause Male Sterility in Rice Plants with Transcriptional Alterations During Pollen Development. *Plant and Cell Physiology*, **50**, 1911-1922.

Fabre J., Dauzat M., Negre V., Wuyts N., Tireau A., Gennari E.(...)Granier C. (2011) PHENOPSIS DB: an Information System for Arabidopsis thaliana phenotypic data in an environmental context. *BMC Plant Biology*, **11**, 77.

Fahlgren N., Gehan M.A. & Baxter I. (2015) Lights, camera, action: high-throughput plant phenotyping is ready for a close-up. *Current Opinion in Plant Biology*, **24**, 93-99.

Famoso A.N., Zhao K., Clark R.T., Tung C.-W., Wright M.H., Bustamante C.(...)McCouch S.R. (2011) Genetic Architecture of Aluminum Tolerance in Rice (*Oryza sativa*) Determined through Genome-Wide Association Analysis and QTL Mapping. *PLoS Genet*, **7**, e1002221.

Fan J., Quan S., Orth T., Awai C., Chory J. & Hu J. (2005) The Arabidopsis PEX12 Gene Is Required for Peroxisome Biogenesis and Is Essential for Development. *Plant Physiology*, **139**, 231-239.

Fang Y. & Xiong L. (2015) General mechanisms of drought response and their application in drought resistance improvement in plants. *Cellular and Molecular Life Sciences*, **72**, 673-689.

Farooq M., Hussain M., Wahid A. & Siddique K.H.M. (2012) Drought Stress in Plants: An Overview. In: *Plant Responses to Drought Stress*, pp. 1-33. Springer Berlin Heidelberg.

Fernandez D.E., Heck G.R., Perry S.E., Patterson S.E., Bleecker A.B. & Fang S.-C. (2000) The Embryo MADS Domain Factor AGL15 Acts Postembryonically: Inhibition of Perianth Senescence and Abscission via Constitutive Expression. *The Plant Cell Online*, **12**, 183-197.

Ferrier T., Matus J.T., Jin J. & Riechmann J.L. (2011) Arabidopsis paves the way: genomic and network analyses in crops. *Current Opinion in Biotechnology*, **22**, 260-270.

Freedman M.L., Monteiro A.N.A., Gayther S.A., Coetzee G.A., Risch A., Plass C.(...)Mills I.G. (2011) Principles for the post-GWAS functional characterization of cancer risk loci. *Nature Genetics*, **43**, 513-518.

Fu Y.X. & Li W.H. (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693-709.

Fuentes S., Pires N. & Østergaard L. (2010) A clade in the QUASIMODO2 family evolved with vascular plants and supports a role for cell wall composition in adaptation to environmental changes. *Plant molecular biology*, **73**, 605-615.

Furbank R.T. & Tester M. (2011) Phenomics – technologies to relieve the phenotyping bottleneck. *Trends in Plant Science*, **16**, 635-644.

Gan Y., Angadi S.V., Cutforth H., Potts D., Angadi V.V. & McDonald C.L. (2004) Canola and mustard response to short periods of temperature and water stress at different developmental stages. *Canadian Journal of Plant Science*, **84**, 697-704.

Gauch Jr H. (1992) *Statistical analysis of regional yield trials: AMMI analysis of factorial designs*. Elsevier Science Publishers.

Gazzani S., Gendall A.R., Lister C. & Dean C. (2003) Analysis of the Molecular Basis of Flowering Time Variation in Arabidopsis Accessions. *Plant Physiology*, **132**, 1107-1114.

Gebert M., Dresselhaus T. & Sprunck S. (2008) F-Actin Organization and Pollen Tube Tip Growth in Arabidopsis Are Dependent on the Gametophyte-Specific Armadillo Repeat Protein ARO1. *The Plant Cell*, **20**, 2798-2814.

Geng Y., Wu R., Wee C.W., Xie F., Wei X., Chan P.M.Y.(...)Dinneny J.R. (2013) A Spatio-Temporal Understanding of Growth Regulation during the Salt Stress Response in Arabidopsis. *The Plant Cell Online*, **25**, 2132-2154.

Giorno F., Wolters-Arts M., Mariani C. & Rieu I. (2013) Ensuring Reproduction at High Temperatures: The Heat Stress Response during Anther and Pollen Development. *Plants*, **2**, 489-506.

Gompertz B. (1825) On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. *Philosophical Transactions of the Royal Society of London*, **115**, 513-583.

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

Gonzalez N., Beemster G.T.S. & Inzé D. (2009) David and Goliath: what can the tiny weed Arabidopsis teach us to improve biomass production in crops? *Current Opinion in Plant Biology*, **12**, 157-164.

Gonzalez N., Vanhaeren H. & Inzé D. (2012) Leaf size control: complex coordination of cell division and expansion. *Trends in Plant Science*, **17**, 332-340.

Gosti F., Bertauche N., Vartanian N. & Giraudat J. (1995) Abscisic-acid-dependent and abscisic-acid-independent regulation of gene-expression by progressive drought in *Arabidopsis thaliana*. *Molecular & General Genetics*, **246**, 10-18.

Gourdji S.M., Sibley A.M. & Lobell D.B. (2013) Global crop exposure to critical high temperatures in the reproductive period: historical trends and future projections. *Environmental Research Letters*, **8**, 024041.

Granier C., Aguirrezabal L., Chenu K., Cookson S.J., Dauzat M., Hamard P.(...)Tardieu F. (2006) PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in Arabidopsis thaliana permitted the identification of an accession with low sensitivity to soil water deficit. *New Phytologist*, **169**, 623-635.

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

Grilli G.V.G., Braz L.T. & Lemos E.G.M. (2007) QTL identification for tolerance to fruit set in tomato by fAFLP markers. *Crop Breeding and Applied Biotechnology*, **7**, 234-241.

Guo M., Yang A., Zhou C. & Liu X. (2011) The new understanding of Arabidopsis thaliana proteins associated with salinity. *Journal of Plant Interactions*, **7**, 348-355.

Guo Y.-L., Todesco M., Hagmann J., Das S. & Weigel D. (2012) Independent FLC Mutations as Causes of Flowering-Time Variation in Arabidopsis thaliana and Capsella rubella. *Genetics*, **192**, 729-739.

Gutha L., Casassa L., Harbertson J. & Naidu R. (2010) Modulation of flavonoid biosynthetic pathway genes and anthocyanins due to virus infection in grapevine (Vitis vinifera L.) leaves. *BMC Plant Biology*, **10**, 187.

Hagenblad J., Tang C., Molitor J., Werner J., Zhao K., Zheng H.(...)Nordborg M. (2004) Haplotype Structure and Phenotypic Associations in the Chromosomal Regions Surrounding Two Arabidopsis thaliana Flowering Time Loci. *Genetics*, **168**, 1627-1638.

Hale K.L., McGrath S.P., Lombi E., Stack S.M., Terry N., Pickering I.J.(...)Pilon-Smits E.A.H. (2001) Molybdenum Sequestration in BrassicaSpecies. A Role for Anthocyanins? *Plant Physiology*, **126**, 1391-1402.

Hannah M.A., Wiese D., Freund S., Fiehn O., Heyer A.G. & Hincha D.K. (2006) Natural Genetic Variation of Freezing Tolerance in Arabidopsis. *Plant Physiology*, **142**, 98-112.

Harada H. & Leigh R.A. (2006) Genetic mapping of natural variation in potassium concentrations in shoots of Arabidopsis thaliana. *Journal of Experimental Botany*, **57**, 953-960.

Harb A., Krishnan A., Ambavaram M.M.R. & Pereira A. (2010) Molecular and Physiological Analysis of Drought Stress in Arabidopsis Reveals Early Responses Leading to Acclimation in Plant Growth. *Plant Physiology*, **154**, 1254–1271.

Harbinson J., Prinzenberg A.E., Kruijer W. & Aarts M.G.M. (2012) High throughput screening with chlorophyll fluorescence imaging and its use in crop improvement. *Current Opinion in Biotechnology*, **23**, 221-226.

Hartmann A., Czauderna T., Hoffmann R., Stein N. & Schreiber F. (2011) HTPheno: An image analysis pipeline for high-throughput plant phenotyping. *BMC Bioinformatics*, **12**, 148.

Hasdai M., Weiss B., Levi A., Samach A. & Porat R. (2006) Differential responses of Arabidopsis ecotypes to cold, chilling and freezing temperatures. *Annals of Applied Biology*, **148**, 113-120.

He H., de Souza Vidigal D., Snoek L.B., Schnabel S., Nijveen H., Hilhorst H. & Bentsink L. (2014) Interaction between parental environment and genotype affects plant and seed performance in Arabidopsis. *Journal of Experimental Botany*, **65**, 6603-6615.

Hedhly A. (2011) Sensitivity of flowering plant gametophytes to temperature fluctuations. *Environmental and Experimental Botany*, **74**, 9-16.

Hedhly A., Hormaza J.I. & Herrero M. (2004) Effect of temperature on pollen tube kinetics and dynamics in sweet cherry, Prunus avium (Rosaceae). *American Journal of Botany*, **91**, 558-564.

Hedhly A., Hormaza J.I. & Herrero M. (2005) The Effect of Temperature on Pollen Germination, Pollen Tube Growth, and Stigmatic Receptivity in Peach. *Plant biol (Stuttg)*, **7**, 476-483.

Herridge R., Day R., Baldwin S. & Macknight R. (2011) Rapid analysis of seed size in Arabidopsis for mutant and QTL discovery. *Plant Methods*, **7**, 3.

Hershey A.D. & Chase M. (1952) Independent functions of viral protein and nucleid acid in growth of bacteriophage. *The Journal of General Physiology*, **36**, 39-56.

Heuven H. & Janss L. (2010) Bayesian multi-QTL mapping for growth curve parameters. *BMC Proceedings*, **4**, S12.

Hincha D.K., Höfner R., Schwab K.B., Heber U. & Schmitt J.M. (1987) Membrane Rupture Is the Common Cause of Damage to Chloroplast Membranes in Leaves Injured by Freezing or Excessive Wilting. *Plant Physiology*, **83**, 251-253.

Hittalmani S., Huang N., Courtois B., Venuprasad R., Shashidhar H.E., Zhuang J.Y.(...)Khush G.S. (2003) Identification of QTL for growth- and grain yield-related traits in rice across nine locations of Asia. *TAG Theoretical and Applied Genetics*, **107**, 679-690.

Hoch W.A., Zeldin E.L. & McCown B.H. (2001) Physiological significance of anthocyanins during autumnal leaf senescence. *Tree Physiology*, **21**, 1-8.

Honys D. & Twell D. (2004) Transcriptome analysis of haploid male gametophyte development in Arabidopsis. *Genome Biology*, **5**, R85.

Houle D., Govindaraju D.R. & Omholt S. (2010) Phenomics: the next challenge. *Nature Reviews Genetics*, **11**, 855-866.

Huang D., Wu W., Abrams S.R. & Cutler A.J. (2008) The relationship of drought-related gene expression in Arabidopsis thaliana to hormonal and environmental factors. *Journal of Experimental Botany*, **59**, 2991-3007.

Huang X., Paulo M.-J., Boer M., Effgen S., Keizer P., Koornneef M. & van Eeuwijk F.A. (2011) Analysis of natural allelic variation in Arabidopsis using a multiparent recombinant inbred line population. *Proceedings of the National Academy of Sciences*, **108**, 4488-4493.

Huang Z., Footitt S. & Finch-Savage W.E. (2014) The effect of temperature on reproduction in the summer and winter annual Arabidopsis thaliana ecotypes Bur and Cvi. *Annals of Botany*, **113**, 921-929.

Huefner N.D., Mizuno Y., Weil C.F., Korf I. & Britt A.B. (2011) Breadth by depth: Expanding our understanding of the repair of transposon-induced DNA double strand breaks via deep-sequencing. *DNA Repair*, **10**, 1023-1033.

Hummel I., Pantin F., Sulpice R., Piques M., Rolland G., Dauzat M.(...)Muller B. (2010) Arabidopsis Plants Acclimate to Water Deficit at Low Cost through Changes of Carbon Usage: An Integrated Perspective Using Growth, Metabolite, Enzyme, and Gene Expression Analysis. *Plant Physiology*, **154**, 357-372.

IBM Corp (2010) IBM SPSS Statistics for Windows, version 19.0. *Armonk, NY: IBM Corp.*

Ilk N., Ding J., Ihnatowicz A., Koornneef M. & Reymond M. (2015) Natural variation for anthocyanin accumulation under high-light and low-temperature stress is attributable to the ENHANCER OF AG-4 2 (HUA2) locus in combination with PRODUCTION OF ANTHOCYANIN PIGMENT1 (PAP1) and PAP2. *New Phytologist*, **206**, 422-435.

International HapMap Consortium T. (2003) The International HapMap Project. *Nature*, **426**, 789-796.

Jansen R.C. (1993) Interval Mapping of Multiple Quantitative Trait Loci. *Genetics*, **135**, 205-211.

Jayashree B., Pradeep R., Kumar A. & Gopal B. (2008) Correlation between the sucrose synthase protein subfamilies, variations in structure and expression in stress-derived expressed sequence tag datasets. *Journal of Proteomics & Bioinformatics*, **1**, 408-423.

Jha D., Shirley N., Tester M. & Roy S.J. (2010) Variation in salinity tolerance and shoot sodium accumulation in Arabidopsis ecotypes linked to differences in the natural expression levels of transporters involved in sodium transport. *Plant, Cell & Environment*, **33**, 793-804.

Jiang C., Gao X., Liao L., Harberd N.P. & Fu X. (2007) Phosphate Starvation Root Architecture and Anthocyanin Accumulation Responses Are Modulated by the Gibberellin-DELLA Signaling Pathway in Arabidopsis. *Plant Physiology*, **145**, 1460-1470.

Jiménez-Gómez J.M. (2011) Next generation quantitative genetics in plants. *Frontiers in Plant Science*, **2**.

Jiménez-López J.C., Rodríguez-García M.I. & Alché J. (2011) Systematic and phylogenetic analysis of the Ole e 1 pollen protein family members in plants. *Systems and computational biology-bioinformatics and computational modeling, InTech (Ed.)*, 245-260.

Joo J., Choi H., Lee Y., Kim Y.-K. & Song S. (2013) A transcriptional repressor of the ERF family confers drought tolerance to rice and regulates genes preferentially located on chromosome 11. *Planta*, **238**, 155-170.

Joosen R.V.L., Kodde J., Willems L.A.J., Ligterink W., van der Plas L.H.W. & Hilhorst H.W.M. (2010) germinator: a software package for high-throughput scoring and curve fitting of Arabidopsis seed germination. *The Plant Journal*, **62**, 148-159.

Joosen R.V.L., Willems L.A.J., Lajo Morgan G., Kruijer W., Keurentjes J.J.B., Ligterink W. & Hilhorst H.W.M. (2013) Comparing Genome Wide Association and Linkage analysis for seed traits *Thesis, Wageningen University*, 127-138.

Juenger T.E., McKay J.K., Hausmann N., Keurentjes J.J.B., Sen S., Stowe K.A.(...)Richards J.H. (2005) Identification and characterization of QTL underlying whole-plant physiology in Arabidopsis thaliana: δ13C, stomatal conductance and transpiration efficiency. *Plant, Cell & Environment*, **28**, 697-708.

Kang H.M., Sul J.H., Service S.K., Zaitlen N.A., Kong S.-y., Freimer N.B.(...)Eskin E. (2010) Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, **42**, 348-354.

Keurentjes J.J.B. (2009) Genetical metabolomics: closing in on phenotypes. *Current Opinion in Plant Biology*, **12**, 223-230.

Keurentjes J.J.B., Fu J.Y., Terpstra I.R., Garcia J.M., van den Ackerveken G., Snoek L.B.(...)Jansen R.C. (2007) Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 1708-1713.

Khurana N., Chauhan H. & Khurana P. (2012) Expression analysis of a heat-inducible, Myo-inositol-1-phosphate synthase (MIPS) gene from wheat and the alternatively spliced variants of rice and Arabidopsis. *Plant Cell Rep.*, **31**, 237-251.

Kim S., Plagnol V., Hu T.T., Toomajian C., Clark R.M., Ossowski S.(...)Nordborg M. (2007) Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nature Genetics*, **39**, 1151-1155.

Kim S.Y., Hong C.B. & Lee I. (2001) Heat Shock Stress Causes Stage-specific Male Sterility in *Arabidopsis thaliana*. *Journal of Plant Research*, **114**, 301-307.

Kimura M., Yamamoto Y.Y., Seki M., Sakurai T., Sato M., Abe T.(...)Matsui M. (2003) Identification of Arabidopsis Genes Regulated by High Light–Stress Using cDNA Microarray. *Photochemistry and Photobiology*, **77**, 226-233.

Kleinboelting N., Huep G., Kloetgen A., Viehoever P. & Weisshaar B. (2012) GABI-Kat SimpleSearch: new features of the Arabidopsis thaliana T-DNA mutant database. *Nucleic Acids Research*, **40**, D1211-D1215.

Kloth K.J., Thoen M.P.M., Bouwmeester H.J., Jongsma M.A. & Dicke M. (2012) Association mapping of plant resistance to insects. *Trends in Plant Science*, **17**, 311-319.

Komeda Y., Takahashi T. & Hanzawa Y. (1998) Development of inflorescences in Arabidopsis thaliana. *Journal of Plant Research*, **111**, 283-288.

Koornneef M., Alonso-Blanco C. & Vreugdenhil D. (2004) Naturally occurring genetic variation in Arabidopsis thaliana. *Annual Review of Plant Biology*, **55**, 141-172.

Korte A. & Farlow A. (2013) The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*, **9**, 29.

Korte A., Vilhjalmsson B.J., Segura V., Platt A., Long Q. & Nordborg M. (2012) A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nature Genetics*, **44**, 1066-1071.

Kötting O., Santelia D., Edner C., Eicke S., Marthaler T., Gentry M.S.(...)Zeeman S.C. (2009) STARCH-EXCESS4 Is a Laforin-Like Phosphoglucan Phosphatase Required for Starch Degradation in Arabidopsis thaliana. *The Plant Cell Online*, **21**, 334-346.

Kover P.X., Valdar W., Trakalo J., Scarcelli N., Ehrenreich I.M., Purugganan M.D.(...)Mott R. (2009) A Multiparent Advanced Generation Inter-Cross to Fine-Map Quantitative Traits in *Arabidopsis thaliana*. *PLoS genetics*, **5**, e1000551.

Kowalski S., Lan T.-H., Feldmann K. & Paterson A. (1994) QTL mapping of naturally-occurring variation in flowering time of Arabidopsis thaliana. *Molecular Genetics and Genomics*, **245**, 548-555.

Kurosaki H., Yumoto S. & Matsukawa I. (2004) Correlation of cold-weather tolerance with pubescence color and flowering time in yellow hilum soybeans in Hokkaido. *Breeding Science*, **54**, 303-311.

Laubinger S., Zeller G., Henz S., Sachsenberg T., Widmer C., Naouar N.(...)Weigel D. (2008) At-TAX: a whole genome tiling array resource for developmental expression analysis and transcript identification in Arabidopsis thaliana. *Genome Biology*, **9**, R112.

Lavagi I., Estelle M., Weckwerth W., Beynon J. & Bastow R.M. (2012) From Bench to Bountiful Harvests: A Road Map for the Next Decade of Arabidopsis Research. *The Plant Cell Online*, **24**, 2240-2247.

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

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

Lempe J., Lachowiec J., Sullivan A.M. & Queitsch C. (2013) Molecular mechanisms of robustness in plants. *Current Opinion in Plant Biology*, **16**, 62-69.

Li N., Shi J., Wang X., Liu G. & Wang H. (2014a) A combined linkage and regional association mapping validation and fine mapping of two major pleiotropic QTLs for seed weight and silique length in rapeseed (Brassica napus L.). *BMC Plant Biology*, **14**, 114.

Li P., Filiault D., Box M.S., Kerdaffrec E., van Oosterhout C., Wilczek A.M.(...)Dean C. (2014b) Multiple FLC haplotypes defined by independent cis-regulatory variation underpin life history diversity in Arabidopsis thaliana. *Genes & development*, **28**, 1635-1640.

Li S.F., Milliken O.N., Pham H., Seyit R., Napoli R., Preston J.(...)Parish R.W. (2009a) The Arabidopsis MYB5 Transcription Factor Regulates Mucilage Synthesis, Seed Coat Development, and Trichome Morphogenesis. *The Plant Cell Online*, **21**, 72-89.

Li X.-F., Li Y.-J., An Y.-H., Xiong L.-J., Shao X.-H., Wang Y. & Sun Y. (2009b) AKINβ1 is Involved in the Regulation of Nitrogen Metabolism and Sugar Signaling in Arabidopsis. *Journal of Integrative Plant Biology*, **51**, 513-520.

Li Y., Cheng R., Spokas K.A., Palmer A.A. & Borevitz J.O. (2013) Genetic Variation for Life History Sensitivity to Seasonal Warming in Arabidopsis thaliana. *Genetics*, **196**, 569-577.

Li Y., Huang Y., Bergelson J., Nordborg M. & Borevitz J.O. (2010) Association mapping of local climate-sensitive quantitative trait loci in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences*, **107**, 21199-21204.

Li Y.B., Wu C.J., Jiang G.H., Wang L.Q. & He Y.Q. (2007) Dynamic analyses of rice blast resistance for the assessment of genetic and environmental effects. *Plant Breeding*, **126**, 541-547.

Li Z., Palmer W.M., Martin A.P., Wang R., Rainsford F., Jin Y.(...)Ruan Y.-L. (2012) High invertase activity in tomato reproductive organs correlates with enhanced sucrose import into, and heat tolerance of, young fruit. *Journal of Experimental Botany*, **63**, 1155-1166.

Liang Q., Lu X., Jiang L., Wang C., Fan Y. & Zhang C. (2010) EMB1211 is required for normal embryo development and influences chloroplast biogenesis in Arabidopsis. *Physiologia Plantarum*, **140**, 380-394.

Librado P. & Rozas J. (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451-1452.

Licausi F., Ohme-Takagi M. & Perata P. (2013) APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: mediators of stress responses and developmental programs. *New Phytologist*, **199**, 639-649.

Lisec J., Meyer R.C., Steinfath M., Redestig H., Becher M., Witucka-Wall H.(...)Willmitzer L. (2008) Identification of metabolic and biomass QTL in Arabidopsis thaliana in a parallel analysis of RIL and IL populations. *The Plant Journal*, **53**, 960-972.

Lothier J., Gaufichon L., Sormani R., Lemaître T., Azzopardi M., Morin H.(...)Masclaux-Daubresse C. (2011) The cytosolic glutamine synthetase GLN1;2 plays a role in the control of plant growth and ammonium homeostasis in Arabidopsis rosettes when nitrate supply is not limiting. *Journal of Experimental Botany*, **62**, 1375-1390.

Loudet O., Chaillou S., Merigout P., Talbotec J. & Daniel-Vedele F. (2003) Quantitative Trait Loci Analysis of Nitrogen Use Efficiency in Arabidopsis. *Plant Physiology*, **131**, 345-358.

Loukehaich R., Wang T., Ouyang B., Ziaf K., Li H., Zhang J.(...)Ye Z. (2012) SpUSP, an annexin-interacting universal stress protein, enhances drought tolerance in tomato. *Journal of Experimental Botany*, **63**, 5593-5606.

Lucas J.R., Courtney S., Hassfurder M., Dhingra S., Bryant A. & Shaw S.L. (2011) Microtubule-Associated Proteins MAP65-1 and MAP65-2 Positively Regulate Axial Cell Growth in Etiolated Arabidopsis Hypocotyls. *The Plant Cell Online*, **23**, 1889-1903.

Lucas J.R. & Shaw S.L. (2012) MAP65-1 and MAP65-2 promote cell proliferation and axial growth in Arabidopsis roots. *The Plant Journal*, **71**, 454-463.

Lucas M., Ehlers J., Huynh B.-L., Diop N.-N., Roberts P. & Close T. (2013) Markers for breeding heat-tolerant cowpea. *Molecular breeding*, **31**, 529-536.

Ma C.-X., Casella G. & Wu R. (2002) Functional Mapping of Quantitative Trait Loci Underlying the Character Process: A Theoretical Framework. *Genetics*, **161**, 1751-1762.

Ma C.-X., Wu R. & Casella G. (2004) FunMap: functional mapping of complex traits. *Bioinformatics*, **20**, 1808-1811.

Ma J., Jiang Q.-T., Wei L., Yang Q., Zhang X.-W., Peng Y.-Y.(...)Zheng Y.-L. (2014a) Conserved structure and varied expression reveal key roles of phosphoglucan phosphatase gene starch excess 4 in barley. *Planta*, **240**, 1179-1190.

Ma X., Sukiran N., Ma H. & Su Z. (2014b) Moderate drought causes dramatic floral transcriptomic reprogramming to ensure successful reproductive development in Arabidopsis. *BMC Plant Biology*, **14**, 164.

Mackay T.F.C. (2004) Complementing complexity. *Nature Genetics*, **36**, 1145-1147.

Mackay T.F.C., Stone E.A. & Ayroles J.F. (2009) The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics*, **10**, 565-577.

Malosetti M., Ribaut J.-M. & van Eeuwijk F.A. (2013) The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. *Frontiers in Physiology*, **4**.

Malosetti M., Visser R.G.F., Celis-Gamboa C. & Eeuwijk F.A. (2006) QTL methodology for response curves on the basis of non-linear mixed models, with an illustration to senescence in potato. *Theor Appl Genet.*, **113**, 288-300.

Mansfield S.G. & Briarty L.G. (1991) Early embryogenesis in Arabidopsis thaliana. II. The developing embryo. *Canadian Journal of Botany*, **69**, 461-476.

Marcelis L.F.M., Heuvelink E., Baan Hofman-Eijer L.R., Den Bakker J. & Xue L.B. (2004) Flower and fruit abortion in sweet pepper in relation to source and sink strength. *Journal of Experimental Botany*, **55**, 2261-2268.

Maruyama K., Takeda M., Kidokoro S., Yamada K., Sakuma Y., Urano K.(...)Yamaguchi-Shinozaki K. (2009) Metabolic Pathways Involved in Cold Acclimation Identified by Integrated Analysis of Metabolites and Transcripts Regulated by DREB1A and DREB2A. *Plant Physiology*, **150**, 1972-1980.

Massonnet C., Dauzat M., Bédiée A., Vile D. & Granier C. (2015) Individual Leaf Area of Early Flowering Arabidopsis Genotypes Is More Affected by Drought than Late Flowering Ones: A Multi-Scale Analysis in 35 Genetically Modified Lines. *American Journal of Plant Sciences*, **6**, 955-971.

Massonnet C., Vile D., Fabre J., Hannah M.A., Caldana C., Lisec J.(...)Granier C. (2010) Probing the Reproducibility of Leaf Growth and Molecular Phenotypes: A Comparison of Three Arabidopsis Accessions Cultivated in Ten Laboratories. *Plant Physiology*, **152**, 2142-2157.

Masuda H.P., Cabral L.M., De Veylder L., Tanurdzic M., de Almeida Engler J., Geelen D.(...)Hemerly A.S. (2008) *ABAP1 is a novel plant Armadillo BTB protein involved in DNA replication and transcription* (vol. 27).

Matías-Hernández L., Aguilar-Jaramillo A.E., Marín-González E., Suárez-López P. & Pelaz S. (2014) RAV genes: regulation of floral induction and beyond. *Annals of Botany*, **114**, 1459-1470.

McKay J.K., Richards J.H. & Mitchell-Olds T. (2003) Genetics of drought adaptation in Arabidopsis thaliana: I. Pleiotropy contributes to genetic correlations among ecological traits. *Molecular Ecology*, **12**, 1137-1151.

McKay J.K., Richards J.H., Nemali K.S., Sen S., Mitchell-Olds T., Boles S.(...)Juenger T.E. (2008) Genetics of drought adaptation in Arabidopsis Thaliana II. QTL analysis of a new mapping population, Kas-1 x Tsu-1. *Evolution*, **62**, 3014-3026.

McMullen M.D., Kresovich S., Villeda H.S., Bradbury P., Li H., Sun Q.(...)Buckler E.S. (2009) Genetic Properties of the Maize Nested Association Mapping Population. *Science*, **325**, 737-740.

Meijon M., Satbhai S.B., Tsuchimatsu T. & Busch W. (2014) Genome-wide association study using cellular traits identifies a new regulator of root development in Arabidopsis. *Nature Genetics*, **46**, 77-81.

Melzer S., Lens F., Gennen J., Vanneste S., Rohde A. & Beeckman T. (2008) Flowering-time genes modulate meristem determinacy and growth form in Arabidopsis thaliana. *Nature Genetics*, **40**, 1489-1492.

Metzker M.L. (2010) Sequencing technologies - the next generation. *Nature Reviews Genetics*, **11**, 31-46.

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

Michaels S.D. & Amasino R.M. (2001) Loss of FLOWERING LOCUS C Activity Eliminates the Late-Flowering Phenotype of FRIGIDA and Autonomous Pathway Mutations but Not Responsiveness to Vernalization. *The Plant Cell Online*, **13**, 935-941.

Mickelbart M.V., Hasegawa P.M. & Bailey-Serres J. (2015) Genetic mechanisms of abiotic stress tolerance that translate to crop yield stability. *Nature Reviews Genetics*, **advance online publication**.

Mitchell-Olds T. (1996) Genetic Constraints on Life-History Evolution: Quantitative-Trait Loci Influencing Growth and Flowering in Arabidopsis thaliana. *Evolution*, **50**, 140-145.

Mitchell-Olds T. (2010) Complex-trait analysis in plants. *Genome Biology*, **11**, 113.

Mitchell-Olds T. (2013) Selection on QTL and complex traits in complex environments. *Molecular Ecology*, **22**, 3427-3429.

Mizoi J., Shinozaki K. & Yamaguchi-Shinozaki K. (2012) AP2/ERF family transcription factors in plant abiotic stress responses. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, **1819**, 86-96.

Mohammadkhani N. & Heidari R. (2008) Drought-induced accumulation of soluble sugars and proline in two maize varieties. *World Applied Sciences Journal*, **3**, 448-453.

Mollet J.-C., Leroux C., Dardelle F. & Lehner A. (2013) Cell Wall Composition, Biosynthesis and Remodeling during Pollen Tube Growth. *Plants*, **2**, 107-147.

Moore C.R., Johnson L.S., Kwak I.-Y., Livny M., Broman K.W. & Spalding E.P. (2013) High-Throughput Computer Vision Introduces the Time Axis to a Quantitative Trait Map of a Plant Growth Response. *Genetics*, **195**, 1077-1086.

Morris E.R., Chevalier D. & Walker J.C. (2006) DAWDLE, a Forkhead-Associated Domain Gene, Regulates Multiple Aspects of Plant Development. *Plant Physiology*, **141**, 932-941.

Morrison G.D. & Linder C.R. (2014) Association Mapping of Germination Traits in Arabidopsis thaliana Under Light and Nutrient Treatments: Searching for G×E Effects. *G3: Genes|Genomes|Genetics*, **4**, 1465-1478.

Mu J., Tan H., Hong S., Liang Y. & Zuo J. (2013) Arabidopsis Transcription Factor Genes NF-YA1, 5, 6, and 9 Play Redundant Roles in Male Gametogenesis, Embryogenesis, and Seed Development. *Molecular Plant*, **6**, 188-201.

Nakashima K., Takasaki H., Mizoi J., Shinozaki K. & Yamaguchi-Shinozaki K. (2012) NAC transcription factors in plant abiotic stress responses. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, **1819**, 97-103.

Nakashima K., Yamaguchi-Shinozaki K. & Shinozaki K. (2014) The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses including drought, cold and heat. *Frontiers in Plant Science*, **5**.

Neff M.M. & Chory J. (1998) Genetic Interactions between Phytochrome A, Phytochrome B, and Cryptochrome 1 during Arabidopsis Development. *Plant Physiology*, **118**, 27-35.

Nemhauser J.L., Hong F. & Chory J. (2006) Different Plant Hormones Regulate Similar Processes through Largely Nonoverlapping Transcriptional Responses. *Cell*, **126**, 467-475.

Nordborg M., Borevitz J.O., Bergelson J., Berry C.C., Chory J., Hagenblad J.(...)Weigel D. (2002) The extent of linkage disequilibrium in *Arabidopsis thaliana*. *Nature Genetics*, **30**, 190-193.

Normanly J., Yazdanbakhsh N. & Fisahn J. (2012) High-Throughput Phenotyping of Root Growth Dynamics. In: *High-Throughput Phenotyping in Plants*, pp. 21-40. Humana Press.

Norton G.J., Douglas A., Lahner B., Yakubova E., Guerinot M.L., Pinson S.R.M.(...)Price A.H. (2014) Genome Wide Association Mapping of Grain Arsenic, Copper, Molybdenum and Zinc in Rice (*Oryza sativa* L.) Grown at Four International Field Sites. *PLoS ONE*, **9**, e89685.

Olsen K.M., Halldorsdottir S.S., Stinchcombe J.R., Weinig C., Schmitt J. & Purugganan M.D. (2004) Linkage disequilibrium mapping of Arabidopsis CRY2 flowering time alleles. *Genetics*, **167**, 1361-1369.

Oñate-Sánchez L. & Vicente-Carbajosa J. (2008) DNA-free RNA isolation protocols for Arabidopsis thaliana, including seeds and siliques. *BMC Research Notes*, **1**, 1-7.

Ooijen J.W. (1992) Accuracy of mapping quantitative trait loci in autogamous species. *TAG Theoretical and Applied Genetics*, **84**, 803-811.

Østergaard L., Lauvergeat V., Næsted H., Mattsson O. & Mundy J. (2001) Two differentially regulated Arabidopsis genes define a new branch of the DFR superfamily. *Plant Science*, **160**, 463-472.

Pang P.P. & Meyerowitz E.M. (1987) Arabidopsis Thaliana: A Model System for Plant Molecular Biology. *Nature Biotechnology*, **5**, 1177-1181.

Park W., Han K.-H. & Ahn S.-J. (2012) Differences in root-to-shoot Cd and Zn translocation and by HMA3 and 4 could influence chlorophyll and anthocyanin content in Arabidopsis Ws and Col-0 ecotypes under excess metals. *Soil Science and Plant Nutrition*, **58**, 334-348.

Patel D. & Franklin K.A. (2009) Temperature-regulation of plant architecture. *psb*, **4**, 577-579.

Pennisi E. (2008) The Blue Revolution, Drop by Drop, Gene by Gene. *Science*, **320**, 171-173.

Petroni K. & Tonelli C. (2011) Recent advances on the regulation of anthocyanin synthesis in reproductive organs. *Plant Science*, **181**, 219-229.

Platt A., Horton M., Huang Y.S., Li Y., Anastasio A.E., Mulyati N.W.(...)Borevitz J.O. (2010) The Scale of Population Structure in *Arabidopsis thaliana*. *PLoS Genet*, **6**, e1000843.

Polanski K., Rhodes J., Hill C., Zhang P., Jenkins D.J., Kiddle S.J.(...)Denby K.J. (2014) Wigwams: identifying gene modules co-regulated across multiple biological conditions. *Bioinformatics*, **30**, 962-970.

Poxleitner M., Rogers S.W., Lacey Samuels A., Browse J. & Rogers J.C. (2006) A role for caleosin in degradation of oil-body storage lipid during seed germination. *The Plant Journal*, **47**, 917-933.

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

Quesada V.c., Garcí a-Martí nez S., Piqueras P., Ponce M.a.R. & Micol J.L. (2002) Genetic Architecture of NaCl Tolerance in Arabidopsis. *Plant Physiology*, **130**, 951-963.

R Development Core Team (2012) R: A language and environment for statistical computing. *R Foundation for Stasticial computing, Vienna, Austria*.

Radauer C. & Breiteneder H. (2006) Pollen allergens are restricted to few protein families and show distinct patterns of species distribution. *Journal of Allergy and Clinical Immunology*, **117**, 141-147.

Ral J.-P., Bowerman A.F., Li Z., Sirault X., Furbank R., Pritchard J.R.(...)Morell M.K. (2012) Down-regulation of Glucan, Water-Dikinase activity in wheat endosperm increases vegetative biomass and yield. *Plant Biotechnology Journal*, **10**, 871-882.

Rédei G. (1962) Single locus heterosis. *Zeitschrift für Vererbungslehre*, **93**, 164-170.

Redillas M.F.R., Park S.-H., Lee J., Kim Y., Jeong J., Jung H.(...)Kim J.-K. (2012) Accumulation of trehalose increases soluble sugar contents in rice plants conferring tolerance to drought and salt stress. *Plant Biotechnology reports*, **6**, 89-96.

Reymond M., Muller B. & Tardieu F. (2004) Dealing with the genotype×environment interaction via a modelling approach: a comparison of QTLs of maize leaf length or width with QTLs of model parameters. *Journal of Experimental Botany*, **55**, 2461-2472.

Richard C., Hickey L., Fletcher S., Jennings R., Chenu K. & Christopher J. (2015) High-throughput phenotyping of seminal root traits in wheat. *Plant Methods*, **11**, 13.

Risch N. & Merikangas K. (1996) The future of genetic studies of complex human diseases. *Science-AAAS-Weekly Paper Edition*, **273**, 1516-1517.

Rizhsky L., Liang H., Shuman J., Shulaev V., Davletova S. & Mittler R. (2004) When Defense Pathways Collide. The Response of Arabidopsis to a Combination of Drought and Heat Stress. *Plant Physiology*, **134**, 1683-1696.

Rong W., Qi L., Wang A., Ye X., Du L., Liang H.(...)Zhang Z. (2014) The ERF transcription factor TaERF3 promotes tolerance to salt and drought stresses in wheat. *Plant Biotechnology Journal*, **12**, 468-479.

Ruijter J.M., Ramakers C., Hoogaars W.M.H., Karlen Y., Bakker O., van den Hoff M.J.B. & Moorman A.F.M. (2009) Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Research*, **37**, e45.

Sadras V.O., Reynolds M.P., de la Vega A.J., Petrie P.R. & Robinson R. (2009) Phenotypic plasticity of yield and phenology in wheat, sunflower and grapevine. *Field Crops Research*, **110**, 242-250.

Saini H.S. (1997) Effects of water stress on male gametophyte development in plants. *Sexual Plant Reproduction*, **10**, 67-73.

Saini H.S. & Aspinall D. (1982) Abnormal Sporogenesis in Wheat (Triticum aestivum L.) Induced by Short Periods of High Temperature. *Annals of Botany*, **49**, 835-846.

Sakata T., Oshino T., Miura S., Tomabechi M., Tsunaga Y., Higashitani N.(...)Higashitani A. (2010) Auxins reverse plant male sterility caused by high temperatures. *Proceedings of the National Academy of Sciences*, **107**, 8569-8574.

Salehi H., Ransom C.B., Oraby H.F., Seddighi Z. & Sticklen M.B. (2005) Delay in flowering and increase in biomass of transgenic tobacco expressing the Arabidopsis floral repressor gene FLOWERING LOCUS C. *Journal of Plant Physiology*, **162**, 711-717.

Samizadeh H., Yazdi-Samadi B., Bihamta M., Taleii A. & Stringam G. (2010) Study of pod length trait in doubled haploid brassica napus population by molecular markers. *Journal of Agricultural Science and Technology*, **9**, 129-136.

Sanders G.J. & Arndt S.K. (2012) Osmotic adjustment under drought conditions. In: *Plant Responses to Drought Stress*, pp. 199-229. Springer.

Sanders P.M., Bui A.Q., Weterings K., McIntire K.N., Hsu Y.C., Lee P.Y.(...)Goldberg R.B. (1999) Anther developmental defects in Arabidopsis thaliana male-sterile mutants. *Sexual Plant Reproduction*, **11**, 297-322.

Sato S., Kamiyama M., Iwata T., Makita N., Furukawa H. & Ikeda H. (2006) Moderate Increase of Mean Daily Temperature Adversely Affects Fruit Set of *Lycopersicon esculentum* by Disrupting Specific Physiological Processes in Male Reproductive Development. *Annals of Botany*, **97**, 731-738.

Sato S., Peet M.M. & Thomas J.F. (2002) Determining critical pre- and post-anthesis periods and physiological processes in Lycopersicon esculentum Mill. exposed to moderately elevated temperatures. *Journal of Experimental Botany*, **53**, 1187-1195.

Schachtman D.P. & Goodger J.Q.D. (2008) Chemical root to shoot signaling under drought. *Trends in Plant Science*, **13**, 281-287.

Schmalenbach I., Zhang L., Reymond M. & Jiménez-Gómez J.M. (2014) The relationship between flowering time and growth responses to drought in the Arabidopsis Landsberg erecta x Antwerp-1 population. *Frontiers in Plant Science*, **5**, 609.

Schmid M., Davison T.S., Henz S.R., Pape U.J., Demar M., Vingron M.(...)Lohmann J.U. (2005) A gene expression map of Arabidopsis thaliana development. *Nature Genetics*, **37**, 501-506.

Schmittgen T.D. & Livak K.J. (2008) Analyzing real-time PCR data by the comparative CT method. *Nature Protocols*, **3**, 1101-1108.

Schmuths H., Bachmann K., Weber W.E., Horres R. & Hoffman M.H. (2006) Effects of Preconditioning and Temperature During Germination of 73 Natural Accessions of Arabidopsis thaliana. *Annals of Botany*, **97**, 623-634.

Schneider C.A., Rasband W.S. & Eliceiri K.W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, **9**, 671-675.

Schneitz K., Hülskamp M. & Pruitt R.E. (1995) Wild-type ovule development in Arabidopsis thaliana: a light microscope study of cleared whole-mount tissue. *The Plant Journal*, **7**, 731-749.

Schroeder J.I., Kwak J.M. & Allen G.J. (2001) Guard cell abscisic acid signalling and engineering drought hardiness in plants. *Nature*, **410**, 327-330.

Schulz E., Tohge T., Zuther E., Fernie A.R. & Hincha D.K. (2015) Natural variation in flavonol and anthocyanin metabolism during cold acclimation in Arabidopsis thaliana accessions. *Plant, Cell & Environment*, n/a-n/a.

Segura V., Vilhjalmsson B.J., Platt A., Korte A., Seren U., Long Q. & Nordborg M. (2012) An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature Genetics*, **44**, 825-830.

Semagn K., Bjørnstad Å. & Ndjiondjop M.N. (2006) An overview of molecular marker methods for plants. *African Journal of Biotechnology*, **5**, 2540-2568.

Seren Ü., Vilhjálmsson B.J., Horton M.W., Meng D., Forai P., Huang Y.S.(...)Nordborg M. (2012) GWAPP: A Web Application for Genome-Wide Association Mapping in Arabidopsis. *The Plant Cell Online*, **24**, 4793-4805.

Sharma M., Singh A., Shankar A., Pandey A., Baranwal V., Kapoor S.(...)Pandey G.K. (2014) Comprehensive Expression Analysis of Rice Armadillo Gene Family During Abiotic Stress and Development. *DNA Research*, **21**, 267-283.

Sharova E.I. (2007) Expansins: Proteins involved in cell wall softening during plant growth and morphogenesis. *Russian Journal of Plant Physiology*, **54**, 713-727.

Shindo C., Aranzana M.J., Lister C., Baxter C., Nicholls C., Nordborg M. & Dean C. (2005) Role of FRIGIDA and FLOWERING LOCUS C in Determining Variation in Flowering Time of Arabidopsis. *Plant Physiology*, **138**, 1163-1173.

Shindo C., Lister C., Crevillen P., Nordborg M. & Dean C. (2006) Variation in the epigenetic silencing of FLC contributesto natural variationin Arabidopsis vernalization response. *Genes & development*, **20**, 3079-3083.

Silva-Correia J., Freitas S., Tavares R., Lino-Neto T. & Azevedo H. (2014) Phenotypic analysis of the Arabidopsis heat stress response during germination and early seedling development. *Plant Methods*, **10**, 7.

Simpson G.G., Dijkwel P.P., Quesada V., Henderson I. & Dean C. (2003) FY Is an RNA 3′ End-Processing Factor that Interacts with FCA to Control the Arabidopsis Floral Transition. *Cell*, **113**, 777-787.

Skirycz A., Claeys H., De Bodt S., Oikawa A., Shinoda S., Andriankaja M.(...)Inzé D. (2011a) Pause-and-Stop: The Effects of Osmotic Stress on Cell Proliferation during Early Leaf Development in Arabidopsis and a Role for Ethylene Signaling in Cell Cycle Arrest. *The Plant Cell Online*, **23**, 1876-1888.

Skirycz A. & Inzé D. (2010) More from less: plant growth under limited water. *Current Opinion in Biotechnology*, **21**, 197-203.

Skirycz A., Vandenbroucke K., Clauw P., Maleux K., De Meyer B., Dhondt S.(...)Inze D. (2011b) Survival and growth of Arabidopsis plants given limited water are not equal. *Nature Biotechnology*, **29**, 212-214.

Smith A.M. & Stitt M. (2007) Coordination of carbon supply and plant growth. *Plant, Cell & Environment*, **30**, 1126-1149.

Smyth D.R., Bowman J.L. & Meyerowitz E.M. (1990) Early flower development in Arabidopsis. *The Plant Cell Online*, **2**, 755-767.

Solfanelli C., Poggi A., Loreti E., Alpi A. & Perata P. (2006) Sucrose-Specific Induction of the Anthocyanin Biosynthetic Pathway in Arabidopsis. *Plant Physiology*, **140**, 637-646.

Songnuan W. (2013) Wind-pollination and the roles of pollen allergenic proteins. *Asian Pac J Allergy Immunol*, **31**, 261-270.

Sperdouli I. & Moustakas M. (2012) Interaction of proline, sugars, and anthocyanins during photosynthetic acclimation of Arabidopsis thaliana to drought stress. *Journal of Plant Physiology*, **169**, 577-585.

Stamatiou G., Vidaurre D.P., Shim I., Tang X., Moeder W., Bonetta D. & McCourt P. (2013) Forward Genetic Screening for the Improved Production of Fermentable Sugars from Plant Biomass. *PLoS ONE*, **8**, e55616.

Stinchcombe J.R., Weinig C., Ungerer M., Olsen K.M., Mays C., Halldorsdottir S.S.(...)Schmitt J. (2004) A latitudinal cline in flowering time in Arabidopsis thaliana modulated by the flowering time gene FRIGIDA. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 4712-4717.

Stracke R., Werber M. & Weisshaar B. (2001) The R2R3-MYB gene family in Arabidopsis thaliana. *Current Opinion in Plant Biology*, **4**, 447-456.

Streb S. & Zeeman S.C. (2012) Starch Metabolism in Arabidopsis. *The Arabidopsis Book*, e0160.

Sugiyama T., Iwahori S. & Takahashi K. (1965) Effect of high temperature on fruit setting of tomato under cover. *Acta Horticulturae*, **4**, 63-69.

Sulpice R., Flis A., Ivakov A.A., Apelt F., Krohn N., Encke B.(...)Stitt M. (2014) Arabidopsis Coordinates the Diurnal Regulation of Carbon Allocation and Growth across a Wide Range of Photoperiods. *Molecular Plant*, **7**, 137-155.

Sulpice R., Pyl E.-T., Ishihara H., Trenkamp S., Steinfath M., Witucka-Wall H.(...)Stitt M. (2009) Starch as a major integrator in the regulation of plant growth. *Proceedings of the National Academy of Sciences*, **106**, 10348-10353.

Suzuki K., Tsukaguchi T., Takeda H. & Egawa Y. (2001) Decrease of Pollen Stainability of Green Bean at High Temperatures and Relationship to Heat Tolerance. *Journal of the American Society for Horticultural Science*, **126**, 571-574.

Swarup K., Alonso-Blanco C., Lynn J.R., Michaels S.D., Amasino R.M., Koornneef M. & Millar A.J. (1999) Natural allelic variation identifies new genes in the Arabidopsis circadian system. *The Plant Journal*, **20**, 67-77.

Tajima F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585-595.

Teng S., Keurentjes J., Bentsink L., Koornneef M. & Smeekens S. (2005) Sucrose-Specific Induction of Anthocyanin Biosynthesis in Arabidopsis Requires the MYB75/PAP1 Gene. *Plant Physiology*, **139**, 1840-1852.

Tessmer O.L., Jiao Y., Cruz J.A., Kramer D.M. & Chen J. (2013) Functional approach to high-throughput plant growth analysis. *BMC systems biology*, **7**, S17.

Tester M. & Langridge P. (2010) Breeding Technologies to Increase Crop Production in a Changing World. *Science*, **327**, 818-822.

Thirunavukkarasu N., Hossain F., Arora K., Sharma R., Shiriga K., Mittal S.(...)Gupta H. (2014) Functional mechanisms of drought tolerance in subtropical maize (Zea mays L.) identified using genome-wide association mapping. *Bmc Genomics*, **15**, 1182.

Tian D., Traw M.B., Chen J.Q., Kreitman M. & Bergelson J. (2003) Fitness costs of R-gene-mediated resistance in Arabidopsis thaliana. *Nature*, **423**, 74-77.

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

Tisné S., Serrand Y., Bach L., Gilbault E., Ben Ameur R., Balasse H.(...)Loudet O. (2013) Phenoscope: an automated large-scale phenotyping platform offering high spatial homogeneity. *The Plant Journal*, **74**, 534-544.

Todesco M., Balasubramanian S., Hu T.T., Traw M.B., Horton M., Epple P.(...)Lanz C. (2010) Natural allelic variation underlying a major fitness trade-off in Arabidopsis thaliana. *Nature*, **465**, 632-636.

Tuberosa R. (2012) Phenotyping for drought tolerance of crops in the genomics era. *Frontiers in Physiology*, **3**.

Turnbull C. (2011) Long-distance regulation of flowering time. *Journal of Experimental Botany*.

Valdivia E.R., Wu Y., Li L.-C., Cosgrove D.J. & Stephenson A.G. (2007) A Group-1 Grass Pollen Allergen Influences the Outcome of Pollen Competition in Maize. *PLoS ONE*, **2**, e154.

Vallejo A.J., Yanovsky M.J. & Botto J.F. (2010) Germination variation in Arabidopsis thaliana accessions under moderate osmotic and salt stresses. *Annals of Botany*, **106**, 833-842.

van Rooijen R., Aarts M.G.M. & Harbinson J. (2015) Natural genetic variation for acclimation of photosynthetic light use efficiency to growth irradiance in Arabidopsis thaliana. *Plant Physiology*.

Vanderauwera S., Zimmermann P., Rombauts S., Vandenabeele S., Langebartels C., Gruissem W.(...) Van Breusegem F. (2005) Genome-Wide Analysis of Hydrogen Peroxide-Regulated Gene Expression in Arabidopsis Reveals a High Light-Induced Transcriptional Cluster Involved in Anthocyanin Biosynthesis. *Plant Physiology*, **139**, 806-821.

Verhulst P.F. (1838) Notice sur la loi que la population suit dans son acroissement. *Correspondence Mathématique et Physique*, **10**, 113–121.

Verslues P.E. & Juenger T.E. (2011) Drought, metabolites, and Arabidopsis natural variation: a promising combination for understanding adaptation to water-limited environments. *Current Opinion in Plant Biology*, **14**, 240-245.

Verslues P.E., Lasky J.R., Juenger T.E., Liu T.-W. & Kumar M.N. (2014) Genome-Wide Association Mapping Combined with Reverse Genetics Identifies New Effectors of Low Water Potential-Induced Proline Accumulation in Arabidopsis. *Plant Physiology*, **164**, 144-159.

Veselov D.S., Mustafina A.R., Sabirjanova I.B., Akhiyarova G.R., Dedov A.V., Veselov S.U. & Kudoyarova G.R. (2002) Effect of PEG-treatment on the leaf growth response and auxin content in shoots of wheat seedlings. *Plant Growth Regulation*, **38**, 191-194.

Via S., Gomulkiewicz R., De Jong G., Scheiner S.M., Schlichting C.D. & Van Tienderen P.H. (1995) Adaptive phenotypic plasticity: consensus and controversy. *Trends in Ecology & Evolution*, **10**, 212-217.

Vile D., Pervent M., Belluau M., Vasseur F., Bresson J., Muller B.(...)Simonneau T. (2012) Arabidopsis growth under prolonged high temperature and water deficit: independent or interactive effects? *Plant, Cell & Environment*, **35**, 702-718.

Visscher Peter M., Brown Matthew A., McCarthy Mark I. & Yang J. (2012) Five Years of GWAS Discovery. *The American Journal of Human Genetics*, **90**, 7-24.

Vogel J.P., Schuerman P., Woeste K., Brandstatter I. & Kieber J.J. (1998) Isolation and Characterization of Arabidopsis Mutants Defective in the Induction of Ethylene Biosynthesis by Cytokinin. *Genetics*, **149**, 417-427.

VSN International (2013) GenStat for Windows,16th Edition. *VSN International, Hemel Hempstead, UK*, Web page: GenStat.co.uk.

Wahid A., Gelani S., Ashraf M. & Foolad M.R. (2007) Heat tolerance in plants: An overview. *Environmental and Experimental Botany*, **61**, 199-223.

Wang X., Feng S., Nakayama N., Crosby W.L., Irish V., Deng X.W. & Wei N. (2003) The COP9 Signalosome Interacts with SCFUFO and Participates in Arabidopsis Flower Development. *The Plant Cell Online*, **15**, 1071-1082.

Wang Y., Yang Q. & Wang Z. (2015) The Evolution of Nanopore Sequencing. *Frontiers in Genetics*, **5**.

Warner R.M. & Erwin J.E. (2005) Naturally occurring variation in high temperature induced floral bud abortion across Arabidopsis thaliana accessions. *Plant, Cell & Environment*, **28**, 1255-1266.

Waters E.R. (2012) The evolution, function, structure, and expression of the plant sHSPs. *Journal of Experimental Botany*, **64**, 391-403.

Watson J.D. & Crick F.H. (1953) Molecular structure of nucleic acids. *Nature*, **171**, 737-738.

Weigel D. (2012) Natural Variation in Arabidopsis: From Molecular Genetics to Ecological Genomics. *Plant Physiology*, **158**, 2-22.

Weise S.E., Aung K., Jarou Z.J., Mehrshahi P., Li Z., Hardy A.C.(...)Sharkey T.D. (2012) Engineering starch accumulation by manipulation of phosphate metabolism of starch. *Plant Biotechnology Journal*, **10**, 545-554.

Westoby M., Falster D.S., Moles A.T., Vesk P.A. & Wright I.J. (2002) Plant Ecological Strategies: Some Leading Dimensions of Variation between Species. *Annual Review of Ecology and Systematics*, **33**, 125-159.

Winefield C., Davies K., Gould K. & Steyn W.J. (2009) Prevalence and Functions of Anthocyanins in Fruits. In: *Anthocyanins*, pp. 86-105. Springer New York.

Wingler A., Purdy S.J., Edwards S.-A., Chardon F. & Masclaux-Daubresse C. (2010) QTL analysis for sugar-regulated leaf senescence supports flowering-dependent and -independent senescence pathways. *New Phytologist*, **185**, 420-433.

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

Winsor C.P. (1932) The Gompertz Curve as a Growth Curve. *Proceedings of the National Academy of Sciences*, **18**, 1-8.

Winter D., Vinegar B., Nahal H., Ammar R., Wilson G.V. & Provart N.J. (2007) An "Electronic Fluorescent Pictograph" Browser for Exploring and Analyzing Large-Scale Biological Data Sets. *PLoS ONE*, **2**, e718.

Wu G. & Poethig R.S. (2006a) Temporal regulation of shoot development in Arabidopsis thaliana by miR156 and its target SPL3. *Development*, **133**, 3539-3547.

Wu R. & Lin M. (2006b) Functional mapping — how to map and study the genetic architecture of dynamic complex traits. *Nature Reviews Genetics*, **7**, 229-237.

Wurschum T., Liu W., Busemeyer L., Tucker M., Reif J., Weissmann E.(...)Maurer H. (2014) Mapping dynamic QTL for plant height in triticale. *BMC Genetics*, **15**, 59.

Yamaguchi A. & Abe M. (2012) Regulation of reproductive development by non-coding RNA in Arabidopsis: to flower or not to flower. *Journal of Plant Research*, **125**, 693-704.

Yamashino T., Matsushika A., Fujimori T., Sato S., Kato T., Tabata S. & Mizuno T. (2003) A Link between Circadian-Controlled bHLH Factors and the APRR1/TOC1 Quintet in Arabidopsis thaliana. *Plant and Cell Physiology*, **44**, 619-629.

Yang R., Li J., Wang X. & Zhou X. (2011) Bayesian functional mapping of dynamic quantitative traits. *Theor Appl Genet.*, **123**, 483-492.

Ye C., Argayoso M.A., Redoña E.D., Sierra S.N., Laza M.A., Dilla C.J.(...)Hernandez J.E. (2012) Mapping QTL for heat tolerance at flowering stage in rice using SNP markers. *Plant Breeding*, **131**, 33-41.

Yi X., Du Z. & Su Z. (2013) PlantGSEA: a gene set enrichment analysis toolkit for plant community. *Nucleic Acids Research*, **41**, W98-W103.

Yoshizumi T., Tsumoto Y., Takiguchi T., Nagata N., Yamamoto Y.Y., Kawashima M.(...)Matsui M. (2006) INCREASED LEVEL OF POLYPLOIDY1, a Conserved Repressor of CYCLINA2 Transcription, Controls Endoreduplication in Arabidopsis. *The Plant Cell Online*, **18**, 2452-2468.

Young L.W., Wilen R.W. & Bonham-Smith P.C. (2004) High temperature stress of Brassica napus during flowering reduces micro- and megagametophyte fertility, induces fruit abortion, and disrupts seed production. *Journal of Experimental Botany*, **55**, 485-495.

Yu B., Bi L., Zheng B., Ji L., Chevalier D., Agarwal M.(...)Chen X. (2008) The FHA domain proteins DAWDLE in Arabidopsis and SNIP1 in humans act in small RNA biogenesis. *Proceedings of the National Academy of Sciences*, **105**, 10073-10078.

Yu T.-S., Kofler H., Häusler R.E., Hille D., Flügge U.-I., Zeeman S.C.(...)Weber A. (2001) The Arabidopsis sex1 Mutant Is Defective in the R1 Protein, a General Regulator of Starch Degradation in Plants, and Not in the Chloroplast Hexose Transporter. *The Plant Cell Online*, **13**, 1907-1918.

Zeeman S.C. & Rees T.A. (1999) Changes in carbohydrate metabolism and assimilate export in starch-excess mutants of Arabidopsis. *Plant, Cell & Environment*, **22**, 1445-1453.

Zeng X.-Q., Chow W.S., Su L.-J., Peng X.-X. & Peng C.-L. (2010) Protective effect of supplemental anthocyanins on Arabidopsis leaves under high light. *Physiologia Plantarum*, **138**, 215-225.

Zhang K.-M., Yu H.-J., Shi K., Zhou Y.-H., Yu J.-Q. & Xia X.-J. (2010) Photoprotective roles of anthocyanins in Begonia semperflorens. *Plant Science*, **179**, 202-208.

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

Zhao F. & Xu S. (2012) Genotype by Environment Interaction of Quantitative Traits: A Case Study in Barley. *G3: Genes|Genomes|Genetics*, **2**, 779-788.

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

Zinn K.E., Tunc-Ozdemir M. & Harper J.F. (2010) Temperature stress and plant sexual reproduction: uncovering the weakest links. *Journal of Experimental Botany*, **61**, 1959-1968.

Summary

All plants are able to respond to changes in their environment by adjusting their morphology and metabolism, but large differences are observed in the effectiveness of these responses in the light of plant fitness. Between and within species large differences are observed in plant responses to drought, heat and other abiotic stresses. This natural variation is partly due to variation in the genetic composition of individuals. Within-species variation can be used to identify and study genes involved in the genetic regulation of plant performance.

Growth of the world population will, in the coming years, lead to an increased demand for food, feed and other natural products. In addition, extreme weather conditions with, amongst others, more and prolonged periods of drought and heat are expected to occur due to climate change. Therefore breeders are challenged to produce stress tolerant cultivars with improved yield under sub-optimal conditions. Knowledge about the mechanisms and genes that underlie tolerance to drought, heat and other abiotic stresses will ease this challenge.

The aim of this thesis was to identify and study the role of genes that are underlying natural variation in plant performance under drought, salt and heat stress. To reach this goal a genome wide association (GWA) mapping approach was taken in the model species *Arabidopsis thaliana*. A population of 350 natural accessions of Arabidopsis, genotyped with 215k SNPs, was grown under control and several stress conditions and plant performance was evaluated by phenotyping one or several plant traits per environment. Genes located in the genomic regions that were significantly associated with plant performance, were studied in more detail.

Plant performance was first evaluated upon osmotic stress (Chapter 2). This treatment resulted not only in a reduced plant size, but also caused the colour of the rosette leaves to change from green to purple-red due to anthocyanin accumulation. The latter was visually quantified and subsequent GWA mapping revealed that a large part of the variation in anthocyanin accumulation could be explained by a small genomic region on chromosome 1. The analysis of re-sequence data allowed us to associate the second most frequent allele of *MYB90* with higher anthocyanin accumulation and to identify the causal SNP. Interestingly *MYB75*, a close relative of MYB90, was not identified by GWA mapping, although causal sequence variation of this gene for anthocyanin accumualtion was identified in the Cvi x L*er* and L*er* x Eri-1 RIL populations. Re-sequence data revealed that one allele of *MYB75* was dominating the population and that the *MYB75* alleles of Cvi and *Ler* were both rare, explaining the lack of association at this locus in GWA mapping. For *MYB90*, two alleles were present in a substantial part of the population, suggesting balancing selection between them.

Next, the natural population was exposed to short-term heat stress during flowering (Chapter 3). This short-term stress has a large impact on seed set, while it hardly affects the vegetative tissues. Natural variation for tolerance against the effect of heat on seed set was evaluated by measuring the length of all siliques along the inflorescence in both heat-treated and control plants. Because the flower that opened during the treatment was tagged, we could analyse the heat response for several developmental stages separately. GWA mapping revealed that the heat response before and after anthesis involved different genes. For the heat response before anthesis strong evidence was gained that *FLC*, a flowering time regulator and *QUL2*, a gene suggested to play a role in vascular tissue development, were causal for two strong associations.

Furthermore, the impact of moderate drought on plant performance was evaluated in the plant phenotyping platform PHENOPSIS. Homogeneous drought was assured by tight regulation of climate cell conditions and the robotic weighing and watering of the pots twice a day. Because plant growth is a dynamic trait it was monitored over time by top-view imaging under both moderate drought and control conditions (Chapter 4 and 5). To characterise growth it was modelled with an exponential function. GWA mapping of temporal growth data resulted in the detection of time-dependent QTLs whereas mapping of model parameters resulted in another set of QTLs related to the entire growth period. Most of these QTLs would not have been identified if plant size had only been determined on a single day. For the QTLs detected under control conditions eight candidate genes with a growth-related mutant or overexpression phenotype were identified (Chapter 4). Genes in the support window of the drought-QTLs were prioritized based on previously reported gene expression data (Chapter 5). Additional validation experiments are needed to confirm causality of the candidate genes.

Next, to search for genes that determine plant size across many environments, biomass accumulation in the natural population was determined in 25 different environments (Chapter 6). Joint analysis of these data by multi-environment GWA mapping resulted in the detection of 106 strongly associated SNPs with significant effects in 7 to 16 environments. Several genes involved in starch metabolism, leaf size control and flowering time determination were located in close proximity of the associated SNPs. Two genes, *RPM1* and *ACD6*, were located in close proximity of SNPs with significant GxE effects. For both genes, alleles have been identified that increase resistance to bacterial infection, but that reduce biomass accumulation. The sign of the allelic effect is therefore dependent on the environmental conditions. Whole genome predictions revealed that most of the GxE interactions observed at the phenotypic level were not the consequence of strong associations with strong QxE effects, but of moderate and weak associations with weak QxE effects.

Finally, in Chapter 7 I discuss the usefulness of GWA mapping in the identification of genes underlying natural variation in plant performance under drought, heat stress and a number of other environments. Strong associations were observed for both environment-specific as well as common plant performance regulators. Some choices in phenotyping and experimental design were crucial for our success, like evaluation of plant performance over time and simplification of the quantification of the phenotype. It is suggested that follow-up work should focus on the functional characterization of the causal genes, because such analyses would be helpful to identify pathways in which the causal genes are involved and to understand why sequence variation results in changes at the phenotype level. Although translation of the findings to applications in crops is challenging, this thesis contributes to the understanding of the genetic regulation of stress response and therefore will likely contribute to the development of stress tolerant and stable yielding crops.

Samenvatting

De groei van de wereldbevolking zal in de komende jaren leiden tot een toename van de vraag naar voedsel, diervoeder en andere natuurlijke producten. Daarnaast wordt, als gevolg van klimaatverandering, een toename verwacht in extreme weersomstandigheden, zoals meer en langere droogte- en hitteperiodes. Plantenveredelaars worden uitgedaagd om nieuwe cultivars te ontwikkelen die beter presteren onder sub-optimale omstandigheden. Kennis van mechanismen en genen die bijdragen aan het verkrijgen van droogte-, hitte- of andere abiotische stresstolerantie, zullen deze uitdaging makkelijker maken.

Planten reageren op veranderingen in hun omgeving door middel van allerlei aanpassingen in, bijvoorbeeld, groeisnelheid, morfologie en metabolisme. Grote verschillen zijn echter waargenomen in de effectiviteit van deze reacties en de gevolgen daarvan voor plantfitness. Zo zijn er verschillen in de reacties van planten – zowel behorende tot verschillende soorten als tussen verschillende genotypes van dezelfde soort - op droogte, hitte en andere sub-optimale omstandigheden. Deze natuurlijke variatie is grotendeels toe te schrijven aan verschillen in de genetische samenstelling van de individuele planten. De natuurlijke variatie binnen een soort kan worden gebruikt om genen te identificeren die een rol spelen in de regulatie van de groei en stresstolerantie van planten.

Het doel van het onderzoek beschreven in dit proefschrift is de identificatie en bestudering van de rol van genen die verantwoordelijk zijn voor natuurlijke variatie in de prestaties van planten onder droogte, hitte en andere stresscondities. Een Genome Wide Association (GWA) mapping benadering in de modelplant Arabidopsis thaliana is gekozen om dit doel te bereiken. Een populatie van 350 natuurlijke ecotypes van Arabidopsis, gegenotypeerd met 215.000 merkers, werd opgekweekt onder controle-omstandigheden en onder een aantal stresscondities. De prestaties van de planten werden geëvalueerd door middel van het fenotyperen van een of meerdere groei-eigenschappen in elke testomgeving. Deze groei-eigenschappen werden gekoppeld aan genomische regio's door middel van GWA mapping. Genen binnen de genomische regio's die significant geassocieerd zijn met de gemeten plant prestaties, werden in meer detail bestudeerd.

De prestatie van planten werd als eerste geëvalueerd in een omgeving met osmotische stress (Hoofdstuk 2). Deze behandeling zorgde niet alleen voor een afname in plantgrootte, maar veroorzaakte ook een verandering in de kleur van de rozet, van groen naar paars-rood. Dit is een gevolg van anthocyaan-ophoping. GWA mapping van deze, met het oog gekwantificeerde, roodverkleuring liet zien dat een groot deel van de natuurlijke variatie in anthocyaan-ophoping kan worden verklaard door variatie in een kleine regio op chromosoom 1. Vergelijking van een aantal genoom sequenties van Arabidopsis ecotypes zorgde er voor dat we het op één-na-meest voorkomende allel van het gen MYB90

konden associëren met hogere anthocyaan-ophoping en dat we de verantwoordelijke SNP (single nucleotide polymorphism) konden identificeren. Opmerkelijk was dat MYB75, een nauwe verwant van MYB90, niet werd geïdentificeerd door de GWA mapping, hoewel in een eerdere studie de variatie in de sequentie van dit gen was aangewezen als verantwoordelijk voor verschillen in anthocyaan-ophoping in de Cvi x Ler and Ler x Eri-1 RIL populaties. Ecotype genoomsequenties lieten zien dat één allel van MYB75 de GWA populatie domineerde en dat de allelen van Cvi en Ler alleen in zeer kleine aantallen voorkwamen. Dit verklaart de afwezigheid van een associatie met dit locus in de GWA mapping. Twee allelen van MYB90 waren in een substantieel deel van de populatie aanwezig, wat suggereert dat er een balans is in de selectie van beiden allelen.

De GWA populatie werd ook blootgesteld aan kortstondige hitte op het moment van bloeien (Hoofdstuk 3). Deze kortstondige stress had een enorme impact op de zaadzetting, terwijl het de vegetatieve weefsels nauwelijks beïnvloedde. Natuurlijke variatie in de hittetolerantie van zaadzetting werd geëvalueerd door het bepalen van de lengte van alle hauwtjes aan de hoofdstengel van controle en hitte-behandelde planten. Omdat de bloem die opende gedurende de hitte-behandeling werd gemarkeerd, konden we de hitte-reactie van verschillende ontwikkelingsstadia in bloemen van dezelfde bloeiwijze afzonderlijk bestuderen. GWA mapping liet zien dat andere genen betrokken zijn bij de hitte-reactie vóór de opening van de bloem dan daarna. Sterk bewijs werd geleverd dat de hitte-reactie vóór de opening van de bloem, onder andere, wordt bepaald door FLC, een gen dat ook de bloeitijd bepaalt en QUL2, een gen dat waarschijnlijk een rol speelt in de ontwikkeling van vaatweefsel.

De invloed van milde droogte op de prestatie van planten werd geëvalueerd met behulp van het plantenfenotyperingsysteem PHENOPSIS in Montpellier. Door precieze regulatie van de omstandigheden in de klimaatcel en door het gebruik van een robot die de potten met plantjes twee keer per dag woog en daarna water gaf, werd een gecontroleerde en homogene uitdroging bereikt. Omdat plantengroei een dynamische eigenschap is werd deze, in zowel controle als droogtecondities, dagelijks geobserveerd door middel van een camera die van bovenaf foto's nam (Hoofdstuk 4 en 5). De toename in de plantgrootte werd gemodelleerd met een exponentiele functie. GWA mapping van de groeigegevens resulteerde in de detectie van tijdsafhankelijke QTLs, terwijl mapping van model parameters resulteerde in een andere set van QTLs gerelateerd aan de totale groeiperiode. De meeste van de QTLs zouden niet geïdentificeerd zijn als de plantgrootte alleen op één tijdsstip was gemeten. Aan de QTLs die gedetecteerd werden in controle condities konden acht kandidaatgenen met een mutant- of overexpressiefenotype worden toegekend (Hoofdstuk 4). Genen in de support-window van de droogte-QTLs werden geprioriteerd op basis van eerder gerapporteerde genexpressie gegevens. Extra bevestigingsexperimenten zijn nodig om de causaliteit van de kandidaatgenen te bewijzen.

Vervolgens deden we onderzoek naar genen die de plantgrootte bepalen onder een reeks van verschillende omstandigheden, De biomassatoename in de GWA populatie werd bepaald in 25 verschillende omgevingen (Hoofdstuk 6), die onder andere verschillenden in daglengte, temperatuur en lichtintensiteit. In de helft van de omgevingen werden de planten ook blootgesteld aan verschillende abiotische stressen. Gezamenlijke analyse van deze gegevens door middel van multi-environment GWA mapping resulteerde in de detectie van 106 sterke associaties met een significant effect in 7 tot 16 omgevingen. Verscheidene genen die een rol spelen in zetmeelmetabolisme, de regulatie van bladgrootte of de bepaling van bloeitijd lagen dichtbij de geassocieerde merkers. Twee genen, RPM1 en ACD6, waren in de nabijheid van merkers met significante genotype-met-omgeving (GxE) interactie gelegen. Voor beiden genen zijn allelen geïdentificeerd die verantwoordelijk zijn voor een toename in de resistentie tegen bacteriële infecties, maar die daarnaast ook verantwoordelijk zijn voor afname in biomassa. De richting van het alleleffect is daarom afhankelijk van omgevingsfactoren. Whole genome prediction liet zien dat de meeste GxE interacties die zichtbaar waren op fenotype-niveau niet veroorzaakt werden door sterke associaties met sterke QTL-met-omgeving (QxE) interacties, maar door matige of zwakke associaties met zwakke QxE effecten.

Als laatste wordt de bruikbaarheid van GWA mapping voor de identificatie van genen die verantwoordelijk zijn voor de prestatie van planten in diverse omgevingen waaronder droogte en hitte bediscussieerd (Hoofdstuk 7). Sterke associaties werden waargenomen voor zowel omgevings-specifieke als ook voor algemene regulators van de prestatie van planten. Enkele keuzes die te maken hadden met de fenotypering en het experimentele ontwerp waren cruciaal voor ons succes, zoals de evaluatie van de groei van planten over tijd en de vereenvoudiging van de kwantificeringsmethode van de fenotypen. Er wordt aangeraden om in vervolgonderzoek te focussen op de functionele karakterisering van de causale genen, omdat deze analyse zal helpen bij de identificatie van de signaaltransductie cascade waarbinnen de causale genen een rol spelen. Ook zullen deze analyses bijdragen aan het begrijpen waarom variatie in de sequentie resulteert in veranderingen op het niveau van het fenotype. Dit proefschrift draagt bij aan het begrijpen van de genetische regulatie van stress-reacties van planten en, hoewel de vertaling van deze resultaten naar toepassingen in gewassen een uitdaging is, het draagt waarschijnlijk ook bij aan de ontwikkeling van stresstolerante gewassen met een stabiele opbrengst.

Acknowledgements

Ik heb er geen seconde spijt van gehad dat ik, na wat omzwervingen, uiteindelijk toch aan mijn promotie ben begonnen. Het bestuderen van het leven is fascinerend. Ik raak steeds weer onder de indruk van al die verschillende, complexe processen die samen resulteren in een weerbare, flexibele en mooie natuur. Het is ook bijzonder om te zien dat we de diversiteit van de natuur kunnen gebruiken om ons leven aangenamer te maken, door bijvoorbeeld het ontwikkelen van gewassen met stabiele en hoge opbrengst of door het kweken van mooie kamerplanten. In al deze biologische processen zie ik telkens weer Gods hand en ik ben hem erg dankbaar dat hij de natuur aan ons heeft gegeven om van te genieten, te onderzoeken en te gebruiken.

Hoewel mijn interesse in de biologie en de voldoening die onderzoek geeft mijn motivatie waren om aan mijn promotietraject te beginnen, zijn het vooral ook mensen geweest die er voor gezorgd hebben dat ik gemotiveerd ben gebleven en dat ik met plezier naar Radix kwam.

Als eerste wil ik mijn begeleiders Joost en Dick noemen. Jullie verschillen van elkaar als dag en nacht. Soms was dat lastig, maar meestal was dit in mijn voordeel. Joost, jij was altijd geïnteresseerd in mijn resultaten en je nam altijd tijd om ze te bespreken, ook als deze niet waren zoals verwacht. Het heeft even geduurd voor onze schrijfstijl 'compatible' was en onze schrijf- en correctie-snelheid op elkaar waren aangepast, maar nu mogen we samen kijken naar een mooie lijst met publicaties. Jouw precieze, soms wat verheven schrijfstijl heeft daar zeker toe bijgedragen. Bedankt voor al je hulp! Dick, in het begin van de promotie was jij vooral de man van de handtekeningen en de officiële verslagen voor STW, EPS etc. Maar naar het einde toe, toen er geschreven moest worden, werd je rol actiever. Je bent meer een man van de grote lijnen dan van de details en het is fijn om af en toe te weten, juist als de details nog niet helemaal kloppen, dat het goed zit met de grote lijn. Bedankt voor je positieve kijk op dingen en je interesse in mij ook op het niet wetenschappelijke vlak. Daarnaast wil ik ook Harro bedanken voor zijn inzet bij het maken van het boekje. Je leessnelheid is onvoorstelbaar en je hebt een wonderbaarlijke hand die met bijv. het omdraaien van twee woorden een zin kan upgraden van economy naar business class. Rina, Wytske en Bas, bedankt voor jullie praktische steun en bereidheid mijn (soms urgente) vragen te beantwoorden.

During my PhD I went to Montpellier for 4 months to perform a big experiment in the PHENOPSIS facility. Christine, thanks a lot for your hospitality. It was amazing how well organised everything was and how smooth the experiments went. Also thanks for your input in the manuscripts and the chance you gave me to come back twice and even bring my husband. Also thanks to all people that helped with sowing, harvesting, grinding, managing etc. Thanks, Myriam, Crispulo (muchas gracias), Alexis, Gaelle, Marie-Claude

and Denis. I will never forget Beatriz and her happy smile, who became my four-month-best friend. Merci Stéphane, pour la trip á Mont Aigoual avec la première neige (writing France is not the same as speaking it ;-)).

I also want to mention the colleagues who also work on natural variation. Rik, je bent een harde werker die ook houdt van gezelligheid. Bedankt voor het meedenken met mijn onderzoek, het meehelpen als er weer een groot aantal planten moest worden gezaaid of geoogst en de gesprekken en muziek ondertussen. Emilie, it was very nice that you joined the project halfway. Not only because it guaranteed that my work would not got covered with dust on a shelf, but also because of your friendly personality. You are a very nice person to work with. I have deep respect for your hard work and your perseverance. I hope we will stay in contact and publish our second article together one day. Frank, jij was de redder in nood, vooral in de eerste jaren. Jij kunt alle vragen beantwoorden en je had vaak tijd om te helpen. Frank bedankt voor je inzet. Also thanks to the other natural variation people: Jurriaan, Ramon, Cris, Myriam, Ana Carolina, Mohamed, Nihal, Padraic, PingPing, Ronny, Xi, Eric, Laurens, Charles, Nelson, Roxanne, Aina, Mark and Colette. I supervised the master theses of Matthijs, Adrian, and Yusong and the bachelor thesis of Marc. I enjoyed teaching you how to perform experiments and, more importantly, how to interpret them. Unfortunately, most of your experiments did not give the expected results, which complicated things. Your research helped me to decide where to go. Thanks for your work!

I was privileged to work together with several people from Biometris. It is very nice to repeatedly note that the research fields of biology and mathematics/statistics are different worlds, but that it is possible to work together if you both respect each other's knowledge. I made a proposition about this subject, so please have a look and start attacking! Joao, Fred, Willem, Sabine, Hans, thanks for you kindness and patience when I got lost. Special thanks to Marcos. I am sorry, if I was pushing you a bit in the last months, but your work is really appreciated and it looks like our efforts will result in a nice paper.

Coffee time was also an important aspect of my PhD. Just to relax and talk about other things than science. Jun, Imran and Farzaneh our desks were close together and we also spend many coffee breaks together. Very nice that you shared your culture, your way of life and even your family with me. Jun, I am honoured that you will be my paranymph! Also thanks to the many other people with whom I spend countless numbers of lunches and coffee breaks. Without you it would sometimes have become boring.

Familie en vrienden zijn een belangrijk deel van mijn leven. Pa, u heeft een bijzondere rol gespeeld bij mijn promotie. U blijft altijd positief en door uw werkkamer op loopafstand was u vaak het luisterend oor of de adviseur. Ma, sorry dat we soms thuis te veel over Radix en WUR praatten. Bedankt ook voor uw meeleven. Ik hou van jullie! Wouter,

hoewel je pas halverwege mijn promotie in beeld kwam, ben je in korte tijd erg belangrijk geworden. Je nuchtere en zakelijke kijk op dingen heeft me zeker beïnvloed. Gelukkig gaat ook na deze promotie ons leven samen verder en ik kijk uit naar het moment om je als vader in actie te zien. Julia, lieve zus, super dat je mijn paranimf wilt zijn. Lianne, we hebben heel wat afgepraat over ons werk, de balans werk en privé en over onze 'mannen'. Hier gaan we na de promotie gewoon mee door. Koenraad, PhD leed, fietstochten en veel andere leuke dingen hebben we gedeeld. Ik ben benieuwd waar de toekomst ons zal brengen. Ik wil ook de andere broers, zussen en vrienden bedanken voor hun meeleven tijdens de promotie. Pauline, superleuk dat je zowel bij de bruiloft als bij mijn promotie in NL kunt zijn. Wie had dat gedacht. Also many people of ICF have given me a lot of joy and happiness. Our ways have split already, but my nice memories will stay.

In conclusion, I am thankful and happy that I was able to start and finish my PhD at Plant Physiology and Genetics. All the best to everyone and God bless you.

This work was performed at the Laboratories of Plant Physiology and Genetics,Wageningen University. This research is part of the 'Learning from Nature' program, supported by the Dutch Technology Foundation (STW grant number 10996), which is part of the Netherlands Organisation for Scientific Research (NWO).

Explanation cover design

This thesis describes the study of plant responses to several environments. The backside of the cover symbolises optimal growing conditions in which the plants are not hindered by any stress. In this environments plants grow well, will bloom and produce plenty of seeds. The frontcover symbolises the growth of plants in less optimal conditions, like drought and heat. Some plants perform well in these conditions, while other plants show for example reduced seed set. This thesis aims to contribute to the development of plants that perform well even in less optimal conditions.

© Johanna A. Bac-Molenaar 2015.