Missing heritability and soft inheritance of morphology and metabolism in Arabidopsis

Rik Kooke

Thesis committee

Promotors

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

Prof. Dr J.J.B. Keurentjes
Assistant 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, Plant Research International, Wageningen University
Dr. A.G. Bovy, Plant Research International, Wageningen University
Dr. K. Verhoeven, Netherlands Institute of Ecology, Wageningen
Dr J.M. Jiménez-Gómez, Max-Planck Institute, Köln, Germany

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

Missing heritability and soft inheritance of morphology and metabolism in Arabidopsis

Rik Kooke

Thesis
submitted in fulfillment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr. M.J. Kropff,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Tuesday 26 August 2014
at 4 p.m. in the Aula.

Rik Kooke

Missing heritability and soft inheritance of morphology and metabolism in Arabibidopsis,
284 pages

PhD thesis Wageningen University, Wageningen, NL (2014)
With references, with summaries in English and Dutch

ISBN 978-94-6257-041-2

Contents

Chapter 1

General introduction

1

The distinction between the stable genotype and the flexible phenotype was first introduced and coined by Wilhelm Johannsen in 1909 and is regarded as the cornerstone of classical genetics (Johannsen, 1909; Roll-Hansen, 2009). A more modern definition of phenotype is 'the visible or otherwise measurable physical and biochemical characteristics of an organism, resulting from the interaction of genotype and environment' (Henderson and Lawrence, 2000). Phenotypic traits can be classified into two groups based on their monogenic or polygenic inheritance pattern: qualitative and quantitative traits. Qualitative traits follow simple Mendelian inheritance and occur in distinct phenotypic classes within a population, in which the classes are defined by (combinations of) different alleles at a single locus. Examples in plants include flower colour and pea seed morphology, first described by Gregor Mendel. The influence of the environment on such traits is usually small and mostly negligible. Quantitative traits also follow Mendelian inheritance, but are defined by multiple genes, gene interactions and a larger effect of the environment, and therefore show a continuous variation in phenotypes. Examples of quantitative traits are numerous in plants, and vary from gene expression to metabolite levels and agronomical important traits, such as biomass, flowering time and reproductive success.

Natural variation within species can be exploited to associate phenotypic traits with genotypic variation. *Arabidopsis thaliana* has been adopted by the scientific plant community as the model species for such research as it combines a small genome size with a short generation cycle and despite being a self-fertiliser, it is suitable for out-crossing. Due to its widespread distribution across the earth, from sea level to high altitudes and from Northern Europe to New Zealand, Arabidopsis has adapted to a broad range of environments (Koornneef et al., 2004; Weigel, 2012). The association between genotype and phenotype can be explored by growing different genotypes in the same environment. The phenotype is heavily dependent on the environment and, therefore, to unravel the genetic regulation of phenotypes, preferably different genotypes are also studied in different environments (Anderson et al., 2014).

The genetic basis of the phenotype

Perhaps the most effective way to study the effect of a gene on a phenotype is through gene mutation, such as gene knockout and silencing, or through gene over-expression. The selective forces of nature have ensured that every gene has a function and each gene knockout should therefore show a phenotype under the appropriate circumstances (Lloyd and Meinke, 2012; Meinke, 2013). However, there are a number of exceptions in which a mutant will not show a phenotype. Gene redundancy or feedback regulation can mask the mutant phenotype, the mutant phenotype can be subtle and almost undistinguishable from wild-type, or the phenotype only becomes apparent under certain circumstances or in a certain genetic background (Lloyd and Meinke, 2012). Sometimes, gene over-expression may overcome such limitations.

Alternatively, natural variation within species can be explored to investigate the genotype-phenotype relationship. To that purpose, many different bi-parental mapping populations have been produced the last decades by crossing two divergent genotypes. Depending on the preferred type of population, repeated back-crossing or self-fertilisation can generate immortal mapping populations, consisting of completely homozygous lines (Doerge, 2002). Although F_2 populations can be valuable, immortal mapping populations such as recombinant inbred lines (RILs), near isogenic lines (NILs) and doubled haploids (DHs) are most often used to associate phenotypic traits with molecular markers in quantitative trait loci (QTL) analyses (Keurentjes et al., 2007). The advantage of immortal mapping populations compared to F_2 populations is that they can be replicated endlessly, greatly reducing the genotyping costs.

Recently, rapid advances in next-generation sequencing have enabled the comparison of hundreds of genotypes at the nucleotide level (Schneeberger and Weigel, 2011). Thirteen years since the release of the first Arabidopsis genome sequence (Kaul et al., 2000), another 1048 Arabidopsis genomes have been resequenced (http://1001genomes.org) (Ossowski et al., 2008; Cao et al., 2011; Schneeberger et al., 2011; Long et al., 2013). The comparison of individual genomic sequences allows the identification of sequence variants between accessions at a single position. These so-called single nucleotide polymorphisms (SNPs) can be surveyed for an association with quantitative traits using genome-wide association (GWA) mapping (Nordborg and Weigel, 2008). GWA mapping is based on linkage disequilibrium (LD), the non-random association of alleles at different loci, and is in many aspects similar to classical (QTL) linkage mapping.

GWA studies have several advantages compared to the use of conventional bi-parental QTL mapping populations, such as recombinant inbred lines (RILs) and near-isogenic lines (NILs) (Keurentjes et al., 2007; Kooke et al., 2012; Korte and Farlow, 2013). First, the creation of RILs and NILs requires several generations of self-fertilisation and/ or backcrossing (in the case of NILs) to obtain homozygous lines. Natural accessions of Arabidopsis are in principle homozygous by nature although some heterozygosity may exist due to infrequent out-crossing (Tang et al., 2007). On the other hand, the time required to generate RILs or doubled haploids (DHs) has been substantially reduced due to the discovery of the haploid inducer mutant *cenh3* that eliminates its own genome (Ravi and Chan, 2010). By crossing a heterozygous F_1 plant – derived from a cross between two divergent accessions - with the *cenh3* mutant, haploid plants are obtained that spontaneously produce DHs in the next generation. The recombination is slightly reduced, but allele frequency in DHs is similar to that in RILs, and they are well suited for QTL mapping (Seymour et al., 2012). Second, GWA studies profit from numerous recombination events that have accumulated during the long evolutionary history of Arabidopsis, providing an almost optimal resolution down to the gene level (Bergelson and Roux, 2010). Although LD decays on average already within 10 kb (Kim et al., 2007), large LD regions, such as those regions that contain a selective

sweep, are present in Arabidopsis. QTL support intervals from RIL populations, however, are much larger, covering hundreds, if not thousands of genes, making candidate gene selection difficult. Third, the variation in GWA studies is based on a large number of accessions, which enables the identification of common alleles that have been established through natural selection. In bi-parental mapping populations the variation is limited to the segregation of alleles present in the two parents, and could thus lead to the identification of rare, large-effect alleles, which may not be relevant in an evolutionary context (Rockman, 2012).

Nonetheless, GWAS suffer from the association of quantitative traits with false-positive SNPs and non-causal alleles, due to population structure within GWA populations. By correcting for genetic relatedness among the accessions, the problem of population structure and false-positive associations is reduced, but could result in the rejection of false-negative (true) associations. Furthermore, genetic or allelic heterogeneity, different genes or alleles leading to the same phenotype, impedes the identification of causal alleles and common variants of small effect (Korte and Farlow, 2013). Moreover, power is substantially reduced in GWAS due to the segregation of multiple alleles at a single locus, while epistatic interactions are difficult to find due to the extremely high number of possible interactions in GWAS (Korte and Farlow, 2013). Nevertheless, GWA studies have contributed to the understanding of the genetic architecture underlying quantitative traits and the identification of genes involved in regulating such traits (Baxter et al., 2010; Brachi et al., 2010; Chan et al., 2010b; Chan et al., 2010a; Li et al., 2010; Chan et al., 2011; Filiault and Maloof, 2012; Verslues et al., 2014).

The role of the environment

Owing to their sessile nature, plants need to adjust their phenotype to fluctuations in the environment to optimize development and reproductive success. Plants are thus strongly influenced by their environment, and by means of mutation, recombination and natural selection, plants become adapted to the local conditions (Anderson et al., 2011; Long et al., 2013). A number of studies have shown adaptive clines of Arabidopsis to environmental conditions. Development and flowering-related traits are correlated with day length, latitude and longitude (Hancock et al., 2011; Samis et al., 2012; Debieu et al., 2013). Moreover, natural variation at a salt-tolerance locus, *HKT1*, associates with saline regions in Europe (Baxter et al., 2010) and natural variation in the defense locus *GS-ELONG* correlates geographically with the relative abundance of two specialist aphids (Zust et al., 2012), suggesting local adaptation through abiotic and biotic interactions. Local populations are then expected to outperform others in their native range, and indeed, in most cases local populations have increased fitness in the local habitat (Orr, 2005; Hereford, 2009; Fournier-Level et al., 2011; Agren and Schemske, 2012).

The interaction between genotype and environment (GxE) may be investigated by studying multiple genotypes in different environments. Also transplant experiments in which

genotypes from at least two environments are exchanged between their native and non-native environment might further assist in the understanding of the genetic regulation of local adaptation (Fournier-Level et al., 2011; Agren and Schemske, 2012; Savolainen et al., 2013). Most studies find that genes involved in local adaptation are conditionally neutral, which means that increased fitness in the local environment does not reduce fitness in other environments (Anderson et al., 2011). However, some studies suggest that there is a trade-off involved in local adaptation and that adapted plants outcompete others in their native range, but are outperformed elsewhere (Scarcelli et al., 2007; Todesco et al., 2010; Fournier-Level et al., 2011; Ågren et al., 2013). This is well illustrated by a study on natural variation at the *ACD6* locus in which allelic variation resulted in a trade-off between pathogen defense and reproductive success (Todesco et al., 2010). It has long been debated whether there is a trade-off between investment in defense and growth, in which it is assumed that investment in defense is costly (Herms and Mattson, 1992). Two recent studies on plant-herbivore interactions, however, show that not investment in metabolite quantity, but rather specific, inexpensive investment in metabolite quality increased plant fitness (Agrawal et al., 2012; Hare, 2012; Zust et al., 2012).

The ability of an organism to adjust its physiology and morphology in response to the biotic and abiotic environment is named phenotypic plasticity (Schlichting, 1986). The opposite of plasticity is called canalization or robustness, when the phenotype is unaltered in response to changes in the environment. Because different genotypes display variation in their degree of plasticity, phenotypic plasticity is thought to have a genetic basis (Ungerer et al., 2003; Lacaze et al., 2009; Tetard-Jones et al., 2011). Increased plasticity may be adaptive in fluctuating environments (Agrawal, 2001). Even though possibly of value in changing environments, phenotypic plasticity might have a fitness cost due to plastic rather than fixed development in stable environments (DeWitt et al., 1998; Van Kleunen and Fischer, 2005).

On top of genetics: epigenetics

The proportion of phenotypic variation in a population that is regulated by genetic effects can be estimated by the broad-sense heritability (H^2). A high H^2 indicates that a large proportion of the phenotypic variation can be ascribed to genetic loci and the interactions between them. Recently, there is an intense debate in human genetics about the disagreement between the high heritability found for human traits and diseases in GWA studies and the low percentage of genetic variation that can be explained by genetic loci (Manolio et al., 2009). Although not generally identified as such in plants (Brachi et al., 2011), missing heritability could potentially be a problem for quantitative traits in plants as well (Weigel, 2012; Korte and Farlow, 2013). A number of studies relate the discrepancy between high H^2 and low explained variance to the complex genetic architecture of quantitative traits. If many genes of small effect or rare variants define the phenotype, they could remain unidentified

in GWAS (Gibson, 2010; Yang et al., 2010; Gibson, 2011; Yang et al., 2011; Zuk et al., 2012; Zuk et al., 2014). Alternatively, epigenetic variation might regulate a large percentage of the phenotypic variation independent of genetic variation (Manolio et al., 2009; Weigel, 2012).

Epigenetic inheritance is defined as the transfer of phenotypic changes to subsequent generations of cells or organisms realized by epigenetic modifications independent of DNA sequence variation (Jablonka and Raz, 2009). Epigenetic variation is most often associated with variation in chromatin marks including DNA methylation, histone modifications and small RNAs. Such epigenetic marks closely interact and together they determine gene expression, DNA repair, chromatin accessibility and chromosome organization (Cedar and Bergman, 2009; Jablonka and Raz, 2009). DNA methylation is the best studied epigenetic mark in plants and concerns the methylation of cytosine nucleotides in predominantly CG, but also in CHG and CHH islands (in which H represents any nucleotide except guanine) (Law and Jacobsen, 2010). DNA methylation in plants occurs mostly on transposons and other repetitive DNA sequences, although CG methylation is also present on single-copy genes (Schmitz et al., 2013b). Loss of methylation may cause transposable element (TE) mobilization and subsequent insertion in other parts of the genome (Bucher et al., 2012; Fedoroff, 2012), but methylation present in *cis*-regulatory regions may also alter gene expression (Richards, 2006). Epigenetic variants that cause phenotypic variation and that can be stably transmitted to future generations have been found in plants showing that epigenetic inheritance occurs in plants (Cubas et al., 1999; Manning et al., 2006; Martin et al., 2009).

To discriminate between genetic and epigenetic variation, three classes of epialleles have been proposed: obligate, facilitated and pure epialleles. Obligate epialleles are dependent on DNA sequence variation, while facilitated epialleles arise through a genetic modification such as TE insertion, but the maintenance is not dependent on genetic variation. Pure epialleles form independent of DNA sequence variation. The epialleles identified thus far are either obligate or facilitated epialleles, and although mitotically pure epialleles exist, meiotically pure epialleles have been suggested to exist but have not yet been confirmed (Schmitz and Ecker, 2012).

Epigenetic variation in DNA methylation can be induced through mutation of certain genes involved in epigenetic regulation, such as *DDM1* and *MET1*, (Kakutani et al., 1996; Saze et al., 2003; Johannes et al., 2009; Reinders et al., 2009). *Met1* mutants show an almost complete loss of DNA methylation and flower much later than wild-type plants (Saze et al., 2003). *Ddm1* mutants have severely reduced levels of methylation and increased TE transcription and mobilisation, and they show clear developmental defects after repeated self-fertilisation (Kakutani et al., 1996; Tsukahara et al., 2009). Both mutants were crossed to Col-0 wild-type, backcrossed and repeatedly self-fertilised to create populations of isogenic

lines, coined epigenetic recombinant inbred lines or epiRILs, with differentially methylated regions across the genome (Johannes et al., 2009; Reinders et al., 2009). The epigenetic variation among the epiRILs was found to cause significant heritable variation in phenotypic traits in different environments (Johannes et al., 2009; Reinders et al., 2009; Latzel et al., 2012; Zhang et al., 2013; Cortijo et al., 2014).

Epigenetic variation is present in nature and recent studies have investigated methylome variation in natural accessions and experimental populations of Arabidopsis, soybean and maize (Vaughn et al., 2007; Lister et al., 2008; Eichten et al., 2013; Schmitz et al., 2013a; Schmitz et al., 2013b). DNA methylome variation was found to be heritable and to exist both independent and dependent of DNA sequence variation (Schmitz et al., 2013b). Moreover, comparison of the methylation level in different Arabidopsis Col-0 lines propagated for 30 generations through single seed descent, revealed numerous differentially methylated regions (DMRs) among the lines (Becker et al., 2011; Schmitz et al., 2011). This suggests that spontaneous epimutations can become fixed and that they may alter transcription and increase phenotypic diversity independent of genetic variation (Schmitz et al., 2011). Because epigenetic changes can regulate heritable phenotypic changes (Johannes et al., 2009; Zhang et al., 2013) and epigenetic variation is present and heritable in nature (Eichten et al., 2013; Schmitz et al., 2013b), epigenetics might have adaptive value.

In contrast to genetic variation, however, epigenetic variation is reversible and it will therefore most likely not contribute to long-term evolutionary processes. It might be more important in phenotypic plasticity in fluctuating environments. Genetic mutations that cause increased plasticity are permanent and might thus reduce fitness in other environments. Loss or gain of DNA methylation, on the other hand, may be rapidly induced or reversed in the case of a changing environment. Indeed, both chemical and experimental demethylation increased phenotypic plasticity under various abiotic and biotic stresses (Tatra et al., 2000; Reinders et al., 2009; Bossdorf et al., 2010; Zhang et al., 2013). The plasticity was found to be highly heritable and profitable in terms of plant height and root:shoot ratio under drought (Zhang et al., 2013). Because genetic QTL studies also found high variation in plasticity among RILs (Ungerer et al., 2003; Lacaze et al., 2009; Tetard-Jones et al., 2011), phenotypic plasticity most likely has a genetic and epigenetic basis. This is further exemplified by a study on natural Arabidopsis accessions where chemically induced DNA hypomethylation caused differential phenotypic plasticity (Bossdorf et al., 2010).

Transgenerational epigenetic inheritance

Numerous studies have shown that epigenetic modifications can be triggered by exposure to different environmental conditions (Steward et al., 2002; Sung and Amasino, 2004; Hashida et al., 2006; Choi and Sano, 2007; Boyko and Kovalchuk, 2008; Chinnusamy and Zhu, 2009; Lang-Mladek et al., 2010; Verhoeven et al., 2010; Grativol et al., 2012; Karan et al.,

2012). A popular hypothesis states that these environmentally induced epigenetic changes can be inherited across generations and that they may be valuable in an evolutionary context, reviving Lamarckian soft inheritance (Richards, 2006). Indicative evidence in favor of this hypothesis was found (Molinier et al., 2006; Luna et al., 2012; Rasmann et al., 2012). Homologous recombination frequency was found to be substantially increased in unstressed progeny for multiple generations after stress exposure of the parental plants (Molinier et al., 2006). Because there was no basis to expect DNA sequence variation between the plants, the signal was hypothesized to be epigenetic. Repetition of the experiment, however, did not confirm these results (Pecinka et al., 2009) and studies since then have not been able to univocally prove stress-induced transgenerational epigenetic inheritance. Most of these studies have focused on the immediate progeny of the stressed plants, and therefore epigenetic modification cannot be distinguished from maternal effects or environmental effects on the pre-meiotic cells (Pecinka and Scheid, 2012). Moreover, most studies did not look into the molecular mechanisms that may explain the variation and can actually not differentiate between DNA sequence variation and epigenetic variation. And finally, epigenetic variation correlating with stress exposure does not necessarily have to be causal, as the epigenetic variation may also be induced spontaneously, and not in response to stress (Becker et al., 2011; Schmitz et al., 2011). Nevertheless, all the above suggests that epigenetic modifications occur in nature, that they can be stable and heritable and that they affect the phenotype. However, it remains elusive whether environmentally induced epigenetic changes are pure epialleles, and not random, and whether they are transmitted to future generations (Paszkowski and Grossniklaus, 2011; Pecinka and Scheid, 2012; Heard and Martienssen, 2014).

Scope of the thesis

The relationship between genotype and phenotype is extremely complex and depends on the interaction between genotype, epi-genotype and the environment. In this thesis, we will use experimental and natural populations of *Arabidopsis thaliana* to gain insight into this complex regulation of the phenotype, which will be studied on the enzyme, metabolite and morphological trait level.

Chapter 2 describes the genetic architecture underlying quantitative morphological traits, such as flowering time, leaf length and plant height in a GWA study. The high heritability in this study did not correspond with the low explained variance by the significant loci, and suggests complex regulation. When the significance threshold was lowered, numerous candidate genes that had been previously linked to the phenotype were found, and suggest that missing heritability is hidden rather than missing.

Chapter 3 reviews the recent literature to determine to what extent temporal (biological clock, source-sink transitions), spatial (different cells, tissues), environmental and genetic

regulation contribute to metabolic organisation in plants.

Chapter 4 describes the genetic regulation of the complex, coordinated primary metabolic network in relation to plant biomass. GWA analysis was performed on biomass and a number of enzyme activities, structural components and metabolites of primary carbon and nitrogen metabolism. Numerous pleiotropic genes were found that regulate enzyme activity, metabolic content and plant biomass with opposite effects on metabolism and biomass, suggesting a trade-off between primary metabolism and biomass.

Chapter 5 describes the epigenetic regulation of morphology and phenotypic plasticity. Ninety-nine epiRILs of the *ddm1*-inherited epiRIL population were studied for morphological traits under control and saline conditions. DMR-based QTL mapping revealed many co-locating QTLs for growth, morphology and plasticity, suggesting pleiotropic regulation via epigenetic modifications.

Chapter 6 describes the epigenetic regulation of plant secondary metabolism in leaves and flowers. A number of DMR-based QTLs were detected for glucosinolates and flavonoids in both flowers and leaves. Metabolic traits could predict plant growth and morphological traits, and many QTLs overlapped, suggesting pleiotropic epigenetic regulation.

Chapter 7 describes transgenerational inheritance of abiotic and biotic stress-induced epigenetic changes that affect gene expression, secondary metabolism, growth and morphology in non-stressed progeny. The abiotic, salt stress-elicited responses deviated distinctly from the biotic, methyljasmonate induced responses and suggest that the transgenerational signals are stress-specific.

Chapter 8 discusses the work presented in this thesis with particular attention for missing heritability, hard versus soft inheritance and trade-offs between metabolism and growth.

References

Agrawal AA (2001) Ecology - Phenotypic plasticity in the interactions and evolution of species. Science 294: 321-326

Agrawal AA, Hastings AP, Johnson MTJ, Maron JL, et al (2012) Insect Herbivores Drive Real-Time Ecological and Evolutionary Change in Plant Populations. Science 338: 113-116

Ågren J, Oakley CG, McKay JK, Lovell JT, et al. (2013) Genetic mapping of adaptation reveals fitness tradeoffs in Arabidopsis thaliana. Proceedings of the National Academy of Sciences 110: 21077-21082

Ågren J, Schemske DW (2012) Reciprocal transplants demonstrate strong adaptive differentiation of the model organism Arabidopsis thaliana in its native range. New Phytologist 194: 1112-1122

Anderson JT, Wagner MR, Rushworth CA, Prasad K, et al (2014) The evolution of quantitative traits in complex environments. Heredity 112: 4-12

Anderson JT, Willis JH, Mitchell-Olds T (2011) Evolutionary genetics of plant adaptation. Trends Genet 27: 258-266

Baxter I, Brazelton JN, Yu D, Huang YS, et al (2010) A coastal cline in sodium accumulation in Arabidopsis thaliana is driven by natural variation of the sodium transporter AtHKT1;1. PLoS Genet 6: e1001193

Becker C, Hagmann J, Muller J, Koenig D, et al (2011) Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. Nature 480: 245-249

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

Bossdorf O, Arcuri D, Richards CL, Pigliucci M (2010) Experimental alteration of DNA methylation affects the phenotypic plasticity of ecologically relevant traits in Arabidopsis thaliana. Evolutionary Ecology 24: 541-553

Boyko A, Kovalchuk I (2008) Epigenetic control of plant stress response. Environ Mol Mutagen 49: 61-72

Brachi B, Faure N, Horton M, Flahauw E, et al (2010) Linkage and Association Mapping of Arabidopsis thaliana Flowering Time in Nature. PLoS genetics 6

Brachi B, Morris GP, Borevitz JO (2011) Genome-wide association studies in plants: the missing heritability is in the field. Genome Biology 12

Bucher E, Reinders J, Mirouze M (2012) Epigenetic control of transposon transcription and mobility in Arabidopsis. Curr Opin Plant Biol 15: 503-510

Cao J, Schneeberger K, Ossowski S, Gunther T, et al (2011) Whole-genome sequencing of multiple Arabidopsis thaliana populations. Nat Genet 43: 956-963

Cedar H, Bergman Y (2009) Linking DNA methylation and histone modification: patterns and paradigms. Nat Rev Genet 10: 295-304

Chan EK, Rowe HC, Corwin JA, Joseph B, et al (2011) Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in Arabidopsis thaliana. PLoS Biol 9: e1001125

Chan EK, Rowe HC, Hansen BG, Kliebenstein DJ (2010) The complex genetic architecture of the metabolome. PLoS genetics 6: e1001198

Chan EK, Rowe HC, Kliebenstein DJ (2010) Understanding the evolution of defense metabolites in Arabidopsis thaliana using genome-wide association mapping. Genetics 185: 991-1007

Chinnusamy V, Zhu JK (2009) Epigenetic regulation of stress responses in plants. Curr Opin Plant Biol 12: 133-139

Choi CS, Sano H (2007) Abiotic-stress induces demethylation and transcriptional activation of a gene encoding a glycerophosphodiesterase-like protein in tobacco plants. Mol Genet Genomics 277: 589-600

Cortijo S, Wardenaar R, Colomé-Tatché M, Gilly A, et al (2014) Mapping the epigenetic basis of complex traits. Science 343: 1145-1148

Cubas P, Vincent C, Coen E (1999) An epigenetic mutation responsible for natural variation in floral symmetry. Nature 401: 157-161

Debieu M, Tang C, Stich B, Sikosek T, et al (2013) Co-Variation between Seed Dormancy, Growth Rate and Flowering Time Changes with Latitude in Arabidopsis thaliana. Plos One 8

DeWitt TJ, Sih A, Wilson DS (1998) Costs and limits of phenotypic plasticity. Trends in Ecology & Evolution 13: 77-81

Doerge RW (2002) Mapping and analyzis of quantitative trait loci in experimental populations. Nat Rev Genet 3: 43-52

Eichten SR, Briskine R, Song J, Li Q, et al (2013) Epigenetic and genetic influences on DNA methylation variation in maize populations. Plant Cell 25: 2783-2797

Fedoroff NV (2012) Transposable Elements, Epigenetics, and Genome Evolution. Science 338: 758-767

Filiault DL, Maloof JN (2012) A genome-wide association study identifies variants underlying the Arabidopsis thaliana shade avoidance response. PLoS Genet 8: e1002589

Fournier-Level A, Korte A, Cooper MD, Nordborg M, et al (2011) A map of local adaptation in Arabidopsis thaliana. Science 334: 86-89

Gibson G (2010) Hints of hidden heritability in GWAS. Nat Genet 42: 558-560

Gibson G (2011) Rare and common variants: twenty arguments. Nat Rev Genet 13: 135-145

Grativol C, Hemerly AS, Ferreira PC (2012) Genetic and epigenetic regulation of stress responses in natural plant populations. Biochim Biophys Acta 1819: 176-185

Hancock AM, Brachi B, Faure N, Horton MW, et al (2011) Adaptation to climate across the Arabidopsis thaliana genome. Science 334: 83-86

Hare JD (2012) How Insect Herbivores Drive the Evolution of Plants. Science 338: 50-51

Hashida SN, Uchiyama T, Martin C, Kishima Y, et al (2006) The temperature-dependent change in methylation of the Antirrhinum transposon Tam3 is controlled by the activity of its transposase. Plant Cell 18: 104-118

Heard E, Martienssen RA (2014) Transgenerational Epigenetic Inheritance: Myths and Mechanisms. Cell 157: 95-109

Henderson WD, Lawrence E (2000) Henderson's dictionary of biological terms. Prentice Hall, Harlow [etc.]

Hereford J (2009) A quantitative survey of local adaptation and fitness trade-offs. Am Nat 173: 579-588

Herms DA, Mattson WJ (1992) The dilemma of plants - to grow or defend. Quarterly Review of Biology 67: 283-335

Jablonka E, Raz G (2009) Transgenerational epigenetic inheritance: Prevalence, mechanisms, and implications for the study of heredity and evolution. Quarterly Review of Biology 84: 131-176

Johannes F, Porcher E, Teixeira FK, Saliba-Colombani V, et al (2009) Assessing the Impact of Transgenerational Epigenetic Variation on Complex Traits. PLoS genetics 5

Johannsen W (1909) Elemente der exakten Erblichkeitslehre. Fischer Jena

Kakutani T, Jeddeloh JA, Flowers SK, Munakata K, et al (1996) Developmental abnormalities and epimutations associated with DNA hypomethylation mutations. Proceedings of the National Academy of Sciences of the United States of America 93: 12406-12411

Karan R, DeLeon T, Biradar H, Subudhi PK (2012) Salt stress induced variation in DNA methylation pattern and its influence on gene expression in contrasting rice genotypes. PLoS One 7: e40203

Kaul S, Koo HL, Jenkins J, et al (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408: 796-815

Keurentjes JJB, Bentsink L, Alonso-Blanco C, Hanhart CJ, et al (2007) Development of a near-isogenic line population of Arabidopsis thaliana and comparison of mapping power with a recombinant inbred line population. Genetics 175: 891-905

Kim S, Plagnol V, Hu TT, Toomajian C, et al (2007) Recombination and linkage disequilibrium in *Arabidopsis thaliana*. Nat Genet 39: 1151-1155

Kooke R, Wijnker E, Keurentjes JB (2012) Backcross Populations and Near Isogenic Lines. In SA Rifkin, ed, Quantitative Trait Loci (QTL), Vol 871. Humana Press, pp 3-16

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

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

Lacaze X, Hayes PM, Korol A (2009) Genetics of phenotypic plasticity: QTL analyzis in barley, Hordeum vulgare. Heredity (Edinb) 102: 163-173

Lang-Mladek C, Popova O, Kiok K, Berlinger M, et al (2010) Transgenerational inheritance and resetting of stress-induced loss of epigenetic gene silencing in Arabidopsis. Mol Plant 3: 594-602

Latzel V, Zhang Y, Moritz KK, Fischer M, et al (2012) Epigenetic variation in plant responses to defence hormones. Annals of Botany 110: 1423-1428

Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet 11: 204-220

Li Y, Huang Y, Bergelson J, Nordborg M, et al (2010) Association mapping of local climate-sensitive quantitative trait loci in Arabidopsis thaliana. Proc Natl Acad Sci U S A 107: 21199-21204

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, et al (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133: 523-536

Lloyd J, Meinke D (2012) A comprehensive dataset of genes with a loss-of-function mutant phenotype in Arabidopsis. Plant Physiol 158: 1115-1129

Long Q, Rabanal FA, Meng D, Huber CD, et al (2013) Massive genomic variation and strong selection in Arabidopsis thaliana lines from Sweden. Nat Genet 45: 884-890

Luna E, Bruce TJA, Roberts MR, Flors V, et al (2012) Next-Generation Systemic Acquired Resistance. Plant Physiology 158: 844-853

Manning K, Tor M, Poole M, Hong Y, et al (2006) A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. Nature Genetics 38: 948-952

Manolio TA, Collins FS, Cox NJ, Goldstein DB, et al (2009) Finding the missing heritability of complex diseases. Nature 461: 747-753

Martin A, Troadec C, Boualem A, Rajab M, et al (2009) A transposon-induced epigenetic change leads to sex determination in melon. Nature 461: 1135-U1237

Meinke DW (2013) A survey of dominant mutations in *Arabidopsis thaliana*. Trends in Plant Science 18: 84-91

Molinier J, Ries G, Zipfel C, Hohn B (2006) Transgeneration memory of stress in plants. Nature 442: 1046-1049

Nordborg M, Weigel D (2008) Next-generation genetics in plants. Nature 456: 720-723

Orr HA (2005) The genetic theory of adaptation: a brief history. Nat Rev Genet 6: 119-127

Ossowski S, Schneeberger K, Clark RM, Lanz C, et al (2008) Sequencing of natural strains of Arabidopsis thaliana with short reads. Genome Research 18: 2024-2033

Paszkowski J, Grossniklaus U (2011) Selected aspects of transgenerational epigenetic inheritance and resetting in plants. Curr Opin Plant Biol 14: 195-203

Pecinka A, Rosa M, Schikora A, Berlinger M, et al (2009) Transgenerational stress memory is not a general response in Arabidopsis. PLoS One 4: e5202

Pecinka A, Scheid OM (2012) Stress-Induced Chromatin Changes: A Critical View on Their Heritability. Plant and Cell Physiology 53: 801-808

Rasmann S, De Vos M, Casteel CL, Tian D, et al (2012) Herbivory in the previous generation primes plants for enhanced insect resistance. Plant Physiol 158: 854-863

Ravi M, Chan SWL (2010) Haploid plants produced by centromere-mediated genome elimination. Nature 464: 615-U180

Reinders J, Wulff BBH, Mirouze M, Mari-Ordonez A, et al (2009) Compromised stability of DNA methylation and transposon immobilization in mosaic Arabidopsis epigenomes. Genes & Development 23: 939-950

Richards EJ (2006) Opinion - Inherited epigenetic variation - revisiting soft inheritance. Nature Reviews Genetics 7: 395-U392

Rockman MV (2012) The QTN program and the alleles that matter for evolution: All that's gold does not glitter. Evolution 66: 1-17

Roll-Hansen N (2009) Sources of Wilhelm Johannsen's Genotype Theory. Journal of the History of Biology 42: 457-493

Samis KE, Murren CJ, Bossdorf O, Donohue K, et al (2012) Longitudinal trends in climate drive flowering time clines in North American Arabidopsis thaliana. Ecol Evol 2: 1162-1180

Savolainen O, Lascoux M, Merila J (2013) Ecological genomics of local adaptation. Nat Rev Genet 14: 807-820

Saze H, Scheid OM, Paszkowski J (2003) Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. Nature Genetics 34: 65-69

Scarcelli N, Cheverud JM, Schaal BA, Kover PX (2007) Antagonistic pleiotropic effects reduce the potential adaptive value of the FRIGIDA locus. Proceedings of the National Academy of Sciences 104: 16986-16991

Schlichting CD (1986) The evolution of phenotypic plasticity. Annual Review of Ecology and Systematics 17: 667-693

Schmitz RJ, Ecker JR (2012) Epigenetic and epigenomic variation in Arabidopsis thaliana. Trends Plant Sci 17: 149-154

Schmitz RJ, He Y, Valdes-Lopez O, Khan SM, et al (2013) Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population. Genome Res 23: 1663-1674

Schmitz RJ, Schultz MD, Lewsey MG, O'Malley RC, et al (2011) Transgenerational Epigenetic Instability Is a Source of Novel Methylation Variants. Science 334: 369-373

Schmitz RJ, Schultz MD, Urich MA, Nery JR, et al (2013) Patterns of population epigenomic diversity. Nature 495: 193-198

Schneeberger K, Ossowski S, Ott F, Klein JD, et al (2011) Reference-guided assembly of four diverse Arabidopsis thaliana genomes. Proceedings of the National Academy of Sciences 108: 10249-10254

Schneeberger K, Weigel D (2011) Fast-forward genetics enabled by new sequencing technologies. Trends in Plant Science 16: 282-288

Seymour DK, Filiault DL, Henry IM, Monson-Miller J, et al (2012) Rapid creation of Arabidopsis doubled haploid lines for quantitative trait locus mapping. Proceedings of the National Academy of Sciences of the United States of America 109: 4227-4232

Steward N, Ito M, Yamaguchi Y, Koizumi N, et al (2002) Periodic DNA methylation in maize nucleosomes and demethylation by environmental stress. J Biol Chem 277: 37741-37746

Sung SB, Amasino RM (2004) Vernalization and epigenetics: how plants remember winter. Current Opinion in Plant Biology 7: 4-10

Tang CL, Toomajian C, Sherman-Broyles S, Plagnol V, et al (2007) The evolution of selfing in Arabidopsis thaliana. Science 317: 1070-1072

Tatra GS, Miranda J, Chinnappa CC, Reid DM (2000) Effect of light quality and 5-azacytidine on genomic methylation and stem elongation in two ecotypes of Stellaria longipes. Physiologia Plantarum 109: 313-321

Tetard-Jones C, Kertesz MA, Preziosi RF (2011) Quantitative trait loci mapping of phenotypic plasticity and genotype-environment interactions in plant and insect performance. Philos Trans R Soc Lond B Biol Sci 366: 1368-1379

Todesco M, Balasubramanian S, Hu TT, Traw MB, et al (2010) Natural allelic variation underlying a major fitness trade-off in Arabidopsis thaliana. Nature 465: 632-U129

Tsukahara S, Kobayashi A, Kawabe A, Mathieu O, et al (2009) Bursts of retrotransposition reproduced in Arabidopsis. Nature 461: 423-426

Ungerer MC, Halldorsdottir SS, Purugganan MD, Mackay TFC (2003) Genotype-Environment Interactions at Quantitative Trait Loci Affecting Inflorescence Development in Arabidopsis thaliana. Genetics 165: 353-365

Van Kleunen M, Fischer M (2005) Constraints on the evolution of adaptive phenotypic plasticity in plants. New Phytologist 166: 49-60

Vaughn MW, Tanurdzic M, Lippman Z, Jiang H, et al (2007) Epigenetic natural variation in Arabidopsis thaliana. Plos Biology 5: 1617-1629

Verhoeven KJ, Jansen JJ, van Dijk PJ, Biere A (2010) Stress-induced DNA methylation changes and their heritability in asexual dandelions. New Phytol 185: 1108-1118

Verslues PE, Lasky JR, Juenger TE, Liu TW, et al (2014) Genome-wide association mapping combined with reverse genetics identifies new effectors of low water potential-induced proline accumulation in Arabidopsis. Plant Physiol 164: 144-159

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

Yang J, Benyamin B, McEvoy BP, Gordon S, et al (2010) Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42: 565-569

Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, et al (2011) Genome partitioning of genetic variation for complex traits using common SNPs. Nat Genet 43: 519-525

Zhang YY, Fischer M, Colot V, Bossdorf O (2013) Epigenetic variation creates potential for evolution of plant phenotypic plasticity. New Phytol 197: 314-322

Zuk O, Hechter E, Sunyaev SR, Lander ES (2012) The mystery of missing heritability: Genetic interactions create phantom heritability. Proceedings of the National Academy of Sciences of the United States of America 109: 1193-1198

Zuk O, Schaffner SF, Samocha K, Do R, et al. (2014) Searching for missing heritability: Designing rare variant association studies. Proceedings of the National Academy of Sciences of the United States of America 111: E455-E464

Zust T, Heichinger C, Grossniklaus U, Harrington R,et al (2012) Natural Enemies Drive Geographic Variation in Plant Defenses. Science 338: 116-119

1

CHAPTER 2

GWA mapping and genomic prediction reveal the genetic architecture of quantitative traits in *Arabidopsis thaliana*

Authors:

Rik Kooke[1,2,3], Frank Becker[2], Ralph Bours[1], André Kuhn[1], Henri van der Geest[3], Willem Kruijer[4], Jaap Buntjer[5], Timo Doeswijk[5], José Guerra[5], Harro J. Bouwmeester[1], Dick Vreugdenhil[1,3] and Joost J.B. Keurentjes[2,3]

[1] Laboratory of Plant Physiology, Wageningen University, the Netherlands
[2] Laboratory of Genetics, Wageningen University, the Netherlands
[3] Centre for Biosystems Genomics, Wageningen Campus, the Netherlands
[4] Biometris, Wageningen University, the Netherlands
[5] Keygene, Wageningen, the Netherlands

Abstract

Missing heritability is a well-known phenomenon in human genome-wide association studies (GWAS) where the genetic information cannot fully explain the observed phenotypic variation. In the model plant species, *Arabidopsis thaliana*, the issue of missing heritability has not yet been addressed in great detail leading to the false assumption that most common phenotypic variation is explained by a relatively low level of common genetic variation. Here, we analysed 350 Arabidopsis accessions and found extensive variation and high heritabilities for different shoot morphological traits. Furthermore, some of the traits showed significant correlations with climatic factors. However, the number of significant genome wide associations was very low. The quantitative nature of these traits suggests that there are many loci contributing to the phenotype, each with a rather small effect which prevents them from exceeding the significance threshold. Genomic selection models were applied and they revealed different genetic architectures for the morphological traits with some traits being more complex than others. The complexity indicates that many loci of small effect contribute to the phenotypes and indeed when the significance threshold was lowered, numerous plausible candidate genes could be identified. These genes were analysed for function and sequence diversity and good indications that natural allelic variation in many of these genes contributes to phenotypic variation were obtained. *ACS11*, an ethylene biosynthesis gene underlying a QTL for the ratio between petiole and leaf length (PL/LL), was differently expressed between accessions and ectopically supplied ethylene abolished the difference in PL/LL between the accessions.

Introduction

The natural phenomena of mutation and recombination that change the genetic code with each generation have given rise to the enormous genetic diversity between and within species. During evolution, plants have accumulated a vast number of molecular polymorphisms that enabled adaptation to a wide range of environments. With the recent innovations in genetic and genomic tools, the nucleotide diversity can be fully surveyed to find causal polymorphisms for many different phenotypes. This will allow the identification of molecular changes that provided evolutionary advantages and facilitate the search for solutions to feed a fast growing human population in a changing climate.

Through variation in performance, plants have adapted to different environments. Plant performance is directly determined by life history traits, such as flowering time, fecundity and growth rate, which in turn depend on genetics, morphology and physiology (Roff, 2007). Understanding the regulation of plant growth and morphology is therefore essential for the complete comprehension of plant performance and productivity. *Arabidopsis thaliana* has adapted to a wide range of environments and displays an extensive variety in morphological and growth-related phenotypes. Its small genome size, the publicly available genome sequences of over 1000 accessions and its short life cycle make *A. thaliana* an excellent model for the study of natural variation (Horton et al., 2012; Weigel, 2012).

Over the last decades, bi-parental mapping populations have been extremely valuable in the detection of QTLs responsible for trait variation between segregating progeny of two divergent parents (Alonso-Blanco and Koornneef, 2000; Koornneef et al., 2004; Alonso-Blanco et al., 2005; Alonso-Blanco et al., 2009). Identifying the underlying genes, however, remains a laborious and time-consuming effort. Moreover, these QTLs explain the variation between two accessions, but this does not guarantee relevance in an evolutionary context as the QTLs might reflect rare alleles, and the full range of natural variation is not covered by such QTLs. Genome-wide association studies (GWAS) profiting from a wide allelic diversity and high resolution were expected to fill the gap between QTLs and candidate genes and at the same time lead to the identification of more evolutionary relevant variation (Atwell et al., 2010; Bergelson and Roux, 2010). GWAS already confirmed many of the previously identified genes in experimental mapping populations and mutant studies, but did not yet lead to the identification of many novel causal genes (Atwell et al., 2010; Brachi et al., 2010; Li et al., 2010; Todesco et al., 2010; Chan et al., 2011; Chao et al., 2012; Filiault and Maloof, 2012; Sterken et al., 2012; Yano et al., 2013).

The lack of detection power in GWAS might be caused by the extreme polygeneity and plasticity of quantitative traits with very small effect sizes contributing to missing heritability, as concluded from human GWAS (Yang et al., 2010; Gibson, 2011; Makowsky et al., 2011). However, the currently available large populations, dense genotyping and the advantage of

homozygous lines in the mainly self-fertilising species Arabidopsis, should greatly enhance the power of studies of polygenic traits.

Here we present the identification of novel candidate genes involved in plant growth and architecture through the application of GWA mapping in a population of 349 densely genotyped natural accessions of Arabidopsis. Relating the observed morphological differences between accessions to climatological data from their site of origin revealed that most traits were to a certain extent adaptive to climate. In line with this, most candidate genes contained non-synonymous single nucleotide polymorphisms (nsSNPs) and were subject to purifying selection. For the petiole to leaf length ratio, a QTL harboring *ACS11*, a gene involved in ethylene biosynthesis, was detected. We show that *ACS11* is expressed in petioles and that nsSNPs in the gene define two distinct haplotypes, suggesting that ethylene is the signal causal for the phenotypic differences between the two haplotypes.

Results

Phenotypic variation

A collection of 349 natural accessions of *Arabidopsis thaliana* assembled to contain maximum genetic diversity and least population structure (Li et al., 2010; Horton et al., 2012) was analysed to assess the extent of natural variation in shoot morphology. The abbreviations for the morphological traits are listed in Supplemental Table 1. All analysed traits showed extensive phenotypic variation, not only globally, but also within geographical classes (Figure 1). The least variation was found on the North American continent, most likely due to the recent introduction of the species and the relatively short period of adaptation (Platt et al., 2010). Northern European accessions, mainly from Sweden, showed the greatest deviation from other geographical classes in terms of flowering time, rosette branching, plant height and petiole length.

Because most traits displayed different distributions in the analysed geographical classes it was examined whether the trait variation could be partly due to adaptation to the local climate. To investigate the relationships, climatological data from the collection site of the accessions was obtained (Kistler et al., 2001; New et al., 2002; Hancock et al., 2011). Because genetic polymorphisms may be strongly correlated with climate simply due to demographic history and genetic relatedness between accessions may bias the correlations, we used a correlation method that allows correction for population structure (Hancock et al., 2011). Specifically, a Mantel correlation matrix based on Spearman's rank that controls for population structure was generated among the morphological traits and between the morphological traits and climatological data (Supplemental table 2 and 3).

First, the correlations among the morphological traits were compared. All inflorescence

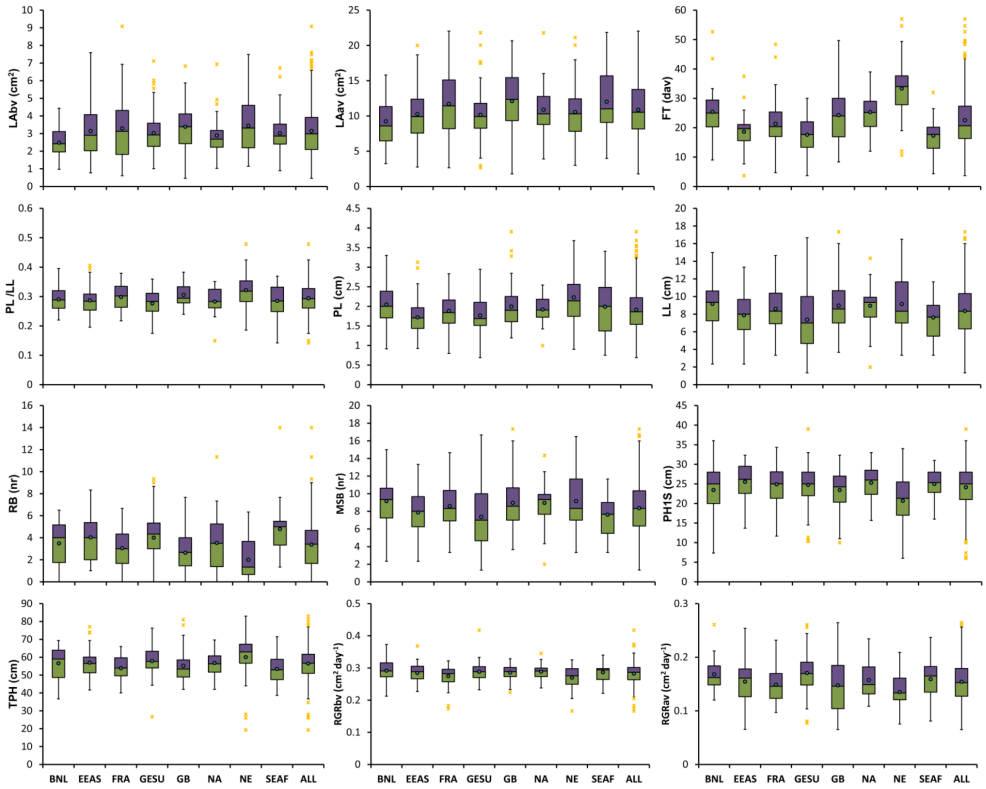


Figure 1. Boxplots showing statistical distribution of morphological traits divided over geographical origin based on the minimum (value > first quartile-1.5*IQR), first quartile, median, third quartile and maximum (value < third quartile+1.5*IQR). The blue dot indicates the mean value, while the orange stars depict suspected outliers (Tukey test). Abbreviations (nr of accessions): BNL, Belgium and The Netherlands (21); EUAS, Eastern Europe and Asia (56); FRA, France (61); GESU, Germany and Suisse (65); GB, Great Britain (48); NA, North America (30); NE, Northern Europe (47); SEAF, Southern Europe and Africa (19); ALL, all accessions (349, incl. 2 accessions from regions outside designated classes); LAbv, leaf area before vernalisation; LAav, leaf area after vernalisation; FT, flowering time; PL/LL, petiole to leaf length ratio; PL, petiole length; LL, leaf length; RB, rosette branching; MSB, main stem branching; PH1S, plant height at 1st silique; TPH, total plant height; RGRbv, relative growth rate before vernalisation; RGRav, relative growth rate after vernalisation.

related traits correlated moderately with flowering time ($0.09 < r > 0.25$, $P < 0.001$), which suggests that flowering time has a considerable impact on most other morphological traits (Supplemental table 2). The correlations are most likely caused by the requirement of long-term vernalisation to initiate flowering for some accessions. Strong correlations were also detected between leaf area and relative growth rate and among the leaf length traits. Comparison between Mantel correlations and uncorrected Spearman correlations revealed

that the Mantel correlations are somewhat lower and even opposite for the correlation between relative growth rate before vernalisation and leaf area after vernalisation (Mantel r = 0.35, Spearman r = -0.36) showing that correction for population structure indeed influences the correlations among morphological traits.

Subsequently, the morphological traits were correlated to the climatological data. Flowering time, main stem branching and plant height at 1st silique correlated significantly with a number of different geographical and climatological factors, such as latitude, number of wet days, temperature, precipitation and ground frost ($P < 0.05$) (Supplemental table 3). Furthermore, day length and wind speed correlated significantly ($P < 0.05$) with a large number of morphological traits (i.e. FT, PL, PL/LL, RB and TPH). The correlations were generally low and when Bonferroni correction was applied, nearly all correlations were insignificant ($P > 0.05$). Uncorrected Spearman correlations were slightly higher than the Mantel correlations, especially for the correlation between latitude and flowering time (Mantel r = 0.09, Spearman r = 0.35), which indicates that population structure also has a major influence on these correlations. Nonetheless, the Mantel correlations were comparable to (Hancock et al., 2011) and are quite substantial in an evolutionary context. This analysis thus suggests that the variation in morphology is to a certain extent resulting from climate adaptations.

Heritability and GWAS

To determine to what extent the morphological variation is defined by the underlying genetic variation, broad-sense heritability (H^2) and the marker-based heritability based on individual plant data (h^2) were estimated. The H^2 and h^2 were calculated using ANOVA and the EMMAX mixed-model used for GWAS, respectively. The heritability estimates were very comparable between the two methods, although h^2 was generally somewhat higher. The small difference in calculations is most likely caused by the sequence data that were lacking for ten accessions, which were therefore excluded from the heritability measurements in the marker-based estimation. h^2 Ranged from 0.42 for relative growth rate before vernalisation to 0.93 for flowering time with the majority of traits having a heritability higher than 0.70, suggesting that most of the variation could be attributed to genetic variation (Supplemental table 4). Next, it was assessed whether significant associations could be detected between a set of approximately 200,000 genome-wide SNPs and the variation in morphological traits, using a linear mixed model that corrects for population structure (EMMAX). Using a stringent Bonferroni corrected threshold of $-\log_{10}(P) = 6.6$, only three significant SNPs could be detected, two for flowering time at the known *flowering locus C* (*FLC*) and one for rosette branching (Figure 2). The discrepancy between high heritability and a very low number of significantly associated SNPs might be due to a number of reasons. The linear mixed model, for instance, does not account for epistatic, GxE and epigenetic effects, which might contribute strongly to phenotypic variation (Gibson, 2011). Moreover, the model might be

too conservative in applying Bonferroni corrected thresholds (Riedelsheimer et al., 2012) or wrongly correct for population structure (Filiault and Maloof, 2012). Furthermore, allelic (or genetic) heterogeneity (Johanson et al., 2000; Atwell et al., 2010; Barboza et al., 2013), a too small sample size (Korte and Farlow, 2013), rare alleles (Gibson, 2011) and high numbers of small-effect loci (Yang et al., 2010) might prevent the significant association of SNPs with variation in trait values. If many genes contribute to a phenotype the isolated effect of each gene is too small to exceed the stringent significance threshold.

Genomic selection

To be able to assign much more of the genetic variation to the observed population variation in phenotypes, a genomic selection model was applied to all traits. Genomic selection (GS) models assign allele substitution effects to all markers for the prediction of trait values (Meuwissen et al., 2001). For all morphological traits, a genomic selection model on the effects of the 100,000 most informative SNPs was built that was capable of predicting the phenotype at high accuracy in cross validation (Supplemental figure 1). The heteroscedastic effect models, which allow the variance to be adjusted at each locus, explain between 49% and 75% of the variation in the various traits (Supplemental figure 1) (Shen et al., 2013). Although GS models are not intended for GWA locus identification, when the allele substitution effects were plotted, a visual pattern was observed emerging from the variance in the effect distribution for expected genetically less complex traits, such as flowering time versus complex traits, such as relative growth rate (Supplemental figure 2). These differences in distribution patterns were quantified and a dendrogram was constructed using UPGMA clustering, demonstrating the relative genetic complexity of all traits (Supplemental figure 3). Individual marker contributions, as deduced from their effect-size, were lowest for the petiole/leaf length ratio and relative growth rate before and after vernalization in the models. Flowering time and total plant height, and to a lesser degree plant height at 1st silique and leaf area after vernalisation, are less complex than all other traits (Supplemental figure 3). Indeed, for flowering time, we found two SNPs above the Bonferroni threshold and a high number of loci above the $-\log_{10}(P) = 4$ threshold. The same was true for total plant height and plant height at 1st silique that have a much higher number of SNPs above the $-\log_{10}(P) = 4$ threshold than all other traits apart from flowering time. A high number of SNPs above the $-\log_{10}(P) = 4$ threshold indicates that a (relatively) large part of the genetic variance can be explained by a small number of SNPs provided that there is an equal amount of false-positives for each trait. The genomic inflation factors give a good indication of the expected false-positives for each trait, and for all traits the inflation factors were close to 1 (between 0.99 and 1.1), which suggests indeed that an equal amount of false-positives can be expected for each trait (de Bakker et al., 2008). Overall, these results imply that much of the phenotypic variation can be explained by genetic diversity, but that the high number of contributing loci and their inherent small effect sizes hamper the significant association of individual loci in GWAS of complex traits.

Candidate genes

The GS models demonstrate that most of the phenotypic variation can be explained by the additive effects of multiple SNPs. To locate the SNPs that cause the phenotypic variation, all SNPs that exceeded the arbitrary $-\log_{10}(P) = 4$ threshold, which is substantially above the "noise" level, were regarded as candidate SNPs. For all these SNPs, the LD support interval ($r^2 > 0.3$) was determined and each gene was analysed within the support intervals for sequence diversity, function and expression profile using publicly available data. Based on this combined information a number of candidate genes for each trait were selected (Figure 2, Supplemental table 5 & 6).

Two candidate genes for flowering time, *FLC* and *DOG1*, have previously been proven to be involved in the phenotype and have allelic variation explaining the phenotypic variation (Koornneef et al. 1994, Bentsink et al. 2006). Both genes were also detected in previous GWA studies (Atwell et al. 2009, Brachi et al. 2010). Analysis of further known flowering time genes yielded weak QTLs for *FRI*, *TOR1*, *SVP* and *PHYB* (Figure 2, Supplemental table 5).

Three candidate genes for leaf area after vernalization, rosette branching and main stem branching, respectively *ASN2*, *BRANCHED1* and *BIL4* were previously shown to display knock-out phenotypes for the respective traits but natural allelic variation within the genes has not been reported before (Figure 2, Supplemental table 6; (Aguilar-Martinez et al., 2007; Yamagami et al., 2009; Gaufichon et al., 2013; Gonzalez-Grandio et al., 2013; Niwa et al., 2013). The recently released re-sequencing data from the Arabidopsis 1001 genomes project enabled the comparison of the gene sequences of 530 accessions (http://signal.salk.edu/atg1001/3.0/gebrowser). For a subset of these accessions (174) phenotypic information was obtained from our study allowing a detailed linkage analysis. Both *ASN2* and *BIL4* have a non-synonymous polymorphism (nsSNP) within the gene which is in high LD ($r^2 > 0.3$) with the most significant GWA SNP, while *BRANCHED1* contains a nsSNP in low LD ($r^2 = 0.11$) with the most significant GWA SNP in the LD region (Supplemental table 5). To determine whether natural selection had acted on these loci, we compared the nucleotide diversity between silent, synonymous sites and non-synonymous sites. For *BRANCHED1* the nucleotide diversity was rather low ($\pi_T = 0.0027$, Supplemental table 7) and the π_{non}/π_{syn} ratio ($\pi_{non}/\pi_{syn} = 0.225$, Supplemental table 7) was considerably lower than the neutral one to one ratio, suggesting that this locus is under purifying selection. This is further illustrated by a significant negative value for Tajima's D ($D_n = -2.187$, $P < 0.01$, Supplemental table 7) at non-synonymous sites, which implies deviation from neutrality (Supplemental table 7). Accessions of the non-Col-0 *BRANCHED1* haplotype developed on average 1.6 rosette branches more than accessions of the Col-0 haplotype. Within the non-Col haplotype, accessions from the Czech Republic (0.39 vs. 0.09) and Spain (0.16 vs. 0.02) are highly overrepresented, indicating strong population structure, possibly due to adaptive mechanisms. Negative values for Tajima's D statistic were also found for *ASN2* and *BIL4*,

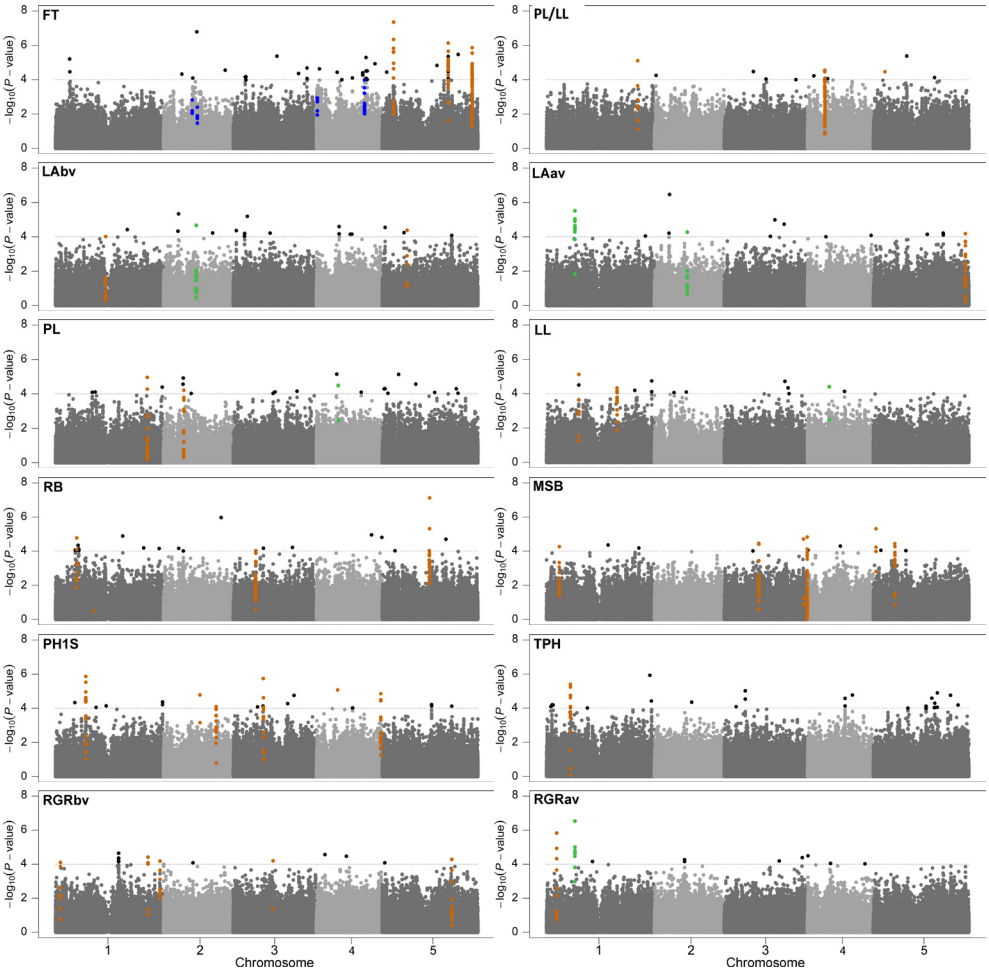


Figure 2. Manhattan plots for morphological traits.
Orange color SNPs indicate SNPs ($r^2 > 0.3$ with candidate SNP) mentioned in Supplemental table 5, green color SNPs are found for multiple traits, blue SNPs are important flowering genes *PHYB*, *SVP*, *FRI* and *TOR1* in respective order on the genome. Abbreviations: LAbv, leaf area before vernalisation; LAav, leaf area after vernalisation; FT, flowering time; PL/LL, petiole to leaf length ratio; PL, petiole length; LL, leaf length; RB, rosette branching; MSB, main stem branching; PH1S, plant height at 1st silique; TPH, total plant height; RGRbv, relative growth rate before vernalisation; RGRav, relative growth rate after vernalisation.

most significantly at silent sites (D_{sil} = -2.294, $P < 0.01$ and D_{sil} = -1.864, $P < 0.05$ for *ASN2* and *BIL4* respectively, Supplemental table 7). These findings suggest that allelic variation at these loci is responsible for the observed phenotypic variation.

For eight other candidate genes strong evidence for purifying selection at non-synonymous

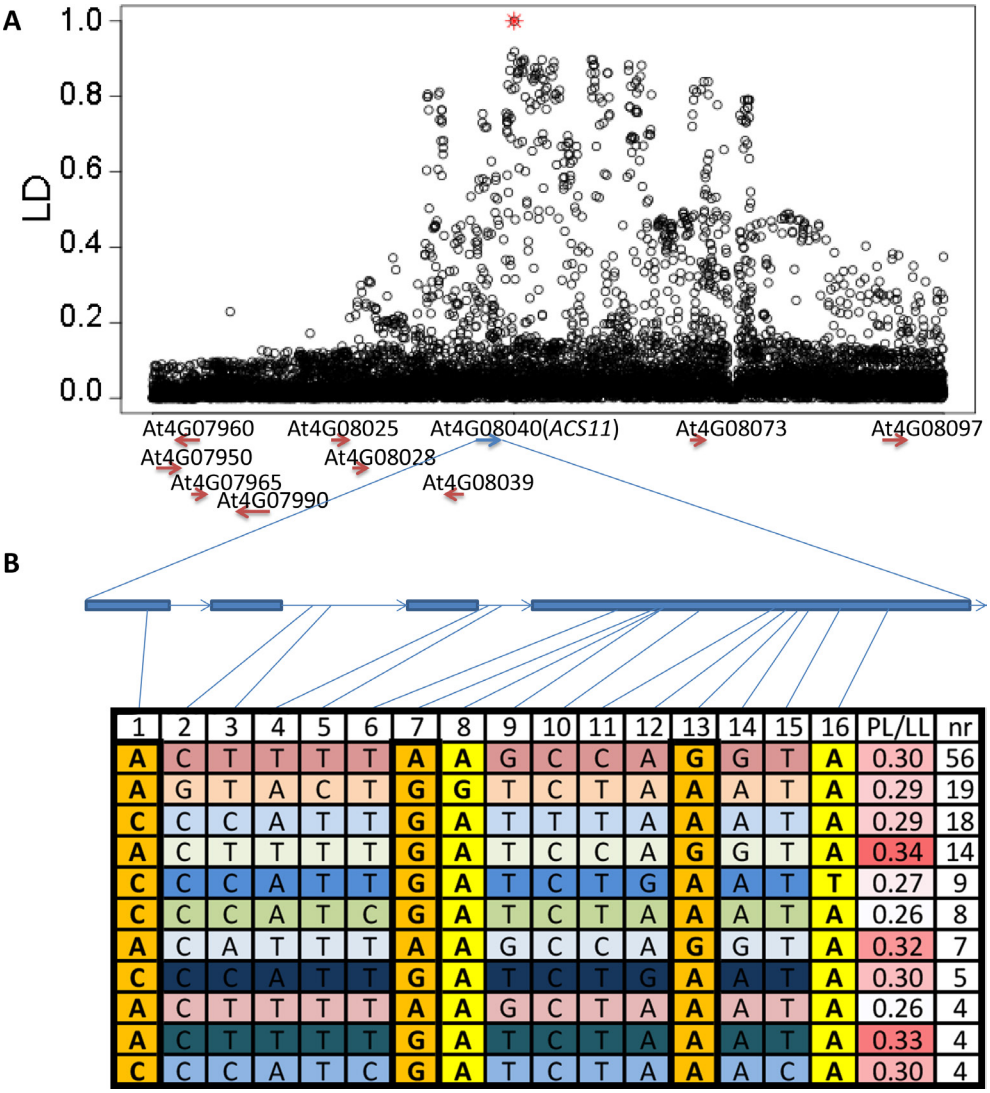
sites was detected (D_n < -2, P < 0.01, Supplemental table 7), while for another eighteen genes, selective pressure was moderate (D_n < -1.7, P <0.05, Supplemental table 7). For fifteen genes, substantial evidence from literature suggests that they play a role in the associated trait, although most of this evidence is circumstantial (Supplemental table 6). Nine of these genes contain one or more nsSNPs within the gene's coding sequence (r^2 > 0.3) and another nine contain polymorphisms in the 1kb upstream promoter region (r^2 > 0.3). Three genes do not contain any nsSNP in LD with the GWA SNP, neither in the gene coding sequence nor in the promoter (Supplemental table 5). Many of the genes displayed signs of purifying selection, having a low π_{non}/π_{syn} ratio and a negative value for Tajima's D statistic (Supplemental table 7). Twenty-nine other selected genes are members of gene families known to be involved in the regulation of the corresponding trait (Supplemental table 6). Two of these genes, *CESA9* and *NI*, were selected for more than one trait. *CESA9* was selected for leaf area before and after vernalisation, while *NI* was selected for leaf area after vernalisation and relative growth rate after vernalisation (Supplemental table 6).

For the majority of candidate genes, either nsSNPs (61% of candidate genes) or SNPs in the 1kb upstream promoter region (51% of candidate genes) in LD (r^2 > 0.3) with the most significant GWA SNP were identified (Supplemental table 5). For only 21% of the candidate genes, no nsSNPs or SNPs in the promoter regions in LD with the most significant GWA SNP could be identified. As SNPs in the UTR, introns and at synonymous sites may still play a role in determining gene expression or gene function, these candidate genes cannot be completely ruled out a priori. In conclusion, the strong enrichment in candidate genes for which independent additional indirect evidence of their involvement in the trait under study is available indicates that the GWA mapping approaches can greatly contribute to the selection of candidate genes. Moreover, these results show that genome-wide thresholds for GWA mapping are in general too stringent for the detection of small-effect loci controlling quantitative traits.

ACS11

To further validate our gene candidate discovery approach, an interesting QTL for the ratio between petiole length and leaf length (PL/LL) on the top of chromosome 4, coinciding with an ethylene biosynthesis gene, *ACS11* (Figure 2), was analysed in more detail. ACS enzymes, forming homo-and heterodimers, are thought to catalyze the rate limiting step in ethylene biosynthesis converting S-adenosyl-L-methionine into 1-aminocyclopropane-1-carboxylate, the precursor of ethylene (Bleecker and Kende, 2000). Analyses of single, double and multiple *acs* mutants revealed that ACS enzymes play essential roles in leaf development, flowering time, disease resistance and ethylene production (Tsuchisaka et al., 2009).

ACS11 and its most significantly associated SNP are in strong LD with a large region of 140 kB in which mainly transposable elements, pseudogenes and a number of small genes are



1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	PL/LL	nr
A	C	T	T	T	T	A	A	G	C	C	A	G	G	T	A	0.30	56
A	G	T	A	C	T	G	G	T	C	T	A	A	A	T	A	0.29	19
C	C	C	A	T	T	G	A	T	T	T	A	A	A	T	A	0.29	18
A	C	T	T	T	T	G	A	T	C	C	A	G	G	T	A	0.34	14
C	C	C	A	T	T	G	A	T	C	T	G	A	A	T	T	0.27	9
C	C	C	A	T	C	G	A	T	C	T	A	A	A	T	A	0.26	8
A	C	A	T	T	T	A	A	G	C	C	A	G	G	T	A	0.32	7
C	C	C	A	T	T	G	A	T	C	T	G	A	A	T	A	0.30	5
A	C	T	T	T	T	A	A	G	C	T	A	A	A	T	A	0.26	4
A	C	T	T	T	T	G	A	T	C	T	A	A	A	T	A	0.33	4
C	C	C	A	T	C	G	A	T	C	T	A	A	A	C	A	0.30	4

Figure 3. Haplotype of *ACS11*.

(A) LD around the most significant SNP associated with PL/LL ratio, close to AT4G08040, *ACS11*. The LD extends up to 140 kB. **(B)** Haplotypes observed in 174 phenotyped accessions within the 530 re-resequenced accessions. The orange coloured SNPs are the most informative in relation to the phenotype. Both the yellow and orange SNPs are non-synonymous polymorphisms. All the other SNPs are either synonymous polymorphisms or located in introns. The nr indicates the total nr of accessions in each haplotype.

located (Figure 3). Extensive LD is indicative of a selective sweep suggesting strong adaptive selection of allelic variation. Indeed, within the *ACS11* coding region, five nsSNPs can be identified of which two are in high LD ($r^2 > 0.3$) with the most significant GWA SNP (Figure 3). The most informative SNP related to the PL/LL phenotype is located in the fourth exon, substituting a nonpolar glycine for a polar serine. Two other nsSNPs associated with a significant effect on the trait are located in the first exon, substituting proline (nonpolar) for threonine (polar) and in the fourth exon, substituting aspartic acid (negatively charged) for glycine (nonpolar). The nucleotide haplotype AGG, representing these three polymorphisms displays the highest PL/LL ratio and differs significantly from all other haplotypes (Figure 3). All accessions in the GWA population belonging to this haplotype originate from Sweden. In addition, further analysis of resequenced accessions not included in the GWA population (http://signal.salk.edu/atg1001/3.0/gebrowser) revealed a similar geographic distribution, with the exception of one accession originating from Finland.

These findings indicate a selective sweep in Scandinavian accessions suggesting an evolutionary advantage of altered PL/LL ratios at higher latitudes (Figure 1). If selection is limited to geographical regions this can result in strong population structure. Correcting for population structure, as done in the applied GWA model, can considerably reduce the GWA likelihoods and lead to false negative associations. Omitting population structure correction in the GWA model increased the significance of the *ACS11* locus, exceeding the Bonferroni threshold (Figure 4A).

To determine the expression profile of *ACS11*, an *ACS11::GUS* reporter line (Tsuchisaka and Theologis, 2004) was analysed, revealing that *ACS11* is indeed expressed in the mid vein of the petiole, and not in the leaf blade (Figure 4B). Subsequently, a subset of accessions, representing different haplotypes, was analysed for differences in gene expression (Supplemental table 8, Supplemental figure 4). Although no significant difference could be observed between the two most distinct haplotypes, variation in gene expression was detected (Supplemental figure 4). Accession var2-1, belonging to the AGG haplotype (Supplemental table 8), showed the highest expression, and accessions belonging to the non-Col haplotype, haplotype 0, at the most significant SNP (Chr. 4: Pos. 4888589, Supplemental table 8) on average expressed *ACS11* at higher levels ($P < 0.1$) (Supplemental figure 4). Although these results indicate a role for transcriptional regulation, the lack of strong expression differences suggests that functional variation is a more likely explanation for the phenotypic differences. A T-DNA knockout of *ACS11* was analysed and a minor but non-significant reduction in the PL/LL ratio was detected (Supplemental table 8). The non-significance might be due to redundancy and/or the Col-0 genetic background. Indeed, the petiole length in an *acs octuple* mutant and in an ethylene insensitive mutant, *ein2-1* (Alonso et al., 1999; Tsuchisaka et al., 2009) was significantly reduced, although concomitant with a similar reduction in leaf length (Supplemental table 8). Moreover, application of ethephon,

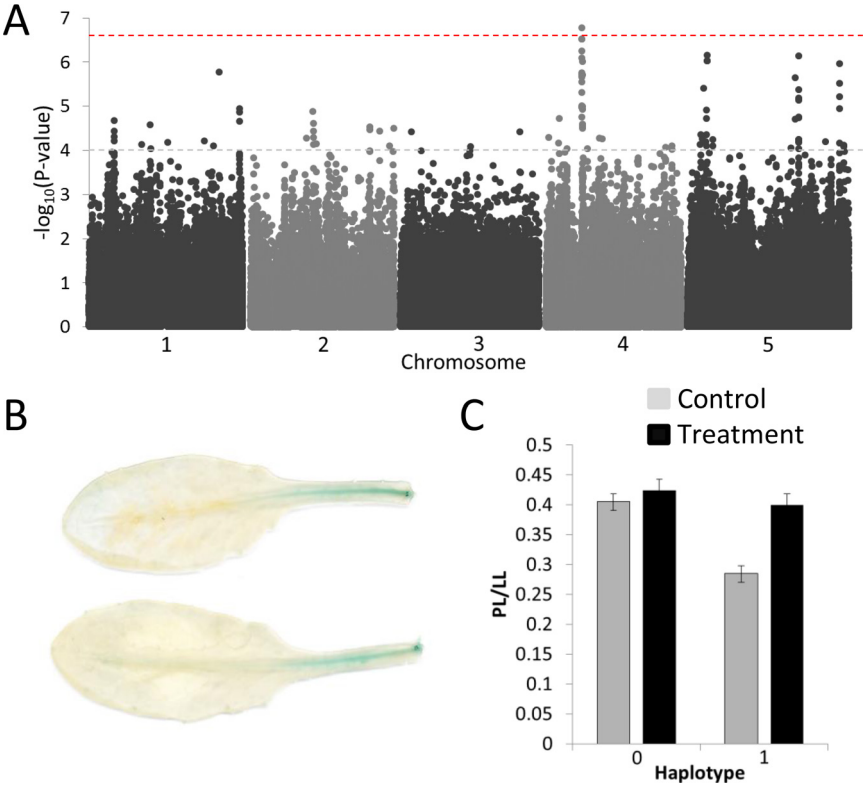


Figure 4. *ACS11* candidate gene confirmation.
(A) Manhattan plot for PL/LL ratio from a linear GWAS model (LM). Grey, dashed bar indicates – $\log_{10}(P) = 4$ threshold; red, dashed bar indicates Bonferroni threshold. **(B)** *ACS11::GUS* expression in 3-week old leaves. **(C)** Ethylene complementation experiment between two haplotypes. Treated plants were sprayed with 0.5mM ethephon. Results are shown 3 days after ethephon treatment. Haplotype 0 denotes the average of 10 accessions with 10 replicates of the G haplotype at the 4th nsSNP. Haplotype 1 denotes the average of 10 accession with 10 replicates of the A haplotype at the 4th nsSNP.

a compound that after application is quickly metabolised into ethylene, to the two most informative haplotypes suggests ethylene to be the signal explaining the variation in PL/LL ratios (Supplemental table 8, Figure 4C). Under controlled conditions, a significant difference ($P < 0.0001$) in PL/LL ratio between the two haplotypes was detected, which disappeared 3 days after ethephon application (Figure 4C). Altogether, these findings suggest that *ACS11* is the causal gene underlying a strong association with PL/LL ratios and that selective pressure favored the proliferation of accessions with high PL/LL ratios at Northern latitudes, either through altered expression or functional diversification.

Discussion

Phenotypic and geographic variation

From this and other studies it is evident that extensive variation for morphological traits is present in natural populations of *Arabidopsis thaliana*. Interestingly, the extent of global natural variation is not much larger than the variation present within geographical regions, especially considering that the number of accessions per geographical class is about one tenth of the global set of accessions (Figure 1). Similar results were obtained for flowering related traits in French local populations compared with a global population (Brachi et al., 2013). This argues for GWAS on regional populations in which the confounding effects of population structure and allelic heterogeneity are substantially reduced. Furthermore, the global population used in this study might not be optimal for gene identification in GWAS. The high number of alleles that segregate in this population with sometimes low allele frequencies cause a low detection power, which can result in relatively low $-\log_{10}(P)$ values. The larger allelic diversity in a global population, however, also has its advantages. It, for instance, allows the comparison with climatic gradients. Moreover, fitness is enhanced through environmental adaptation at loci that are polymorphic in the same environment (Fournier-Level et al., 2011). These loci do not necessarily affect fitness in other environments, suggesting a local genetic basis for adaptation. Such alleles can only be detected by comparing multiple local populations or considering a global scale of adaptation.

When comparing the phenotypic variation between the geographical classes, the North American accessions show the least variation for most of the phenotypes (Figure 1). As the species was introduced only about 300 years ago in North America, it had a much shorter time to migrate and evolve than its Eurasian counterparts (Platt et al., 2010). Although Arabidopsis is common across the entire North American continent, it shows much less haplotype diversity and weaker isolation by distance compared to accessions from Eurasia (Platt et al., 2010). The Northern European accessions showed the greatest deviation from all other classes, which is probably due to the requirement of vernalisation and long days to initiate flowering. Most developmental phenotypes are thought to be very dependent on this transition and *FLC*, a major flowering locus, impacts plant development by regulating a great number of developmental genes that are important throughout the plant's life cycle (Deng et al., 2011). Moreover, it was recently found that the floral integrators *FLC* and *FRI* regulate stem branching in an epistatic manner (Huang et al., 2013). These observations explain to a large extent why most of the phenotypes in this study correlate well with flowering time.

Most of the phenotypes also showed significant correlations with climate variables suggesting that local adaptation is partly driven by climate (Supplemental table 3). There are many positive correlations between latitude, day length (spring) and many of the morphological

traits. This relationship is well explained by the delayed flowering time at Northern latitudes due to the vernalisation requirement and longer day lengths during spring (Hancock et al., 2011). Moreover, the impact of cold temperatures and precipitation on the genetics of important flowering genes has also been reported before (Mendez-Vigo et al., 2013b). The correlation between latitude and day length (spring) with petiole length and the petiole to leaf length ratio is interesting in light of the geographical distribution of *ACS11* haplotypes described here (Figure 3). Our findings suggest that a longer petiole to leaf length ratio is adaptive at higher latitudes. It is worthwhile to note here that mutations in the *ACS* gene family also affect *FLC* expression and flowering time (Tsuchisaka et al. 2009). Some mutations tend to induce flowering while others, concomitant with a decrease in ethylene production, seem to delay flowering. Besides the effect of *ACS* genes on flowering time, the number of branches is significantly reduced and plant height is significantly increased in the *acs* mutants (Tsuchisaka et al., 2009).

Heritability

Most traits analysed thus far with GWAS in Arabidopsis such as flowering time and resistance/ avirulence interactions are to a large extent defined by one or a few genes and the questions of missing or hidden heritability have therefore not been addressed yet (Atwell et al., 2010; Brachi et al., 2010). Moreover, not many novel candidate genes that were not already previously identified in bi-parental populations or mutant studies were identified using GWAS. All phenotypes tested in this study were highly heritable with a heritability ranging from 0.42 for relative growth rate before vernalisation to 0.93 for flowering time (Supplemental table 4). This is similar to or even slightly higher than the heritabilities measured in many bi-parental mapping populations for the same traits (Ungerer et al., 2002; Bandaranayake et al., 2004; Keurentjes et al., 2007). However, only a very limited number of candidate genes could be detected above the stringent Bonferroni threshold (Supplemental table 4, Figure 2). As all phenotypes tested here are likely to be highly polygenic, it appears that most of the effects of the underlying genes are too small to be captured in GWAS. As is the case in human GWAS, it is expected that the heritability is hidden rather than missing (Gibson, 2010). In the human field, most of the variation for human height can be explained by the additive effects of individual SNPs, showing that the common genetic variation is able to explain a large part of the variation for height (Yang et al., 2010). As plant phenotypes, due to the plant's sessile nature, might be even more complex, we reasoned that similar mechanisms might explain the quantitative traits in the present study. Therefore, a genomic selection model explaining most of the genetic variation by a set of the most informative 100,000 SNP markers was adopted. Indeed, the predicted values from the GS model correlated well with the observed phenotypic values with correlations ranging from 0.48 to 0.75 for the heteroscedastic model (Supplemental figure 1). This indicates that a large percentage of the phenotypic variation can be explained by the additive effects of individual SNPs and the various resulting alleles. It does, however,

not give exclusive evidence for an infinitesimal model as part of the missing heritability can still be caused by rare alleles. However, given the strong adaptation to climate and long-standing common genetic variation in Arabidopsis, we consider this option rather unlikely.

To illustrate this point, we lowered the significance threshold to the arbitrary $-\log_{10}(P) = 4$ threshold, enabling the assignment of many strong candidate genes to a majority of QTLs (Supplemental table 5, 6 and 7). For flowering time, two previously confirmed genes were identified, *FLC* and *DOG1*, illustrating the validity of the approach (Koornneef et al., 1994; Bentsink et al., 2006; Atwell et al., 2010; Brachi et al., 2010). Besides these two genes, a number of weak QTLs were found for known flowering time genes. Although it is known that allelic variation exists around the *FRI*, *SVP* and *PHYB* loci, allelic heterogeneity within these loci or the rather small effect compared to other flowering time genes is likely to be the cause of the rather weak QTLs (Halliday et al., 1994; Johanson, 2000; Borevitz et al., 2002; Filiault and Maloof, 2012; Mendez-Vigo et al., 2013a). This further indicates that many genes impacting the phenotype have a too small effect to be captured significantly in GWAS. To overcome these difficulties, one could invest in local or larger GWAS populations to increase mapping power (Korte and Farlow, 2013).

An interesting QTL was found on chromosome 3 for rosette branching, where an association was found with the *BRANCHED1* locus. *BRANCHED1* is a signal integrator controlling bud outgrowth and arrest dependent on different hormonal pathways and important flowering time genes, such as *FT* and *PHYB* (Aguilar-Martinez et al., 2007; Gonzalez-Grandio et al., 2013; Niwa et al., 2013). Homologs of *BRANCHED1*, known as *TB1*, are also present in rice, maize and sorghum and seem to be conserved among the angiosperms (Aguilar-Martinez et al., 2007). Good indications of purifying selection were found at this locus and a nsSNP in the first exon that converts isoleucine (nonpolar amino acid) to threonine (polar amino acid), possibly changing protein structure and function. The nsSNP was also found to be significantly associated with the phenotype, and it was overrepresented in accessions from the Czech Republic and Spain. Follow-up studies should show if the allelic variation at the locus is indeed responsible for the difference in phenotype, and whether it is maintained in local populations across the world.

For the majority of candidate genes significant negative values for Tajima's D statistic were found, supported by a low nucleotide diversity π_T and a low π_{non}/π_{syn} ratio, suggesting purifying selection at the locus. Note, however, that demographic history can dramatically skew natural diversity estimates, especially in Arabidopsis (Nordborg et al., 2005; Schmid et al., 2005). The values for Tajima's D are substantially negatively skewed in this species and the results should therefore be taken with caution.

Interestingly, twenty-nine genes were found for which good indications exist that the gene

or gene family plays a role in the phenotype (Supplemental table 6). These results add to the body of evidence gathered from mutant studies in the reference accession Col-0 as allelic diversity might depend on the genetic background. Certain alleles might display similar functions in Col-0, but act differently in other accessions, playing a significant role in determining the phenotype. For example, it was found that *SPATULA* mutants in a Col-0 background show an opposite phenotype as in a L*er* background (Vaistij et al., 2013). Further studies using transformation, gene knockdown and expression analyses in different accessions are, therefore, needed to determine allelic diversity and gene function.

In depth analysis of a candidate gene: *ACS11*

An ethylene biosynthesis gene from the large *ACS* gene family, *ACS11*, was found to underlie the QTL for petiole to leaf length (PL/LL) ratio. We showed that *ACS11* is expressed in petioles and that ethylene can nullify the differences in the petiole to leaf length ratio between two distinct haplotypes (Figure 4). From previous studies, it is known that ethylene can significantly influence petiole length and angle under different conditions and that ACS enzymes affect ethylene production (Tsuchisaka et al., 2009; van Zanten et al., 2009; Bours et al., 2013). All *ACS* genes are expressed in petioles, but only *ACS2*, *8* and *11* are differentially expressed between petiole and leaf (Bours et al., 2013). A T-DNA knockout of *ACS11* results in a small, but non-significant reduction of the PL/LL ratio (Supplemental table 9). Possibly, other genes in the family take over the function of *ACS11* abolishing the effect of the T-DNA knockout. Moreover, the *ACS11* expression is rather low in Col-0 compared to other accessions making the study of a T-DNA knockout in the Col-0 background difficult. The petiole length in the *acs octuple* mutant and in the *ein2-1* mutant was significantly reduced, although concomitant with a similar reduction in leaf length. Therefore, we could not find a difference in the PL/LL ratio in these mutants. Given that most *ACS* genes are expressed in the leaf blade and the petiole, a knockout of multiple *ACS* genes is expected to reduce the length of both petiole and leaf. Similar results were expected and found for the ethylene insensitive mutant, *ein2-1*. At the haplotype level, an 8% difference in PL/LL ratio between the AGG haplotype and the AAA haplotype was observed (Figure 3). Accessions with the AGG haplotype all originated from Sweden and PL/LL ratio correlated positively with latitude and day length (spring). Longer petioles might enhance light capture and aeration at higher latitudes, and thus provide plants with an adaptive advantage. A non-Col haplotype accession, var2-1, had the highest *ACS11* expression, while accessions with a non-synonymous polymorphism in the last exon had on average a higher *ACS11* expression. Although there is clear variation in *ACS11* transcript abundance, it is likely that other mechanisms also play a significant role in the determination of the phenotype. Sequence variation within the exons gives rise to different amino acids, possibly changing protein structure and/ or function. Given that the ACS enzymes form homo –and heterodimers, the binding between the proteins might also be affected, leading to changes in ethylene production in petioles, and hence in petiole growth (Tsuchisaka et al., 2009).

This study also presents a clear example of the consequences of correcting for population structure. Correcting for population structure does reduce the number of false positives in GWAS, but it could also lead to over-correction and the removal of false negative associations (Bergelson and Roux, 2010). Since the contrasting haplotype of *ACS11* is only found at Northern latitudes, correction for population structure reduced the association significance of the QTL at the *ACS11* locus (Figure 2 and Figure 4). Omitting the correction for population structure increased the QTL likelihood beyond the Bonferroni threshold. This advocates for the careful application of statistical correction models in biological studies as important findings might be overlooked if these are restricted to confined geographical regions.

Material and Methods

Plant growth conditions

Seeds from 350 natural accessions of *Arabidopsis thaliana*, collected worldwide and genotyped with 250K SNPs (Li et al., 2010; Horton et al., 2012), were sown on filter paper with demi water and stratified at 4˚C in dark conditions for 5 d. Subsequently, seeds were transferred to a culture room (16 h LD, 24˚C) to induce seed germination for 42 h. Three replicates per accession were transplanted to wet Rockwool blocks of 4 x 4 cm in a climate chamber (16 h LD, 125 µmol m^{-2} s^{-1}, 70% RH, 20/18˚C day/night cycle). Two control accessions (Col-0 and L*er*-1, each 10 replicates) were transplanted in the middle of the flooding table. All plants were watered daily for 5 min with 1/1000 Hyponex solution (Hyponex, Osaka, Japan). 19 Days after germination, all plants were moved to a cold room (12 h light, 4˚C) for 6 weeks for vernalisation. After the vernalisation period plants were transferred back to the same climate chamber in the same order, but divided over two tables to increase the growth space. For the confirmation of the *ACS11* locus, 21 natural accessions (10 replicates each) were grown, 10 accessions from each haplotype plus Col-0. Ten replicates of *ACS11::GUS*, *acs11-1*, *acs octuple* and *ein2-1* were grown in the same conditions as listed above (Supplemental table 9).

Morphological traits and ethylene complementation

A variety of developmental traits were measured on all individual plants. Rosette photos for leaf area (LA) were taken 15, 19 (LAbv), 63 (LAav) and 68 DAG. Relative growth rate (RGR) was calculated using the following equation: $(\ln(LA_x)-\ln(LA_y))/dt_{(x-y)}$. RGRbv and RGRav were calculated between 19 and 15 DAG and 68 and 63 DAG, respectively. Flowering time (FT) was recorded as the time the first flower opened and photos were taken of individual plants two weeks after flowering. Plant height at the 1st silique (PH1S) was measured two weeks after flowering and total plant height at the end of the growth period (TPH). Branching was measured as the number of branches that were present on the main inflorescence (MSB) two weeks after flowering and as the number of rosette branches (RB) at the end of the growth period. Leaf length (LL) and petiole length (PL) were measured on photographs

taken from the longest leaf two weeks after flowering. For the confirmation of the *ACS11* locus, all plants were phenotyped daily. One leaf (the second leaf of the second whorl) of each plant was marked 3d prior to the start of the experiment without damage to the plant. Both petiole and leaf length were measured from 19 DAG with a caliper for one week. For the ethylene complementation experiment 3wk old rosettes were sprayed with 0.5mM ethephon and 0.005% Tween-20. Control plants were sprayed with mock solutions that lacked the active component. Leaf and petiole length were measured daily for three consecutive days.

Climate Data

Climate data for each accession was obtained from the Climate Research Unit at the University of East Anglia. Data were extracted for 9 climate variables giving the average per month over a 30-year (1961-1990) period (New et al., 2002). From these 9 variables, most other variables were extracted. Day length (spring) and relative humidity (spring) from the origin of 306 accessions were obtained from the NCEP-NCAR climate re-analysis project (Kistler et al., 2001; Hancock et al., 2011) and the FAO GeoNetwork (http://fao.org.geonetworks/srv/en/main.home).

Correlation analyses

A partial Mantel test was used to calculate the Spearman correlations between the morphological traits and between the morphological traits and climate variables correcting for population structure by using a kinship matrix based on the genome-wide SNPs as a covariate in the model (Mantel, 1967). Partial Mantel tests were conducted using the Ecodist package in R (Goslee and Urban, 2007). We assessed the significance by running 1000 permutations on the dependent variable (Smouse et al., 1986).

Descriptive statistics

Coefficient of variation (CV_G) was calculated as $\sigma_G/X*100\%$.

The variance components for all the individual traits were used to calculate the broad-sense heritability, H^2, in analysis of variance (ANOVA) according to the formula

$$H^2 = \sigma^2_G/(\sigma^2_G + \sigma^2_E), \text{ with } \sigma^2_G = (MS(G) - MS(E))/r, \sigma^2_E = MS(E), \quad (1)$$

where r is the number of replicates and MS(G) and MS(E) are the mean sums of squares for genotype and residual error, respectively. Narrow-sense heritability, h^2, is defined as

$$h^2 = \sigma^2_A/(\sigma^2_G + \sigma^2_E), \quad (2)$$

which takes only the additive genetic effects (σ^2_A) in account. Marker-based estimates of narrow-sense heritability can be obtained using the mixed models (3) and (4) defined below, which contain random genetic effects. The covariances between these effects are modeled by a genetic relatedness matrix (GRM) estimated from markers, which is called kinship

matrix K, with elements:

$$Ki,j = \frac{1}{p}\sum_{l=1}^{p}\frac{(x_{i,l} - f_l)(x_{j,l} - f_l)}{4f_l(1-f_l)} \qquad (i,j = 1,\ldots n)$$

where p = 214051 is the total number of markers. The numbers $x_{i,l}$ denote the minor allele count at marker *l* for genotype *i*, and f_l is the minor allele frequency at marker *l*.

A commonly used mixed model for quantitative traits is given by

$$y_i = \mu + G_i + E_i, \quad (i = 1,\ldots,n) \qquad G \sim N(0, \sigma^2_A K),\ E_i \sim N(0, \sigma^2_E), \qquad (3)$$

where n = 350 is the total number of accessions, y_i is the mean phenotypic value of accession *i*, μ is the intercept and $G = (G_1,\ldots,G_n)$ is the vector of random effects, which follows a N(0, $\sigma^2_A K$) distribution. The random error effects E_i follow independent normal distributions with variance σ^2_E.

Model (3) is widely used for marker-based estimation of (narrow-sense) heritability of human traits (Yang et al., 2010) which are usually measured on cohorts of thousands of individuals. However for plant traits phenotyped on only several hundreds of genotypes it has been shown recently that such estimates can be very imprecise, and that accuracy is greatly improved if phenotypic data of genetically identical replicates (rather than means) are included in the mixed model. We therefore considered the following extension of model (3):

$$y_{i,j} = \mu + G_i + E_{i,j}, \quad (i = 1,\ldots,n,\ j = 1,\ldots.,r) \qquad G \sim N(0, \sigma^2_A K),\ E_{i,j} \sim N(0, \sigma^2_E). \qquad (4)$$

r = 3 is the number of replicates, *yi,j* is the phenotypic response of replicate *j* of genotype *i*, μ is the intercept, $G = (G_1,\ldots,G_n)$ is the vector of random genetic effects, and the errors $E_{i,j}$ have independent normal distributions with variance σ^2_E. σ^2_E Is the residual variance for a single individual. Estimates of σ^2_A and σ^2_E are obtained with the method of residual maximum likelihood (REML), and heritability is then estimated by $h^2_r = (\sigma^2_A / \sigma^2_A + \sigma^2_E)$ where σ^2_A and σ^2_E are based on all replicates. Note that in model (4), σ^2_E is the residual variance for a single individual, whereas in model (3), it is the residual variance of a genotypic mean. Since our interest is in individual plant level heritability and not line-heritability, σ^2_E in model (4) is indeed the variance parameter of interest (the use of model (3) would require multiplication of estimated residual variance by r). Both models (3) and (4) can only account for additive genetic effects; hence σ^2_E includes also non-additive genetic effects, and the denominator $\sigma^2_A + \sigma^2_E$ equals the total phenotypic variance.

Genome-wide association mapping

GWA mapping on the morphological traits was performed on between 335 and 339 accessions, because for some accessions we missed genotype data and others were removed before harvest. All accessions were genotyped with 214,051 SNPs of which 199,589 were used for GWA mapping after removal of SNP with MAF < 0.05.

In mixed-model based GWAS, the fixed marker effect $x_{i\beta}$ is added to the model (3) above:

$$\hat{y}_i = \mu + x_{i\beta} + G_i + E_i, \qquad G \sim N(0, \sigma^2_A K), Ei \sim N(0, \sigma^2_E). \tag{5}$$

The term x_i is the marker score, β is the marker effect and the genotypic effects $G = (G_1,\ldots,G_n)$ follow a $N(0, \sigma^2_A K)$ distribution. GWA mapping was performed on the means. The covariances between these effects are modeled by a genetic relatedness matrix (GRM) estimated from markers, which is called K, kinship matrix.

Following the methodology of EMMAX (Kang et al., 2010), we first obtained REML estimates of the variance components σ^2_A and σ^2_E in model (3); given these estimates the significance of the marker effect β in (5) was tested for each SNP-marker in turn, using generalized least squares (GLS). REML estimates of the variance components were obtained with the commercial R-package ASREML (Butler et al., 2007) and for the GLS calculations the command-line program scan_GLS was used (Kruijer et al.,. in prep.). Estimates of narrow sense heritability based on model (4) were obtained with the R-package heritability (Kruijer et al.,. in prep.).

Genomic selection

For the construction of genomic selection models for each trait, the markers were ranked high to low according to their correlation with the quantitative trait data, using custom scripts in R. The 100,000 highest ranking markers were used to construct heteroscedastic models using the bigRR library (Shen et al., 2013). The similarity in the distribution of effect sizes in all models was quantified by calculating pairwise the Euclidian distance over the first five moments (R-library moments). A trait-complexity dendogram was constructed from the distance matrix of the moments using UPGMA clustering (R-library hclust).

Sequence analysis

All sequences from the re-sequenced Arabidopsis accessions were obtained from http://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 a 1 or 0 for compliance or no compliance with Col-0. The resulting data is stored as data frames (.csv file) on disk. In order to calculate the LD, required data is extracted from the .csv files with the gnu program 'cut' in order to slice out the region of interest. The sliced data frame is

read into 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. In order to 'annotate' the genome with SNP polymorphisms we applied the tool Snpeff (Cingolani et al., 2012). With the output of this tool, which is stored in a mySQL database we are able to predict the effect of each mutation. Both the output of this tool and the LD scores are made available to the user via a web interface (in house access only). The user can calculate the LD in any region on the genome and is performed on the fly.

Nucleotide diversity analysis

Nucleotide diversity was measured with Tajima's π (Tajima, 1983) using DnaSP software version 4.0 (Rozas et al., 2003). ∏ was calculated for all sites, synonymous, non-synonymous and silent sites (synonymous plus non-coding sites) for each candidate gene. For deviation from neutrality, we tested using Tajima's D statistic (Tajima, 1989) and Fu & Li's D and F statistic (Fu and Li, 1993) using DnaSP version 4.0 (Rozas et al., 2003).

GUS assays

ACS11::GUS lines were grown for three weeks and complete plants were harvested in 50 ml tubes containing cold acetone (4°C) for 20 minutes, washed twice with rinsing solution (50mM $NaPO_4$, pH 7.2, 0.5mM $K_3Fe(CN)_6$ and 0.5mM $K_4Fe(CN)_6$), and then placed in staining solution (50mM $NaPO_4$, pH 7.2, 0.5mM $K_3Fe(CN)_6$, 0.5mM $K_4Fe(CN)_6$ and 2mM X-Glc. The plants were vacuum-infiltrated twice for 30s and then wrapped in aluminum foil and incubated at 37°C for 24h. Then, the staining solution was removed, plants were rinsed twice with water and then 30% ethanol was added to the plants. To completely remove the chlorophyll, this was followed by washes with graded ethanol series from 30 to 98% ethanol. Then, individual leaves were photographed.

Quantitative Real-Time PCR

RNA extraction was performed as described in (Onate-Sanchez and Vicente-Carbajosa, 2008). Remaining DNA was removed using RNA-free DNase I (Qiagen). cDNA synthesis was performed using the iScript cDNA synthesis Kit (Bio-Rad). For each qPCR, 5 µl of sample, 10 µl of iQ SYBR Green Supermix (Bio-Rad) and 0.5 µl of each primer (10mM) were mixed and MQ was added to a total volume of 20 µl. The RT-PCR was performed on the MyiQ (Bio-Rad). The program was started with a cycle of 95°C for 3 min and then 50 cycles of 15s at 95°C and 1 min at 60°C followed by a cycle of 95°C for 1 min and, one cycle at 55°C for 1 min and then 80 cycles at 55°C for 10s, raising the temperature by 0.5°C each cycle. The primers used are listed in Supplemental table 10.

Normalisation

RefGenes in Genevestigator was used to find neutral reference genes. The top twenty reference genes from leaf material were chosen and then checked for their stability in petiole tissue. The four highest scoring genes were selected. Additionally, UBQ10 was used (Hong et al., 2010). All reference genes are listed in Supplemental table 10, together with the primers used in RT-PCR. Bestkeeper was used to select the most stable reference genes

from the set of 5 genes, and these were then used to normalize the RT-qPCR data.

References

Aguilar-Martinez JA, Poza-Carrion C, Cubas P (2007) Arabidopsis BRANCHED1 acts as an integrator of branching signals within axillary buds. Plant Cell 19: 458-472

Alonso-Blanco C, Aarts MGM, Bentsink L, Keurentjes JJB, et al (2009) What Has Natural Variation Taught Us about Plant Development, Physiology, and Adaptation? The Plant cell 21: 1877-1896

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

Alonso JM, Hirayama T, Roman G, Nourizadeh S, et al (1999) EIN2, a bifunctional transducer of ethylene and stress responses in Arabidopsis. Science 284: 2148-2152

Atwell S, Huang YS, Vilhjalmsson BJ, Willems G, et al. (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature 465: 627-631

Balazadeh S, Parlitz S, Mueller-Roeber B, Meyer RC (2008) Natural developmental variations in leaf and plant senescence in Arabidopsis thaliana. Plant Biol (Stuttg) 10 Suppl 1: 136-147

Bandaranayake CK, Koumproglou R, Wang XY, Wilkes T, et al (2004) QTL analyzis of morphological and developmental traits in the Ler x Cvi population of Arabidopsis thaliana - QTL analyzis in Arabidopsis. Euphytica 137: 361-371

Barboza L, Effgen S, Alonso-Blanco C, Kooke R, et al (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

Bentsink L, Jowett J, Hanhart CJ, Koornneef M (2006) Cloning of DOG1, a quantitative trait locus controlling seed dormancy in Arabidopsis. Proc Natl Acad Sci U S A 103: 17042-17047

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

Bernard DG, Netz DJ, Lagny TJ, Pierik AJ, et al (2013) Requirements of the cytosolic iron-sulfur cluster assembly pathway in Arabidopsis. Philos Trans R Soc Lond B Biol Sci 368: 20120259

Bleecker AB, Kende H (2000) Ethylene: A gaseous signal molecule in plants. Annual Review of Cell and Developmental Biology 16: 1-+

Borevitz JO, Maloof JN, Lutes J, Dabi T, et al (2002) Quantitative trait loci controlling light and hormone response in two accessions of Arabidopsis thaliana. Genetics 160: 683-696

Bours R, Zanten Mv, Pierik R, Bouwmeester HJ, et al (2013) Antiphase Light and Temperature Cycles Affect PHYB-controlled Ethylene Sensitivity and Biosynthesis; Limiting Leaf Movement and Growth of Arabidopsis. Plant Physiology

Brachi B, Faure N, Horton M, Flahauw E, et al (2010) Linkage and Association Mapping of Arabidopsis thaliana Flowering Time in Nature. PLoS genetics 6

Brachi B, Villoutreix R, Faure N, Hautekeete N, et al (2013) Investigation of the geographical scale of adaptive phenological variation and its underlying genetics in Arabidopsis thaliana. Mol Ecol 22: 4222-4240

Butler D, Cullis BR, Gilmour A, Gogel B (2007) ASReml-R reference manual.

Chan EK, Rowe HC, Corwin JA, Joseph B, et al (2011) Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in Arabidopsis thaliana. PLoS Biol 9: e1001125

Chao DY, Silva A, Baxter I, Huang YS, et al (2012) Genome-Wide Association Studies Identify Heavy Metal ATPase3 as the Primary Determinant of Natural Variation in Leaf Cadmium in Arabidopsis thaliana. Plos Genetics 8

Chen IC, Huang IC, Liu MJ, Wang ZG, et al (2007) Glutathione S-transferase interacting with far-red insensitive 219 is involved in phytochrome A-mediated signaling in Arabidopsis. Plant Physiology 143: 1189-1202

Choi G, Kim JI, Hong SW, Shin B, et al (2005) A possible role for NDPK2 in the regulation of auxin-mediated responses for plant growth and development. Plant Cell Physiol 46: 1246-1254

Cingolani P, Platts A, Wang LL, Coon M, et al (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly 6: 80-92

de Bakker PI, Ferreira MA, Jia X, Neale BM, et al (2008) Practical aspects of imputation-driven meta-analyzis of genome-wide association studies. Human molecular genetics 17: R122-R128

Deeks MJ, Fendrych M, Smertenko A, Bell KS, et al (2010) The plant formin AtFH4 interacts with both actin and microtubules, and contains a newly identified microtubule-binding domain. Journal of Cell Science 123: 1209-1215

Deng WW, Ying H, Helliwell CA, Taylor JM, et al (2011) FLOWERING LOCUS C (FLC) regulates development pathways throughout the life cycle of Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America 108: 6680-6685

Feng HZ, Chen QG, Feng J, Zhang J, et al (2007) Functional characterization of the arabidopsis eukaryotic translation initiation factor 5A-2 that plays a crucial role in plant growth and development by regulating cell division, cell growth, and cell death. Plant Physiology 144: 1531-1545

Filiault DL, Maloof JN (2012) A genome-wide association study identifies variants underlying the Arabidopsis thaliana shade avoidance response. PLoS Genet 8: e1002589

Fournier-Level A, Korte A, Cooper MD, Nordborg M, et al (2011) A map of local adaptation in Arabidopsis thaliana. Science 334: 86-89

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. Genetics 133: 693-709

Gaufichon L, Masclaux-Daubresse C, Tcherkez G, Reisdorf-Cren M, et al (2013) Arabidopsis thaliana ASN2 encoding asparagine synthetase is involved in the control of nitrogen assimilation and export during vegetative growth. Plant Cell Environ 36: 328-342

Gibson G (2010) Hints of hidden heritability in GWAS. Nat Genet 42: 558-560

Gibson G (2011) Rare and common variants: twenty arguments. Nat Rev Genet 13: 135-145

Gonzalez-Grandio E, Poza-Carrion C, Sorzano CO, Cubas P (2013) BRANCHED1 Promotes Axillary Bud Dormancy in Response to Shade in Arabidopsis. Plant Cell 25: 834-850

Goslee SC, Urban DL (2007) The ecodist package for dissimilarity-based analyzis of ecological data. Journal of Statistical Software 22: 1-19

Halliday KJ, Koornneef M, Whitelam GC (1994) Phyotchrome B and at least one other phytochrome mediate the accelerated flowering response of Arabidopsis thaliana to low red/far-red ratio. Plant Physiology 104: 1311-1315

Hancock AM, Brachi B, Faure N, Horton MW, et al (2011) Adaptation to climate across the Arabidopsis thaliana genome. Science 334: 83-86

Hong SM, Bahn SC, Lyu A, Jung HS, Ahn JH (2010) Identification and Testing of Superior Reference Genes for a Starting Pool of Transcript Normalization in Arabidopsis. Plant and Cell Physiology 51: 1694-1706

Horton MW, Hancock AM, Huang YS, Toomajian C, et al (2012) Genome-wide patterns of genetic variation in worldwide Arabidopsis thaliana accessions from the RegMap panel. Nat Genet 44: 212-216

Hou H, Erickson J, Meservy J, Schultz EA (2010) FORKED1 encodes a PH domain protein that is required for PIN1 localization in developing leaf veins. Plant J 63: 960-973

Hu J, Mitchum MG, Barnaby N, Ayele BT, et al (2008) Potential sites of bioactive gibberellin production during reproductive growth in Arabidopsis. Plant Cell 20: 320-336

Huang XQ, Ding J, Effgen S, Turck F, et al (2013) Multiple loci and genetic interactions involving flowering time genes regulate stem branching among natural variants of Arabidopsis. New Phytologist 199: 843-857

Huber SC, Huber JL (1996) Role and regulation of sucrose-phosphate synthase in higher plants. Annual Review of Plant Physiology and Plant Molecular Biology 47: 431-444

Huttner S, Veit C, Schoberer J, Grass J, et al (2012) Unraveling the function of Arabidopsis thaliana OS9 in the endoplasmic reticulum-associated degradation of glycoproteins. Plant Mol Biol 79: 21-33

Johanson U (2000) Molecular Analyzis of FRIGIDA, a Major Determinant of Natural Variation in Arabidopsis Flowering Time. Science 290: 344-347

Johanson U, West J, Lister C, Michaels S, et al (2000) Molecular analyzis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. Science 290: 344-347

Kang HM, Sul JH, Service SK, Zaitlen NA, et al (2010) Variance component model to account for sample structure in genome-wide association studies. Nat Genet 42: 348-354

Keurentjes JJB, Bentsink L, Alonso-Blanco C, Hanhart CJ, et al (2007) Development of a near-isogenic line population of Arabidopsis thaliana and comparison of mapping power with a recombinant inbred line population. Genetics 175: 891-905

Kim MJ, Shin R, Schachtman DP (2009) A nuclear factor regulates abscisic acid responses in Arabidopsis. Plant Physiol 151: 1433-1445

Kinoshita T, Haranishimura I, Siraishi H, Okada K, et al (1994) Nucleotide-sequence of a transmembrane protein (TMP-C) cDNA in Arabidopsis thaliana. Plant Physiology 105: 1441-1442

Kistler R, Kalnay E, Collins W, Saha S, et al (2001) The NCEP-NCAR 50-year reanalyzis: Monthly means CD-ROM and documentation. Bulletin of the American Meteorological Society 82: 247-267

Kiyosue T, Abe H, Yamaguchi-Shinozaki K, Shinozaki K (1998) ERD6, a cDNA clone for an early dehydration-induced gene of Arabidopsis, encodes a putative sugar transporter. Biochimica Et Biophysica Acta-Biomembranes 1370: 187-191

Kocabek T, Repkova J, Dudova M, Hoyerova K, et al (2006) Isolation and characterization of a novel semi-lethal Arabidopsis thaliana mutant of gene for pentatricopeptide (PPR) repeat-containing protein. Genetica 128: 395-407

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

Koornneef M, Blankestijndevries H, Hanhart C, Soppe W, et al (1994) The phenotype of some late-flowering mutants is enhanced by a locus on chromosome 5 that is not effective in the landsberg erecta wild-type. Plant Journal 6: 911-919

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

Kubo M, Furuta K, Demura T, Fukuda H, et al (2011) The CKH1/EER4 Gene Encoding a TAF12-Like Protein Negatively Regulates Cytokinin Sensitivity in Arabidopsis thaliana. Plant and Cell Physiology 52: 629-637

Lei G, Shen M, Li ZG, Zhang B, et al (2011) EIN2 regulates salt stress response and interacts with a MA3 domain-containing protein ECIP1 in Arabidopsis. Plant Cell Environ 34: 1678-1692

Lenz HD, Haller E, Melzer E, Kober K, et al (2011) Autophagy differentially controls plant basal immunity to biotrophic and necrotrophic pathogens. Plant Journal 66: 818-830

Li SF, Parish RW (1995) Isolation of two novel myb-like genes from Arabidopsis and studies on the DNA-binding properties of their products. Plant Journal 8: 963-972

Li Y, Huang Y, Bergelson J, Nordborg M, et al (2010) Association mapping of local climate-sensitive quantitative trait loci in Arabidopsis thaliana. Proc Natl Acad Sci U S A 107: 21199-21204

Lu W, Tang X, Huo Y, Xu R, et al (2012) Identification and characterization of fructose 1,6-bisphosphate aldolase genes in Arabidopsis reveal a gene family with diverse responses to abiotic stresses. Gene 503: 65-74

Makowsky R, Pajewski NM, Klimentidis YC, Vazquez AI, et al (2011) Beyond missing heritability: prediction of complex traits. PLoS Genet 7: e1002051

Mantel N (1967) Detection of disease clustering and a generalized regression approach. Cancer Research 27: 209-&

Mendez-Vigo B, Gomaa NH, Alonso-Blanco C, Pico FX (2013) Among- and within-population variation in flowering time of Iberian Arabidopsis thaliana estimated in field and glasshouse conditions. New Phytol 197: 1332-1343

Mendez-Vigo B, Martinez-Zapater JM, Alonso-Blanco C (2013) The flowering repressor SVP underlies a novel Arabidopsis thaliana QTL interacting with the genetic background. PLoS Genet 9: e1003289

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 157: 1819-1829

Millenaar FF, Cox MCH, van Berkel Y, Welschen RAM, et al (2005) Ethylene-induced differential growth of petioles in arabidopsis. Analyzing natural variation, response kinetics, and regulation. Plant Physiology 137: 998-1008

Min MK, Jang M, Lee M, Lee J, et al (2013) Recruitment of Arf1-GDP to Golgi by Glo3p-Type ArfGAPs Is Crucial for Golgi Maintenance and Plant Growth. Plant Physiology 161: 676-691

Mitchum MG, Yamaguchi S, Hanada A, Kuwahara A, et al (2006) Distinct and overlapping roles of two gibberellin 3-oxidases in Arabidopsis development. Plant J 45: 804-818

Nagane T, Tanaka A, Tanaka R (2010) Involvement of AtNAP1 in the regulation of chlorophyll degradation in Arabidopsis thaliana. Planta 231: 939-949

Nakajima K, Furutani I, Tachimoto H, Matsubara H, et al (2004) SPIRAL1 encodes a plant-specific microtubule-localized protein required for directional control of rapidly expanding Arabidopsis cells. Plant Cell 16: 1178-1190

Nejidat A, Zhang GF, Grinberg M, Heimer YM (1997) Increased protein content in transgenic Arabidopsis thaliana over-expressing nitrate reductase activity. Plant Science 130: 41-49

New M, Lister D, Hulme M, Makin I (2002) A high-resolution data set of surface climate over global land areas. Climate Research 21: 1-25

Niwa M, Daimon Y, Kurotani K, Higo A, et al (2013) BRANCHED1 interacts with FLOWERING LOCUS T to repress the floral transition of the axillary meristems in Arabidopsis. Plant Cell 25: 1228-1242

Nordborg M, Hu TT, Ishino Y, Jhaveri J, et al (2005) The pattern of polymorphism in Arabidopsis thaliana. Plos Biology 3: 1289-1299

Okushima Y, Overvoorde PJ, Arima K, Alonso JM, et al (2005) Functional genomic analyzis of the AUXIN RESPONSE FACTOR gene family members in Arabidopsis thaliana: Unique and overlapping functions of ARF7 and ARF19. Plant Cell 17: 444-463

Onate-Sanchez L, Vicente-Carbajosa J (2008) DNA-free RNA isolation protocols for Arabidopsis thaliana, including seeds and siliques. BMC Res Notes 1: 93

Platt A, Horton M, Huang YS, Li Y, et al (2010) The Scale of Population Structure in Arabidopsis thaliana. Plos Genetics 6

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

Reyes JC, Muro-Pastor MI, Florencio FJ (2004) The GATA family of transcription factors in Arabidopsis and rice. Plant Physiol 134: 1718-1732

Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, et al (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nat Genet 44: 217-220

Roff DA (2007) Contributions of genomics to life-history theory. Nat Rev Genet 8: 116-125

Rosloski SM, Jali SS, Balasubramanian S, Weigel D, et al (2010) Natural diversity in flowering responses of *Arabidopsis thaliana* caused by variation in a tandem gene array. Genetics 186: 263-276

Rosloski SM, Singh A, Jali SS, Balasubramanian S, et al (2013) Functional analyzis of splice variant expression of MADS AFFECTING FLOWERING 2 of Arabidopsis thaliana. Plant Mol Biol 81: 57-69

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyzes by the coalescent and other methods. Bioinformatics 19: 2496-2497

Schmid M, Davison TS, Henz SR, Pape UJ, et al (2005) A gene expression map of Arabidopsis thaliana development. Nat Genet 37: 501-506

Shen X, Alam M, Fikse F, Ronnegard L (2013) A Novel Generalized Ridge Regression Method for Quantitative Genetics. Genetics 193: 1255-1268

Shoji T, Narita NN, Hayashi K, Asada J, et al (2004) Plant-specific microtubule-associated protein SPIRAL2 is required for anisotropic growth in Arabidopsis. Plant Physiol 136: 3933-3944

Smouse PE, Long JC, Sokal RR (1986) Multiple regression and correlation extensions of the mantel test of matrix correspondence. Systematic Zoology 35: 627-632

Somerville C (2006) Cellulose synthesis in higher plants. Annu Rev Cell Dev Biol 22: 53-78

Song JT, Lu H, Greenberg JT (2004) Divergent roles in *Arabidopsis thaliana* development and defense of two homologous genes, aberrant growth and death2 and AGD2-LIKE DEFENSE RESPONSE PROTEIN1, encoding novel aminotransferases. Plant Cell 16: 353-366

Song SK, Lee MM, Clark SE (2006) POL and PLL1 phosphatases are CLAVATA1 signaling intermediates required for Arabidopsis shoot and floral stem cells. Development 133: 4691-4698

Sterken R, Kiekens R, Boruc J, Zhang FH, et al (2012) Combined linkage and association mapping reveals CYCD5;1 as a quantitative trait gene for endoreduplication in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America 109: 4678-4683

Stout J, Romero-Severson E, Ruegger MO, Chapple C (2008) Semidominant mutations in reduced epidermal fluorescence 4 reduce phenylpropanoid content in Arabidopsis. Genetics 178: 2237-2251

Suetsugu N, Yamada N, Kagawa T, Yonekura H, et al (2010) Two kinesin-like proteins mediate actin-based chloroplast movement in Arabidopsis thaliana. Proceedings of the National Academy of Sciences of the United States of America 107: 8860-8865

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. Genetics 105: 437-460

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

Todesco M, Balasubramanian S, Hu TT, Traw MB, et al (2010) Natural allelic variation underlying a major fitness trade-off in Arabidopsis thaliana. Nature 465: 632-U129

Tsuchisaka A, Theologis A (2004) Unique and overlapping expression patterns among the Arabidopsis 1-amino-cyclopropane-1-carboxylate synthase gene family members. Plant Physiol 136: 2982-3000

Tsuchisaka A, Yu G, Jin H, Alonso JM, et al (2009) A combinatorial interplay among the 1-aminocyclopropane-1-carboxylate isoforms regulates ethylene biosynthesis in Arabidopsis thaliana. Genetics 183: 979-1003

Ungerer MC, Halldorsdottir SS, Modliszewski JL, Mackay TFC, et al (2002) Quantitative trait loci for inflorescence development in Arabidopsis thaliana. Genetics 160: 1133-1151

Vaistij FE, Gan YB, Penfield S, Gilday AD, et al (2013) Differential control of seed primary dormancy in Arabidopsis ecotypes by the transcription factor SPATULA. Proceedings of the National Academy of Sciences of the United States of America 110: 10866-10871

van Zanten M, Voesenek L, Peeters AJM, Millenaar FF (2009) Hormone- and Light-Mediated Regulation of Heat-Induced Differential Petiole Growth in Arabidopsis. Plant Physiology 151: 1446-1458

Vanlijsebettens M, Vanderhaeghen R, Deblock M, Bauw G, et al (1994) An S18 ribosomal-protein gene copy at the Arabidopsis PFL locus affects plant development by its specific expression in meristems. Embo Journal 13: 3378-3388

Versaw WK (2002) A Chloroplast Phosphate Transporter, PHT2;1, Influences Allocation of Phosphate within the Plant and Phosphate-Starvation Responses. The Plant Cell Online 14: 1751-1766

Vierstra R (2003) The ubiquitin/26S proteasome pathway, the complex last chapter in the life of many plant proteins. Trends in Plant Science 8: 135-142

Wang Z, Meng P, Zhang X, Ren D, et al (2011) BON1 interacts with the protein kinases BIR1 and BAK1 in modulation of temperature-dependent plant growth and cell death in Arabidopsis. Plant J 67: 1081-1093

Watahiki MK, Yamamoto KT (1997) The massugu1 mutation of Arabidopsis identified with failure of auxin-induced growth curvature of hypocotyl confers auxin insensitivity to hypocotyl and leaf. Plant Physiology 115: 419-426

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

Werner T, Motyka V, Laucou V, Smets R, et al (2003) Cytokinin-deficient transgenic Arabidopsis plants show multiple developmental alterations indicating opposite functions of cytokinins in the regulation of shoot and root meristem activity. Plant Cell 15: 2532-2550

Xiang L, Le Roy K, Bolouri-Moghaddam MR, Vanhaecke M, et al (2011) Exploring the neutral invertase-oxidative stress defence connection in Arabidopsis thaliana. Journal of Experimental Botany 62: 3849-3862

Yamada K, Osakabe Y, Mizoi J, Nakashima K, et al (2010) Functional analyzis of an Arabidopsis thaliana abiotic stress-inducible facilitated diffusion transporter for monosaccharides. J Biol Chem 285: 1138-1146

Yamagami A, Nakazawa M, Matsui M, Tujimoto M, et al (2009) Chemical Genetics Reveal the Novel Transmembrane Protein BIL4, Which Mediates Plant Cell Elongation in Brassinosteroid Signaling. Bioscience Biotechnology and Biochemistry 73: 415-421

Yang J, Benyamin B, McEvoy BP, Gordon S, et al (2010) Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42: 565-569

Yano R, Takebayashi Y, Nambara E, Kamiya Y, et al (2013) Combining association mapping and transcriptomics identify HD2B histone deacetylase as a genetic factor associated with seed dormancy in Arabidopsis thaliana. Plant J 74: 815-828

Zeeman SC, Thorneycroft D, Schupp N, Chapple A, et al (2004) Plastidial alpha-glucan phosphorylase is not required for starch degradation in arabidopsis leaves but has a role in the tolerance of abiotic stress. Plant Physiology 135: 849-858

Zhang JY, Vankova R, Malbeck J, Dobrev PI, et al (2009) AtSOFL1 and AtSOFL2 Act Redundantly as Positive Modulators of the Endogenous Content of Specific Cytokinins in Arabidopsis. Plos One 4

Zhang Y, Feng SH, Chen FF, Chen HD, et al (2008) Arabidopsis DDB1-CUL4 ASSOCIATED FACTOR1 forms a nuclear E3 ubiquitin ligase with DDB1 and CUL4 that is involved in multiple plant developmental processes. Plant Cell 20: 1437-1455

Supplemental table 1. Trait abbreviations.

Trait	Description
LAbv	Leaf are before vernalisation, 19 days after germination
LAav	Leaf area 1 day after vernalisation, 63 days after germination
FT	Flowering time
PL/LL	Petiole length/ leaf length ratio
PL	Petiole length
LL	Leaf length
RB	Rosette branching
MSB	Main stem branching
PH1S	Plant height 1st silique
TPH	Total plant height
RGRbv	Relative growth rate before vernalisation
RGRav	Relative growth rate after vernalisation

Supplemental table 2. Mantel correlation matrix based on Spearman (top-right) and their respective *P*-values (left-bottom) between morphological traits with kinship as a covariate in the model. Numbers shown in the heatmap are partial correlation coefficients (x100%). LAbv, leaf area before vernalisation; LAav, leaf area after vernalisation; FT, flowering time; PL/LL, petiole to leaf length ratio; PL, petiole length; LL, leaf length; RB, rosette branching; MSB, main stem branching; PH1S, plant height at 1st silique; TPH, total plant height; RGRbv, relative growth rate before vernalisation; RGRav, relative growth rate after vernalisation.

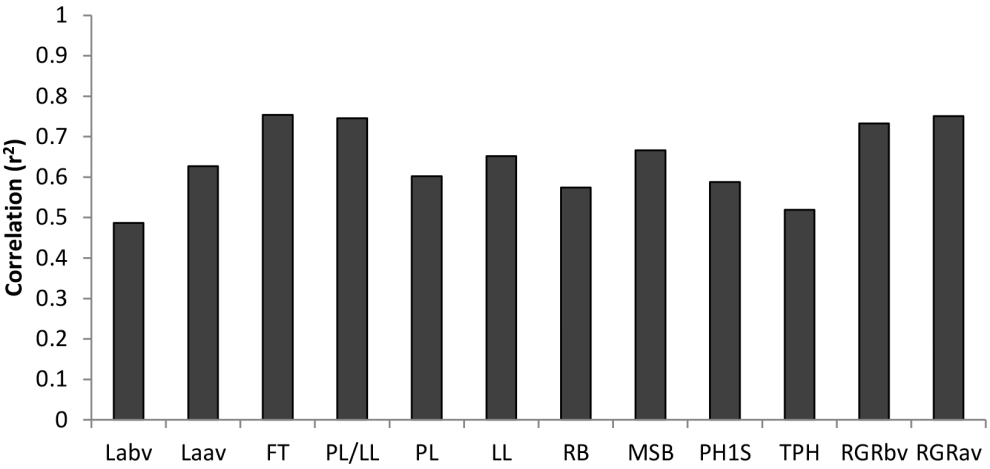
r^2	LAbv	LAav	FT	PL/LL	PL	LL	RB	MSB	PH1S	TPH	RGRav	RGRbv
LAbv	x	0.61	0.05	0.04	0.05	0.07	0.00	0.03	0.01	0.04	0.04	0.26
LAav	0.001	x	0.04	0.05	0.05	0.07	0.04	0.01	0.02	0.02	0.08	0.35
FT	0.016	0.042	x	0.10	0.24	0.25	0.13	0.22	0.09	0.16	0.02	0.03
PL/LL	0.044	0.01	0.001	x	0.55	0.11	-0.01	0.01	0.01	0.07	0.02	0.04
PL	0.009	0.008	0.001	0.001	x	0.56	0.01	0.07	0.02	0.15	0.01	0.05
LL	0.003	0.002	0.001	0.001	0.001	x	0.01	0.09	0.02	0.14	0.04	0.11
RB	1.000	0.017	0.001	0.456	0.545	0.446	x	0.08	0.05	-0.01	-0.01	0.08
MSB	0.101	0.45	0.001	0.779	0.001	0.001	0.001	x	0.08	0.03	-0.02	0.01
PH1S	0.574	0.331	0.001	0.437	0.464	0.31	0.003	0.001	x	0.17	0.00	0.06
TPH	0.037	0.223	0.001	0.001	0.001	0.001	0.721	0.131	0.001	x	0.00	0.04
RGRav	0.074	0.001	0.346	0.366	0.582	0.136	0.761	0.279	0.922	0.999	x	0.06
RGRbv	0.001	0.001	0.165	0.021	0.017	0.001	0.001	0.579	0.004	0.056	0.008	x

Supplemental table 3. Mantel correlation matrix based on Spearman (top) and their respective *P*-values (bottom) between climate variables (nr's) and morphological traits with kinship as a covariate in the model. Numbers shown in the heatmap are partial correlation coefficients. Climate variables are: 1, latitude; 2, longitude; 3, number of wet days/year; 4, temperature in coldest month; 5, temperature in warmest month; 6, mimimal sunshine (month with least hours of sunshine/ daylength; 7, maximal sunshine (month with most hours of sunshine/ daylength; 8, Precipitation in driest month; 9, precipitation in wettest month; 10, windspeed in m/s; 11, number of ground frost days/ year; 12, number of ground frost days in coldest month; 13, relative humidity in spring; 14, Daylength in spring. LAbv, leaf area before vernalisation; LAav, leaf area after vernalisation; FT, flowering time; PL/LL, petiole to leaf length ratio; PL, petiole length; LL, leaf length; RB, rosette branching; MSB, main stem branching; PH1S, plant height at 1st silique; TPH, total plant height; RGRbv, relative growth rate before vernalisation; RGRav, relative growth rate after vernalisation.

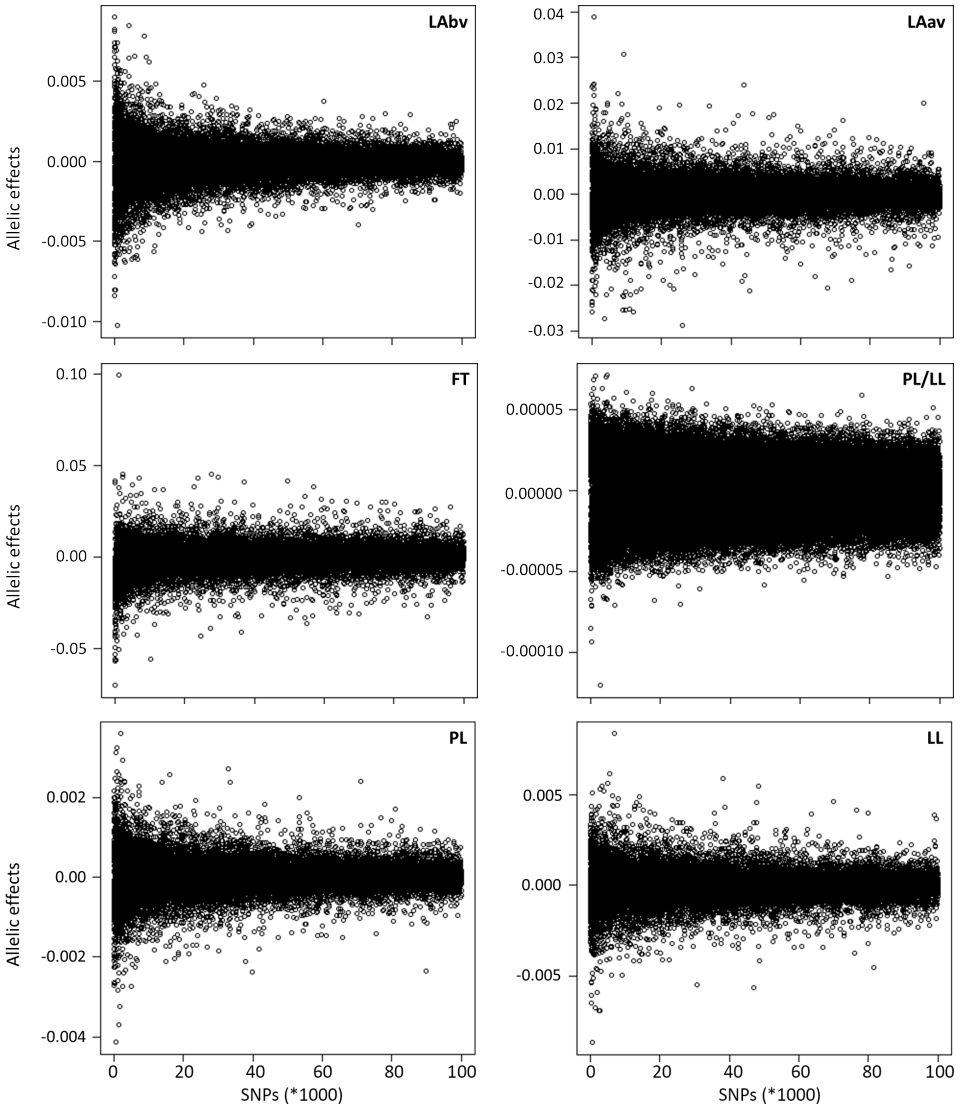
r^2	1	2	3	4	5	6	7	8	9	10	11	12	13	14
LAbv	0.04	0.00	-0.02	0.01	0.00	0.01	0.01	0.04	-0.03	0.02	0.00	-0.01	0.04	0.04
LAav	0.01	0.02	-0.02	0.03	0.01	0.01	0.04	0.04	0.00	0.00	0.00	0.02	0.00	-0.01
FT	0.09	-0.04	-0.05	-0.06	-0.03	0.00	-0.04	-0.01	-0.02	0.10	-0.04	-0.05	0.14	0.06
PL/LL	0.08	0.00	-0.01	0.00	0.00	0.03	-0.03	0.03	0.02	0.06	-0.01	-0.02	0.13	0.12
PL	0.05	-0.02	0.00	-0.02	0.01	0.02	0.01	0.03	0.02	0.06	-0.02	-0.03	0.10	0.06
LL	0.00	-0.03	-0.01	-0.03	-0.01	-0.01	0.00	-0.01	-0.01	0.05	-0.04	-0.04	0.02	0.00
RB	0.03	0.03	-0.01	0.01	-0.01	0.04	-0.01	0.00	0.03	0.03	-0.01	0.00	0.02	0.02
MSB	-0.01	-0.04	-0.05	-0.04	-0.05	-0.04	-0.05	-0.02	-0.05	0.01	-0.04	-0.04	-0.02	-0.02
PH1S	-0.01	-0.06	-0.07	-0.06	-0.06	-0.04	-0.06	-0.05	-0.02	0.02	-0.05	-0.05	0.01	0.02
TPH	0.06	-0.01	-0.02	0.00	-0.02	0.01	-0.05	-0.01	-0.01	0.04	0.00	-0.01	0.08	0.03
RGRbv	0.01	-0.01	-0.01	0.01	-0.01	0.01	-0.01	0.01	0.01	0.00	0.00	-0.01	0.02	-0.02
RGRav	0.00	-0.07	0.00	-0.03	0.00	-0.01	-0.02	0.05	0.02	0.03	-0.01	-0.03	0.03	0.01
P	1	2	3	4	5	6	7	8	9	10	11	12	13	14
LAbv	0.085	0.892	0.375	0.635	0.991	0.582	0.632	0.126	0.231	0.347	0.857	0.724	0.125	0.116
LAav	0.748	0.373	0.435	0.116	0.706	0.613	0.123	0.14	0.869	0.981	0.951	0.432	0.906	0.71
FT	0.001	0.096	0.043	0.018	0.306	0.896	0.164	0.64	0.578	0.001	0.078	0.019	0.001	0.018
PL/LL	0.002	0.962	0.709	0.866	0.899	0.228	0.193	0.228	0.429	0.002	0.632	0.396	0.001	0.001
PL	0.015	0.542	0.971	0.495	0.792	0.335	0.691	0.292	0.358	0.001	0.369	0.202	0.001	0.03
LL	0.954	0.293	0.687	0.169	0.584	0.623	0.907	0.738	0.611	0.005	0.038	0.067	0.473	0.985
RB	0.111	0.252	0.785	0.531	0.762	0.05	0.736	0.882	0.175	0.048	0.468	0.95	0.316	0.253
MSB	0.636	0.148	0.035	0.052	0.04	0.105	0.039	0.321	0.049	0.755	0.078	0.014	0.444	0.489
PH1S	0.639	0.03	0.006	0.009	0.022	0.13	0.009	0.026	0.369	0.182	0.029	0.005	0.538	0.502
TPH	0.015	0.737	0.455	0.914	0.379	0.722	0.078	0.697	0.698	0.028	0.966	0.641	0.002	0.244
RGRav	0.897	0.021	0.97	0.185	0.907	0.869	0.481	0.106	0.463	0.117	0.793	0.197	0.369	0.796
RGRbv	0.804	0.565	0.714	0.675	0.532	0.827	0.707	0.554	0.653	0.875	0.814	0.666	0.341	0.523

Supplemental table 4. Genetic parameters for the population of 350 accessions.
The overall average (AVG), minimum (MIN), maximum (MAX), median, broad-sense heritability (H^2), marker-based heritability on individual plant data (h^2) and coefficient of variation (CV) are given for all morphological traits. LAbv, leaf area before vernalisation; LAav, leaf area after vernalisation; FT, flowering time; PL/LL, petiole to leaf length ratio; PL, petiole length; LL, leaf length; RB, rosette branching; MSB, main stem branching; PH1S, plant height at 1st silique; TPH, total plant height; RGRbv, relative growth rate before vernalisation; RGRav, relative growth rate after vernalisation.

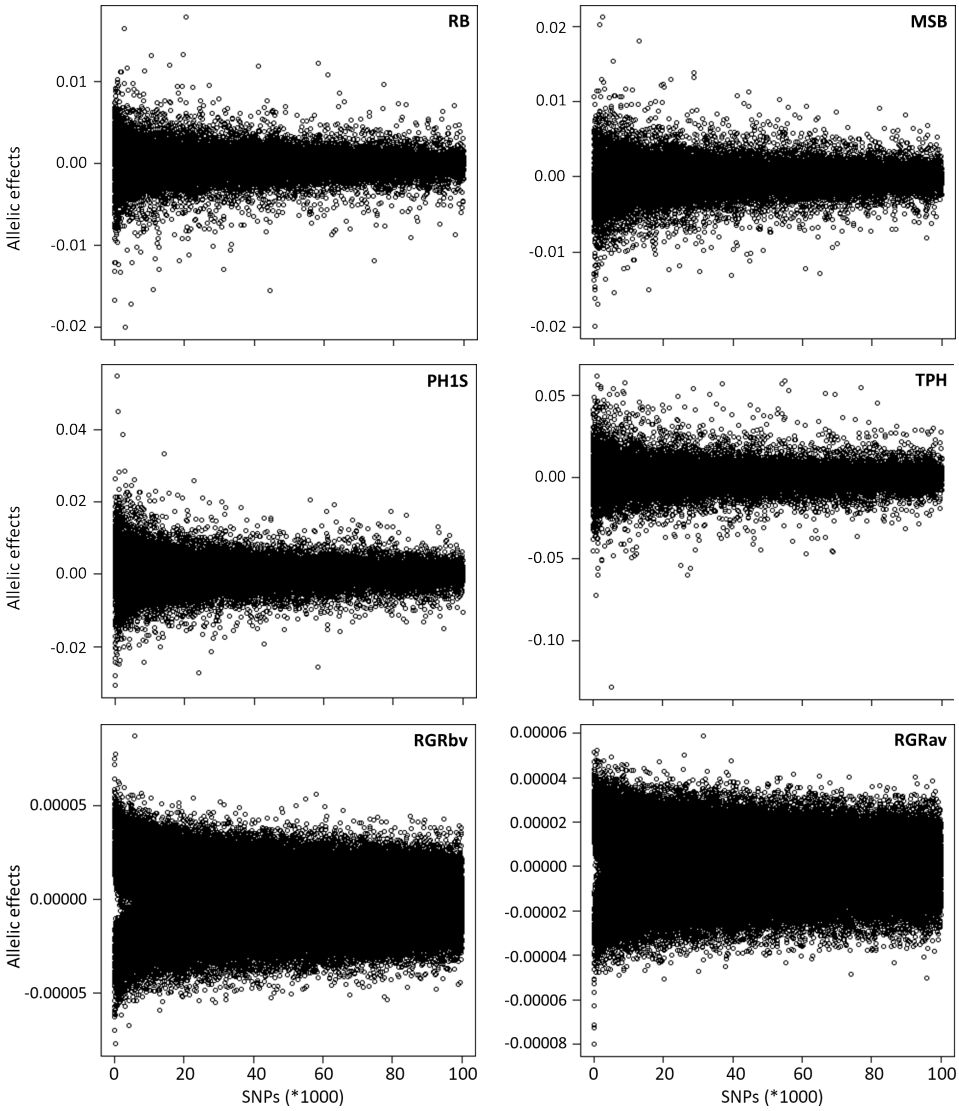
Trait	AVG	MIN	MAX	Median	H^2	h^2	CV
LAbv	3.1	0.5	9.1	3.0	0.63	0.65	45.1
LAav	10.9	1.8	22.0	10.5	0.68	0.70	38.7
FT	22.5	3.7	57.0	20.7	0.95	0.93	41.0
PL/LL	0.3	0.1	0.5	0.3	0.60	0.61	16.6
PL	1.9	0.7	3.9	1.9	0.72	0.73	28.2
LL	6.4	3.5	11.0	6.4	0.73	0.73	17.2
RB	3.4	0.0	14.0	3.4	0.75	0.76	64.4
MSB	8.4	1.3	17.3	8.3	0.80	0.81	34.8
PH1S	24.1	6.0	39.0	25.0	0.75	0.76	23.1
TPH	56.5	19.3	83.0	56.7	0.81	0.83	16.0
RGRbv	0.3	0.2	0.4	0.3	0.40	0.42	10.8
RGRav	0.2	0.1	0.3	0.2	0.79	0.81	24.9



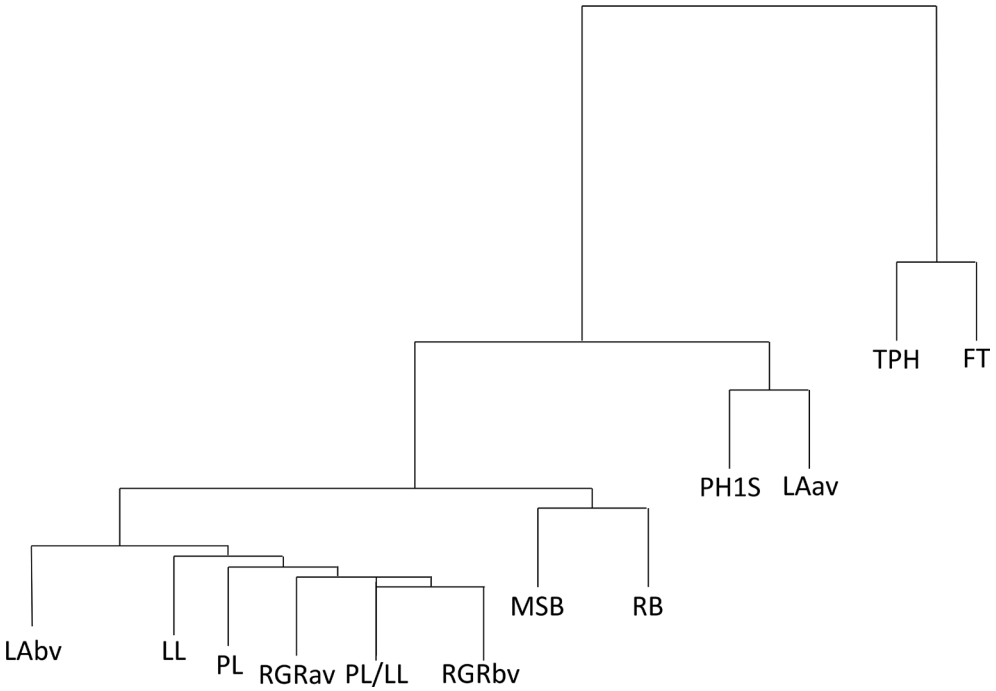
Supplemental figure 1. Coefficient of determination (r2) between predicted and observed values in repeated internal cross-validation for all morphological phenotypes in heteroscedastic genomic selection model based on 100,000 highest correlating SNPs. LAbv, leaf area before vernalisation; LAav, leaf area after vernalisation; FT, flowering time; PL/LL, petiole to leaf length ratio; PL, petiole length; LL, leaf length; RB, rosette branching; MSB, main stem branching; PH1S, plant height at 1st silique; TPH, total plant height; RGRbv, relative growth rate before vernalisation; RGRav, relative growth rate after vernalisation.



Supplemental figure 2. Allelic effect sizes of the 100,000 most correlating SNPs in the heteroscedastic genomic selection model. LAbv, leaf area before vernalisation; LAav, leaf area after vernalisation; FT, flowering time; PL/LL, petiole to leaf length ratio; PL, petiole length; LL, leaf length.



Supplemental figure 2 continued. Allelic effect sizes of the 100,000 most correlating SNPs in the heteroscedastic genomic selection model. RB, rosette branching; MSB, main stem branching; PH1S, plant height at 1st silique; TPH, total plant height; RGRbv, relative growth rate before vernalisation; RGRav, relative growth rate after vernalisation.



Supplemental figure 3. Trait-complexity dendogram constructed from the Euclidian distance matrix over the first five moments using UPGMA clustering. More complex traits branch off later. LAbv, leaf area before vernalisation; LAav leaf area after vernalisation; FT, flowering time; PL/LL, petiole to leaf length ratio; PL, petiole length; LL, leaf length; RB, rosette branching; MSB. main stem branching; PH1S, plant height at 1st silique; TPH, total plant height; RGRbv, relative growth rate before vernalisation; RGRav, relative growth rate after vernalisation.

Supplemental table 5. List of candidate genes from GWAS for different morphological traits. The LD is calculated as the distance between the first and last marker where r2 > 0.3. AF=allele frequency, nsSNP=non-synonymous single nucleotide polymorphism. LAbv, leaf area before vernalisation; LAav leaf area after vernalisation; FT, flowering time; PL/LL, petiole to leaf length ratio; PL, petiole length; LL, leaf length; RB, rosette branching; MSB. main stem branching; PH1S, plant height at 1st silique; TPH, total plant height; RGRbv, relative growth rate before vernalisation; RGRav, relative growth rate after vernalisation.

Trait	Locus	Gene Abbreviation	Chr.	Position	LD (in kB)	AF	LOD	SNPs in gene	nsSNPs in gene	SNPs $r^2 > 0.3$	nsSNP $r^2 > 0.3$	SNPs in promoter (1kb)	SNPs in promoter $r^2 > 0.3$
Labv	At1G37130	NR2	1	14154860	24807	0.09	4.02	84	3	1	1	34	4
Labv	AT2G21770	CESA9	2	9286277	40311	0.16	4.67	96	25	38	5	25	5
Labv	AT5G20730	ARF7	5	7018204	43998	0.17	4.38	111	15	4	1	38	0
Laav	At1G22650	NI	1	8015537	14330	0.33	5.51	62	5	17	1	2	2
Laav	AT2G21770	CESA9	2	9286277	40311	0.16	4.27	96	25	38	5	25	5
Laav	AT5G65010	ASN2	5	25981971	149300	0.17	4.19	18	3	1	1	10	0
	AT5G65050	MAF2						59	3	12	1	25	3
FT	AT2G18790	PHYB	2	8130188	29370	0.35	2.83	100	11	2	0	28	3
FT	At2g22540	SVP	2	9581605	24693	0.17	2.40	79	0	7	0	30	8
FT	AT4G00650	FRI	4	275349	27034	0.23	2.94	24	13	2	1	18	0
FT	AT4G27060	TOR1	4	13581775	57929	0.35	3.99	97	27	75	22	37	12
FT	AT5G10140	FLC	5	3188327	39950	0.21	7.36	105	0	4	0	28	0
FT	At5G45830	DOG1	5	18603055	19856	0.42	6.14	53	17 (4)	6	1	19	0
	At5G45890	SAG12						29	7	0	0	30	0
	At5G45900	AFG7						21	4	0	0	6	0
FT	AT5G63190	none	5	25359325	75411	0.16	5.87	35	3	20	2	8	2
	AT5G63195	other RNA						7	7	4	4	0	0
	AT5G63310	NDPK2						13	0	1	0	7	0
	AT5G63320	NX1						18	3 (1)	3	1	0	0
PL/LL	AT1G68550	CRF10	1	25721918	6	0.29	5.11	23	5	0	0	0	0
PL/LL	AT4G08040	ACS11	4	4912600	141074	0.49	4.55	29	5	5	2	28	2
PL/LL	AT5G10470	KCA1	5	3302831	14967	0.06	4.47	60	18(4)	2	1	11	0
PL	AT1G68870	SOFL2	1	25883368	50707	0.49	4.96	13	2	6	1	1	0
PL	AT2G13810	ALD1	2	5769361	3062	0.17	4.21	157	4	24	2	55	4
PL	AT4G09800	RPS18C	4	6170856	21574	0.44	4.49	26	0	2	0	78	2
LL	AT1G26355	SPIL1	1	9129489	15025	0.43	5.13	10	4(1)	0	0	12	0
LL	AT1G53330	PPR family	1	19897982	5080	0.49	4.34	50	23	22	9	12	6
LL	AT4G09800	RPS18C	4	6170856	21574	0.44	4.41	26	0	2	0	78	2
RB	AT1G17440	CKX1/ EER4	1	5994488	96629	0.12	4.78	63	8(3)	3	0	7	1
RB	At3g18550	BRANCHED1	3	6390880	2529	0.14	4.03	51	12(1)	0	0	29	0
RB	AT5G35080	OS9	5	13355009	66821	0.1	7.13	27	4	7	2	55	0
MSB	AT1G10870	AGD4	1	3625529	18161	0.08	4.27	164	8	1	0	9	0
MSB	AT3G26570	PHT2;1	3	9757577	14292	0.48	4.47	64	12	31	3	50	5
MSB	AT3G60530	GATA4	3	22371590	14042	0.2	4.71	25	5	0	0	0	0
MSB	AT3G63300	FORKED1	3	23419198	103295	0.25	4.83	29	6	20	5	14	8
	AT3G63310	BIL4						7	1	4	1	9	1
	AT3G63440	CKX6						6	0	0	0	4	1

	AT3G63445	none						20	5(1)	8	2 (1)	0	0
MSB	AT5G03250	none	5	775881	7806	0.09	5.32	96	7	7	1	49	2
MSB	AT5G18400	DRE2	5	6093741	7388	0.4	4.28	66	6	39	5	6	0
PH1S	AT1G24150	FH4	1	8546658	960	0.27	5.87	38	11	0	0	26	0
PH1S	AT1G80330	GA3OX4	1	30197048	25003	0.12	4.37	21	6	8	3	48	8
	AT1G80340	GA4H						15	9	2	0	21	2
PH1S	AT2G35350	PLL1	2	14872965	35659	0.17	4.10	38	10	7	4	20	6
PH1S	AT3G23590	RFR1	3	8475104	83923	0.07	5.75	56	10(3)	5	3(3)	42	2
PH1S	AT4G09460	MYB6	4	5994574	NO LD	0.33	5.08	51	8	0	0	1	1
PH1S	AT4G38970	FBA2	4	18166777	8909	0.12	4.86	42	0	1	0	15	2
	AT4G38990	GH9B16						77	19(1)	25	7	40	2
TPH	AT1G19485	WD40	1	6746695	48625	0.08	5.41	74	13	36	7	10	5
RGRbv	AT1G04860	UBP2	1	1393031	50242	0.1	4.11	17	3	1	1	2	0
	AT1G04920	SPS3F						13	3	1	0	1	0
RGRbv	AT1G69410	ELF5A-3	1	26089498	11550	0.18	4.42	20	0	4	0	3	0
RGRbv	AT1G78370	GSTU20	1	29484933	51656	0.07	4.19	42	4	4	1	37	4
RGRbv	AT3G29320	PHS1	3	11252472	9221	0.38	4.21	189	33	2	1	91	3
RGRbv	AT4G04770	ABC1/ LAF6	4	2428754	2643	0.23	4.57	47	8	19	4	75	0
RGRbv	AT5G48380	BIR1	5	19605705	83140	0.15	4.28	49	8	2	0	22	0
RGRav	AT1G08920	ESL1	1	2862409	130039	0.13	5.83	70	4	14	0	13	0
	AT1G08930	ERD6						77	1	15	0	14	0
RGRav	At1G22650	NI	1	8015537	14330	0.33	6.53	62	5	17	1	2	2
RGRav	AT4G00430	PIP1;4	4	190004	296	0.37	4.49	73	0	0	0	75	0

Supplemental table 6. Candidate genes from GWAS.
Table gives information on gene function, published natural variation (allelic diversity) for the gene, proof of the phenotype in mutant analyzes (1 = phenotype is proven, 2 = phenotype indicated, 3 = phenotype related gene proven, 4 = phenotype related gene indicated) and references to published papers. LAbv, leaf area before vernalisation; LAav leaf area after vernalisation; FT, flowering time; PL/LL, petiole to leaf length ratio; PL, petiole length; LL, leaf length; RB, rosette branching; MSB. main stem branching; PH1S, plant height at 1st silique; TPH, total plant height; RGRbv, relative growth rate before vernalisation; RGRav, relative growth rate after vernalisation.

Trait	Locus	Gene Abbreviation	Function	Published allelic diveristy	Phe-notype Proof	References
Labv	At1G37130	NR2	Involved in nitrate assimilation. Over-expression increased protein content.	-	2	(Nejidat et al., 1997)
Labv	AT2G21770	CESA9	Cellulose synthase, related to CESA6	-	4	(Somerville, 2006)
Labv	AT5G20730	ARF7	Encodes an auxin-regulated transcriptional activator. Mutant has epinastic rosette leaves. Double mutant arf7 arf19 has small and epinastic leaves	-	2	(Watahiki and Yamamoto, 1997), (Okushima et al., 2005)
Laav	At1G22650	NI	Plant neutral invertase family protein	-	4	(Xiang et al., 2011)
Laav	AT2G21770	CESA9	Cellulose synthase, related to CESA6	-	4	(Somerville, 2006)
Laav	AT5G65010	ASN2	Encodes asparagine synthetase	-	1	(Gaufichon et al., 2013)
	AT5G65050	MAF2	One of a group of MADS box genes involved in control of flowering time.	+	-	(Rosloski et al., 2010; Rosloski et al., 2013)
FT	AT2G18790	PHYB	Red/far-red photoreceptor involved in the regulation of de-etiolation.	-	1	(Halliday et al., 1994)
FT	At2g22540	SVP	Encodes a nuclear protein that acts as a floral repressor and that functions within the thermosensory pathway	+	1	(Mendez-Vigo et al., 2013a)
FT	AT4G00650	FRI	Encodes a major determinant of natural variation in Arabidopsis flowering time	+	1	(Johanson, 2000)
FT	AT4G27060	TOR1	Encodes a novel, plant-specific microtubule-associated protein that regulates the orientation of cortical microtubules and the direction of organ growth	-	2	(Shoji et al., 2004)
FT	AT5G10140	FLC	MADS-box protein encoded by FLOWERING LOCUS C - transcription factor that functions as a repressor of floral transition and contributes to temperature compensation of the circadian clock	+	1	(Koornneef et al., 1994)
FT	At5G45830	DOG1	Encodes DOG1 (DELAY OF GERMINATION 1). A quantitative trait locus involved in the control of seed dormancy.	+	-	(Bentsink et al., 2006)
	At5G45890	SAG12	Senescence-associated gene 12 (SAG12) encoding a cysteine protease influenced by cytokinin, auxin, and sugars	+	-	(Balazadeh et al., 2008)
	At5G45900	AFG7	Component of autophagy conjugation pathway. Required for proper senescence	-	-	(Lenz et al., 2011)
FT	AT5G63190	none	MA3 domain-containing protein, auxin-mediated signaling pathway. Homolog of ECIP1	-	-	(Lei et al., 2011)

FT	AT5G63195	other RNA	Potential natural antisense gene, locus overlaps with AT5G63190	-	-	-
	AT5G63310	NDPK2	Maintains intracellular dNTP levels except ATP. Involved in phytochrome-mediated light signaling. Participates in auxin-regulated processes, partly through the modulation of auxin transport.	-	-	(Choi et al., 2005)
	AT5G63320	NPX1	Encodes NPX1 (Nuclear Protein X1), a nuclear factor regulating abscisic acid responses.	-	-	(Kim et al., 2009)
PL/LL	AT1G68550	CRF10	CYTOKININ RESPONSE FACTOR 10, encodes a member of the ERF (ethylene response factor) subfamily B-6 of ERF/AP2 transcription factor family	-	4	(Millenaar et al., 2005)
PL/LL	AT4G08040	ACS11	encodes an aminotransferase that belongs to ACC synthase gene family structurally	-	3	(Tsuchisaka et al., 2009)
PL/LL	AT5G10470	KCA1	Kinesin that binds cyclin-dependent kinase CDKA;1 as homodimer or as heterodimer with KCA2.	-	4	(Suetsugu et al., 2010)
PL	AT1G68870	SOFL2	AtSOFL1 acts redundantly with AtSOFL2 as positive regulator of cytokinin levels and cytokinin-mediated development.	-	2	(Zhang et al., 2009)
PL	AT2G13810	ALD1	AGD2-like defense response protein 1 (ALD1)	-	2	(Song et al., 2004)
PL	AT4G09800	RPS18C	encodes a ribosomal protein S18C, a constituent of the small subunit of the ribosomal complex	-	2	(Vanlijsebettens et al., 1994)
LL	AT1G26355	SPIL1	SPIRAL1-LIKE1 belongs to a six-member gene family in Arabidopsis	-	3	(Nakajima et al., 2004)
LL	AT1G53330	PPR family	encodes a member of the pentatricopeptide repeat (PPR) gene family	-	2	(Kocabek et al., 2006)
LL	AT4G09800	RPS18C	encodes a ribosomal protein S18C, a constituent of the small subunit of the ribosomal complex	-	2	(Vanlijsebettens et al., 1994)
RB	AT1G17440	CKX1/ EER4	Encodes one of two Arabidopsis proteins with similarity to the TBP-associated factor TAF12. Loss of function mutants show enhanced response to ethylene	-	4	(Kubo et al., 2011)
RB	At3g18550	BRANCHED1	Encodes a TCP transcription factor, closely related to teosinte branched1, arrests axillary bud development and prevents axillary bud outgrowth	-	1	(Gonzalez-Grandio et al., 2013; Niwa et al., 2013)
RB	AT5G35080	OS9	Encodes a protein involved in the endoplasmic reticulum-associated degradation of glycoproteins	-	-	(Huttner et al., 2012)
MSB	AT1G10870	AGD4	A member of ARF GAP domain (AGD)	-	-	(Min et al., 2013)
MSB	AT3G26570	PHT2;1	low affinity phosphate transporter	-	-	(Versaw, 2002)
MSB	AT3G60530	GATA4	Encodes a member of the GATA factor family of zinc finger transcription factors.	-	-	(Reyes et al., 2004)
MSB	AT3G63300	FORKED1	Encodes a pleckstrin homology domain- and DUF828-containing protein. Proposed to be a key component of the auxin canalization pathway.	-	-	(Hou et al., 2010)
	AT3G63310	BIL4	Mediates cell elongation in brassinosteroid signaling	-	1	(Yamagami et al., 2009)
	AT3G63440	CKX6	It encodes a protein whose sequence is similar to cytokinin oxidase/dehydrogenase, which catalyzes the degradation of cytokinins.	-	4	(Werner et al., 2003)

	AT3G63445	none	Potential natural antisense gene, locus overlaps with AT3G63440	-	-	-
MSB	AT5G03250	none	Phototropic-responsive NPH3 family protein	-	-	-
MSB	AT5G18400	DRE2	Cytokine-induced anti-apoptosis inhibitor 1, Fe-S biogenesis	-	-	(Bernard et al., 2013)
PH1S	AT1G24150	FH4	Encodes a group I formin. Localized to cell junctions. Polymerizes actin. Binds profilin.	-	2	(Deeks et al., 2010)
PH1S	AT1G80330	GA3OX4	Encodes a protein with gibberellin 3-oxidase activity. The enzyme, expressed and purified in E.coli, was shown to catalyze the 3β-hydroxylation of GA20 into GA29	-	2,3	(Hu et al., 2008; Barboza et al., 2013)
	AT1G80340	GA4H	Encodes a protein with gibberellin 3 β-hydroxylase activity. The protein was heterologously expressed in E. coli and shown to catalyze the hydroxylation of both GA9 and GA20.	-	2	(Mitchum et al., 2006)
PH1S	AT2G35350	PLL1	Encodes a protein most similar to the POLTERGEIST locus. Double mutant analyzis of loss of function alleles indicate PLL1 functions redundantly with POL to regulate meristem size and pedicel length.	-	-	(Song et al., 2006)
PH1S	AT3G23590	RFR1	Encodes a protein shown to physically associate with the conserved transcriptional coregulatory complex, Mediator, and is involved in the regulation of phenylpropanoid homeostasis	-	3	(Stout et al., 2008)
PH1S	AT4G09460	MYB6	Encodes myb6 DNA-binding protein	-	-	(Li and Parish, 1995)
PH1S	AT4G38970	FBA2	Protein is tyrosine-phosphorylated and its phosphorylation state is modulated in response to ABA in Arabidopsis thaliana seeds.	-	-	(Lu et al., 2012))
	AT4G38990	GH9B16	glycosyl hydrolase 9B16	-	-	-
TPH	AT1G19485	WD40-repeat	Transducin/WD40 repeat-like superfamily protein	-	4	(Zhang et al., 2008b)
RGRbv	AT1G04860	UBP2	Encodes a ubiquitin-specific protease.	-	4	(Vierstra, 2003)
	AT1G04920	SPS3F	Encodes a protein with putative sucrose-phosphate synthase activity.	-	4	(Huber and Huber, 1996)
RGRbv	AT1G69410	ELF5A-3	Encodes eIF5A-2, a putative eukaryotic translation initiation factor.	-	2	(Feng et al., 2007)
RGRbv	AT1G78370	GSTU20	Encodes glutathione transferase belonging to the tau class of GSTs.	-	2	(Chen et al., 2007)
RGRbv	AT3G29320	PHS1	Encodes a plastidic alpha-glucan phosphorylase.	-	2	(Zeeman et al., 2004)
RGRbv	AT4G04770	ABC1/ LAF6	Encodes an iron-stimulated ATPase. A member of the NAP subfamily of ABC transporters.	-	2	(Nagane et al., 2010)
RGRbv	AT5G48380	BIR1	Encodes a BAK1-interacting receptor-like kinase named BIR1. Negatively regulates multiple plant resistance signaling pathways, one of which is the SOBIR1(AT2G31880)-dependent pathway.	-	2	(Wang et al., 2011)
RGRav	AT1G08920	ESL1	Encodes ESL1, a transporter for monosaccharides.	-	-	(Yamada et al., 2010)
	AT1G08930	ERD6	encodes a putative sucrose transporter whose gene expression is induced by dehydration and cold.	-	-	(Kiyosue et al., 1998; Xiang et al., 2011)
RGRav	At1G22650	NI	Plant neutral invertase family protein	-	4	Xiang et al. (2011)
RGRav	AT4G00430	PIP1;4	a member of the plasma membrane intrinsic protein subfamily PIP1.	-	-	(Kinoshita et al., 1994)

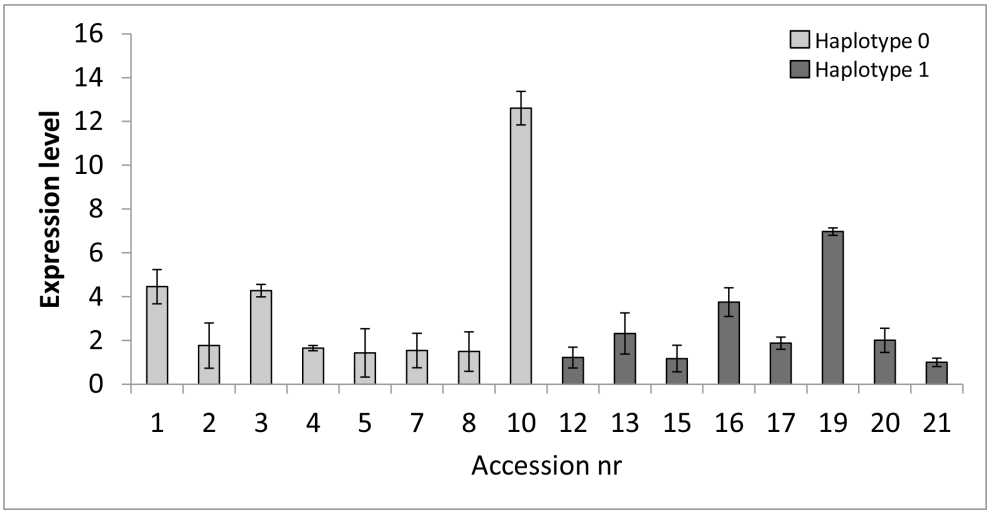
Supplemental table 7. Nucleotide diversity analyzis. Π_T = total nucleotide diversity, π_{ns} =nucleotide diversity at non-synonymous sites, π_s = nucleotide diversity at synonymous sites and π_{sil} = nucleotide diversity at silent sites. LAbv, leaf area before vernalisation; LAav leaf area after vernalisation; FT, flowering time; PL/LL, petiole to leaf length ratio; PL, petiole length; LL, leaf length; RB, rosette branching; MSB. main stem branching; PH1S, plant height at 1st silique; TPH, total plant height; RGRbv, relative growth rate before vernalisation; RGRav, relative growth rate after vernalisation.

Trait	Locus	Gene	π_T	π_{ns}	π_s	π_{sil}	π_{ns}/π_s ratio	Tajima's D_T	Tajima's D_{ns}	Tajima's D_s	Tajima's D_{sil}	Tajima's D_{ns}/D_s	D (Fu & Li)	F (Fu & Li)
Labv	At1G37130	NR2	0.005	0.000	0.017	0.008	0.017	-1.5	-1.9	-0.8	-1.5	2.3	-5.0	-3.6
Labv	AT2G21770	CESA9	0.004	0.001	0.007	0.005	0.162	-1.8	-1.8	-1.2	-1.8	1.5	-3.5	-3.0
Labv	AT5G20730	ARF7	0.004	0.001	0.008	0.007	0.116	-1.0	-1.7	-0.9	-0.9	1.9	-2.1	-1.8
Laav	At1G22650	NI	0.005	0.001	0.022	0.007	0.046	-1.4	-1.2	0.6	-1.3	-2.1	-3.8	-3.0
Laav	AT2G21770	CESA9	0.004	0.001	0.007	0.005	0.162	-1.8	-1.8	-1.2	-1.8	1.5	-3.5	-3.0
Laav	AT5G65010	ASN2	0.001	0.000	0.000	0.001	0.837	-2.3	-1.4	-2.0	-2.3	0.7	-6.1	-4.9
Laav	AT5G65050	MAF2	0.003	0.001	0.001	0.003	0.875	-1.9	-2.4	-1.5	-1.8	1.6	-6.8	-4.8
FT	AT2G18790	PHYB	0.005	0.001	0.017	0.009	0.040	-1.0	-1.7	-0.2	-0.8	8.5	-1.5	-1.4
FT	At2g22540	SVP	0.005	0.000	0.004	0.005	0.024	-1.4	-1.8	-1.1	-1.4	1.6	-6.0	-4.0
FT	AT4G00650	FRI	0.003	0.003	0.001	0.002	2.338	-2.2	-1.9	-2.2	-2.2	0.9	-4.9	-4.0
FT	AT4G27060	TOR1	0.008	0.005	0.015	0.009	0.368	-0.1	0.6	1.9	-0.3	0.3	-3.7	-2.0
FT	AT5G10140	FLC	0.004	0.000	0.002	0.004	0.110	-1.6	-2.0	-1.0	-1.6	2.1	-5.4	-3.8
FT	At5G45830	DOG1	0.005	0.005	0.011	0.005	0.415	-1.7	-1.8	-0.7	-1.7	2.4	-4.0	-3.2
FT	At5G45890	SAG12	0.004	0.002	0.007	0.005	0.335	-1.5	-1.1	-1.6	-1.5	0.7	-3.2	-2.7
FT	At5G45900	AFG7	0.002	0.000	0.001	0.003	0.286	-2.1	-2.1	-1.9	-2.0	1.1	-6.7	-5.0
FT	AT5G63190	none	0.003	0.001	0.003	0.005	0.188	-1.9	-1.8	-1.8	-1.9	1.0	-5.5	-4.0
FT	AT5G63310	NDPK2	0.001	0.000	0.003	0.002	0.080	-2.0	-2.2	-1.2	-1.9	1.7	-5.9	-4.8
FT	AT5G63320	NPX1	0.001	0.000	0.002	0.002	0.186	-2.4	-2.3	-1.6	-2.3	1.5	-9.3	-6.4
PL/LL	AT1G68550	CRF10	0.002	0.000	0.007	0.002	0.073	-2.2	-2.1	-1.1	-2.1	1.8	-5.0	-4.2
PL/LL	AT4G08040	ACS11	0.006	0.002	0.006	0.007	0.344	-1.7	-0.8	-1.1	-1.8	0.8	-3.2	-2.8
PL/LL	AT5G10470	KCA1	0.002	0.001	0.002	0.002	0.585	-1.5	-1.2	-1.5	-1.6	0.8	-6.2	-4.4
PL	AT1G68870	SOFL2	0.005	0.002	0.012	0.006	0.144	-0.1	-1.9	1.3	0.1	-1.4	-2.9	-1.8
PL	AT2G13810	ALD1	0.011	0.001	0.006	0.013	0.146	-0.8	-2.0	-1.8	-0.7	1.1	-3.2	-2.2
PL	AT4G09800	RPS18C	0.006	0.000	0.002	0.007	0.015	-2.1	-1.2	-1.9	-2.1	0.6	-4.8	-3.9
LL	AT1G26355	SPIL1	0.005	0.002	0.000	0.006	5.484	-1.2	-1.7	-1.1	-1.1	1.6	-3.8	-2.9
LL	AT1G53330	PPR family	0.011	0.006	0.027	0.013	0.227	-0.5	-0.6	0.1	-0.4	-5.5	-2.8	-1.8
LL	AT4G09800	RPS18C	0.006	0.000	0.002	0.007	0.015	-2.1	-1.2	-1.9	-2.1	0.6	-4.8	-3.9
RB	AT1G17440	CKX1/ EER4	0.004	0.001	0.009	0.006	0.113	-0.8	-0.7	0.3	-0.8	-2.3	-3.3	-2.4
RB	At3g18550	BRANCHED1	0.003	0.001	0.004	0.003	0.225	-2.0	-2.2	-1.6	-1.9	1.4	-2.2	-2.5
RB	AT5G35080	OS9	0.005	0.002	0.003	0.006	0.700	-1.6	-1.3	-1.9	-1.6	0.7	-4.4	-3.3
MSB	AT1G10870	AGD4	0.007	0.001	0.010	0.009	0.129	0.1	-1.5	0.4	0.3	-3.5	-4.8	-2.4
MSB	AT3G26570	PHT2;1	0.009	0.002	0.024	0.013	0.104	0.1	0.0	1.7	0.1	0.0	-2.4	-1.2
MSB	AT3G60530	GATA4	0.006	0.001	0.014	0.007	0.096	-1.4	-1.4	-0.6	-1.4	2.4	-4.8	-3.5
MSB	AT3G63300	FORKED1	0.004	0.002	0.005	0.005	0.442	-1.5	-1.1	-1.3	-1.6	0.9	-9.0	-5.6
MSB	AT3G63310	BIL4	0.002	0.001	0.001	0.002	0.899	-1.8	-0.7	-1.6	-1.9	0.4	-6.0	-4.8
MSB	AT3G63440	CKX6	0.002	0.002	0.001	0.002	1.306	-1.8	-0.6	-1.8	-2.0	0.3	-4.6	-3.8
MSB	AT5G03250	none	0.010	0.001	0.032	0.016	0.043	-0.4	-1.4	0.8	-0.2	-1.8	-1.3	-0.9
MSB	AT5G18400	DRE2	0.008	0.003	0.024	0.009	0.134	0.7	0.8	3.4	0.6	0.2	-1.1	-0.2
PH1S	AT1G04860	UBP2	0.001	0.000	0.004	0.002	0.098	-2.1	-2.3	-1.4	-1.8	1.7	-5.4	-4.4

PH1S	AT1G19485	WD40	0.005	0.002	0.007	0.007	0.332	-0.9	-1.3	-0.5	-0.8	2.6	-2.8	-2.0
PH1S	AT1G24150	FH4	0.004	0.001	0.009	0.006	0.150	-1.3	-1.8	-0.7	-1.1	2.6	-3.7	-2.8
PH1S	AT1G80330	GA3OX4	0.005	0.001	0.007	0.006	0.189	-1.7	-1.5	-0.1	-1.7	21.4	-4.3	-3.5
PH1S	AT1G80340	GA4H	0.005	0.003	0.004	0.006	0.667	-2.2	-1.6	-1.4	-2.2	1.2	-9.3	-5.8
PH1S	AT2G35350	PLL1	0.004	0.001	0.010	0.005	0.106	-1.0	-1.3	0.0	-1.0	-201.1	-5.0	-3.4
PH1S	AT3G23590	RFR1	0.002	0.000	0.002	0.003	0.189	-2.4	-2.5	-2.2	-2.4	1.1	-8.3	-5.5
PH1S	AT4G09460	MYB6	0.007	0.003	0.018	0.008	0.189	-2.0	-1.8	-0.9	-2.0	2.0	-3.7	-3.2
PH1S	AT4G38970	FBA2	0.004	0.000	0.003	0.005	0.010	-1.0	-1.8	-1.4	-0.9	1.3	-4.8	-3.3
PH1S	AT4G38990	GH9B16	0.016	0.009	0.024	0.018	0.350	-1.0	-0.9	-0.3	-1.0	3.0	-8.8	-4.8
TPH	AT1G04920	SPS3F	0.001	0.000	0.002	0.001	0.120	-2.4	-2.4	-2.1	-2.4	1.1	-9.4	-6.7
RGRbv	AT1G69410	ELF5A-3	0.003	0.000	0.001	0.003	0.007	-2.2	-1.1	-1.6	-2.2	0.7	-6.1	-4.7
RGRbv	AT1G78370	GSTU20	0.009	0.002	0.026	0.012	0.064	-1.4	-1.4	0.4	-1.4	-3.1	-3.8	-2.8
RGRbv	AT3G29320	PHS1	0.007	0.002	0.014	0.010	0.154	-2.0	-1.9	-1.4	-2.0	1.4	-6.2	-4.2
RGRbv	AT4G04770	ABC1/ LAF6	0.011	0.001	0.021	0.017	0.069	-1.8	-1.1	-0.5	-1.8	2.4	-6.6	-4.3
RGRbv	AT5G48380	BIR1	0.005	0.002	0.017	0.007	0.102	-1.7	-1.1	0.5	-1.8	-2.1	-8.7	-5.4
RGRav	AT1G08920	ESL1	0.005	0.001	0.009	0.007	0.145	-1.7	-1.4	-0.2	-1.7	8.0	-9.3	-5.6
RGRav	AT1G08930	ERD6	0.005	0.000	0.009	0.006	0.034	-1.8	-1.7	-0.6	-1.7	3.0	-8.7	-5.4
RGRav	At1G22650	NI	0.005	0.001	0.022	0.007	0.046	-1.4	-1.2	0.6	-1.3	-2.1	-3.8	-3.0
RGRav	AT4G00430	PIP1;4	0.013	0.001	0.020	0.015	0.031	-1.7	-1.7	0.0	-1.7	-190.8	-7.6	-4.7

Supplemental table 8. Accessions used in the ethylene complementation experiment. LL and PL were measured 22 DAG. Haplotypes based on initial analysis and highest candidate SNP. 4887214, 48883127 and 4888589 indicate the position of the SNP on chromosome 4 in the *ACS11* gene.

Accession details				**Leaf measurements**			**Marker**		
NR	**CS nr**	**Accession**	**Haplotype**	**LL**	**PL**	**PL/LL**	**4887214**	**4888317**	**4888589**
1	CS28013	Alst-1	0	28.30	9.15	0.32	A	A	G
2	CS28759	Ting-1	0	23.70	9.07	0.38	A	A	G
3	CS28780	Tsu-0	0	32.00	13.70	0.43	A	A	G
4	CS28809	Wag-4	0	28.90	11.00	0.38			
5	CS76093	Ba1-2	0	24.50	8.18	0.33	A	G	G
6	CS76101	Br-0	0	26.70	10.30	0.39	A	A	G
7	CS76265	TOU-I-6	0	26.90	10.50	0.39			
8	CS76292	UKSW06-202	0	26.80	11.60	0.43			
9	CS76296	Uod-7	0	21.60	9.44	0.44	A	A	G
10	CS76298	Var2-1	0	23.80	9.53	0.40	A	G	G
11	CS28128	Ca-0	1	25.30	6.23	0.25	A	G	A
12	CS28133	Cha-0	1	22.10	5.13	0.23			
13	CS28252	Fi-1	1	28.10	8.01	0.29			
14	CS28583	Old-1	1	27.30	7.46	0.27	C	G	A
15	CS28713	RRS-7	1	28.80	6.82	0.24	A	A	A
16	CS76116	Cvi-0	1	25.30	6.23	0.25	A	A	G
17	CS76125	Eden-2	1	22.10	5.13	0.23	C	G	A
18	CS76214	Pro-0	1	25.50	6.67	0.26	C	G	A
19	CS76293	Ull2-3	1	28.90	6.89	0.24	C	G	A
20	CS76301	Wei-0	1	27.00	6.77	0.25	C	G	A
21	CS76113	Col-0	1	21.30	6.81	0.32	A	G	A



Supplemental figure 4. Expression level of different Arabidopsis accessions.
Expression levels were normalized to the best 3 reference genes in BestKeeper, and finally normalized to Col-0 (nr. 21). Light grey = accessions belonging to haplotype 0 and dark grey = accessions belonging to haplotype 1. Accession nr's are listed in Supplemental table 8.

Supplemental table 9. Differences in leaf length (LL), petiole length (PL) and the petiole to leaf length ratio (PL/LL) between wild type (WT), *acs11-1* single mutant, *acs octuple* mutant and *ein2-1*.

Genotype	LL	PL	PL/LL
WT	34.25	13.50	0.39
acs11-1	34.14	12.84	0.38
acs octuple	22.31	7.82	0.35
ein2-1	27.51	9.71	0.35

Supplemental table 10. Primers used for RT-qPCR.

Gene	Locus ID	Forward 5'-3'	T_m	Reverse 5'-3'	T_m
ACS11	AT4G08040	CGACCGTGTTAAAGGTTTTG	60˚C	GGAAAGCTTGGAGACCCATT	60˚C
GUT1	AT5G61840	CTCAGCAGCAAACATGTGGT	60˚C	GAGCTGCGTTTTGATCTTCC	60˚C
GOX	AT1G18610	AAGAGGCCAGCTATGTCGAA	60˚C	GTTAACACCCAGAGGGTTGC	60˚C
HOP	AT3G18380	ATACGCGGCATGGAAGATAC	60˚C	GGTAACGGAGATGGAAGCAA	60˚C
RZF	AT1G11020	GGAAGCTTGTTTCGCTCTTG	60˚C	GGGGAGTGTAACTTCCGTGA	60˚C
UBQ10	AT4G05320	GGTTTGTGTTTTGGGGCCTTG	60˚C	CGAAGCGATGATAAAGAAGAAGTTCG	60˚C

CHAPTER 3

Multi-dimensional regulation of metabolic networks shaping plant development and performance

Authors:

Rik Kooke[1,2,3] and Joost J.B. Keurentjes[2,3]

[1] Laboratory of Plant Physiology, Wageningen University, the Netherlands
[2] Laboratory of Genetics, Wageningen University, the Netherlands
[3] Centre for Biosystems Genomics, Wageningen Campus, the Netherlands

Journal of Experimental Botany (2012) 63(9), pp 3353-3365
doi: 10.1093/jxb/err373

Abstract

The metabolome is an integral part of a plant's life cycle and determines for a large part its external phenotype. It is the final, internal product of chemical interactions, obtained through developmental, genetic and environmental input, and as such, it defines the state of a plant in terms of development and performance. Understanding its regulation will provide knowledge and new insights into the biochemical pathways and genetic interactions that shape the plant and its surroundings. In this review, we will focus on four dimensions that contribute to the huge diversity of metabolomes and we will illustrate how this diversity shapes the plant in terms of development and performance: (I) temporal regulation: the metabolome is extremely dynamic and temporal changes in the environment can have an immense impact on its composition; (II) spatial regulation: metabolites can be very specific, in both quantitative and qualitative terms, to specialized organs, tissues and cell types; (III) environmental regulation: the metabolic profile of plants is highly dependent on environmental signals, such as light, temperature and nutrients, and very susceptible to biotic and abiotic stresses; (IV) the biosynthesis, structure and accumulation of metabolites have a genetic origin, and there is quantitative and qualitative variation for metabolomes within a species. We will address the contribution of these dimensions to the wide diversity of metabolomes, and we will highlight how the multi-dimensional regulation of metabolism defines the plant's phenotype.

Introduction

Metabolism is an integral part of plant growth, development and performance. The many functions of metabolites include the formation of building blocks for the cell wall machinery, the regulation of flux between source and sink tissues, the attraction of pollinators and the defense against pathogens and herbivores (Roitsch, 1999; Cosgrove, 2005; Allwood et al., 2008). Although the distinction is not always clear, most metabolites and their accompanying functions are usually assigned to primary or secondary metabolism. Metabolites that play a major role in cell maintenance, development and reproduction, such as amino acids, carbohydrates and organic acids, are generally referred to as primary metabolites (Kliebenstein, 2004; Fernie, 2007). Because these are essential to ensure proper development of the sessile plant, their biochemical diversity has been constrained during evolution. Yet, they are structurally extremely complex, under redundant gene regulation and fine-tuned by feedback mechanisms to ascertain plant survival in a wide range of circumstances (Matsuda et al., 2010). Secondary metabolites, on the other hand, have a wider range of chemical diversity but are less complex regulated and show a wider pattern of variation in chemical profiles, both in quantitative and qualitative terms (Keurentjes et al., 2006). It is thought that this large metabolic variation in secondary metabolites can be explained by the variety of interactions that different plants, or specific organs or tissues, have with their environment (Allwood et al., 2008). Their chemical structure can more easily be modified without drastic detrimental effects leading to newly formed compounds that could be advantageous in certain conditions. Natural selection, which is strong for the essential primary metabolites and more relaxed for secondary metabolites may thus have shaped the biochemical composition of plants. The secondary metabolites can be subdivided into 5 major groups based on their biosynthetic origin: polyketides, isoprenoids, alkaloids, phenylpropanoids and flavonoids (Oksman-Caldentey and Inze, 2004). These compounds have diverse roles, including those involved in biotic and abiotic stress responses, hormonal regulation, and to a lesser extent in development (D'Auria and Gershenzon, 2005).

Due to recent advances in detection and annotation of metabolites, the field of plant metabolomics has seen a tremendous increase in popularity over the last decade. The huge, but relatively unknown natural diversity in plant metabolomes has found its way in applications as wide as medicinal science, breeding and food quality research. Nonetheless, only a very small part of the estimated more than 400.000 unique metabolites that the plant kingdom hosts has been discovered thus far (Oksman-Caldentey and Inze, 2004; Saito and Matsuda, 2010). This enormous source of unknown metabolites is currently being explored for many different quality traits (e.g. bioactive agents, fortified foods, commodity chemicals). With the advances made in mass spectrometry and NMR technologies, these molecules can be detected and identified at progressively increasing speed and sensitivity. Even though this has tremendously increased our fundamental knowledge of plant metabolites, we still have very little understanding of their functional roles, let alone their regulation.

The metabolome is the final product of genetic and developmental control and environmental interactions, and as such provides a blueprint for plant growth and performance (Figure 1). Understanding the regulation of the metabolome enables a better understanding of the biochemical pathways that control development. In this review, we will outline several different dimensions of metabolic regulation and link these to control of plant development and performance: (I) temporal regulation: the metabolic profile is highly dynamic and temporal changes, ranging from diurnal rhythms to transitional and seasonal changes can greatly affect this composition; this enables plants to adapt rapidly and often reversibly to changing conditions; (II) spatial regulation: like variation in time, metabolic accumulation may also vary spatially over developmental stages or be specific to certain organs, tissues or cell types; as such, they often determine the very specialized functions of organs; (III) environmental regulation: the flexibility in metabolic content is further illustrated by the dependency on environmental cues, such as light, temperature, nutrient availability, and biotic and abiotic stresses; in addition, many metabolites can act as internal or external signaling molecules and thus enable interactive communication with the environment; (IV) genetic regulation: finally, the metabolic biosynthesis routes are encoded in the plant genome and as such under genetic control; many instances are known where metabolic profiles are determined by the interaction of the local environment with the genetic make-up of plants.

We will address the multi-dimensional regulation of plant metabolism in relation to growth, development and performance and highlight the distinct roles of both primary and secondary metabolism in these processes.

Temporal regulation of metabolism

The temporal regulation of plant metabolism is, besides the intrinsic developmental program, largely dependent on external factors. As the plant's sessile nature impedes its escape from unfavorable environments, it has evolved a highly dynamic and flexible metabolism to cope with changing conditions. Some of these adaptive mechanisms are essential for the completion of the plant's life cycle and have an autonomous or genetic origin. These include transitional changes, e.g. from a vegetative to a reproductive stage, and specific responses to long periods of extreme circumstances, e.g. drought, cold or low light. Others become entrained by the environment through repeated cycling, such as seasonal changes and diurnal rhythms of light and temperature. Importantly, many interactions between the different regulatory mechanisms can be observed. In this section, we will highlight the role and regulation of the metabolome during these temporal fluctuations.

As plants depend on external energy and structural resources like carbohydrates their growth and development is largely tied to their photosynthetic capacity. The formation of many primary metabolites directly depends on the availability of light and their abundance

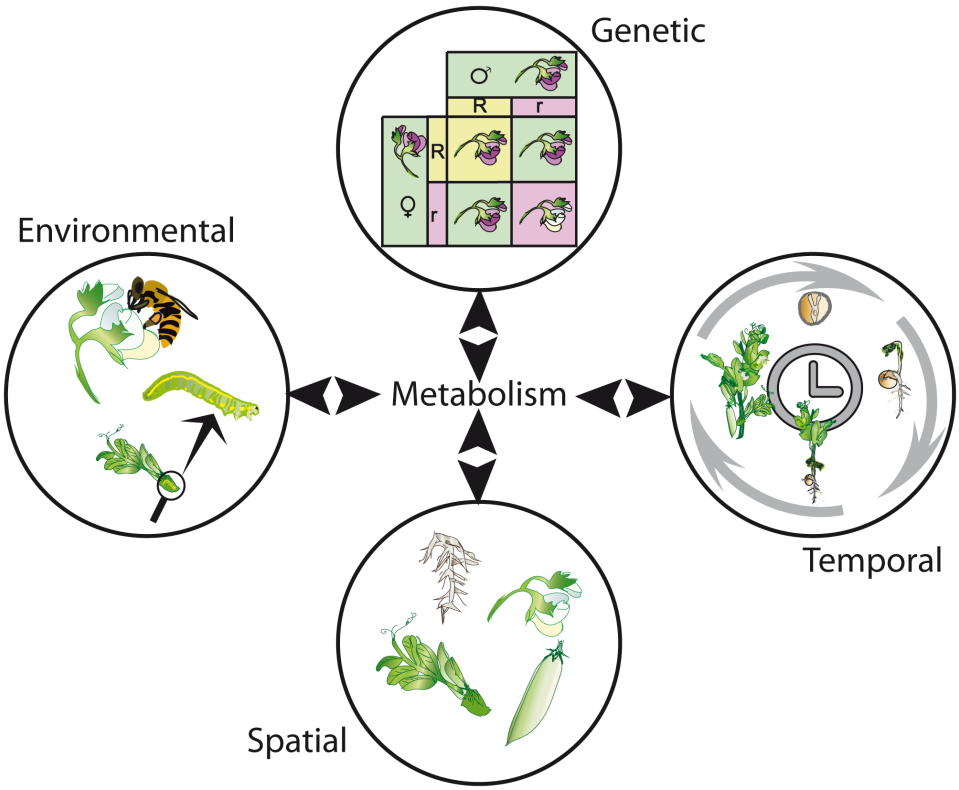


Figure 1. The multi-dimensional regulation of plant metabolism.
The interplay between temporal, spatial, environmental and genetic factors determines the quantitative and qualitative profile of a plant's metabolome.

therefore cycles over day and night periods. Such diurnal cycles are often mistaken with circadian rhythms but their regulation is mechanistically different. Circadian rhythms are, although entrained by day-night differences, controlled by the internal biological clock and maintain cycling even in the absence of diurnal rhythms. The separation of a day in periods of light and darkness, following Earth's rotation and consequent facing towards the sun, has led to the evolution of biological clocks that phase metabolism and growth with these repetitive cycles. Although the circadian clock shows strong resemblance in regulation and features among all organismal kingdoms, it has most likely evolved independently, suggesting major adaptational benefits (Young and Kay, 2001). Indeed, the accurate corresponding between the internal circadian clock and the external diurnal rhythms increases photosynthesis, vegetative growth and fitness in *Arabidopsis thaliana* (Michael et al., 2003; Dodd et al., 2005).

Although there are strong indications that many diurnal oscillating metabolites are regulated

through circadian function (Harmer et al., 2000), detailed studies on the temporal regulation of large sets of metabolites are scarce. In a study on cold acclimation in Arabidopsis, most sugars, but also many organic and amino acids, showed diurnal rhythms in their abundance with the majority of these compounds being circadian regulated (Espinoza et al., 2010). The precise timing between the circadian clock and diurnal rhythms, related to higher growth rates, might be achieved through complex biochemical networks, providing major adaptational benefits (Dodd et al., 2005; Espinoza et al., 2010). Exemplary, starch, the major storage molecule in the plant kingdom, is linearly degraded during the night to ascertain the continuation of vegetative growth in dark conditions. At dawn, 95% of the total starch reserves is, each day, broken down to support metabolism and normal growth (Smith and Stitt, 2007). Interestingly, a sudden, significant decrease in the length of the light period is followed by immediate modifications in the rate of starch breakdown as such that there is a constant supply of starch during the unexpected, prolonged night (Graf et al., 2010) suggesting that plants hold a timing mechanism, anticipated on the previous dawn, to optimize plant growth and productivity (Graf and Smith, 2011). The significant role that the clock plays in these processes is even better exemplified by a recent study in which the light-dark (LD) cycle was either extended or shortened (Graf et al., 2010). When the LD cycle was prolonged to 28 h, starch was exhausted too soon, while a reduced LD cycle of 20 h caused significant remains of starch at dawn. In both situations, plants maintained a 24 h growth cycle, which was most likely controlled through the circadian clock. Additionally, the authors showed that the unbalance between the LD cycle and the circadian cycle caused a significant reduction in growth. Moreover, in a wide range of Arabidopsis accessions, plant biomass is negatively correlated with starch content at the end of the day and at the end of the night (Cross et al., 2006; Sulpice et al., 2009). This implies a complex interplay between carbon allocation and the clock, such that the photosynthetic products are optimally used during the day, and starch during the night, to acquire the highest growth rates.

Many genes encoding enzymes for primary carbon metabolism are controlled by the circadian clock as their rhythmic expression continued under continuous light cycles (Harmer et al., 2000; Lu et al., 2005). It is, therefore, tempting to speculate that the optimal allocation of carbon during day and night, and hence maximum growth rates, are controlled by the circadian clock. Illustratively, both maltose and sucrose are highly correlated with the rate of starch degradation during different day/night regimes and especially maltose sustained an oscillating pattern under continuous light conditions (Lu et al., 2005). Interestingly, recent findings suggest that the circadian clock not only regulates metabolism, but is in turn regulated by metabolic changes (Dalchau et al., 2011). Exogenous supplied sucrose affects the functioning of the circadian clock, most profoundly in dark conditions when the endogenous sucrose content is low. Moreover, exogenous sucrose modulates the period of the clock and has a minor effect on the expression of circadian clock associated genes (Knight et al., 2008). Additionally, a study comparing polymorphic loci in circadian oscillation and metabolomics

showed that *AOP2*, which encodes a biosynthetic glucosinolate enzyme, altered circadian clock regulation (Kerwin et al., 2011), suggesting a bidirectional relationship between metabolism and clock oscillation. These results altogether show the interconnectivity and complex regulation between metabolism and the clock to optimize carbon allocation and growth. The circadian control and the metabolic feedback mechanisms highlight the importance of its accurate phasing with diurnal cycles for optimal plant performance.

Although quite different from the reversible diurnal patterns, transitional changes in plants, such as seed germination and flowering, are thought to be controlled by the clock through the sensing of shifts in day length and temperature to match the timing of the response with most optimal conditions (Gould et al., 2006; McWatters and Devlin, 2011). Transitions are accompanied by major changes in metabolic composition, of which particularly the role of primary metabolites again is important and best studied. A first and major transition in a plant's life cycle is germination. The transformation of a dormant seed into a photosynthetic active seedling is accompanied by the conversion of storage compounds into essential building blocks and energy carriers. Upon imbibition, there is a strong increase in hexose sugars followed by a large change in gene transcription (Fait et al., 2006; Howell et al., 2009). Subsequently, most primary metabolites acquire a stable metabolic state which is more or less enhanced for 24 h after exposure to light and inductive temperature (Allen et al., 2010). Thereafter, there is a major metabolic switch in which many metabolites are being consumed to form building blocks and energy carriers. Sucrose, for example, was found to decrease throughout the developmental period, and interestingly, many transcripts were highly correlated with sucrose levels, indicating metabolic regulation of transcript abundance. The causality in these cases is difficult to prove, and it is likely that there is a very complex feedback regulation between metabolic status, gene transcription and the environment to maximize germination vigor. Many of the distinct phases in a germinating seed can be characterized by these transient metabolic profiles (Fait et al., 2006), indicative of the flexibility of the metabolome towards different temporal needs (Figure 2).

The development from a vegetative to a reproductive state also leads to massive changes in metabolic content for which it is not always clear whether this is causal for, or a result of the transition (Giakountis and Coupland, 2008). The cell division in the shoot apical meristem, indicative of the transition to flowering, is preceded by an essential shift in the level of sucrose, induced by a long or displaced short day (Lejeune et al., 1993; Corbesier et al., 1998). Moreover, many late and early flowering mutants, have a flowering time similar to wild type when grown on media containing sucrose (Araki and Komeda, 1993; Roldan et al., 1999). Additionally, mutants defective in trehalose-6-phosphate synthase are unable to flower (van Dijken et al., 2004). In contrast, flowering is delayed in plants growing on medium with added glucose or sucrose (Zhou et al., 1998; Ohto et al., 2001; Gibson, 2005). These studies indicate a significant role for sugars and related metabolites in floral transition

which is likely dependent on circadian timing.

Because of the strong interconnectivity of metabolic networks, feedback mechanisms to optimally canalize resources, are often observed (Arsenault et al., 2010), explaining the comprehensive metabolic shift of reproductive transitions. An important but often neglected transition is the shift of tissues form sink to source or vice versa. A clear example is the formation of tap roots in biennial plants, which act as sinks in their first year and as source in their second (Godt and Roitsch, 2006). The preferential allocation of resources to a specific tissue can have large consequences for other plant parts and can even affect overall growth and development (van Heerden et al., 2010). More subtle is the gradual transition of leaves from sink to source tissues (Jeong et al., 2004). In annual species, this is often accompanied with progressive senescence and a re-allocation of substrates to reproductive organs. In perennial species, however, leaves function as sinks only during their development, but upon maturity serve as sources which remain vivid throughout the growing season. In contrast to annuals, which are driven much more by their developmental program, perennials react predominantly on environmental cues (Brenes-Arguedas et al.,

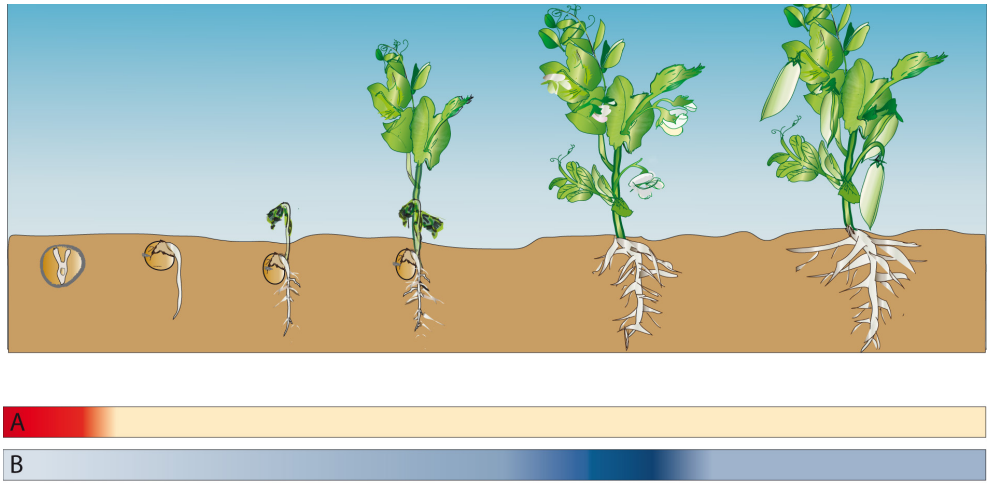


Figure 2. Temporal accumulation patterns of metabolites.
Many metabolites show transient or discrete accumulation patterns during plant development. The intensity gradients indicate the extent of accumulation within a specific developmental stage. (A) Qualitatively tissue-specific accumulation of metabolites in seeds; (B) Quantitatively development-specific accumulation of metabolites during the transition to flowering. Similarly, metabolic accumulation could be quantitatively tissue-specific or qualitatively development-specific, or due to an interaction of tissue and developmental specificity (e.g. only in leaves at a specific developmental stage).

2006). Seasonal changes in light quality and temperature tremendously influence diurnal rhythms and have a major impact on metabolic changes in perennial species (Hoffman et al., 2010). These metabolic changes underlie much of the differences in growth between seasons (Richardson et al., 2009), even in evergreens (Ceusters et al., 2010). Nonetheless, many transcriptional regulators of developmental control are conserved over annual and perennial species, and it remains to be seen how this relates to their metabolic signatures (Brunner and Nilsson, 2004; Zhang et al., 2011). Plants show a remarkable adaptation of their metabolic composition to temporal changing conditions and needs. Some of these are abrupt and irreversible, while others are reoccurring and reversible. It is evident, however, that the different responses are tightly interlinked and should be considered in relation to each other.

Spatial regulation of metabolism

Plants are modular organisms that consist of a wide variety of organs, tissues and cell types. Each of these different entities is characterized by a unique and specific developmental program which is also reflected in their metabolic composition. Due to technical constraints however, metabolic studies have thus far mainly focused on convoluted plant parts such as shoots and roots, neglecting the differences in tissues and cell types within these plant organs. Fortunately, technological advances in micro-dissection, fractionation techniques, analytical sensitivity and bioinformatics have allowed the focus of metabolism to shift towards the detection of metabolites in specialized organs and cell types (Fernie, 2007).

A number of studies have reported on relationships between plant development and metabolic status of different plant organs such as leaves, roots, flowers, seeds and fruits. In relatively few studies, different organs were also compared to each other (Brown et al., 2003; Desbrosses et al., 2005; Velasco et al., 2008; Matsuda et al., 2009; Malik et al., 2010; Matsuda et al., 2010; Moing et al., 2011). These studies illustrated that plants are able to synthesize large numbers of different metabolites of which some are more general and accumulating in many tissues, while others are very specific for distinct tissues. The particular accumulation of metabolites suggests specialized functions in different stages of a plant's development (Sergeeva et al., 2004). Primary metabolites, for instance, can be traced back in most tissues with mainly quantitative differences, indicating essential functions of carbohydrates. This is consistent with their major role in plant growth and development making them essential constituents of every cell type (Matsuda et al., 2010). Secondary metabolites, on the other hand, have a much wider chemical diversity and range of functions, and are often very specific to certain plant tissues. For example, benzoyloxylated glucosinolates and proanthocyanids are only found in seeds, while certain terpenes are specific for flower tissues (Reichelt et al., 2002; Chen, 2003; Debeaujon et al., 2003). The cost effective accumulation of metabolites in a tissue-dependent manner enables the plant to invest valuable resources economically in growth and development and as such contributes to increased fitness and competitive

ability (Brown et al., 2003; Kliebenstein et al., 2005).

In this respect, it is not surprising that leaves, serving general supportive functions from simple basic structures contain the most basal metabolic expression profile, while flowers accumulate the largest number of specific metabolites (Matsuda et al., 2010). Flowers, as the complex prime organs for reproductive success, serve more specialized functions to improve plant fitness, like attracting pollinators and securing anthesis from external influences. Even the different organs within flowers (e.g. sepal, petal, stamen, pistil and receptacle) show a distinctive secondary metabolism, highlighting the importance of the metabolic profile in supporting organ structure and function (Hanhineva et al., 2008). Apparently, the developmental stage of plants and the state of their tissues is reflected in the metabolic signature of its compartments. This is effectively demonstrated by the accumulation and degradation of metabolites following tissue specialization during fruit development and ripening (Moco et al., 2007; Moing et al., 2011). That said, metabolic composition does not always depend on de novo biosynthesis and catabolism but can also result from re-allocation of compounds. Young leaves, for example, have a very distinct glucosinolate profile compared to senescent, older leaves, presumably because of re-allocation to ensure protection of the inner rosette from feeding by herbivores (Brown et al., 2003).

On a deeper level, metabolic differences can also be observed between specific cell types and even organelles in isolated tissues and organs (Schad et al., 2005; Holscher and Schneider, 2007). Laser micro-dissection is an effective technique to in situ separate different cell types and organelles. Using laser micro-dissection, the accumulation of terpenoids was found to be differently regulated in cortical resin ducts and cambial zone tissue in white spruce (Abbott et al., 2010), while two phenolic compounds, effective in bark beetle defense, were found in stone cells of Norway spruce (Li et al., 2006). In Arabidopsis, laser micro-dissection revealed metabolic compositions of the cytosol, vacuole, mitochondria and plastid. Consistent with earlier findings, many secondary metabolites were predominantly localized in the vacuole (storage) and cytosol (synthesis), while primary metabolites, crucial for biochemical pathway regulation, were detected in all compartments (Krueger et al., 2011).

Overall, spatial regulation of metabolism appears to have a great influence on the metabolic profile of different plant species. Moreover, even within individuals, the metabolic constitution mirrors developmental stage and function of the plant's parts list. When describing a plant's metabolic status caution should therefore be taken with respect to different organs and cell types within the studied species.

Environmental regulation of metabolism

The sessile nature of plants obligates it to adapt to its natural local environment. Plants depend on the radiation from the sun and nutrients from the soil for the process of

photosynthesis and the consequent acquirement of energy and resources to continue growth and development. The slightest change in environmental circumstances requires the re-arrangement of photosynthetic assimilates and metabolites to re-establish the plant in its new environment. Cold acclimation, for example, leads to immediate modifications in carbon metabolism towards the protection of proteins and membranes from freezing damage (Hannah et al., 2006; Kaplan et al., 2007). Likewise, plants anticipate rapidly on changes in nutrient status and light quality. Phosphorus deficiency, for instance, reprograms carbohydrate metabolism towards the more efficient organic acids and amino acids to reduce phosphorus consumption (Huang et al., 2008) while changes in the R:FR light ratio, possibly due to overshadowing of neighboring plants, increase the content of soluble metabolites and cell wall constituents (Mazzella et al., 2008). The accumulation of plant metabolites, and especially secondary metabolites, is strongly affected by environmental variation in light, nutrients, temperature and biotic and abiotic stresses and has been extensively reviewed in recent years (Kliebenstein, 2004; Allwood et al., 2008; Guy et al., 2008; Sanchez et al., 2008; Amtmann and Armengaud, 2009; Bundy et al., 2009; Dicke et al., 2009). We will, therefore, emphasize the general patterns observed during metabolic studies and the environmental influence on the spatiotemporal regulation of metabolism.

Environmental cues can shift the delicate balance between primary and secondary metabolites or steer their formation in certain directions. For the biosynthesis of secondary metabolites, products of primary metabolism, such as amino acids, often function as their substrates or as co-factor or ligand in enzymatic reactions (Logemann et al., 2000; Broeckling et al., 2005). An increased activation of secondary metabolism due to environmental changes will thus almost certainly modify the primary metabolism. Upon nitrogen deprivation, for example, a reduction in transcript accumulation of major genes in photosynthesis and chlorophyll synthesis occurs while transcripts involved in secondary metabolism increase (Scheible et al., 2004). Likewise, potassium (K^+) deficiency leads to an increase in the levels of oxylipins and glucosinolates, possibly for the protection against herbivores that prey on the herbivore-attractive sugar and amino acid-rich, K^+-deficient leaves (Troufflard et al., 2010). Additionally, drought stress decreases both growth and carbon assimilation, while increasing secondary metabolism (Hale et al., 2005).

The most well-known examples shifting the balance between primary and secondary metabolism, however, can be observed with plant-herbivory and plant-pathogen interactions. Feeding by various herbivores can induce different volatiles, and leaf consumption leads to temporal expression profiles of different metabolites (Thaler et al., 2002; Kant et al., 2004). The interactions between plants and their pests consequently result in the biosynthesis of many different secondary compounds, such as glucosinolates, toxins and volatile compounds (Pichersky and Gershenzon, 2002; Wittstock and Gershenzon, 2002; Arany et al., 2007). Plant competition for light on the other hand, results in increased carbon allocation towards

the growing parts of the plant to enable rapid growth, and would thus favor an increase in primary metabolites (Kozuka et al., 2005). Interestingly, the metabolic changes during plant competition have only rarely been studied (Mazzella et al., 2008). Nevertheless, different environmental circumstances evoke different effects on the balance between primary and secondary metabolism. In nature, however, plants are constantly competing with neighbors for light and are simultaneously stressed by herbivores and pathogens, which leads to a trade-off in metabolic investment between growth and defense (Herms and Mattson, 1992). The evolutionary consequences of this trade-off are summarized in Box 1.

Notwithstanding the above, it should be noted that most changes in primary and secondary metabolism are restricted to certain tissues and are not a general response within the plant. For example, in response to feeding, the trichomes on leaves of *Tithonia diversifolia* become, in contrast to other cell types, very rich in sesquiterpene lactones which are repellent to patch larvae (Ambrosio et al., 2008). Moreover, secondary metabolites (mainly salicylic acid, phytoalexins and antioxidants) accumulate to much higher levels in infected than in healthy leaf tissue (Simon et al., 2010). Similar results were obtained from a study of the symbiotic relationship between ryegrass and *Neotyphodium lolii* with specific increases in peramine, mannitol and oligopeptides in infected cells (Cao et al., 2008). That tissue specific responses are not limited to biotic interactions is shown by a significant difference in carbon allocation between the upper and lower section of expanding internodes of sunflower under low R:FR concomitant with an increase in growth in the upper part (Mazzella et al., 2008).

The detection and identification of metabolites specifically expressed in plant tissues during exposure to different stresses holds great promises for their beneficial use in human nutrition and health (Dixon and Sumner, 2003; Hall et al., 2008). A number of studies have reported on the qualitative and/or quantitative increase in nutritional compounds upon perturbation. Health promoting glucosinolates effectively arrest herbivore feeding and have accumulated to large amounts and chemical diversity in specific species due to the constant evolutionary arms race between insects and plants (Kliebenstein, 2004). Furthermore, isoflavonoids accumulate in response to microbial or insect compounds and have anti-carcinogenic and antioxidant activities (Dixon and Sumner, 2003). The production of metabolic compounds can be actively controlled by invoking a desired response after application of a certain treatment. This is demonstrated by growth at high temperature which significantly decreases the isoflavanoid content in soybean seeds (Tsukamoto et al., 1995). In contrast, exposure of St John's wort to high temperature increases, among other secondary metabolites, the hyperforin concentration in leaves, which was shown to have a calming effect on depressions (Zobayed et al., 2005). The use of untargeted metabolomics and the subsequent identification of molecules in environmental metabolomics are, therefore, expected to increase the number of beneficial metabolites for human nutrition and health as well as a wide range of other purposes.

BOX. 1

Trade-off between growth and defense

Plants invest most of their resources in growth related metabolism to maximize reproductive success and out-compete neighboring plants but at the same time have to maintain necessary defenses to ensure survival in presence of pathogens and herbivores. The trade-off between these two mutually excluding processes is known as the growth differentiation balance hypothesis (Herms and Mattson, 1992). When plants invest in secondary metabolism and defense, those resources are diverted from the production of vegetative tissues, and the expected growth is decreased. This hypothesis is supported by the universal down-regulation of photosynthesis-related genes upon biotic attack from different agents (Bilgin et al., 2010). Moreover, when plants are challenged by competition with neighbors and herbivores simultaneously they invest more resources towards fast growth (and competition) than to disease prevention (Izaguirre et al., 2006; Ballare, 2009). When plants sense a shift in the R:FR ratio of the light, due to competition with neighbors, a signal-transduction pathway is activated which enhances the shade-avoidance syndrome (Smith and Whitelam, 1997) resulting in increased growth of petioles and internodes. At the same time, the inactivation of phytochromes upon low R:FR signals causes a strong reduction in sensitivity for jasmonate, an important hormone in response to disease defense (Moreno et al., 2009). Plants that display the shade avoidance syndrome are less resistant to insects, have fewer trichomes and a higher C:N ratio indicating fast growth (McGuire and Agrawal, 2005).

Notwithstanding the afore mentioned, nutrient and water deprivation have a larger impact on growth than on the process of photosynthesis and in some environments more resources can thus be invested in differentiation, without a direct cost on growth (Stamp, 2004). Another strategy to ensure survival at lower costs is to make use of induced defenses upon biotic interaction (Baldwin, 1998; Cipollini, 2004). Indeed, fast growers show lower constitutive defense, and higher induced defense (Van Zandt, 2007). Moreover, trade-offs can be observed between plant competition, growth rate and constitutive defense reactions (Kempel et al., 2011). Better competitors, but not faster growers, show lower constitutive and higher induced defense responses. Especially plants from nutrient-limiting or stressful environments, grow slower, invest more in constitutive defense, and are less attractive to herbivores (Endara and Coley, 2011). Additionally, priming of induced defenses has a relatively low cost on plant growth rate compared to induced direct defense, but significantly increases disease resistance, such that priming maximizes fitness in disease prone areas (van Hulten et al., 2006). These studies show that fast growth does not necessarily increase disease risks but that plants have evolved different strategies to cope with multiple stresses, which most likely contributed to species diversity.

Another interesting observation is that the timing of the environmental cue affects the metabolic response of the plant both qualitatively and quantitatively. The impact of many environmental signals, from cold to shade avoidance appears to be gated by the circadian clock (Salter et al., 2003; Fowler et al., 2005). Benzene, isoprene (e.g. terpene) and fatty acid derived volatile emissions are all controlled by the circadian clock and their accumulation pattern is dependent on environmental signals (Loughrin et al., 1994; Loivamaki et al., 2007; Roeder et al., 2007; Arimura et al., 2008). Herbivore-injured cotton plants emit terpenoids signals to attract enemies of the herbivorous insect. Interestingly, cotton plants have a varying volatile blend during the day with some compounds following a circadian rhythm and possibly herbivore activity, while others being continuously released during the day, independent of the timing of herbivore feeding (Loughrin et al., 1994). Remarkably, the temporally varying volatile blends of plants also affects the hiding behavior of nocturnal larvae (Shiojiri et al., 2006). Exposure of caterpillars in the light to the nocturnal volatile blend of corn plants resulted in larvae behavior as under dark conditions. Similarly, benzenoids, monoterpenes and sesquiterpenes are emitted in the night to attract nocturnal pollinators (Roeder et al., 2007), while isoprenes accumulate during the day to prevent thermal and/or oxidative stress (Loivamaki et al., 2007). Additionally, methyl benzoate accumulation in petal tissue of snapdragon is regulated by the circadian clock to ensure high emission in the presence of bumblebees at day time, while in tobacco and petunia, which are pollinated by nocturnal insects, accumulation peaks during the night (Kolosova et al., 2001). This nicely shows the interaction between the environment, the temporal and the spatial accumulation of metabolites in different plant species.

Genetic regulation of metabolism

An often-overlooked aspect of plant metabolism is genetic diversity in biosynthesis and the regulation thereof. As mentioned, the diversity in plant metabolites is immense. This is true for individual plants, but even more so for the whole plant kingdom. Each of the estimated 400.000 plant species is predicted to have at least one or two unique metabolites in addition to the more common compounds, rendering the metabolic diversity in the plant kingdom inherently huge (Saito and Matsuda, 2010). Because the biosynthesis and accumulation of metabolites has a genetic origin, it is not surprising that quantitative and qualitative variation in these metabolites can be observed between but also within species (Keurentjes, 2009). This natural genetic diversity represents the phenotypic output of historical adaptations in plant molecular physiology and behavior towards environmental constraints (Rockman, 2008). Since plant metabolism is heavily entwined with plant performance, unraveling the genetic basis of metabolism will provide a better understanding of the biochemical nature and regulation of the metabolome, and eventually plant development, adaptation and growth.

Although some successes to elucidate and modify biochemical pathways for metabolic

engineering of beneficial properties have been reported, many attempts were hampered by a lack of knowledge on the control of biochemical pathways (Trethewey, 2004). Following these efforts, experimental genetic mapping populations are used for a targeted analysis of particular metabolic traits, such as glucosinolates and soluble oligosaccharides in Arabidopsis seeds and flavonoid biosynthesis in Poplar (Bentsink et al., 2000; Kliebenstein et al., 2001; Morreel et al., 2006). The use of natural variation within species in such targeted analyzes enables the identification of candidate genes for rate-limiting enzymatic steps in biochemical pathways or the genetic regulation thereof. Nonetheless, the majority of the metabolome is still unexplored and many metabolites still need to be identified, let alone their function in plant performance. Because many of the relationships between metabolites and plant development have yet to be established and hence the uncertainty of which compounds to target, untargeted approaches have become in vogue (Keurentjes et al., 2006; Schauer et al., 2006). The objective of the untargeted approach is to find genetic factors with regulatory functions in biochemical pathways without a priori knowledge of the metabolic composition of the sampled population. These genetical metabolomics studies can be combined with transcriptomics and proteomics to investigate the intermediate steps in the relationship between genotype and phenotype (Wentzell et al., 2007; Keurentjes et al., 2008; Fu et al., 2009). As exemplified in the previous sections, a huge variation exists in metabolic profiles at different developmental stages, tissues or cell types, and conditions and natural variation in genotypic interactions with each of these factors has been observed (Kliebenstein et al., 2001; Wentzell et al., 2008; Wentzell and Kliebenstein, 2008; Eduardo et al., 2010; Schilmiller et al., 2010).

As an extension of analysing natural variation in experimental populations other approaches such as genome-wide association (GWA) mapping are currently being explored (Keurentjes et al., 2011). GWA mapping is rapidly developing as a new tool in plant science to associate phenotypic characteristics with genetic markers to elucidate chromosomal regions explaining variation in complex traits (Atwell et al., 2010). It makes use of the long evolutionary history and the wide variation of species and its accumulated recombination events within natural populations to an extent that enables mapping at the gene level (For a review see Bergelson and Roux, 2010). To date, only a very limited number of studies applied GWA mapping on metabolites. In Arabidopsis, a targeted analyzis of 43 glucosinolates in 96 wild accessions revealed a number of significant associations (Chan et al., 2010a), including two major loci previously identified in experimental populations (Kliebenstein et al., 2001; Keurentjes et al., 2006). Illustratively, not all previously identified causal genes were detected, demonstrating the complementary value of the two approaches as was also observed for other traits and species (Buckler et al., 2009; Brachi et al., 2010). In an untargeted approach using the same GWA population in different environments the most profound findings were the validation of the complex, quantitative nature of the metabolome and the strong environmental component in its genetic regulation (Chan et al., 2010b). Although many metabolites

correlated over the two environments, only one was significantly associated to the same genomic region in both environments, suggesting a very strong environmental impact on the genetic regulation of metabolism. This is in line with previous findings and favors studies from different environments, or from field conditions with multiple signals perceived by the plant simultaneously, to provide a more robust understanding of genetical metabolomics (Bergelson and Roux, 2010). Strikingly, the phenotypic diversity in metabolites was smaller in the set of 96 natural accessions than in an experimental population derived from a cross between two accessions indicating that, due to the lack of selective forces, the control on metabolism is loosened in artificial populations (Chan et al., 2010b).

Finally we note on the natural epigenetic variation in species such as differential DNA methylation, chromatin structure and histone modification (Vaughn et al., 2007; Zhang et al., 2008a). As both epigenetic regulation and the metabolome are prone to strong environmental regulation, epigenetic regulation could prove to be a very important determinant of metabolic status (Sung and Amasino, 2004; Chan et al., 2010b). Epigenetic regulation can be effectively studied in experimental mapping populations exploiting epigenetic differences in population individuals (Johannes et al., 2009; Reinders et al., 2009), but thus far no such studies have been reported for metabolic analyzes.

The multi-dimensional regulation of metabolism in relation to plant development and performance

As the metabolome specifies the state of the plant in terms of development and performance, its predictive power can be used in breeding strategies. In a number of studies this lowdown has been used to determine the relationship between growth rate, biomass and metabolism in different natural variants (Cross et al., 2006; Schauer et al., 2006; Meyer et al., 2007; Lisec et al., 2008; Sulpice et al., 2010). Although individual metabolites occasionally show a weak correlation between accumulation and plant biomass (Sulpice et al., 2009), the predictive power of a group of metabolites is much higher (Meyer et al., 2007). Most of these metabolites are associated with central carbon metabolism and stress responses, both of great importance to the regulation of plant growth (See BOX 1). The majority of the metabolites in central carbon metabolism have a negative correlation with growth, implying that these metabolite pools are consumed at high growth rates. This is supported by the observation of positive correlations between enzyme activities in primary metabolism and rosette biomass, indicating that higher flux rates increase the formation of structural end products to enhance growth (Cross et al., 2006). Interestingly, high correlations between enzyme activities and biomass, do not always match with metabolite accumulation (Sulpice et al., 2010), illustrating the complexity of the regulation of plant metabolism. Additional evidence for the link between growth and metabolism comes from genetic analyses in which a strong co-regulation of biomass formation and metabolite accumulation was observed (Lisec et al., 2008).

As illustrated in the previous sections, plant metabolism is extremely complex and highly dynamic, and can only be fully understood by taking into account all the components that contribute to its complexity. The metabolome is highly variable on a spatiotemporal scale, strongly influenced by the environment and to a great extent determined by genetic constraints (Macel et al., 2010). The different impact of all these factors on the plant's metabolism is best exemplified by studies on glucosinolates (Wentzell and Kliebenstein, 2008; Burow et al., 2010). Glucosinolate accumulation depends on developmental stage, tissue type, planting density and genotype (Wentzell and Kliebenstein, 2008). Moreover, significant interactions between these different factors are abound. This suggests that the multi-dimensional regulation of glucosinolate accumulation modifies defense strategies to maximize the efficiency towards changing risks of herbivory (in time and space) and to increase the overall fitness of the plant (Wentzell and Kliebenstein, 2008; Burow et al., 2010). Since the investments in defense (mainly secondary metabolites) suggest a negative effect while investments in central carbon metabolism (primary metabolites) have a strong positive effect on plant growth and fitness, this argues for comparative studies of primary and secondary metabolism and their relation to plant growth rate and fitness in different environments. In this respect, differences might be expected between species with different growth strategies (see BOX1).

Acknowledgements

We acknowledge support from the Netherlands Organization for Scientific Research (JK) and the Centre for Biosystems Genomics (RK, JK). We thank W.F.H. Kooke for composing the figures.

References

Abbott E, Hall D, Hamberger B, Bohlmann J (2010) Laser microdissection of conifer stem tissues: isolation and analyzis of high quality RNA, terpene synthase enzyme activity and terpenoid metabolites from resin ducts and cambial zone tissue of white spruce (Picea glauca). BMC plant biology 10: 106.

Allen E, Moing A, Ebbels TM, Maucourt M, et al(2010) Correlation Network Analyzis reveals a sequential reorganization of metabolic and transcriptional states during germination and gene-metabolite relationships in developing seedlings of Arabidopsis. BMC systems biology 4

Allwood JW, Ellis DI, Goodacre R (2008) Metabolomic technologies and their application to the study of plants and plant-host interactions. Physiologia plantarum 132: 117-135.

Ambrosio SR, Oki Y, Heleno VC, Chaves JS, et al (2008) Constituents of glandular trichomes of Tithonia diversifolia: relationships to herbivory and antifeedant activity. Phytochemistry 69: 2052-2060.

Amtmann A, Armengaud P (2009) Effects of N, P, K and S on metabolism: new knowledge gained from multi-level analyzis. Current Opinion in Plant Biology 12: 275-283.

Araki T, Komeda Y (1993) Flowering in darkness in Arabidopsis thaliana. Plant Journal 4: 801-811.

Arany AM, Jong TJ, Kim HK, Dam NM, et al (2007) Glucosinolates and other metabolites in the leaves of Arabidopsis thaliana from natural populations and their effects on a generalist and a specialist herbivore. Chemoecology 18: 65-71.

Arimura G, Kopke S, Kunert M, Volpe V, et al (2008) Effects of feeding Spodoptera littoralis on lima bean leaves: IV. Diurnal and nocturnal damage differentially initiate plant volatile emission. Plant physiology 146: 965-973.

Arsenault PR, Vail D, Wobbe KK, Erickson K, et al (2010) Reproductive Development Modulates Gene Expression and Metabolite Levels with Possible Feedback Inhibition of Artemisinin in Artemisia annua. Plant physiology 154: 958-968.

Atwell S, Huang YS, Vilhjalmsson BJ, Willems G, et al (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature 465: 627-631.

Baldwin IT (1998) Jasmonate-induced responses are costly but benefit plants under attack in native populations. Proceedings of the National Academy of Sciences of the United States of America 95: 8113-8118.

Ballare CL (2009) Illuminated behaviour: phytochrome as a key regulator of light foraging and plant anti-herbivore defence. Plant, cell & environment 32: 713-725.

Bentsink L, Alonso-Blanco C, Vreugdenhil D, Tesnier K, et al (2000) Genetic analyzis of seed-soluble oligosaccharides in relation to seed storability of Arabidopsis. Plant physiology 124: 1595-1604.

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

Bilgin DD, Zavala JA, Zhu J, Clough SJ, et al (2010) Biotic stress globally downregulates photosynthesis genes. Plant, cell & environment 33: 1597-1613.

Brachi B, Faure N, Horton M, Flahauw E, et al (2010) Linkage and Association Mapping of Arabidopsis thaliana Flowering Time in Nature. PLoS genetics 6

Brenes-Arguedas T, Horton MW, Coley PD, Lokvam J, et al (2006) Contrasting mechanisms of secondary metabolite accumulation during leaf development in two tropical tree species with different leaf expansion strategies. Oecologia 149: 91-100.

Broeckling CD, Huhman DV, Farag MA, Smith JT, et al (2005) Metabolic profiling of Medicago truncatula cell cultures reveals the effects of biotic and abiotic elicitors on metabolism. Journal of experimental botany 56: 323-336.

Brown PD, Tokuhisa JG, Reichelt M, Gershenzon J (2003) Variation of glucosinolate accumulation among different organs and developmental stages of Arabidopsis thaliana. Phytochemistry 62: 471-481.

Brunner AM, Nilsson O (2004) Revisiting tree maturation and floral initiation in the poplar functional genomics era. New Phytologist 164: 43-51.

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, et al (2009) The Genetic Architecture of Maize Flowering Time. Science 325: 714-718.

Bundy JG, Davey MP, Viant MR (2009) Environmental metabolomics: a critical review and future perspectives. Metabolomics 5: 3-21.

Burow M, Halkier BA, Kliebenstein DJ (2010) Regulatory networks of glucosinolates shape Arabidopsis thaliana fitness. Current Opinion in Plant Biology 13: 348-353.

Cao M, Koulman A, Johnson LJ, Lane GA, et al (2008) Advanced data-mining strategies for the analyzis of direct-infusion ion trap mass spectrometry data from the association of perennial ryegrass with its endophytic fungus, Neotyphodium lolii. Plant physiology 146: 1501-1514.

Ceusters J, Borland AM, Ceusters N, Verdoodt V, et al (2010) Seasonal influences on carbohydrate metabolism in the CAM bromeliad Aechmea 'Maya': consequences for carbohydrate partitioning and growth. Annals of Botany 105: 301-309.

Chan EK, Rowe HC, Hansen BG, Kliebenstein DJ (2010a) The complex genetic architecture of the metabolome. PLoS genetics 6: e1001198.

Chan EK, Rowe HC, Kliebenstein DJ (2010b) Understanding the evolution of defense metabolites in Arabidopsis thaliana using genome-wide association mapping. Genetics 185: 991-1007.

Chen F (2003) Biosynthesis and Emission of Terpenoid Volatiles from Arabidopsis Flowers. The Plant Cell Online 15: 481-494.

Cipollini D (2004) Stretching the limits of plasticity: Can a plant defend against both competitors and herbivores? Ecology 85: 28-37.

Corbesier L, Lejeune P, Bernier G (1998) The role of carbohydrates in the induction of flowering in Arabidopsis thaliana: comparison between the wild type and a starchless mutant. Planta 206: 131-137.

Cosgrove DJ (2005) Growth of the plant cell wall. Nature reviews. Molecular cell biology 6: 850-861.

Cross JM, von Korff M, Altmann T, Bartzetko L, et al (2006) Variation of enzyme activities and metabolite levels in 24 Arabidopsis accessions growing in carbon-limited conditions. Plant physiology 142: 1574-1588.

D'Auria JC, Gershenzon J (2005) The secondary metabolism of Arabidopsis thaliana: growing like a weed. Current Opinion in Plant Biology 8: 308-316.

Dalchau N, Baek SJ, Briggs HM, Robertson FC, et al (2011) The circadian oscillator gene GIGANTEA mediates a long-term response of the Arabidopsis thaliana circadian clock to sucrose. Proceedings of the National Academy of Sciences of the United States of America 108: 5104-5109.

Debeaujon I, Nesi N, Perez P, Devic M, et al (2003) Proanthocyanidin-accumulating cells in Arabidopsis testa: regulation of differentiation and role in seed development. The Plant cell 15: 2514-2531.

Desbrosses GG, Kopka J, Udvardi MK (2005) Lotus japonicus metabolic profiling. Development of gas chromatography-mass spectrometry resources for the study of plant-microbe interactions. Plant physiology 137: 1302-1318.

Dicke M, van Loon JJ, Soler R (2009) Chemical complexity of volatiles from plants induced by multiple attack. Nature chemical biology 5: 317-324.

Dixon RA, Sumner LW (2003) Legume natural products: understanding and manipulating complex pathways for human and animal health. Plant physiology 131: 878-885.

Dodd AN, Salathia N, Hall A, Kevei E, et al (2005) Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. Science 309: 630-633.

Eduardo I, Chietera G, Bassi D, Rossini L, et al (2010) Identification of key odor volatile compounds in the essential oil of nine peach accessions. Journal of the science of food and agriculture 90: 1146-1154.

Endara M-J, Coley PD (2011) The resource availability hypothesis revisited: a meta-analyzis. Functional Ecology 25: 389-398.

Espinoza C, Degenkolbe T, Caldana C, Zuther E, et al (2010) Interaction with Diurnal and Circadian Regulation Results in Dynamic Metabolic and Transcriptional Changes during Cold Acclimation in Arabidopsis. Plos One 5

Fait A, Angelovici R, Less H, Ohad I, et al (2006) Arabidopsis seed development and germination is associated with temporally distinct metabolic switches. Plant physiology 142: 839-854.

Fernie AR (2007) The future of metabolic phytochemistry: larger numbers of metabolites, higher resolution, greater understanding. Phytochemistry 68: 2861-2880.

Fowler SG, Cook D, Thomashow MF (2005) Low temperature induction of Arabidopsis CBF1, 2, and 3 is gated by the circadian clock. Plant physiology 137: 961-968.

Fu J, Keurentjes JJB, Bouwmeester H, America T, et al (2009) System-wide molecular evidence for phenotypic buffering in Arabidopsis. Nature genetics 41: 166-167.

Giakountis A, Coupland G (2008) Phloem transport of flowering signals. Current Opinion in Plant Biology 11: 687-694.

Gibson SI (2005) Control of plant development and gene expression by sugar signaling. Current Opinion in Plant Biology 8: 93-102.

Godt D, Roitsch T (2006) The developmental and organ specific expression of sucrose cleaving enzymes in sugar beet suggests a transition between apoplasmic and symplasmic phloem unloading in the tap roots. Plant Physiology and Biochemistry 44: 656-665.

Gould PD, Locke JCW, Larue C, Southern MM, et al (2006) The molecular basis of temperature compensation in the Arabidopsis circadian clock. The Plant cell 18: 1177-1187.

Graf A, Schlereth A, Stitt M, Smith AM (2010) Circadian control of carbohydrate availability for growth in Arabidopsis plants at night. Proceedings of the National Academy of Sciences of the United States of America 107: 9458-9463.

Graf A, Smith AM (2011) Starch and the clock: the dark side of plant productivity. Trends in plant science 16: 169-175.

Guy C, Kaplan F, Kopka J, Selbig J, et al (2008) Metabolomics of temperature stress. Physiologia plantarum 132: 220-235.

Hale BK, Herms DA, Hansen RC, Clausen TP, et al (2005) Effects of drought stress and nutrient availability on dry matter allocation, phenolic glycosides, and rapid induced resistance of poplar to two lymantriid defoliators. Journal of chemical ecology 31: 2601-2620.

Hall RD, Brouwer ID, Fitzgerald MA (2008) Plant metabolomics and its potential application for human nutrition. Physiologia plantarum 132: 162-175.

Hanhineva K, Rogachev I, Kokko H, Mintz-Oron S, et al (2008) Non-targeted analyzis of spatial metabolite composition in strawberry (Fragariaxananassa) flowers. Phytochemistry 69: 2463-2481.

Hannah MA, Wiese D, Freund S, Fiehn O, et al (2006) Natural genetic variation of freezing tolerance in Arabidopsis. Plant physiology 142: 98-112.

Harmer SL, Hogenesch LB, Straume M, Chang HS, et al (2000) Orchestrated transcription of key pathways in Arabidopsis by the circadian clock. Science 290: 2110-2113.

Herms DA, Mattson WJ (1992) The dilemma of plants - to grow or defend. Quarterly Review of Biology 67: 283-335.

Hoffman DE, Jonsson P, Bylesjo M, Trygg J, et al (2010) Changes in diurnal patterns within the Populus transcriptome and metabolome in response to photoperiod variation. Plant Cell and Environment 33: 1298-1313.

Holscher D, Schneider B (2007) Laser microdissection and cryogenic nuclear magnetic resonance spectroscopy: an alliance for cell type-specific metabolite profiling. Planta 225: 763-770.

Howell KA, Narsai R, Carroll A, Ivanova A, et al (2009) Mapping Metabolic and Transcript Temporal Switches during Germination in Rice Highlights Specific Transcription Factors and the Role of RNA Instability in the Germination Process. Plant physiology 149: 961-980.

Huang CY, Roessner U, Eickmeier I, Genc Y, et al (2008) Metabolite profiling reveals distinct changes in carbon and nitrogen metabolism in phosphate-deficient barley plants (Hordeum vulgare L.). Plant & cell physiology 49: 691-703.

Izaguirre MM, Mazza CA, Biondini M, Baldwin IT, et al(2006) Remote sensing of future competitors: impacts on plant defenses. Proceedings of the National Academy of Sciences of the United States of America 103: 7170-7174.

Jeong ML, Jiang H, Chen HS, Tsai CJ, et al (2004) Metabolic profiling of the sink-to-source transition in developing leaves of quaking aspen. Plant physiology 136: 3364-3375.

Johannes F, Porcher E, Teixeira FK, Saliba-Colombani V, et al (2009) Assessing the Impact of Transgenerational Epigenetic Variation on Complex Traits. PLoS genetics 5

Kant MR, Ament K, Sabelis MW, Haring MA, et al (2004) Differential timing of spider mite-induced direct and indirect defenses in tomato plants. Plant physiology 135: 483-495.

Kaplan F, Kopka J, Sung DY, Zhao W, et al (2007) Transcript and metabolite profiling during cold acclimation of Arabidopsis reveals an intricate relationship of cold-regulated gene expression with modifications in metabolite content. The Plant journal : for cell and molecular biology 50: 967-981.

Kempel A, Schadler M, Chrobock T, Fischer M, et al (2011) Tradeoffs associated with constitutive and induced plant resistance against herbivory. Proceedings of the National Academy of Sciences of the United States of America 108: 5685-5689.

Kerwin RE, Jimenez-Gomez JM, Fulop D, Harmer SL, et al (2011) Network quantitative trait loci mapping of circadian clock outputs identifies metabolic pathway-to-clock linkages in Arabidopsis. The Plant cell 23: 471-485.

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

Keurentjes JJB, Fu J, de Vos CH, Lommen A, et al (2006) The genetics of plant metabolism. Nature genetics 38: 842-849.

Keurentjes JJB, Sulpice R, Gibon Y, Steinhauser MC, et al (2008) Integrative analyzes of genetic variation in enzyme activities of primary carbohydrate metabolism reveal distinct modes of regulation in Arabidopsis thaliana. Genome biology 9: R129.

Keurentjes JJB, Willems G, van Eeuwijk F, Nordborg M, et al (2011) A comparison of population types used for QTL mapping in Arabidopsis thaliana. Plant Genetic Resources 9: 185-188.

Kliebenstein DJ (2004) Secondary metabolites and plant/environment interactions: a view through Arabidopsis thaliana tinged glasses. Plant Cell and Environment 27: 675-684.

Kliebenstein DJ, Kroymann J, Brown P, Figuth A,et al (2001) Genetic control of natural variation in Arabidopsis glucosinolate accumulation. Plant physiology 126: 811-825.

Kliebenstein DJ, Kroymann J, Mitchell-Olds T (2005) The glucosinolate-myrosinase system in an ecological and evolutionary context. Current Opinion in Plant Biology 8: 264-271.

Knight H, Thomson AJW, McWatters HG (2008) Sensitive to freezing6 integrates cellular and environmental inputs to the plant circadian clock. Plant physiology 148: 293-303.

Kolosova N, Gorenstein N, Kish CM, Dudareva N (2001) Regulation of circadian methyl benzoate emission in diurnally and nocturnally emitting plants. The Plant cell 13: 2333-2347.

Kozuka T, Horiguchi G, Kim GT, Ohgishi M, et al (2005) The different growth responses of the Arabidopsis thaliana leaf blade and the petiole during shade avoidance are regulated by photoreceptors and sugar. Plant & cell physiology 46: 213-223.

Krueger S, Giavalisco P, Krall L, Steinhauser MC, et al (2011) A Topological Map of the Compartmentalized Arabidopsis thaliana Leaf Metabolome. Plos One 6

Lejeune P, Bernier G, Requier MC, Kinet JM (1993) Sucrose increase during floral induction in the phloem sap collected at the apical part of the shoot of the long-day plant Sinapis alba L. Planta 190: 71-74.

Li P, Sioson A, Mane SP, Ulanov A, et al (2006) Response diversity of Arabidopsis thaliana ecotypes in elevated [CO2] in the field. Plant molecular biology 62: 593-609.

Lisec J, Meyer RC, Steinfath M, Redestig H, et al (2008) Identification of metabolic and biomass QTL in Arabidopsis thaliana in a parallel analyzis of RIL and IL populations. The Plant journal : for cell and molecular biology 53: 960-972.

Logemann E, Tavernaro A, Schulz WG, Somssich IE, et al (2000) UV light selectively coinduces supply pathways from primary metabolism and flavonoid secondary product formation in parsley. Proceedings of the National Academy of Sciences of the United States of America 97: 1903-1907.

Loivamaki M, Louis S, Cinege G, Zimmer I, et al (2007) Circadian rhythms of isoprene biosynthesis in grey poplar leaves. Plant physiology 143: 540-551.

Loughrin JH, Manukian A, Heath RR, Turlings TCJ, et al (1994) Diurnal cycle of emission of induced volatile terpenoids herbivore-injured cotton plants. Proceedings of the National Academy of Sciences of the United States of America 91: 11836-11840.

Lu Y, Gehan JP, Sharkey TD (2005) Daylength and circadian effects on starch degradation and maltose metabolism. Plant physiology 138: 2280-2291.

Macel M, Van Dam NM, Keurentjes JJB (2010) Metabolomics: the chemistry between ecology and genetics. Molecular ecology resources 10: 583-593.

Malik MS, Riley MB, Norsworthy JK, Bridges W, Jr (2010) Glucosinolate profile variation of growth stages of wild radish (Raphanus raphanistrum). Journal of agricultural and food chemistry 58: 3309-3315.

Matsuda F, Hirai MY, Sasaki E, Akiyama K, et al (2010) AtMetExpress development: a phytochemical atlas of Arabidopsis development. Plant physiology 152: 566-578.

Matsuda F, Yonekura-Sakakibara K, Niida R, Kuromori T, et al (2009) MS/MS spectral tag-based annotation of non-targeted profile of plant secondary metabolites. The Plant journal : for cell and molecular biology 57: 555-577.

Mazzella MA, Zanor MI, Fernie AR, Casal JJ (2008) Metabolic responses to red/far-red ratio and ontogeny show poor correlation with the growth rate of sunflower stems. Journal of experimental botany 59: 2469-2477.

McGuire R, Agrawal AA (2005) Trade-offs between the shade-avoidance response and plant resistance to herbivores? Tests with mutant Cucumis sativus. Functional Ecology 19: 1025-1031.

McWatters HG, Devlin PF (2011) Timing in plants--a rhythmic arrangement. FEBS letters 585: 1474-1484.

Meyer RC, Steinfath M, Lisec J, Becher M, et al (2007) The metabolic signature related to high plant growth rate in Arabidopsis thaliana. Proceedings of the National Academy of Sciences of the United States of America 104: 4759-4764.

Michael TP, Salome PA, Yu HJ, Spencer TR, et al (2003) Enhanced fitness conferred by naturally occurring variation in the circadian clock. Science 302: 1049-1053.

Moco S, Capanoglu E, Tikunov Y, Bino RJ, et al (2007) Tissue specialization at the metabolite level is perceived during the development of tomato fruit. Journal of experimental botany 58: 4131-4146.

Moing A, Aharoni A, Biais B, Rogachev I, et al (2011) Extensive metabolic cross-talk in melon fruit revealed by spatial and developmental combinatorial metabolomics. The New phytologist 190: 683-696.

Moreno JE, Tao Y, Chory J, Ballare CL (2009) Ecological modulation of plant defense via phytochrome control of jasmonate sensitivity. Proceedings of the National Academy of Sciences of the United States of America 106: 4935-4940.

Morreel K, Goeminne G, Storme V, Sterck L, et al (2006) Genetical metabolomics of flavonoid biosynthesis in Populus: a case study. The Plant journal : for cell and molecular biology 47: 224-237.

Ohto M, Onai K, Furukawa Y, Aoki E, et al (2001) Effects of sugar on vegetative development and floral transition in arabidopsis. Plant physiology 127: 252-261.

Oksman-Caldentey KM, Inze D (2004) Plant cell factories in the post-genomic era: new ways to produce designer secondary metabolites. Trends in plant science 9: 433-440.

Pichersky E, Gershenzon J (2002) The formation and function of plant volatiles: perfumes for pollinator attraction and defense. Current Opinion in Plant Biology 5: 237-243.

Reichelt M, Brown PD, Schneider B, Oldham NJ, et al (2002) Benzoic acid glucosinolate esters and other glucosinolates from Arabidopsis thaliana. Phytochemistry 59: 663-671.

Reinders J, Wulff BBH, Mirouze M, Mari-Ordonez A, et al (2009) Compromised stability of DNA methylation and transposon immobilization in mosaic Arabidopsis epigenomes. Genes & Development 23: 939-950.

Richardson AD, Hollinger DY, Dail DB, Lee JT, et al (2009) Influence of spring phenology on seasonal and annual carbon balance in two contrasting New England forests. Tree physiology 29: 321-331.

Rockman MV (2008) Reverse engineering the genotype-phenotype map with natural genetic variation. Nature 456: 738-744.

Roeder S, Hartmann AM, Effmert U, Piechulla B (2007) Regulation of simultaneous synthesis of floral scent terpenoids by the 1,8-cineole synthase of Nicotiana suaveolens. Plant molecular biology 65: 107-124.

Roitsch T (1999) Source-sink regulation by sugar and stress. Current Opinion in Plant Biology 2: 198-206.

Roldan M, Gomez-Mena C, Ruiz-Garcia L, Salinas J, et al (1999) Sucrose availability on the aerial part of the plant promotes morphogenesis and flowering of Arabidopsis in the dark. Plant Journal 20: 581-590.

Saito K, Matsuda F (2010) Metabolomics for functional genomics, systems biology, and biotechnology. Annual review of plant biology 61: 463-489.

Salter MG, Franklin KA, Whitelam GC (2003) Gating of the rapid shade-avoidance response by the circadian clock in plants. Nature 426: 680-683.

Sanchez DH, Siahpoosh MR, Roessner U, Udvardi M, et al (2008) Plant metabolomics reveals conserved and divergent metabolic responses to salinity. Physiologia plantarum 132: 209-219.

Schad M, Mungur R, Fiehn O, Kehr J (2005) Metabolic profiling of laser microdissected vascular bundles of Arabidopsis thaliana. Plant methods 1: 2.

Schauer N, Semel Y, Roessner U, Gur A, et al (2006) Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nature biotechnology 24: 447-454.

Scheible WR, Morcuende R, Czechowski T, Fritz C, et al (2004) Genome-wide reprogramming of primary and secondary metabolism, protein synthesis, cellular growth processes, and the regulatory infrastructure of Arabidopsis in response to nitrogen. Plant physiology 136: 2483-2499.

Schilmiller A, Shi F, Kim J, Charbonneau AL, et al (2010) Mass spectrometry screening reveals widespread diversity in trichome specialized metabolites of tomato chromosomal substitution lines. The Plant journal : for cell and molecular biology 62: 391-403.

Sergeeva LI, Vonk J, Keurentjes JJB, van der Plas LH, et al (2004) Histochemical analyzis reveals organ-specific quantitative trait loci for enzyme activities in Arabidopsis. Plant physiology 134: 237-245.

Shiojiri K, Ozawa R, Takabayashi J (2006) Plant volatiles, rather than light, determine the nocturnal behavior of a caterpillar. Plos Biology 4: 1044-1047.

Simon C, Langlois-Meurinne M, Bellvert F, Garmier M, et al (2010) The differential spatial distribution of secondary metabolites in Arabidopsis leaves reacting hypersensitively to Pseudomonas syringae pv. tomato is dependent on the oxidative burst. Journal of experimental botany 61: 3355-3370.

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

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

Stamp N (2004) Can the growth-differentiation balance hypothesis be tested rigorously? Oikos 107: 439-448.

Sulpice R, Pyl ET, Ishihara H, Trenkamp S, et al (2009) Starch as a major integrator in the regulation of plant growth. Proceedings of the National Academy of Sciences of the United States of America 106: 10348-10353.

Sulpice R, Trenkamp S, Steinfath M, Usadel B, et al (2010) Network analyzis of enzyme activities and metabolite levels and their relationship to biomass in a large panel of Arabidopsis accessions. The Plant cell 22: 2872-2893.

Sung SB, Amasino RM (2004) Vernalisation and epigenetics: how plants remember winter. Current Opinion in Plant Biology 7: 4-10.

Thaler JS, Farag MA, Pare PW, Dicke M (2002) Jasmonate-deficient plants have reduced direct and indirect defences against herbivores. Ecology Letters 5: 764-774.

Trethewey RN (2004) Metabolite profiling as an aid to metabolic engineering in plants. Current Opinion in Plant Biology 7: 196-201.

Troufflard S, Mullen W, Larson TR, Graham IA, et al (2010) Potassium deficiency induces the biosynthesis of oxylipins and glucosinolates in Arabidopsis thaliana. BMC plant biology 10: 172.

Tsukamoto C, Shimada S, Igita K, Kudou S, et al (1995) Factors affecting isoflavone content in soybean seeds - changes in isoflavones, saponins, and composition of fatty-acids at different temperatures duringseed development. Journal of agricultural and food chemistry 43: 1184-1192.

van Dijken AJ, Schluepmann H, Smeekens SC (2004) Arabidopsis trehalose-6-phosphate synthase 1 is essential for normal vegetative growth and transition to flowering. Plant physiology 135: 969-977.

van Heerden PDR, Donaldson RA, Watt DA, Singels A (2010) Biomass accumulation in sugarcane: unravelling the factors underpinning reduced growth phenomena. Journal of experimental botany 61: 2877-2887.

van Hulten M, Pelser M, van Loon LC, Pieterse CM, et al (2006) Costs and benefits of priming for defense in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America 103: 5602-5607.

Van Zandt PA (2007) Plant defense, growth, and habitat: A comparative assessment of constitutive and induced resistance. Ecology 88: 1984-1993.

Vaughn MW, Tanurdzic M, Lippman Z, Jiang H, et al (2007) Epigenetic natural variation in Arabidopsis thaliana. Plos Biology 5: 1617-1629.

Velasco P, Soengas P, Vilar M, Cartea ME, et al (2008) Comparison of glucosinolate profiles in leaf and seed tissues of different Brassica napus crops. Journal of the American Society for Horticultural Science 133: 551-558.

Wentzell AM, Boeye I, Zhang ZY, Kliebenstein DJ (2008) Genetic Networks Controlling Structural Outcome of Glucosinolate Activation across Development. PLoS genetics 4

Wentzell AM, Kliebenstein DJ (2008) Genotype, age, tissue, and environment regulate the structural outcome of glucosinolate activation. Plant physiology 147: 415-428.

Wentzell AM, Rowe HC, Hansen BG, Ticconi C, et al (2007) Linking metabolic QTLs with network and cis-eQTLs controlling biosynthetic pathways. PLoS genetics 3: 1687-1701.

Wittstock U, Gershenzon J (2002) Constitutive plant toxins and their role in defense against herbivores and pathogens. Current Opinion in Plant Biology 5: 300-307.

Young MW, Kay SA (2001) Time zones: A comparative genetics of circadian clocks. Nature Reviews Genetics 2: 702-715.

Zhang JZ, Ai XY, Sun LM, Zhang DL, et al (2011) Transcriptome profile analyzis of flowering molecular processes of early flowering trifoliate orange mutant and the wild-type [Poncirus trifoliata (L.) Raf.] by massively parallel signature sequencing. BMC genomics 12: 63.

Zhang X, Shiu S, Cal A, Borevitz JO (2008) Global analyzis of genetic, epigenetic and transcriptional polymorphisms in Arabidopsis thaliana using whole genome tiling arrays. PLoS genetics 4

Zhou L, Jang JC, Jones TL, Sheen J (1998) Glucose and ethylene signal transduction crosstalk revealed by an Arabidopsis glucose-insensitive mutant. Proceedings of the National Academy of Sciences of the United States of America 95: 10294-10299.

Zobayed SMA, Afreen F, Kozai T (2005) Temperature stress can alter the photosynthetic efficiency and secondary metabolite concentrations in St. John's wort. Plant Physiology and Biochemistry 43: 977-984.

CHAPTER 4

GWA mapping reveals pleiotropic regulation of plant primary metabolism and biomass formation

Authors:

Rik Kooke[1,2,3], Corina Fusari[4], Willem Kruijer[5], Armin Schlereth[4], Ronan Sulpice[4,6], Dick Vreugdenhil[1,3], Mark Stitt[4] and Joost J.B. Keurentjes[2,3]

[1] Laboratory of Plant Physiology, Wageningen University, the Netherlands
[2] Laboratory of Genetics, Wageningen University, the Netherlands
[3] Centre for Biosystems Genomics, Wageningen Campus, the Netherlands
[4] Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany
[5] Biometris, Wageningen University, the Netherlands
[6] National University of Galway, Plant Systems Biology, Galway, Ireland

Abstract

Plant primary metabolism is a coordinated, complex and flexible network of processes that are regulated at multiple levels to integrate signals from different tissues, developmental cues and the environment. The genetic basis of this network can be explored by analysing the abundance of its individual components, such as chlorophyll, protein and amino acids, organic acids and non-structural carbohydrates and the enzymes that catalyse metabolic reactions leading to these products. Here, we report strong connectivity among different components together with moderate to high variation and heritability. Genome-wide association (GWA) mapping was applied to unravel the genetic architecture of each of the traits and to find genes involved in regulating specific and multiple pathways of primary metabolism. Among the candidate regulators, a number of structural genes for enzymes involved in metabolic conversions were identified. The metabolic analyses were integrated with plant biomass formation to identify pleiotropic regulators of enzyme activity, metabolite abundance and plant growth. Two genes involved in stress tolerance were identified with opposite effects on biomass formation and accumulation of primary metabolites suggesting that slow growing plants maintain high levels of non-structural carbohydrates and proteins as a resource that can rapidly be used in the case of stress. Another pleiotropic gene, *XXT2*, with opposite effects on primary metabolism and biomass formation was identified and is involved in xyloglucan formation and cell expansion, again indicating that slower growth is related to a higher or less efficient investment in primary metabolism.

Introduction

Primary carbon and nitrogen metabolism are directly associated with plant growth as they provide the essential building blocks for cell division, maintenance and expansion. Plant growth is determined by interconnected transcriptional and biochemical networks that are fine-tuned by the environment, the circadian clock and developmental cues. To disentangle this complex, polygenic regulation the individual components of central carbon and nitrogen metabolism, such as chlorophyll, protein and amino acids, organic acids and sugars, and the enzymes involved in this metabolic network need to be studied in concert. In contrast to transcriptional programs, protein content and enzyme activity are less susceptible to short term environmental changes and are the outcome of gene translation, post-translational modifications and protein turnover, and as such might be better predictors for plant growth (Gibon et al., 2004; Stitt and Gibon, 2013). Unravelling the regulation of primary metabolism will, therefore, further improve the understanding of the regulation of plant growth.

Natural variation within species has been widely explored to decipher the genotype-phenotype relationship in model and crop species. *Arabidopsis thaliana* has for decades been an outstanding model species for the study of natural variation due to its genetic adaptation to several natural habitats worldwide and its extensive variation in morphology, metabolism and growth (Alonso-Blanco et al., 2009). Moreover, there is abundant natural variation in Arabidopsis for the activities of enzymes and levels of primary metabolites (Causse et al., 1995; Mitchell-Olds and Pedersen, 1998; Sergeeva et al., 2004; Cross et al., 2006; Meyer et al., 2007; Keurentjes et al., 2008; Sulpice et al., 2009; Sulpice et al., 2010). These studies provided important insights into the interrelatedness of the primary metabolic network and growth. Starch turnover and total protein content are integrated components of the metabolic status and are strongly connected with biomass formation (Sulpice et al., 2009). A strong negative correlation between starch accumulation at the end of the day and biomass formation suggests that fast growing accessions use their carbohydrate pools more efficiently for growth (Cross et al., 2006; Sulpice et al., 2009). The significant negative correlation between total protein content and biomass formation is related to the investment of protein into enzymes, mainly Rubisco, for the photosynthetic machinery (Sulpice et al., 2010). Although there is no strong correlation between the levels of individual enzymes and biomass, the proportion of total protein invested in enzymes is significantly positively correlated with biomass (Sulpice et al., 2010).

Strong relationships have been detected among enzymes and among metabolites and between carbon and nitrogen metabolism, but weak connectivity was found between enzymes and metabolites (Cross et al., 2006; Sulpice et al., 2009; Sulpice et al., 2010). A lack of strong connectivity between metabolite content and enzyme activities suggests that coordinated enzyme activities mainly determine the pathway flux, leaving the intermediary metabolites levels unaltered and that maximum enzyme activities are not necessary to

enhance flux and metabolite levels (Kacser and Acerenza, 1993). It may also reflect the complex interaction between enzyme activity, metabolite levels and metabolic flux, in which the impact of an increase in enzyme activity on the level of a given metabolite depends on pathway topology and the thermodynamic structure of the pathways (Sulpice et al., 2010).

To analyse the genetic regulation of enzyme levels, metabolite abundance and biomass in more detail, natural variation can be used to identify QTLs that explain part of the variation. Moreover, it may reveal whether the different metabolic components are regulated pleiotropically and whether the activity of enzymes is regulated *in cis* or *in trans*. A number of studies have analysed recombinant inbred lines (RILs) to establish a link between primary metabolism and biomass. Biomass QTLs significantly associate with a higher number of metabolic QTLs than expected by chance, providing evidence for the close link between metabolism and growth (Lisec et al., 2008). Moreover, enzyme activity appears to be regulated at multiple loci, both *in cis* and *in trans*, and complex regulatory networks can be constructed between genes, enzymes and metabolites (Keurentjes et al., 2008). The identification of genes implicated in the regulatory networks has, however, been limited by a dearth of recombination in experimental populations and the strenuousness of fine-mapping. In addition, the variation in bi-parental populations is restricted to only two genotypes and, importantly, increased variation due to segregation of parental alleles in the RIL population provides little evidence for natural selection in an evolutionary context.

Genome-wide association studies (GWAS) profit from the long evolutionary history of Arabidopsis. Through both out-crossing and self-fertilization, this species combines a high allelic diversity with low levels of linkage disequilibrium (LD) over the entire genome. High-density genotyping and sequencing has resulted in a mapping resolution close to the gene, or even nucleotide level (Bergelson and Roux, 2010; Korte and Farlow, 2013). A number of recent studies employed GWAS to analyse both primary and secondary metabolite accumulation (Chan et al., 2010; Chan et al., 2010; Chan et al., 2011; Riedelsheimer et al., 2012; Verslues et al., 2014). These studies confirmed the involvement of many genes previously identified in reverse genetics or QTL studies but also resulted in the detection of many additional gene candidates involved in the regulation of metabolic variation.

In the present study, the levels of chlorophyll, protein, amino acids, organic acids, sugars and 17 enzymes of primary metabolism were analysed together with the formation of plant biomass in 350 natural accessions of Arabidopsis. The metabolites and enzymes showed strong connectivity within and between metabolic pathways, suggesting a highly coordinated regulation of primary metabolism that is strongly integrated with plant growth regulation. The main objective of this study is to identify novel genes involved in primary metabolism that influence plant growth, and to further elucidate the underlying genetic architecture of complex traits. GWA analysis on all individual traits confirmed the polygenic regulation

of primary metabolism and identified a number of variable structural genes and many pleiotropic regulators. The pleiotropic regulation was not confined to individual metabolic pathways, but, as suggested by the strong connectivity between pathways, was involved in the regulation of enzyme and metabolite levels and biomass. Many pleiotropic genes had opposite effects on biomass formation and primary metabolism and were involved in resistance to stress or cell expansion.

Results

Natural variation in primary metabolism

A global population of 350 natural accessions of *Arabidopsis thaliana* was grown under optimal conditions to determine the extent of variation in a large number of structural components, metabolites and enzymes involved in nitrogen and carbon metabolism (see supplemental table 1 for a full list of traits). All traits showed a normal distribution of trait values divided over multiple phenotypic classes (Figure 1). The level of variation in the population, given by the genetic coefficient of variation (CV_G) was rather small (< 20%) for total amino acids (AA) and the structural components chlorophyll A and B (ChlA, ChlB) and total protein (Prot), but larger for sugars (AVG = 31.9%), organic acids (AVG = 29.5%) and enzyme activities (AVG = 32.9%) (Table 1). This pattern is consistent with previous studies that estimated variation in the same traits in natural and experimental populations (Cross et al., 2006; Keurentjes et al., 2008; Sulpice et al., 2010; Sulpice et al., 2013). Although CV_G values are generally larger in the present study, probably because of the larger population size, all studies detected low variation for structural components and amino acids and a larger variation in sugars and organic acids, and especially enzyme activities.

To give a rough estimate of the variation that can be ascribed to genetic factors and to additive genetic factors, the marked-based narrow-sense heritability, h^2, was calculated (Table 1). The h^2 was extremely low for the enzymes glucose-6-phosphate 1-dehydrogenase (G6PDH) and glucokinase (GK) which illustrates that most of the detected variation for these enzymes is most likely due to biological variation. This is in strong contrast to previous studies which reported high heritabilities for these enzymes (Keurentjes et al., 2008; Sulpice et al., 2010). For the other enzymes, however, the h^2 averaged at 0.24 and compared well to previous analyses (Keurentjes et al., 2008; Sulpice et al., 2010). For biomass, structural components, sugars and organic acids, h^2 values ranged from 0.14 for glucose (Glu) to 0.77 for malate (Mal) with most traits displaying a h^2 higher than 0.4.

For the structural components and amino acids, the CV_G was rather small, which suggests that the level of these compounds are relatively well conserved and similarly regulated in different accessions. The h^2 for these traits, however, was moderate to high, indicating that the low level of variation that is present is most likely caused by genetic factors. Biomass,

organic acids, sugars and most enzymes were found to have moderate to high variation and heritability, suggesting that natural genetic variation is a major causal factor for phenotypic variation in these traits.

Table 1. Descriptive statistics.
The population average (AVG), standard deviation (SD), minimum (Min), maximum (Max), genetic coefficient of variation [CVG] and broad-sense heritabilities (H2) are shown for all traits. Trait abbreviations can be found in supplemental table 1.

Classification	Trait	Units	AVG	SD	Min	Max	[CV_G]	H^2	h^2
Biomass	FW	g.plant^{-1}	0.3	0.1	0.1	0.7	42.2	0.57	0.56
Structural components	ChlA	mg.g^{-1} FW	1.3	0.1	1.0	1.7	9.3	0.39	0.40
	ChlB	mg.g^{-1} FW	0.3	0.0	0.3	0.4	7.2	0.16	0.17
	Prot	mg.g^{-1} FW	11.8	1.5	7.7	16.3	12.4	0.34	0.31
AA	NO3	µmol.g^{-1} FW	123.8	13.7	81.2	161.9	11.0	0.46	0.46
	AA	µmol.g^{-1} FW	25.1	4.9	14.8	44.3	19.5	0.51	0.49
Sugars	Fru	µmol.g^{-1} FW	2.9	1.0	1.0	6.8	34.2	0.18	0.22
	Glu	µmol.g^{-1} FW	4.3	1.9	1.4	12.3	43.4	0.30	0.31
	G6P	µmol.g^{-1} FW	243.6	44.1	131.4	380.4	18.1	0.30	0.30
	Suc	µmol.g^{-1} FW	1.9	0.5	0.4	3.5	28.5	0.14	0.14
	Starch	µmol.g^{-1} FW	53.0	18.7	11.1	129.5	35.3	0.41	0.40
Organic acids	Fum	µmol.g^{-1} FW	13.6	4.5	2.8	24.6	33.2	0.67	0.67
	Mal	µmol.g^{-1} FW	9.3	2.4	4.7	18.0	25.9	0.76	0.77
Sucrose breakdown	FK	nmol.min^{-1}.g^{-1} FW	208.1	94.4	21.9	646.6	45.3	0.09	0.08
	GK	nmol.min^{-1}.g^{-1} FW	232.4	132.5	5.2	967.2	57.0	0.13	0.05
	Ainv	nmol.min^{-1}.g^{-1} FW	679.8	187.4	112.4	1520.1	27.6	0.56	0.56
	Ninv	nmol.min^{-1}.g^{-1} FW	169.7	52.9	52.0	482.5	31.2	0.27	0.30
Glycolysis and Respiration	G6PDH	nmol.min^{-1}.g^{-1} FW	588.1	297.1	65.3	2802.9	50.5	0.00	0.04
	FBP	nmol.min^{-1}.g^{-1} FW	178.5	83.4	2.9	572.5	46.7	0.08	0.10
	FUMA	nmol.min^{-1}.g^{-1} FW	3266.9	1337.1	506.1	7324.7	40.9	0.10	0.14
Calvin-Benson Cycle	iRUB	nmol.min^{-1}.g^{-1} FW	3955.4	960.5	1122.1	6471.0	24.3	0.13	0.18
	mRUB	nmol.min^{-1}.g^{-1} FW	4608.3	925.2	2506.7	7306.0	20.1	0.06	0.09
Sucrose synthesis	UGP	nmol.min^{-1}.g^{-1} FW	7128.4	1410.2	2894.2	12356.0	19.8	0.43	0.43
	PGM	nmol.min^{-1}.g^{-1} FW	7645.5	1724.8	1526.0	11403.9	22.6	0.11	0.15
	cPGI	nmol.min^{-1}.g^{-1} FW	692.1	172.3	162.9	1735.6	24.9	0.25	0.28
	tPGI	nmol.min^{-1}.g^{-1} FW	1037.5	229.5	157.7	1612.7	22.1	0.19	0.23
Starch synthesis	AGP	nmol.min^{-1}.g^{-1} FW	1354.0	290.2	329.7	2450.9	21.4	0.34	0.36
	pPGI	nmol.min^{-1}.g^{-1} FW	346.3	126.2	0.0	754.4	36.4	0.05	0.10
Nitrogen metabolism	NRVm	nmol.min^{-1}.g^{-1} FW	423.1	96.1	161.4	744.9	22.7	0.25	0.27
	NRVs	nmol.min^{-1}.g^{-1} FW	242.2	60.8	86.9	429.4	25.1	0.33	0.34

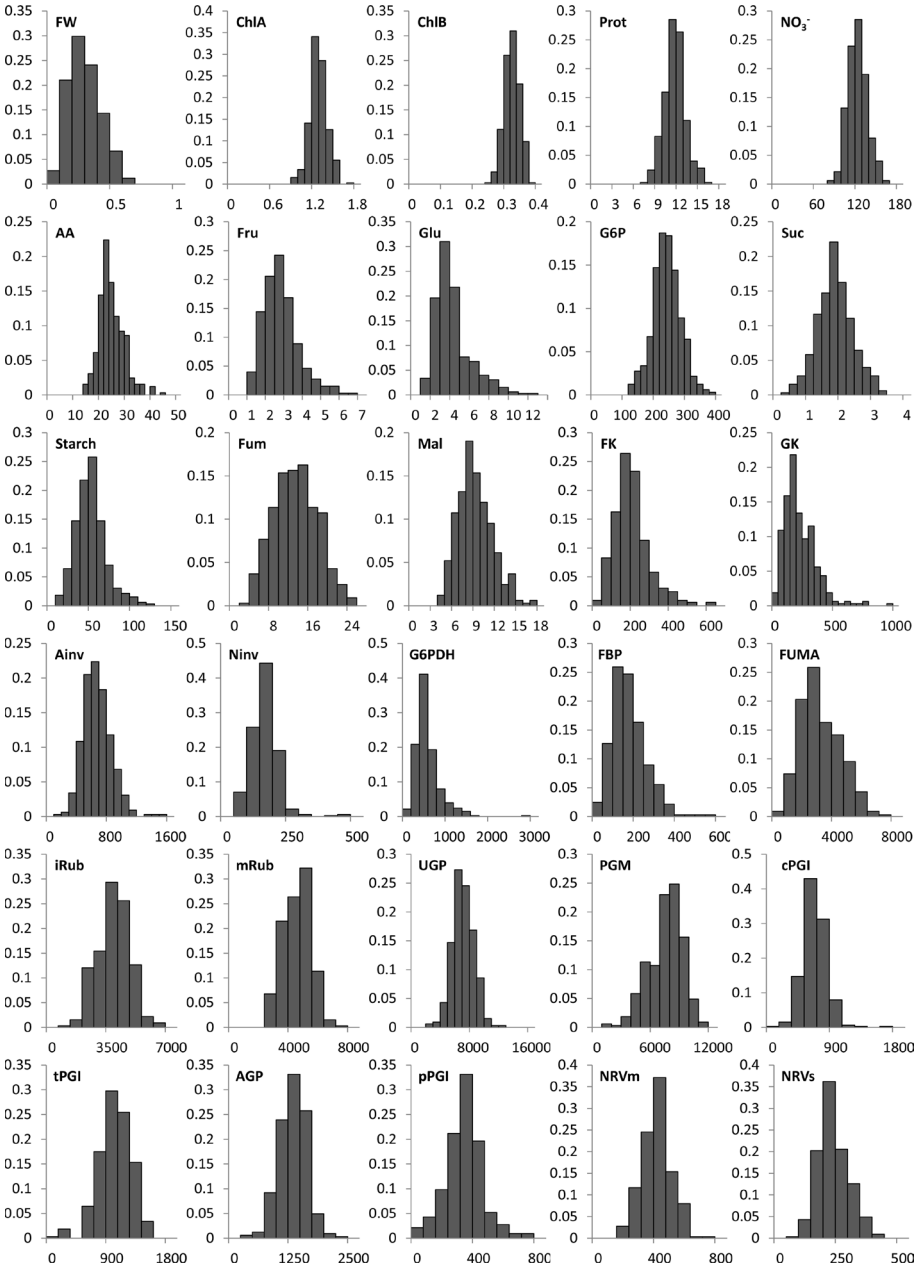


Figure 1. Frequency distribution of trait values for all lines.

The unit for the structural components (ChlA, ChlB, Prot) is $mg.g^{-1}$ FW. Amino acids (AA), NO3, sugars and organic acids are expressed in $\mu mol.g^{-1}$ FW. Enzyme activities are expressed in nmol.min-$1.g^{-1}$ FW. Abbreviations for all traits are listed in Supplemental table 1.

Primary metabolism is a highly coordinated network

To evaluate the connectivity between traits a Spearman rank correlation matrix was built. First, the relationship between biomass and the primary metabolic traits was investigated (Figure 2A). Biomass correlated strongly and negatively with all enzyme activities, which is in disagreement with a previous study in which only a few weak correlations between enzymes and biomass were detected (Sulpice et al. 2010). This may be partly caused by the far higher population size of the present study (350 vs. 129 accessions). The weakest correlation in our study was between biomass and fructose-1,6-biphosphate phosphatase (FBP) (r_s = -0.18, P < 0.05), while the strongest correlation was between biomass and nitrate reductase (NRVs) (r_s = -0.51, P < 0.0001) (Figure 2A). Especially, enzymes related to nitrogen metabolism, such as the nitrate reductase, and sucrose synthesis, such as UDP-glucose pyrophosphorylase (UGP), phosphoglucomutase (PGM) and phosphoglucose isomerase (PGI), correlated strongly with biomass (Figure 2A). Besides enzyme activities, total leaf protein, total amino acids, chlorophyll A, glucose-6-phosphate (G6P), starch, sucrose (Suc) and malate correlated negatively with biomass (P < 0.05). Several of these metabolites, and in particular starch, also correlated negatively with biomass in a smaller panel of 129 accessions (Sulpice et al., 2009). Nitrate (NO_3) and fumarate (Fum) (P < 0.05) correlated positively with biomass. The strong correlations between biomass, metabolites and enzymes provides further evidence that primary metabolism is tightly linked to plant growth. The correlations further suggests that the fast conversion of most primary metabolites enhances plant development in terms of biomass and that optimisation of these primary metabolic pathways may substantially enhance plant growth.

All correlations between the activities of different enzymes were positive and most were highly significant (P < 0.0001) (Figure 2D). This confirms, in a much larger population, the conclusion of Sulpice et al., 2010. Very strong correlations were observed among the enzymes involved in sucrose synthesis, the Calvin-Benson cycle, starch synthesis and nitrogen metabolism (0.32 < r < 0.85, P < 0.0001) suggesting a highly coordinated enzyme network connecting different metabolic pathways. Interestingly, these enzymes also correlated strongly with the major end products of the primary metabolic pathways: total protein, total amino acids, G6P, starch and sucrose (Figure 2C). The correlations among metabolites were

Figure 2. Spearman's rank correlations of trait values.
(A) Correlations (r) and corresponding adjusted *P*-values using Bonferroni correction between biomass and all analysed traits in primary metabolism. (B) Correlations between structural components, amino acids, sugars and organic acids (right top) and corresponding adjusted *P*-values using Bonferroni corrections (bottom left). (C) Correlations between structural components, amino acids, sugars and organic acids (horizontal) and enzyme activities (vertical). (D) Correlations between enzyme activities (right top) and corresponding adjusted *P*-values using Bonferroni corrections (bottom left). Trait abbreviations are listed in Supplemental table 1.

A

Classification	Trait	r	P-value
Structural components	ChlA	-0.19	0.02
	ChlB	0.11	1
	Prot	-0.63	0.00
AA	NO_3	0.24	0.00
	AA	-0.65	0.00
Sugars	Fru	-0.02	1
	Glu	0.05	1
	G6P	-0.45	0.00
	Starch	-0.49	0.00
	Suc	-0.35	0.00
Organic acids	Fum	0.30	0.00
	Mal	-0.24	0.00
Sucrose breakdown	FK	-0.31	0.00
	GK	-0.26	0.00
	Ainv	-0.30	0.00
	Ninv	-0.41	0.00
Glycolysis and Respiration	G6PDH	-0.32	0.00
	FBP	-0.18	0.04
	FUMA	-0.21	0.01
Calvin-Benson Cycle	iRub	-0.33	0.00
	mRub	-0.26	0.00
Sucrose synthesis	UGP	-0.50	0.00
	PGM	-0.39	0.00
	cPGI	-0.47	0.00
	tPGI	-0.50	0.00
Starch synthesis	AGP	-0.39	0.00
	pPGI	-0.35	0.00
Nitrogen metabolism	NRVm	-0.48	0.00
	NRVs	-0.51	0.00

B

Trait	ChlA	ChlB	Prot	NO_3	AA	Fru	Glu	G6P	Starch	Suc	Fum	Mal
ChlA	*	0.56	0.36	-0.20	0.25	0.31	0.41	0.11	0.28	0.25	0.30	0.31
ChlB	0.00	*	-0.01	-0.03	-0.05	0.06	0.19	0.00	-0.12	0.03	0.12	0.15
Prot	0.00	1	*	-0.27	0.53	0.18	0.04	0.41	0.37	0.38	-0.21	0.37
NO_3	0.02	1	0.00	*	-0.10	-0.32	-0.26	-0.34	-0.25	-0.09	-0.21	-0.37
AA	0.00	1	0.00	1	*	0.13	0.02	0.38	0.41	0.29	-0.22	0.41
Fru	0.00	1	0.07	0.00	1	*	0.40	0.18	0.24	0.10	0.16	0.22
Glu	0.00	0.03	1	0.00	1	0.00	*	-0.04	0.29	0.13	0.46	0.12
G6P	1	1	0.00	0.00	1	0.10	1	*	0.28	0.42	-0.21	0.46
Sta	0.00	1	0.00	0.00	0.00	0.00	0.00	0.00	*	0.34	0.19	0.25
Suc	0.00	1	0.00	1.00	0.00	1	1	0.00	0.00	*	-0.06	0.23
Fum	0.00	1	0.01	0.01	0.00	0.30	0.00	0.01	0.04	1	*	-0.09
Mal	0.00	0.35	0.00	0.00	0.00	0.00	1	0.00	0.00	0.00	1	*

C

Trait	ChlA	ChlB	Prot	NO_3	AA	Fru	Glu	G6P	Starch	Suc	Fum	Mal
FK	0.04	-0.06	0.29	-0.22	0.27	0.16	0.07	0.19	0.22	0.18	-0.04	0.17
GK	0.04	-0.07	0.27	-0.10	0.25	0.04	0.04	0.09	0.14	0.10	-0.08	0.15
A-INV	0.02	-0.07	0.24	-0.12	0.25	0.07	0.03	0.31	0.23	0.18	-0.11	0.19
N-INV	0.22	-0.01	0.41	-0.23	0.42	0.21	0.05	0.35	0.29	0.26	-0.13	0.37
G6PDH	0.14	-0.05	0.21	-0.20	0.33	0.27	0.02	0.33	0.21	0.22	-0.07	0.28
FBP	0.00	-0.05	0.09	-0.03	0.17	-0.02	-0.01	0.12	0.13	0.09	-0.08	0.07
FUMA	0.12	-0.04	0.27	-0.15	0.22	0.07	-0.03	0.23	0.10	0.22	0.04	0.09
iRub	0.25	-0.09	0.39	-0.16	0.31	0.16	0.07	0.42	0.33	0.29	0.06	0.24
mRub	0.22	-0.04	0.33	-0.15	0.19	0.15	0.08	0.43	0.32	0.32	0.14	0.16
UGP	0.25	-0.07	0.42	-0.15	0.45	0.17	0.10	0.37	0.41	0.31	0.02	0.22
PGM	0.21	-0.12	0.43	-0.22	0.33	0.25	0.05	0.57	0.32	0.43	0.01	0.21
cPGI	0.29	-0.06	0.46	-0.26	0.40	0.20	0.02	0.50	0.34	0.39	-0.01	0.29
tPGI	0.32	-0.05	0.49	-0.31	0.38	0.29	0.07	0.59	0.37	0.43	0.01	0.27
AGP	0.34	-0.07	0.40	-0.35	0.29	0.27	0.25	0.31	0.47	0.25	0.16	0.26
pPGI	0.27	0.01	0.34	-0.27	0.22	0.21	0.12	0.46	0.29	0.33	0.05	0.17
NRVm	0.14	-0.06	0.42	-0.13	0.36	0.01	-0.15	0.54	0.23	0.34	-0.12	0.28
NRVs	0.16	-0.07	0.42	-0.11	0.37	0.01	-0.14	0.47	0.28	0.33	-0.07	0.22

D

Trait	FK	GK	A-INV	N-INV	G6PDH	FBP	FUMA	iRub	mRub	UGP	PGM	cPGI	tPGI	AGP	pPGI	NRVm	NRVs
FK	*	0.28	0.08	0.15	0.29	0.10	0.14	0.14	0.15	0.31	0.22	0.25	0.28	0.27	0.23	0.13	0.11
GK	0.00	*	0.10	0.11	0.21	0.11	0.03	0.13	0.08	0.15	0.18	0.18	0.19	0.11	0.15	0.13	0.06
A-INV	1	1.00	*	0.53	0.17	0.07	0.11	0.30	0.21	0.28	0.28	0.26	0.29	0.15	0.23	0.22	0.23
N-INV	1	1	0.00	*	0.32	0.23	0.18	0.36	0.29	0.41	0.42	0.41	0.44	0.36	0.31	0.35	0.32
G6PDH	0.00	0.04	0.38	0.00	*	0.16	0.14	0.28	0.28	0.29	0.41	0.37	0.42	0.26	0.30	0.33	0.29
FBP	1	1	1	0.00	0.68	*	0.03	0.18	0.07	0.17	0.12	0.19	0.15	0.17	0.09	0.13	0.10
FUMA	1	1.00	1	0.12	1	1	*	0.25	0.27	0.28	0.37	0.32	0.38	0.22	0.25	0.38	0.40
iRub	1	1	0.00	0.00	0.00	0.20	0.00	*	0.68	0.42	0.66	0.58	0.63	0.45	0.43	0.53	0.52
mRub	0.91	1	0.02	0.00	0.00	1.00	0.00	0.00	*	0.37	0.70	0.53	0.61	0.41	0.48	0.58	0.55
UGP	0.00	1	0.00	0.00	0.00	0.38	0.00	0.00	0.00	*	0.54	0.54	0.57	0.50	0.36	0.41	0.44
PGM	0.01	0.21	0.00	0.00	0.00	1	0.00	0.00	0.00	0.00	*	0.73	0.83	0.51	0.62	0.63	0.54
cPGI	0.00	0.22	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.00	*	0.85	0.53	0.36	0.60	0.55
tPGI	0.00	0.11	0.00	0.00	0.00	0.91	0.00	0.00	0.00	0.00	0.00	0.00	*	0.58	0.74	0.65	0.58
AGP	0.00	1	1	0.00	0.00	0.37	0.01	0.00	0.00	0.00	0.00	0.00	0.00	*	0.43	0.34	0.32
pPGI	0.01	1	0.01	0.00	0.00	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	*	0.45	0.39
mV-NR	1	1	0.01	0.00	0.00	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	*	0.82
sV-NR	1	1	0.00	0.00	0.00	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	*

weaker than among enzymes, but nitrate strongly correlated negatively with most other metabolites (Figure 2B). Interestingly, nitrate also correlated negatively, and in most cases significantly, with enzyme activities (Figure 2C). This suggests either that fast conversion of nitrate results in the accumulation of protein, amino acids, sucrose, starch and G6P, or that rapid accumulation of protein, amino acids and sugars results in faster assimilation of nitrate. Although stronger correlations were detected between enzyme activities and metabolite accumulation, the emerging picture of a highly coordinated, integrated metabolic network shows very strong resemblance to previously reported results (Sulpice et al., 2010).

Genome-wide association analysis of primary metabolism

To identify the genes involved in primary carbohydrate metabolism that are causal for the observed variation, the individual trait values were subjected to genome-wide association (GWA) mapping of 215k SNP markers using a mixed model that accounts for population structure (Figure 3). All SNPs that were above the $-\log_{10}(P) = 3$ threshold were considered as possible candidate SNPs.

In general, associations were much weaker than those observed for gene-for-gene interactions in disease resistance or for other monogenic traits (Atwell et al., 2010; Baxter et al., 2010), suggesting polygenic regulation of primary metabolism. However, a strong association with UDP-glucose pyrophosphorylase activity (UGP) was identified on the top of chromosome three, accompanied by numerous weaker associations, indicating quantitative genetic regulation at multiple loci (Figure 3). These associations are in strong agreement with a previous QTL analysis in a RIL population and another GWA study on primary metabolism (Keurentjes et al., 2008; Chan et al., 2010).

The differences in h^2 and CV observed for the various traits were reflected in the GWA profiles. Traits with higher h^2 and CV_G values displayed on average a higher number of significant associations. Structural components, for example, showed low variation and the lowest number of significant associations, while glucose-6-phosphate 1-dehydrogenase matched extremely low h^2 values with a low number of significant associations (Supplemental table 2). Across all traits, h^2 and CV_G were positively correlated with the number of SNPs above $-\log_{10}(P) = 3$ and $-\log_{10}(P) = 4$ (Supplemental table 3).

Based on GWA significance, minor allele frequency, linkage disequilibrium (LD) and gene function, 46 genes were selected as candidate genes explaining part of the observed variation (Table 2).

Identification of structural genes for enzymes and metabolites

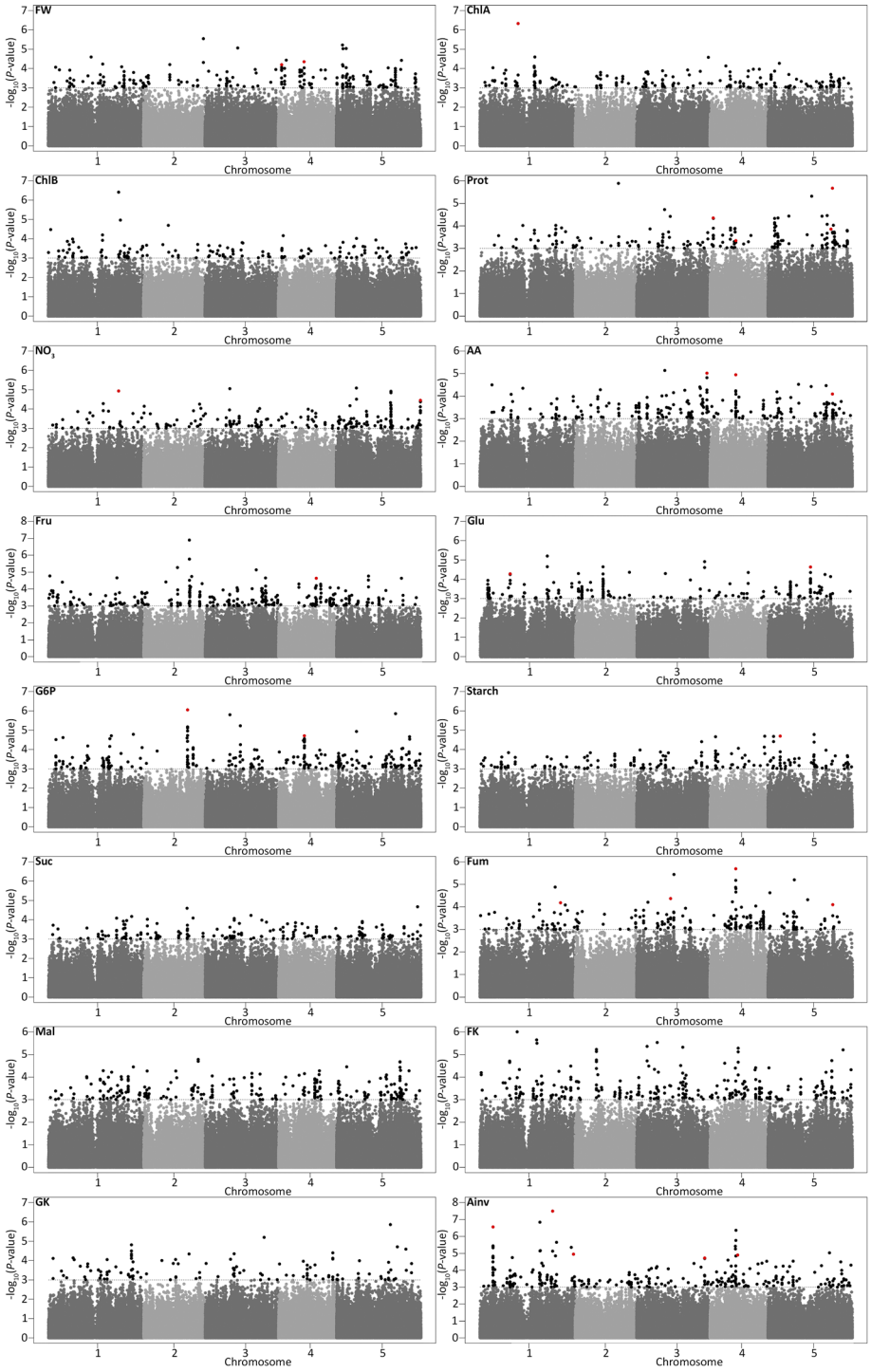
In a number of cases, strong evidence suggests that natural variation in structural genes contributes to variation in enzyme activity, although in general indirect regulation was

observed. Among the selected genes in our study, *UGP1* (AT3G03250) is located within the LD interval of the strongest association for UDP-glucose pyrophosphorylase activity, while *ATBETAFRUCT4* (AT1G12240) is located within the LD interval of an association with acid invertase (Ainv) activity on chromosome 1. Furthermore, *FUM1* (AT2G47510) and *FUM2* (AT5G50950) were selected as candidates explaining variation in fumarase (FUMA) activity and fumarate, respectively (Table 2). This is in good agreement with a QTL analysis in an experimental population in which a number of structural genes were suggested to underlie the detected QTLs (Keurentjes et al., 2008).

The structural gene for UDP-glucose pyrophosphorylase, *UGP1* (AT3G03250) was previously reported in a QTL analysis of a L*er* x Cvi RIL population and simultaneous expression QTL analysis indicated that *cis*-regulatory variation in the promoter regulates the variation in UDP-glucose pyrophosphorylase activity (Keurentjes et al., 2008). The most strongly associated SNP for UDP-glucose pyrophosphorylase activity in our GWA analysis is located in the promoter region, 1 kB upstream of the *UGP1* coding sequence (CDS). The recently released re-sequence data of Arabidopsis (http://1001genomes.org/) revealed no non-synonymous polymorphisms (nsSNP) within the exons of the *UGP1* gene, further indicating that *cis*-regulatory variation in the *UGP1* promoter is likely causal for variation in UDP-glucose pyrophosphorylase activity.

For acid invertase activity, the most strongly associated SNP on top of chromosome 1 was 1 kB upstream of the first exon of the structural gene *ATBETAFRUCT4* (AT1G12240). Although one nsSNP was in LD (r^2 = 0.36) with this associated SNP, possibly causing altered protein function, gene expression variation in earlier studies (Keurentjes et al., 2008) and the high association within the promoter region strongly suggest that cis regulatory variation in expression is also causal for the variation in acid invertase activity.

FUM2, which encodes the cytosolic fumarase (Pracharoenwattana et al., 2010) was previously selected as a candidate gene explaining differences in fumarate accumulation in an mQTL analysis of a RIL population between Col-0 and C24 (Lisec et al., 2008) and *cis*-regulatory expression variation was suggested to be the most likely cause for trait variation (Brotman et al., 2011). This is supported by the high amount of polymorphisms in the *FUM2* promoter and the observation that no nsSNPs are in strong LD with the most significant SNP for fumarate. In our study, the most significantly associated SNP for fumarase was detected 11 kB upstream of the *FUM1* gene, but was in LD (r^2 > 0.4) with two nsSNPs at the second and third amino acid of the first exon. Amino acids 2 to 8 in the first exon are predicted to form a beta sheet (http://ppopen.rostlab.org/) in secondary protein structure and the amino acid polymorphisms might, therefore, change the function of the protein. Although expression variation cannot be ruled out a priori, protein structure modification due to non-synonymous polymorphisms seems more likely causal for the natural variation in fumarase


FW
ChlA
ChlB
Prot
NO3
AA
Fru
Glu
G6P
Starch
Suc
Fum
Mal
FK
GK
Ainv
-log10(P-value)
Chromosome

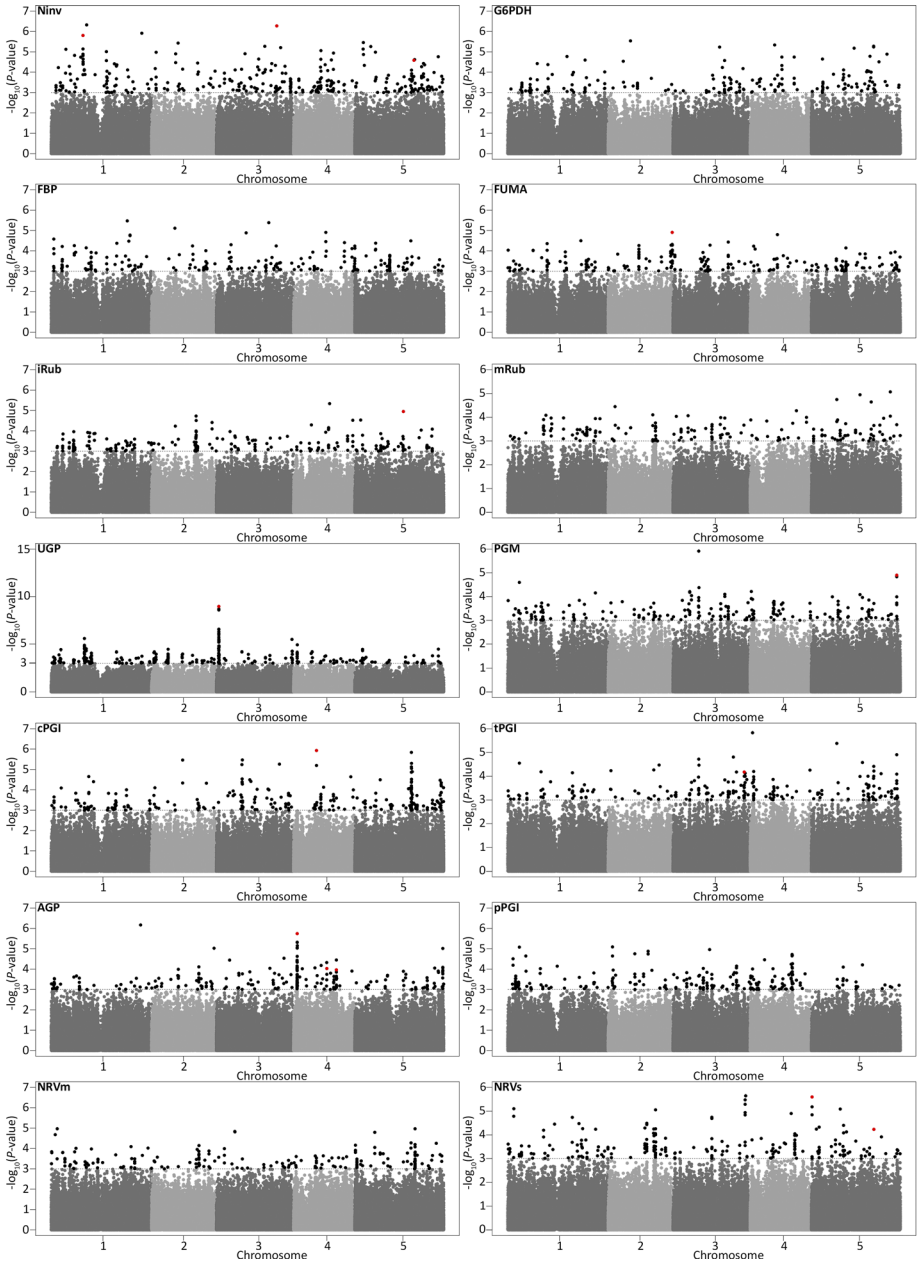


Figure 3. GWA profiles for all traits.

Manhattan plots for all traits with light and dark grey indicating different chromosomes. X-axis displays the chromosomes y-axis displays $-\log_{10}(P\text{-value})$,. SNPs with $-\log_{10}(P\text{-value}) > 3$ are indicated in black and SNPs that are listed in Table 2 (candidate gene list) are depicted in red. Trait abbreviations are listed in Supplemental table 1.

activity. Moreover, the observation that *FUM1* was not identified in any expression study further indicates regulation at the protein level.

Table 2. Candidate gene list.
Selected candidate genes for one or multiple traits. Abbreviations used: ID, alias name (TAIR); SNP, most significantly associated single nucleotide polymorphism (marker); Chr, chromosome number; Pos, position (TAIR10); trait, trait for which LOD score is provided; –log10(P-value); FDR, false discovery rate; MAF, minor allele frequency.

Trait	Gene	ID	Description	SNP	Chr	Pos	Trait	LOD	FDR	MAF
UGP	AT3G03250	UGP1	Putative UTP--glucose-1-phosphate uridylyltransferase 2	m81737	3	748476	UGP	8.96	0.00	0.33
Ainv	AT1G12240	ATBETA-FRUCT4	Beta-fructofuranosidase	m7011	1	4152530	Ainv	6.56	0.01	0.49
	AT1G62710	BETA-VPE	Vacuolar-processing enzyme beta-isozyme	m38695	1	23227304	Ainv	7.50	0.00	0.25
	AT1G79550	PGK	cytosolic phosphoglycerate kinase	m51104	1	29926013	Ainv	4.95	0.12	0.06
	AT3G58940		F-box/rni-like superfamily protein	m121114	3	21781567	Ainv	4.73	0.14	0.25
	AT4G15530	PPDK	pyruvate orthophosphate dikinase	m142924	4	8873185	Ainv	4.90	0.13	0.31
Ninv	AT5G44560	VPS2.2	Vacuolar protein sorting-associated protein 2-2	m196411	5	17949858	Ninv	4.59	0.24	0.45
FUMA	AT2G47510	FUM1	Fumarate hydratase 1 (mitochondrial)	m80197	2	19499948	FUMA	4.04	0.70	0.25
AGP	AT4G18240	SS4	starch synthase 4	m145997	4	10084281	AGP	4.03	0.57	0.4
tPGI	AT3G58530	-	F-box protein	m120898	3	21647418	tPGI	4.00	0.78	0.33
	AT3G58560	CCR4.1	Carbon catabolite repressor protein 4-like 1	m120905	3	21650806	tPGI	4.18	0.78	0.32
NO3	AT1G61100	-	disease resistance protein (TIR class)	m36904	1	22511771	NO3	4.93	0.57	0.25
	near to AT5G67420	LBD37	Encodes a LOB-domain protein involved in N metabolism and affecting leaf morphogenesis	m213929	5	26920288	NO3	4.45	0.76	0.10
	near to AT5G67500	VDAC2	Voltage-dependent anion channel	m213999	5	26944169	NO3	3.90	0.76	0.10
ChlA	AT1G33590	-	leucine-rich repeat (LRR) family protein	m21733	1	12179085	ChlA	6.34	0.09	0.35
Prot	AT5G49630	AAP6	Amino Acid Permease 6	m201708	5	20143984	Prot	3.86	0.84	0.23
AA	AT3G60860	DHDPS1	dihydropicolinate synthase involved in lysine biosynthesis	m122359	3	22487340	AA	5.02	0.85	0.2
Glu	AT1G27340	-	F-box only protein 6	m16354	1	9497278	Glu	4.28	0.70	0.11
Starch	AT5G12080	MSL10	mechanosensitive channel of small conductance-like 10	m167893	5	3907429	Starch	4.69	0.87	0.10

Trait	Gene	ID	Description	SNP	Chr	Pos	Trait	LOD	FDR	MAF
Fum	AT1G68600	-	Aluminum activated malate transporter family protein	m44349	1	25759849	Fum	4.18	0.91	0.23
	AT3G28860	ABCB19	ATP binding cassette B19	m100994	3	10877824	Fum	4.37	0.85	0.15
	AT5G50950	FUM2	Fumarate hydratase 2	m202682	5	20719861	Fum	4.10	0.91	0.30
NRVs NRVm	AT5G01540	LECR-KA4.1	Lectin-domain containing receptor kinase A4.1	m161250	5	217923	NRVs	5.60	0.22	0.35
	AT5G01550	LECR-KA4.2	Lectin-domain containing receptor kinase A4.2							
	AT5G01560	LECR-KA4.3	Lectin-domain containing receptor kinase A4.3							
Ninv Ainv	AT1G27720	TAF4B	TBP-associated factor 4B	m16610	1	9643865	Ninv	5.81	0.16	0.42
cPGI tPGI	AT4G11460	CRK30	putative cysteine-rich receptor-like protein kinase 30	m137560	4	6966308	cPGI	5.93	0.14	0.06
iRub mRub PGM cPGI tPGI NRVm	AT5G37260	CIR1	MYB fam transcription factor Circadian 1	m188153	5	14754311	iRub	4.95	0.90	0.27
mRub cPGI PGM tPGI NRVs	AT5G64813	LIP1	light insensitive period 1	m211982	5	25912160	PGM	4.84	0.91	0.49
Glu Fru	AT5G35360	CAC2	biotin carboxylase subunit	m185972	5	13586045	Glu	4.64	0.70	0.35
Fru Ninv Ainv	AT3G49430	SRp34a	Putative Pre-mRNA splicing factor SF2	m115042	3	18332175	Ninv	6.27	0.11	0.13
Fru Ninv cPGI	AT4G23060	IQD22	A negative regulator of GA response, plays a role in the regulatory network among the GA, calcium and auxin pathways	m150426	4	12086182	Fru	4.63	0.52	0.19
AGP Fum	AT4G25420	GA20OX1	Gibberellin 20 oxidase 1	m152026	4	12995479	AGP	3.97	0.59	0.20
G6P AA Ninv cPGI tPGI	AT2G33150	"KAT2 PKT3"	3-ketoacyl-coa thiolase 3 (peroxisomal)	m72459	2	14050281	G6P	6.05	0.10	0.13
AA FW Prot Starch G6P UGP	AT5G50720	HVA22E	HVA22-like protein e/related to ABA responses	m202544	5	20633709	Prot	5.67	0.43	0.18

Trait	Gene	ID	Description	SNP	Chr	Pos	Trait	LOD	FDR	MAF
FW Prot mRub NRVs NRVm PGI	AT5G46740	UBP21	ubiquitin-specific protease 21	m199255	5	18968187	NRVs	4.23	0.43	0.07
FW Prot AA G6P Mal Fum Ainv Ninv FK GK cPGI tPGI	AT4G14140	DMT2	DNA methyltransferase 2	m141079	4	8150011	Ainv	5.43	0.08	0.06
	AT4G14165	-	F-box family protein-like protein	m141133	4	8175186	Ainv	5.26	0.08	0.08
	AT4G14342	SF3B5	splicing factor 3B subunit 5	m141306	4	8254521	Fum	4.26	0.28	0.06
	AT4G14368	RCC1	regulator of chromosome condensation repeat-containing protein	m141380	4	8274507	Ainv	8.44	0.00	0.08
	AT4G14400	ACD6	ankyrin repeat-containing protein	m141496	4	8297892	Fum	4.66	0.52	0.12
	AT4G14420	-	HR-like lesion-inducing protein-like protein	m141532	4	8303999	Ainv	5.76	0.06	0.10
	AT4G14430	IBR10	indole-3-butyric acid response 10	m141541	4	8306682	Fum	4.82	0.59	0.09
	AT4G14440	HCD1	3-hydroxyacyl-CoA dehydratase 1							
	AT4G14530	-	hypothetical protein	m141623	4	8343754	Ainv	6.36	0.02	0.07
"FW / Prot AGP / AA cPGI/ tPGI PGM UGP"	AT4G02500	XXT2	xyloglucan 6-xylosyltransferase (XXT2)	m126422	4	1095217	AGP	5.75	0.16	0.49

Pleiotropic regulation of primary metabolism

Because of the high interconnectivity between the components of the primary metabolic network, further analysis aimed at identifying central pleiotropic regulators affecting enzyme activity, metabolite content and biomass. Among the list of 46 candidate genes for specific components, 24 genes were selected for multiple traits with sometimes opposite effects between traits (Table 2, Supplemental table 4). The large majority of these genes, 18 out of 24, were found for metabolic traits in multiple primary metabolic pathways. Sixteen candidate genes in eight LD intervals were selected explaining both variation in metabolite accumulation and enzyme activities and four of these eight LD intervals also contained SNPs highly associated with variation in biomass, suggesting extensive pleiotropic regulation (Supplemental table 4). One of the proposed pleiotropic genes is the abscisic acid (ABA) and stress-inducible *HVA22E* gene (Chen et al., 2002) that exerted opposite effects on biomass formation and protein content, G6P, starch and UGP activity (Supplemental table 4). Higher expression or functional diversity of *HVA22E* might induce abiotic stress

resistance concomitant with a decrease in biomass. The major pleiotropic QTL at the center of chromosome 4 was highly associated ($-\log_{10}(P) > 3$) with biomass, total leaf protein, total amino acids, G6P, malate, fumarate and seven enzymes (fructokinase, glucokinase, neutral inverstase, acid invertase, UDP-glucose pyrophosphorylase, cytosolic phosphoglucose isomerase and nitrate reductase), and moderately associated ($-\log_{10}(P) > 2$) with sucrose, phosphoglucomutase and total phosphoglucose isomerase. This QTL is most likely explained by the recently described and validated *ACCELERATED CELL DEATH 6* (*ACD6*) locus. Natural variation at this locus results in pleiotropic effects on pathogen and herbivore defense and vegetative growth and fitness (Todesco et al., 2010). It is conceivable that through its large effect on plant growth, the metabolic status of the plant is also altered. Resources for primary metabolism and growth might be redirected towards plant defense in accessions with different natural variants of this locus. As in our study biomass was previously reported to be significantly reduced in accessions with the non Col-0 haplotype (Todesco et al., 2010). Indicative of a selective sweep, LD around the strongest associated polymorphism extends as far as 192 kB, and numerous nsSNPs were detected within the *ACD6* locus and other genes within the LD interval (http://signal.salk.edu/atg1001/3.0/gebrowser.php). While it cannot be excluded that polymorphisms that are in LD with *ACD6* are responsible for the widespread changes, the simplest explanation is that the widespread changes are a pleiotropic response to variation in *ACD6*. The large amount of nsSNPs in the *ACD6* gene suggests that changes in the protein sequence, rather than changes in the promoter, explain the differences in trait values between the haplotypes. This is supported by the earlier study in which the coding sequence of Est-1 (non Col-0 haplotype) was fused to the Col-0 promoter. Subsequent transformation of the *acd6-2* mutant with this chimeric construct revealed an Est-1-like phenotype (Todesco et al., 2010). While biomass and fumarate content were significantly reduced in our study, all other traits displayed higher values in accessions of the non Col-0 haplotype (Supplemental table 4). This result complies with the correlation matrix across all traits and suggests that plants with low biomass invest more in enzyme activities and metabolites. This suggests that primary metabolites are maintained at high levels in plant cells as a carbon source that can be metabolized rapidly into secondary metabolites in the presence of pathogens (Bolton, 2009). Further studies investigating how secondary metabolism is affected in these plants and how infection by pathogens affects both primary and secondary metabolism are needed to address questions on the regulation and function of *ACD6* in plant metabolism, defense and growth.

A xyloglucan xylosyltransferase is involved in the regulation of plant biomass formation

Another major pleiotropic QTL explained observed variation in biomass formation, protein content, and ADP-glucose pyrophosphorylase (AGP), UGP and total PGI activity with a maximum $-\log_{10}(P)$ value of 5.75 for AGP activity (Table 2, Figure 4, Supplemental table 4). This QTL is located on the top of chromosome 4 in close proximity of the *XXT2* locus, encoding a xyloglucan xylosyltransferase that is important for xyloglucan biosynthesis (Figure

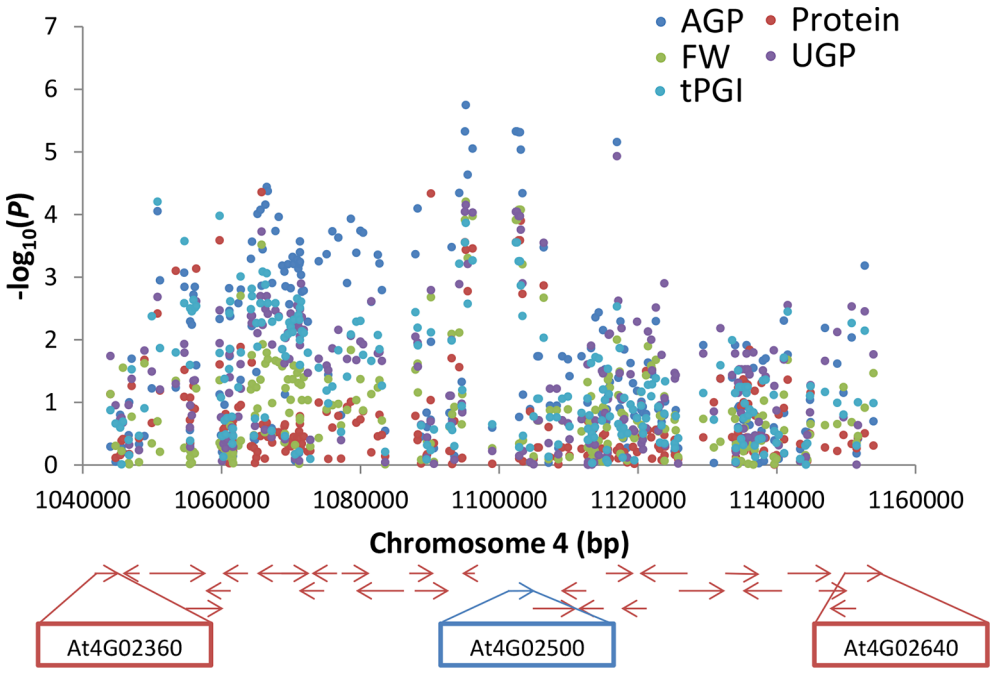


Figure 4. GWA SNPs around XXT2 locus at chr 4.
Enlarged association (-log10(P)) between SNPs at chr 4 (1044010-1153954 bp) and AGP (dark blue), FW (green), tPGI (light blue), Protein (red) and UGP (purple). The LD interval on chr 4 is based on LD information from resequenced accessions. A schematic representation of the genes in the chr 4 interval is represented below the x-axis. XXT2 (AT4G025000) is encircled with a blue box, the other two genes encircled with a red box indicate the genes on the boundary of the LD interval.

4) (Cavalier et al., 2008). Xyloglucans are embedded between the cellulose microfibrils to strengthen the cell wall, but they are also involved in loosening of the cell wall during expansion (Park and Cosgrove, 2012). Therefore, xyloglucan xylosyltransferases are thought to play a significant role in plant growth. Interestingly, biomass was significantly increased in the non Col-0 haplotype accessions while the enzyme activity levels and total protein content were significantly reduced (Supplemental table 4). A region of extensive LD (110 Kb; $r^2 > 0.3$) surrounding the *XXT2* locus contains 28 other genes and two transposable elements (Figure 4, Supplemental table 5). Knock-out studies of *XXT2* and its paralogue *XXT1* have previously confirmed their function in xyloglucan formation, and biomass formation was significantly reduced in the *xxt1/xxt2* double mutant (Cavalier et al., 2008). No nsSNPs could be identified in the coding sequence of the *XXT2* gene (http://signal.salk.edu/atg1001/3.0/gebrowser.php) suggesting that *cis*-regulatory variation in the promoter explains the difference between the haplotypes. Indeed, very strong LD ($r^2 > 0.8$) was observed between the most significantly associated SNP and three polymorphisms in the 1 kB upstream

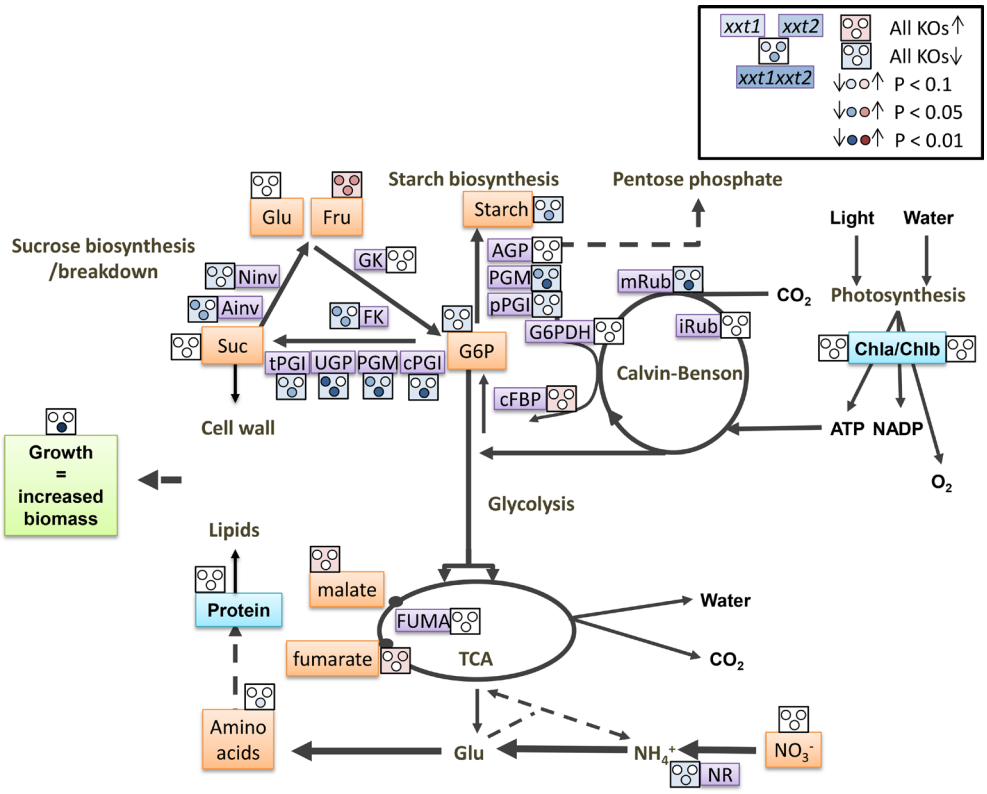


Figure 5. The effect of XXT knockout mutants on primary metabolism.
The orange boxes indicate amino acids, organic acids and sugars, the light blue boxes indicate the structural components, the green box indicates biomass and the purple boxes indicate the enzyme activities. The squared boxes with three circles indicate the effects of the single and double mutants on the attached trait. If all knockouts exert a similar effect on the trait compared to the WT, the box is colored light red (increased values compared to WT) or light blue (decreased values compared to WT). The left top circle in each squared box indicates whether the effect of the xxt1 mutant is significantly different from WT, the right top circle indicates the significance of the effect of the xxt2 mutant and the bottom circle indicates the significance of the effect of the xxt1xxt2 mutant. Darker red or blue indicates increased significance. Blue is decreased compared to WT, red is increased compared to WT. Trait abbreviations are listed in Supplemental table 1.

promoter region, possibly modifying transcription factor binding sites, and consequently gene expression. To determine whether natural selection acts on this locus, nucleotide diversity and Tajima's D statistic were assessed on all available resequenced accessions in Arabidopsis (http://1001genomes.org/). Nucleotide diversity for this locus is extremely low (π_T = 0.00835, π_{ns} = 0.00004 and π_s = 0.024) as is the ratio between non-synonymous and synonymous sites (π_{ns}/π_s = 0.002), which suggests that this locus is under purifying

selection. Moreover, Tajima's D statistic was significantly negative at non-synonymous sites (T_D = -2.05, P < 0.05) further suggesting purifying selection at the *XXT2* locus.

Because the detected QTL explained variation in biomass formation, protein content, and the activity of a number of enzymes in this study, the effect of single and double knockout mutants of *XXT1* and *XXT2* was studied on all the previously analysed metabolites, enzymes and biomass. A double knockout of *XXT1* and *XXT2* exerted the largest effect on all traits, followed by the single knockout of *XXT1* (Figure 5, Supplemental table 6 and 7). The activity of enzymes in sucrose synthesis was heavily reduced in all mutants, but most pronounced in the double mutant coinciding with a strong reduction in biomass. Most enzymes in sucrose breakdown and starch synthesis also displayed significantly lower activities in the double mutant and the *xxt1* single mutant. Furthermore, levels of total amino acids, maximal Rubisco and nitrate reductase activity were reduced, while fructose-1,6-biphosphate phosphatase activity, fructose, malate and fumarate content were increased in the mutants. Due to the knock-out of both xyloglucan xylosyltransferases, the activity of the majority of enzymes was coordinately down-regulated and this suggests lower investment in xyloglucan for the cell wall, and consequently reduced growth.

Discussion

Genetic regulation of primary metabolism

Genome-wide associating mapping has rapidly become a standard procedure in genetic studies but was so far not applied in a systematic analysis of enzyme activities, structural components and important metabolites of primary metabolism in relation to plant growth. The study reported here contributes to the substantial body of research that has investigated these relationships in natural and experimental populations leading to the identification of a number of (possible) regulator genes of central metabolism (Meyer et al., 2007; Keurentjes et al., 2008; Lisec et al., 2008; Sulpice et al., 2009; Sulpice et al., 2010; Brotman et al., 2011; Sulpice et al., 2013; Sulpice et al., 2014). The present study benefited tremendously from the rapid identification of candidate genes due to highly improved resolution in GWAS compared with QTL mapping studies (Bergelson and Roux, 2010). Moreover, the recent re-sequencing of hundreds of Arabidopsis accessions greatly increases the ability to locate candidate causal SNPs. It further enables the identification of pleiotropic regulators as different traits associate directly with unique SNPs instead of large QTL regions with possibly multiple different regulators. Finally, the associations are based on natural variation within a global population, established through mutation, recombination and adaptation providing strong evidence for natural selection and biological relevance.

The regulation of enzyme activity and metabolite accumulation appears to be complex and our results strongly suggest polygenic regulation of primary metabolism, in agreement

with previous studies (Keurentjes et al., 2008; Chan et al., 2010). This includes polygenic regulation of enzyme activities, which using the methods of this study, are a proxy for protein abundance. For most traits, unique and shared associations could be identified, suggesting specific and pleiotropic regulation of enzyme activity, metabolite accumulation and biomass, respectively. Compelling evidence indicates that natural variation in the activity of certain enzymes is partly regulated by *cis*-variation in structural genes, in strong agreement with previous studies (Keurentjes et al., 2008; Brotman et al., 2011). A strong association with UDP-glucose pyrophosphorylase activity was detected at the *UGP1* locus. Because no amino acid changing polymorphisms could be identified in the coding sequence, this suggests transcriptional regulation of *UGP1* activity, which is in agreement with previous UDP-glucose pyrophosphorylase activity QTL and *UGP* expression QTL (eQTL) analyses (Keurentjes et al., 2008). Moreover, a QTL for starch content was detected at the same position in the Bay-0 x Sha RIL population, for which *UGP1* was proposed to be the underlying regulator (Calenge et al., 2006). Nonetheless, we did not observe a high association with starch accumulation at the *UGP1* locus and Bay-0 and Sha share the same haplotype for the SNP most strongly associated with UDP-glucose pyrophosphorylase activity, suggesting that there must be a different regulator of starch content. In addition to *UGP1*, also *UGP2* encodes a structural gene of UDP-glucose pyrophosphorylase, but *UGP2* transcription is almost undetectable in mature leaves and very low in young leaves compared with *UGP1* (Meng et al., 2009). In addition, in two recent knock-out studies, the *ugp1* single mutant caused a much larger reduction in UDP-glucose pyrophosphorylase activity than in the *ugp2* mutant, 74% and 15%, respectively, whereas UDP-glucose pyrophosphorylase activity was further reduced in the *ugp1/ugp2* double mutant (Meng et al., 2009; Park et al., 2010). Moreover, although both single mutants did not show abnormal vegetative or reproductive development, the *ugp1/ugp2* double mutant showed reduced plant growth and male sterility, providing strong support for a redundant and important function of the *UGP* genes in plant growth and development (Park et al., 2010). The low contribution of *UGP2* in most accessions might have prevented the detection of significant associations at this locus. The redundancy in function could imply that *UGP2* plays a more important role in accessions with low expression of *UGP1*. This was indeed demonstrated in a QTL study in a RIL population in which a very strong *UGP1* QTL was accompanied by a weaker *UGP2* QTL with opposite effect, suggesting that *UGP2* is up-regulated if *UGP1* is lower or not expressed (Keurentjes et al., 2008).

Structural gene variation also associated with variation in enzyme activities of acid invertase and fumarase and the metabolite fumarate. Natural variation for acid invertase activity associated with the structural gene *ATBETAFRUCT4*, which was identified earlier in two QTL studies on the Ler x Cvi RIL population (Sergeeva et al., 2006; Keurentjes et al., 2008). In the first of the latter two studies, mutant analysis of the *ATBETAFRUCT4* gene confirmed a QTL with pleiotropic effects on vacuolar acid invertase activity and root length (Sergeeva et al., 2006). In the second study, the detection of a *cis*-eQTL suggested that variation in

the promoter region altering gene expressions was responsible for the variation in enzyme activity (Keurentjes et al., 2008). Interestingly, the structural gene *FUM2*, encoding the biosynthesis enzyme cytosolic fumarase, was associated with variation in the accumulation of the metabolite fumarate, but not with variation in activity of fumarase itself. This might be because the *FUM2* QTL involves subtle changes in substrate specificity or because the cytosolic enzyme FUM2 represents only a small part of the total fumarase activity in Arabidopsis (Pracharoenwattana et al., 2010). Indeed, a moderate non-significant but suggestive association could be detected for fumarase activity at the *FUM2* locus, coinciding with moderate suggestive associations with total amino acids, biomass, G6P, starch, malate and several enzymes (acid invertase, glucose-6-phosphate 1-dehydrogenase, UDP-glucose pyrophosphorylase, phosphoglucomutase and phosphoglucose isomerase). This is in agreement with a knock-out study of *FUM2*, in which fumarase activity and also amino acids, malate and starch were significantly altered in the *fum2* knock-out mutant (Pracharoenwattana et al., 2010). QTL analysis in the Col-0 x C24 RIL population also identified a QTL for fumarate at the *FUM2* locus (Lisec et al., 2008) and mutant and expression studies showed that transcriptional variation is most likely causal for the variation in fumarate content (Brotman et al., 2011). *FUM1* was not indicated previously in studies on natural variation as a candidate gene for the regulation of fumarase activity and attempts to isolate homozygous *FUM1* knock-outs have failed (Pracharoenwattana et al., 2010). Our study suggests, however, that natural variation in fumarase activity is partly regulated by *cis* polymorphisms in the coding sequence of *FUM1*.

Pleiotropic regulation of primary metabolism

After GWA analyzis of enzymes and metabolites, several examples of pleiotropic regulation within the same and different metabolic pathways were observed. Interestingly, natural variation in a cell wall-modifying enzyme, *XXT2* was found to have opposite effects on biomass and primary metabolism. Particularly, activities of enzymes involved in sucrose and starch synthesis were strongly altered in accessions of different haplotype at this locus. Xyloglucan xylosyltransferases (XXTs) are required for the transfer of xylose from UDP-xylose to xyloglucan oligosaccharides during the biosynthesis of xyloglucan (Faik et al., 2002). Xyloglucans are important for strengthening and loosening plant cell walls during cell expansion (Park and Cosgrove, 2012) and are therefore thought to play a significant role in plant growth (Cosgrove, 2005). Previous studies of the *XXT* gene family, with five functional members in Arabidopsis, have shown that xylosyltransferases are important for the formation of xyloglucan and that two members of this family, *XXT1* and *XXT2* have redundant functions (Cavalier et al., 2008; Vuttipongchaikij et al., 2012). The *xxt1* and *xxt2* single mutants exhibit a slight decrease in xyloglucan content, but the double mutant lacks detectable levels of xyloglucan and is significantly impaired in its mechanical properties (Cavalier et al., 2008).

Because of the redundancy between *XXT1* and *XXT2*, we pursued the GWA analysis with a knockout study on *xxt1*, *xxt2* and the *xxt1/xxt2* double mutant (all in the Col-0 background). The *xxt2* knockout contained significantly increased levels of fumarate and fructose, and reduced activity of PGM. Although a QTL for PGM activity was detected at the position of the *XXT2* locus in the GWA study, no QTLs were detected for fructose or fumarate. The *xxt1* mutant exhibited a phenotype which was more in line with the results from the GWA study, displaying reduced activity of many enzymes in sucrose and starch metabolism. The *xxt1/xxt2* double mutant, however, showed a much more pronounced phenotype with significant modifications in activity of most of the enzymes involved in sucrose and starch metabolism but also of nitrate reductase and Rubisco. Furthermore, starch and amino acid content were significantly reduced, while fructose increased, as was the case for the single mutants. Biomass was significantly reduced in the double mutant, confirming observations in a previous study (Cavalier et al., 2008).

Given the redundancy of the *XXT* genes, it is clear from both the GWA and the knockout study that there is a strong relationship between the two genes involved in xyloglucan formation, plant growth and sucrose and starch metabolism. These findings suggest that feedback regulation exists between the genetic regulators of non-structural carbohydrate metabolites, such as fructose, sucrose and starch, and structural carbohydrate metabolites and cell wall-synthesizing enzymes that are involved in cell wall physical properties and may influence cell expansion.

Integration of the primary metabolic pathways and plant growth

Unexpectedly, strong negative correlations between enzyme activities and plant biomass were observed. From previous studies on primary enzyme metabolism, it was anticipated that higher enzyme activities would better catalyse the metabolic conversions and hence result in increased plant growth (Sulpice et al., 2010). Large accessions were previously suggested to more rapidly metabolise the intermediate metabolic products of nitrogen and carbon metabolism into structural carbohydrates and protein to support growth (Poorter et al., 2013). In agreement with this, strong negative correlations between biomass and sucrose, starch, G6P, protein and amino acid contents were reported (Sulpice et al., 2009; Sulpice et al., 2010; Sulpice et al., 2013). The strong negative correlations between enzyme activities and biomass in the present study appear at first sight to differ from these earlier findings. They might be the result of the analysis of different genetic resources, but also increased resolution due to the larger population size. This would suggest that the relationship between plant metabolism and growth is highly plastic and context dependent (Sulpice et al., 2013). In particular, the negative correlation between biomass and protein content is much stronger (r_s = -0.63) than in previous studies (r_s = -0.31) (Sulpice et al., 2010).

The negative correlations between enzyme activities and biomass formation in the present

study might be a result of the very strong negative correlation between biomass and total leaf protein. Protein synthesis and turnover are energy-demanding processes that represent a major cost during plant growth (Hachiya et al., 2007), and reducing maintenance respiration may substantially increase plant biomass (Amthor, 1984). Maintaining low leaf protein levels has been suggested to allow plants to use their starch reserves more efficiently for growth (Stitt et al., 2010). Both leaf protein and amino acid content correlate strongly with most enzyme activities, and a shift of investment from other proteins to enzyme synthesis is positively correlated with biomass (Sulpice et al., 2010). When the enzyme activities were expressed relative to the protein level, the negative correlations with biomass were strongly reduced, lost or even became positive (Supplemental table 8). Most interestingly, the negative correlation with Rubisco, representing approximately 30 to 40% of total leaf protein (Farquhar et al., 2001), became positive, indicating that investment of a relatively larger portion of total leaf protein in Rubisco can enhance growth. This resembles the trend seen in earlier studies for photosynthesis proteins (Sulpice et al., 2010). Alternatively, slow growing plants might have over-invested resources in enzymes such that it does not increase pathway flux, and leads to a decrease in growth (Kacser and Acerenza, 1993) .

Higher enzyme activities might be maintained in smaller plants to sustain higher levels of sugars, protein and amino acids in fluctuating or adverse environments suggesting a trade-off between the rate of protein turnover in changing conditions and plant growth (Bolton, 2009) (Stitt et al., 2010). In this light, it is interesting to discuss the stress-related genes *ACD6* and *HVA22E*, which were both significantly associated to variation in biomass, total leaf protein, amino acids, G6P and the activity of enzymes in sucrose metabolism. *ACD6* is involved in resistance to a broad range of pathogens (Todesco et al., 2010), while *HVA22E* is involved in responses to abiotic stresses, such as cold and salinity (Chen et al., 2002). Investment in higher protein and sugar levels due to polymorphisms in these genes might lead to enhanced tolerance at the expense of rapid growth. Interestingly, from the accessions that have the highest protein levels and lowest biomass in our study, most are originating from harsh environments such as Scandinavia and Northern-Russia. This indicates that these accessions most likely experience stressful situations in which slow growth and increased defense might be preferred. Further studies are needed to test different haplotypes and knockout mutants of both genes for plant growth, defense, primary metabolism and fitness in different biotic and abiotic environments.

An alternative explanation for the negative correlation between plant biomass and enzyme activity might be much simpler, and related to the time of harvest. Larger plants might have reduced relative growth rates compared to smaller plants at the developmental stage they were harvested and, therefore, they might have down-regulated their central metabolism leading to reduced enzyme activities.

Overall, the results correspond well with previous extensive studies on primary metabolism in natural and experimental populations (Meyer et al., 2007; Keurentjes et al., 2008; Lisec et al., 2008; Sulpice et al., 2010). Our study confirmed the tight link between primary metabolism and plant growth (Meyer et al., 2007; Lisec et al., 2008; Sulpice et al., 2010) as both enzyme activities and metabolic content correlated strongly with plant biomass. Moreover, coordinated changes were observed in the enzyme activities with strong correlations within and between metabolic pathways, suggesting metabolic flux regulation through concerted regulation of the primary metabolic pathways instead of regulation through rate-limiting steps (Sweetlove et al., 2008; Stitt et al., 2010). This view is supported by genetic studies in which individual "rate-limiting" enzymes were found to contribute little to metabolic flux (Morandini, 2009), as well as more detailed studies of enzymes involved in photosynthesis showing that control is distributed between enzymes (reviewed in Stitt et al., 2010).

Material & Methods

Plant growth conditions

Seeds from 350 natural accessions of Arabidopsis thaliana, collected worldwide (Li et al., 2010; Horton et al., 2012), were sown on filter paper with demi water and stratified at 4˚C in dark conditions for 5 d. Seeds were then transferred to a culture room (16 h LD, 24˚C) to induce seed germination for 42 h. Six replicates per accession were transplanted to wet Rockwool blocks of 4 x 4 cm in a climate chamber (10 h SD, 125 µmol m^{-2} s^{-1}, 20°C day/18°C night, 70% RH). Three replicates were planted in consecutive order on the left table, the other three on the opposite end of the right table. All plants were watered daily for 5 min with 1/1000 Hyponex solution (Hyponex, Osaka, Japan). Plants were weighed (FW) and harvested in two replicate pools of three plants after 37 d within 2 h time at the end of the light period. The pooled samples were used as two replicate samples for metabolic and enzymatic assays.

Enzyme and metabolite assays

Chemicals were purchased as described by (Gibon et al., 2004). Total protein was assayed using the Bradford method (Bradford, 1976). Starch, Glu, Fru, Suc, and total amino acids were determined by enzymatic assays in ethanolic extracts of 20 mg frozen plant material as described by Cross et al., 2006 and Mal and Fum as described by Nunes-Nesi et al., 2007. Assays were performed in 96-well microplates using a Janus pipetting robot (Perkin-Elmer). Absorbances were determined using a Synergy, an ELX-800, or an ELX-808 microplate reader (Bio-Tek). For all the assays, two technical replicates were determined per biological replicate. Samples were randomized within and between plates and reference material was included in each plate as an internal control.

For enzyme measurements, 20 mg of powdered frozen material were extracted by mixing

with extraction buffer (Nunes-Nesi et al., 2007). AGP, SPS, fumarase, INV, GK, FK, FBP, G6PDH and NR were determined as described by Gibon et al., 2004, Rubisco as described by (Sulpice et al., 2007), cytosolic PGI and plastidic PGI as described by Weeden and Gottlieb, 1982, PGM was assayed as described by Manjunath et al., 1998. UGP was assayed as described Keurentjes et al., 2008. Enzyme activities were expressed on a fresh weight basis (nmol min-1 g-1 FW).

Descriptive statistics

Histograms of trait values were made using EXCEL. Spearman's rho and Pearson's r correlation coefficients were determined using SPSS 21 using a two-tailed significance test.
Coefficient of variation (CV_G) was calculated as $\sigma_G/X*100\%$.
The variance components for all the individual traits were used to calculate the broad-sense heritability, H^2, in analysis of variance (ANOVA) according to the formula

$$H^2 = \sigma^2_G/(\sigma^2_G + \sigma^2_E), \text{ with } \sigma^2_G = (MS(G) - MS(E))/r, \sigma^2_E = MS(E), \quad (1)$$

where r is the number of replicates and MS(G) and MS(E) are the mean sums of squares for genotype and residual error, respectively. Narrow-sense heritability, h^2, is defined as

$$h^2 = \sigma^2_A/(\sigma^2_G + \sigma^2_E), \quad (2)$$

which takes only the additive genetic effects (σ^2_A) in account. Marker-based estimates of narrow-sense heritability can be obtained using the mixed models (3) and (4) defined below, which contain random genetic effects. The covariances between these effects are modeled by a genetic relatedness matrix (GRM) estimated from markers, which is called kinship matrix K, with elements:

$$Ki,j = \frac{1}{p}\sum_{l=1}^{p}\frac{(x_{i,l} - f_l)(x_{j,l} - f_l)}{4f_l(1-f_l)} \qquad (i,j = 1,\ldots n)$$

where p = 214051 is the total number of markers. The numbers $x_{i,l}$ denote the minor allele count at marker *l* for genotype *i*, and f_l is the minor allele frequency at marker *l*.

A commonly used mixed model for quantitative traits is given by

$$y_i = \mu + G_i + E_i, \quad (i = 1,\ldots,n) \qquad G \sim N(0, \sigma^2_A K), E_i \sim N(0, \sigma^2_E), \quad (3)$$

where n = 350 is the total number of accessions, y_i is the mean phenotypic value of accession *i*, μ is the intercept and $G = (G_1,\ldots,G_n)$ is the vector of random effects, which follows a N(0,

$\sigma^2_A K$) distribution. The random error effects E_i follow independent normal distributions with variance σ^2_E.

Model (3) is widely used for marker-based estimation of (narrow-sense) heritability of human traits (Yang et al., 2010) which are usually measured on cohorts of thousands of individuals. However for plant traits phenotyped on only several hundreds of genotypes it has been shown recently that such estimates can be very imprecise, and that accuracy is greatly improved if phenotypic data of genetically identical replicates (rather than means) are included in the mixed model. We therefore considered the following extension of model (3):

$$y_{i,j} = \mu + G_i + E_{i,j}, \quad (i = 1,\ldots,n,\ j = 1,\ldots,r) \qquad G \sim N(0, \sigma^2_A K),\ E_{i,j} \sim N(0, \sigma^2_E). \qquad (4)$$

r = 3 is the number of replicates, *yi,j* is the phenotypic response of replicate *j* of genotype *i*, μ is the intercept, $G = (G_1,\ldots,G_n)$ is the vector of random genetic effects, and the errors $E_{i,j}$ have independent normal distributions with variance σ^2_E. σ^2_E Is the residual variance for a single individual. Estimates of σ^2_A and σ^2_E are obtained with the method of residual maximum likelihood (REML), and heritability is then estimated by $h^2_r = (\sigma^2_A / \sigma^2_A + \sigma^2_E)$ where σ^2_A and σ^2_E are based on all replicates. Note that in model (4), σ^2_E is the residual variance for a single individual, whereas in model (3), it is the residual variance of a genotypic mean. Since our interest is in individual plant level heritability and not line-heritability, σ^2_E in model (4) is indeed the variance parameter of interest (the use of model (3) would require multiplication of estimated residual variance by r). Both models (3) and (4) can only account for additive genetic effects; hence σ^2_E includes also non-additive genetic effects, and the denominator $\sigma^2_A + \sigma^2_E$ equals the total phenotypic variance.

Genome-wide association mapping

GWA mapping was performed on 328 accessions for FW and between 321 and 326 accessions for metabolites and enzymes, because for some accessions we missed genotype data and others were removed before harvest. All accessions were genotyped with 214,051 SNPs of which 199,589 were used for GWA mapping after removal of SNP with MAF < 0.05. In mixed-model based GWAS, the fixed marker effect $x_{i\beta}$ is added to the model (3) above:

$$\hat{y}_i = \mu + x_{i\beta} + G_i + E_i, \qquad G \sim N(0, \sigma^2_A K),\ Ei \sim N(0, \sigma^2_E). \qquad (5)$$

The term x_i is the marker score, β is the marker effect and the genotypic effects $G = (G_1,\ldots,G_n)$ follow a $N(0, \sigma^2_A K)$ distribution. GWA mapping was performed on the means. The covariances between these effects are modeled by a genetic relatedness matrix (GRM) estimated from markers, which is called K, kinship matrix.

Following the methodology of EMMAX (Kang et al., 2010), we first obtained REML estimates of the variance components σ^2_A and σ^2_E in model (3); given these estimates the significance of the marker effect β in (5) was tested for each SNP-marker in turn, using generalized least squares (GLS). REML estimates of the variance components were obtained with the commercial R-package ASREML (Butler et al., 2007) and for the GLS calculations the command-line program scan_GLS was used (Kruijer et al.,. in prep.). Estimates of narrow sense heritability based on model (4) were obtained with the R-package heritability (Kruijer et al.,. in prep.).

Sequence analysis

All sequences from the re-sequenced Arabidopsis accessions were obtained from http://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 a 1 or 0 for compliance or no compliance with Col-0. The resulting data is stored as data frames (.csv file) on disk. In order to calculate the LD, required data is extracted from the .csv files with the gnu program 'cut' in order to slice out the region of interest. The sliced data frame is read into 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. In order to 'annotate' the genome with SNP polymorphisms we applied the tool Snpeff (Cingolani et al., 2012). With the output of this tool, which is stored in a mySQL database we are able to predict the effect of each mutation. Both the output of this tool and the LD scores are made available to the user via a web interface (in house access only). The user can calculate the LD in any region on the genome and is performed on the fly.

Nucleotide diversity analysis

All sequences from the re-sequenced Arabidopsis accessions were obtained from http://signal.salk.edu/atg1001/3.0/gebrowser.php. Nucleotide diversity was measured with Tajima's π (Tajima, 1983) using DnaSP software version 4.0 (Rozas et al., 2003). $\prod$ was calculated for all sites, synonymous, non-synonymous and silent sites (synonymous plus non-coding sites) for each candidate gene. For deviation from neutrality, we tested using Tajima's D statistic (Tajima, 1989) using DnaSP version 4.0 (Rozas et al., 2003).

Acknowledgements

We thank K. Keegstra for kindly providing the seeds for the *xxt1/xxt2* double knockout mutant.

References

Alonso-Blanco C, Aarts MGM, Bentsink L, Keurentjes JJB, et al (2009) What Has Natural Variation Taught Us about Plant Development, Physiology, and Adaptation? The Plant cell 21: 1877-1896

Amthor JS (1984) The role of maintenance respiration in plant growth. Plant Cell and Environment 7: 561-569

Atwell S, Huang YS, Vilhjalmsson BJ, Willems G, et al (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature 465: 627-631

Baxter I, Brazelton JN, Yu D, Huang YS, et al (2010) A coastal cline in sodium accumulation in Arabidopsis thaliana is driven by natural variation of the sodium transporter AtHKT1;1. PLoS Genet 6: e1001193

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

Bolton MD (2009) Primary Metabolism and Plant Defense-Fuel for the Fire. Molecular Plant-Microbe Interactions 22: 487-497

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical biochemistry 72: 248-254

Brotman Y, Riewe D, Lisec J, Meyer RC, et al (2011) Identification of enzymatic and regulatory genes of plant metabolism through QTL analyzis in Arabidopsis. Journal of Plant Physiology 168: 1387-1394

Butler D, Cullis BR, Gilmour A, Gogel B (2007) ASReml-R reference manual.

Calenge F, Saliba-Colombani V, Mahieu S, Loudet O, et al (2006) Natural variation for carbohydrate content in Arabidopsis. Interaction with complex traits dissected by quantitative genetics. Plant Physiology 141: 1630-1643

Causse M, Rocher JP, Henry AM, Charcosset A, et al (1995) Genetic dissection of the relationship between carbon metabolism and early growth in maize with emphasis on key enzyme loci. Molecular Breeding 1: 259-272

Cavalier DM, Lerouxel O, Neumetzler L, Yamauchi K, et al (2008) Disrupting two Arabidopsis thaliana xylosyltransferase genes results in plants deficient in xyloglucan, a major primary cell wall component. Plant Cell 20: 1519-1537

Chan EK, Rowe HC, Corwin JA, Joseph B, et al (2011) Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in Arabidopsis thaliana. PLoS Biol 9: e1001125

Chan EK, Rowe HC, Hansen BG, Kliebenstein DJ (2010) The complex genetic architecture of the metabolome. PLoS genetics 6: e1001198

Chan EK, Rowe HC, Kliebenstein DJ (2010) Understanding the evolution of defense metabolites in Arabidopsis thaliana using genome-wide association mapping. Genetics 185: 991-1007

Chen CN, Chu CC, Zentella R, Pan SM, et al (2002) AtHVA22 gene family in Arabidopsis: phylogenetic relationship, ABA and stress regulation, and tissue-specific expression. Plant Molecular Biology 49: 633-644

Cingolani P, Platts A, Wang LL, Coon M, et al (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly 6: 80-92

Cosgrove DJ (2005) Growth of the plant cell wall. Nature reviews. Molecular cell biology 6: 850-861

Cross JM, von Korff M, Altmann T, Bartzetko L, et al (2006) Variation of enzyme activities and metabolite levels in 24 Arabidopsis accessions growing in carbon-limited conditions. Plant physiology 142: 1574-1588

Faik A, Price NJ, Raikhel NV, Keegstra K (2002) An Arabidopsis gene encoding an alpha-xylosyltransferase involved in xyloglucan biosynthesis. Proceedings of the National Academy of Sciences of the United States of America 99: 7797-7802

Farquhar GD, von Caemmerer S, Berry JA (2001) Models of photosynthesis. Plant Physiology 125: 42-45

Gibon Y, Blaesing OE, Hannemann J, Carillo P, et al (2004) A Robot-based platform to measure multiple enzyme activities in Arabidopsis using a set of cycling assays: comparison of changes of enzyme activities and transcript levels during diurnal cycles and in prolonged darkness. Plant Cell 16: 3304-3325

Hachiya T, Terashima I, Noguchi K (2007) Increase in respiratory cost at high growth temperature is attributed to high protein turnover cost in Petunia x hybrida petals. Plant Cell and Environment 30: 1269-1283

Horton MW, Hancock AM, Huang YS, Toomajian C, et al (2012) Genome-wide patterns of genetic variation in worldwide Arabidopsis thaliana accessions from the RegMap panel. Nat Genet 44: 212-216

Kacser H, Acerenza L (1993) A universal method for achieving increases in metabolite production. European Journal of Biochemistry 216: 361-367

Kang HM, Sul JH, Service SK, Zaitlen NA, et al (2010) Variance component model to account for sample structure in genome-wide association studies. Nat Genet 42: 348-354

Keurentjes JJ, Sulpice R, Gibon Y, Steinhauser MC, et al (2008) Integrative analyzes of genetic variation in enzyme activities of primary carbohydrate metabolism reveal distinct modes of regulation in Arabidopsis thaliana. Genome biology 9: R129

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

Li Y, Huang Y, Bergelson J, Nordborg M, et al (2010) Association mapping of local climate-sensitive quantitative trait loci in Arabidopsis thaliana. Proc Natl Acad Sci U S A 107: 21199-21204

Lisec J, Meyer RC, Steinfath M, Redestig H, et al (2008) Identification of metabolic and biomass QTL in Arabidopsis thaliana in a parallel analyzis of RIL and IL populations. The Plant journal : for cell and molecular biology 53: 960-972

Manjunath S, Lee C-HK, Van Winkle P, Bailey-Serres J (1998) Molecular and biochemical characterization of cytosolic phosphoglucomutase in maize expression during development and in response to oxygen deprivation. Plant Physiology 117: 997-1006

Meng M, Geisler M, Johansson H, Harholt J, et al (2009) UDP-Glucose Pyrophosphorylase is not Rate Limiting, but is Essential in Arabidopsis. Plant and Cell Physiology 50: 998-1011

Meyer RC, Steinfath M, Lisec J, Becher M, et al(2007) The metabolic signature related to high plant growth rate in Arabidopsis thaliana. Proceedings of the National Academy of Sciences of the United States of America 104: 4759-4764

Mitchell-Olds T, Pedersen D (1998) The molecular basis of quantitative genetic variation in central and secondary metabolism in Arabidopsis. Genetics 149: 739-747

Morandini P (2009) Rethinking metabolic control. Plant Science 176: 441-451

Nunes-Nesi A, Carrari F, Gibon Y, Sulpice R, et al (2007) Deficiency of mitochondrial fumarase activity in tomato plants impairs photosynthesis via an effect on stomatal function. Plant Journal 50: 1093-1106

Park JI, Ishimizu T, Suwabe K, Sudo K, et al (2010) UDP-Glucose Pyrophosphorylase is Rate Limiting in Vegetative and Reproductive Phases in Arabidopsis thaliana. Plant and Cell Physiology 51: 981-996

Park YB, Cosgrove DJ (2012) Changes in cell wall biomechanical properties in the xyloglucan-deficient xxt1/xxt2 mutant of Arabidopsis. Plant Physiol 158: 465-475

Poorter H, Anten NPR, Marcelis LFM (2013) Physiological mechanisms in plant growth models: do we need a supra-cellular systems biology approach? Plant Cell and Environment 36: 1673-1690

Pracharoenwattana I, Zhou WX, Keech O, Francisco PB, et al (2010) Arabidopsis has a cytosolic fumarase required for the massive allocation of photosynthate into fumaric acid and for rapid plant growth on high nitrogen. Plant Journal 62: 785-795

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

Riedelsheimer C, Lisec J, Czedik-Eysenberg A, Sulpice R, et al (2012) Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. Proceedings of the National Academy of Sciences of the United States of America 109: 8872-8877

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyzes by the coalescent and other methods. Bioinformatics 19: 2496-2497

Sergeeva LI, Keurentjes JJ, Bentsink L, Vonk J, et al (2006) Vacuolar invertase regulates elongation of Arabidopsis thaliana roots as revealed by QTL and mutant analyzis. Proc Natl Acad Sci U S A 103: 2994-2999

Sergeeva LI, Vonk J, Keurentjes JJ, van der Plas LH, et al (2004) Histochemical analyzis reveals organ-specific quantitative trait loci for enzyme activities in Arabidopsis. Plant Physiol 134: 237-245

Stitt M, Gibon Y (2013) Why measure enzyme activities in the era of systems biology? Trends in plant science

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

Stitt M, Sulpice R, Keurentjes J (2010) Metabolic networks: how to identify key components in the regulation of metabolism and growth. Plant Physiol 152: 428-444

Sulpice R, Flis A, Ivakov AA, Apelt F, et al (2014) Arabidopsis Coordinates the Diurnal Regulation of Carbon Allocation and Growth across a Wide Range of Photoperiods. Molecular Plant 7: 137-155

Sulpice R, Nikoloski Z, Tschoep H, Antonio C, et al (2013) Impact of the carbon and nitrogen supply on relationships and connectivity between metabolism and biomass in a broad panel of Arabidopsis accessions. Plant Physiol 162: 347-363

Sulpice R, Pyl ET, Ishihara H, Trenkamp S, et al (2009) Starch as a major integrator in the regulation of plant growth. Proceedings of the National Academy of Sciences of the United States of America 106: 10348-10353

Sulpice R, Trenkamp S, Steinfath M, Usadel B, et al (2010) Network analyzis of enzyme activities and metabolite levels and their relationship to biomass in a large panel of Arabidopsis accessions. The Plant cell 22: 2872-2893

Sulpice R, Tschoep H, Von Korff M, Bussis D, et al (2007) Description and applications of a rapid and sensitive non-radioactive microplate-based assay for maximum and initial activity of D-ribulose-1,5-bisphosphate carboxylase/oxygenase. Plant Cell and Environment 30: 1163-1175

Sweetlove LJ, Fell D, Fernie AR (2008) Getting to grips with the plant metabolic network. Biochemical Journal 409: 27-41

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. Genetics 105: 437-460

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

Todesco M, Balasubramanian S, Hu TT, Traw MB, et al (2010) Natural allelic variation underlying a major fitness trade-off in Arabidopsis thaliana. Nature 465: 632-U129

Verslues PE, Lasky JR, Juenger TE, Liu TW, et al (2014) Genome-wide association mapping combined with reverse genetics identifies new effectors of low water potential-induced proline accumulation in Arabidopsis. Plant Physiol 164: 144-159

Vuttipongchaikij S, Brocklehurst D, Steele-King C, Ashford DA, et al (2012) Arabidopsis GT34 family contains five xyloglucan alpha-1,6-xylosyltransferases. New Phytol 195: 585-595

Weeden NF, Gottlieb LD (1982) Dissociation, reassociation, and purification of plastid and cytosolic phosphoglucose isomerase isozymes. Plant physiology 69: 717-723

Yang J, Benyamin B, McEvoy BP, Gordon S, et al (2010) Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42: 565-569

Supplemental table 1. Trait abbreviations

Classification	Trait	Description
Biomass	FW	Fresh weight
Structural components	ChlA	Chlorophyll A
	ChlB	Chlorophyll B
	Prot	Protein
Amino acids	NO3	Nitrate
	AA	Amino acid
Sugars	Fru	Fructose
	Glu	Glucose
	G6P	Glucose-6-phosphate
	Suc	Sucrose
	Starch	Starch
Organic acids	Fum	Fumurate
	Mal	Malate
Sucrose breakdown	FK	Fructokinase
	GK	Glucokinase
	Ainv	Acid invertase
	Ninv	Neutral invertase
Glycolysis and Respiration	G6P-DH	Glucose-6-phosphate-dehydrogenase
	FBP	cytosolic Fructose biphosphatase
	FUMA	Fumarase
Calvin-Benson Cycle	iRUB	initial Rubisco
	mRUB	maximum Rubisco
Sucrose synthesis	UGP	UDP glucose pyrophosphorylase
	PGM	Phosphoglycerate mutase
	cPGI	cytosolic Phosphoglucose isomerase
	tPGI	total Phosphoglucose isomerase
Starch synthesis	AGP	ADP-glucose pyrophosphorylase
	pPGI	plastid Phosphoglucose isomerase
Nitrogen metabolism	NRVm	nitrate reductase Velocity maximal
	NRVs	nitrate reductase Velocity selective

Supplemental table 2. Descriptive statistics.

Genetic parameters for the population of 350 accessions. The classification in different metabolic classes, the trait abbreviations, the genetic coefficient of variation (CV_G), marker-based heritability (h^2), number of SNPs above a certain -$\log_{10}$(*P*-value) and the maximum -$\log_{10}$(*P*-value) are given for each trait.

				nr of SNPs > -$\log_{10}$(*P*)						
Classification	**Trait**	**[CV_G]**	**h^2**	**>3**	**>4**	**>5**	**>6**	**>7**	**>8**	**MAX**
Biomass	FW	42.2	0.56	218	26	5	0	0	0	5.55
Structural components	ChlA	9.3	0.40	195	6	1	1	0	0	6.34
	ChlB	7.2	0.17	141	6	1	1	0	0	6.39
	Prot	12.4	0.31	177	18	2	0	0	0	5.67
AA	NO3	11.0	0.46	198	23	2	0	0	0	5.09
	AA	19.5	0.49	225	20	1	0	0	0	5.02
Sugars	Fru	34.2	0.22	259	27	2	1	0	0	6.89
	Glu	43.4	0.31	195	20	0	0	0	0	4.92
	G6P	18.1	0.30	257	36	7	1	0	0	6.05
	Suc	28.5	0.14	176	7	0	0	0	0	4.67
	Starch	35.3	0.40	180	10	0	0	0	0	4.78
Organic acids	Fum	33.2	0.67	212	18	4	0	0	0	5.69
	Mal	25.9	0.77	249	26	0	0	0	0	4.79
Sucrose breakdown	FK	45.3	0.08	263	30	5	0	0	0	5.51
	GK	57.0	0.05	232	31	2	0	0	0	5.94
	Ainv	27.6	0.56	245	52	17	5	2	1	8.44
	Ninv	31.2	0.30	274	53	9	1	0	0	6.27
Glycolysis and Respiration	G6PDH	50.5	0.04	164	17	4	0	0	0	5.54
	FBP	46.7	0.10	201	18	3	0	0	0	5.48
	FUMA	40.9	0.14	218	24	0	0	0	0	4.91
Calvin-Benson Cycle	iRUB	24.3	0.18	216	15	1	0	0	0	5.34
	mRUB	20.1	0.09	168	10	1	0	0	0	5.07
Sucrose synthesis	UGP	19.8	0.43	253	54	18	9	4	4	8.96
	PGM	22.6	0.15	198	11	1	0	0	0	5.92
	cPGI	24.9	0.28	226	30	8	0	0	0	5.93
	tPGI	22.1	0.23	230	24	1	0	0	0	5.83
Starch synthesis	AGP	21.4	0.36	234	34	11	0	0	0	5.75
	pPGI	36.4	0.10	234	27	1	0	0	0	5.05
Nitrogen metabolism	NRVm	22.7	0.27	192	17	0	0	0	0	4.97
	NRVs	25.1	0.34	227	32	7	0	0	0	5.64

Supplemental table 3. Spearman correlation between genetic coefficient of variation, marker-based heritability and nr of SNPs above -$\log_{10}$(*P*-value) > 3 or 4, and their corresponsing *P*-values.

				nr of SNPs above -$\log_{10}$(*P*)	
		[CV_G]	h^2	> 3	> 4
	[CV_G]	*	-.306	.185	.131
	h^2	.100	*	.245	.245
nr of SNPs above -$\log_{10}$(*P*)	> 3	.328	.192	*	.804
	> 4	.490	.192	.000	*

Supplemental table 4. Pleiotropic regulator genes.

For each pleiotropic gene is given the gene identifier (ID from TAIR), the locus, the most significantly associated SNP (marker(s)), the chromome number and position of the marker (TAIR10) and the -log10(P-value) (LOD) and effect size at the significant marker for all traits. The sign of the effect indicates whether the trait in accessions with the Col-0 allele at the specific marker is up-regulated (positive sign) or down-regulated (negative sign).

	Gene ID	**XXT2**		**ACD6**				**KAT2**	
	Locus	**AT4G02500**		**AT4G14400**				**AT2G33150**	
	marker	126422	126422	141380	141380	141513	141513	72459	72459
	Chromsome	4	4	4	4	4	4	2	2
	Position	1095217	1095217	8274507	8274507	8301059	8301059	14050281	14050281
Classification	Trait	$-\log_{10}(P)$	Effect	$-\log_{10}(P)$	Effect	$-\log_{10}(P)$	Effect	$-\log_{10}(P)$	Effect
Biomass	FW	4.21	-0.03	4.36	0.05	3.40	0.03	2.01	0.03
Structural components	ChlA	0.92	0.01	0.14	0.00	0.15	0.00	0.10	0.00
	ChlB	0.65	0.00	0.03	0.00	0.01	0.00	0.47	0.00
	Prot	3.43	0.30	3.27	-0.52	3.35	-0.34	2.56	-0.37
AA	NO3	1.56	-1.73	0.64	1.72	0.85	1.33	2.16	3.13
	AA	2.23	0.79	4.95	-2.17	3.33	-1.10	3.13	-1.40
Sugars	Fru	0.25	0.03	0.63	-0.12	0.57	-0.07	1.74	-0.20
	Glu	1.13	0.19	0.41	0.17	0.53	0.13	1.37	-0.32
	G6P	1.45	5.45	4.56	-19.08	4.50	-11.98	6.05	-18.41
	Suc	1.45	0.07	1.47	-0.12	2.44	-0.10	2.16	-0.12
	Starch	1.57	2.44	0.33	-1.40	1.93	-3.07	0.88	-2.42
Organic acids	Fum	0.03	-0.02	5.18	2.05	3.31	1.01	0.04	-0.05
	Mal	0.98	0.23	1.57	-0.55	3.20	-0.54	1.79	-0.50
Sucrose breakdown	FK	0.32	3.76	3.83	-37.01	1.47	-13.05	1.45	-16.50
	GK	0.00	-0.06	1.79	-33.11	4.02	-33.46	0.05	1.66
	Ainv	0.00	-0.11	8.44	-113.14	2.33	-34.64	0.83	-23.29
	Ninv	0.62	3.72	4.64	-23.14	3.23	-12.01	3.90	-17.51
Glycolysis and Respiration	G6PDH	0.91	25.46	0.64	-37.15	1.33	-38.24	0.93	-38.34
	FBP	0.78	6.76	0.08	1.91	0.28	3.49	0.00	0.09
	FUMA	0.72	102.90	1.00	-226.75	1.02	-146.03	0.35	-87.22
Calvin-Benson Cycle	iRUB	2.52	167.51	0.55	-109.12	1.08	-110.30	1.43	-172.98
	mRUB	2.03	133.05	0.09	-22.52	1.06	-102.79	1.80	-184.03
Sucrose synthesis	UGP	4.15	330.22	2.11	-389.34	3.21	-315.83	2.31	-341.76
	PGM	3.27	348.99	1.09	-315.47	2.76	-355.88	2.30	-416.26
	cPGI	2.81	32.60	3.67	-65.57	2.79	-35.49	3.08	-49.65
	tPGI	3.86	52.09	2.44	-69.47	2.65	-46.19	3.18	-67.60
Starch synthesis	AGP	5.75	81.52	0.22	-15.77	0.29	-12.51	0.59	-28.62
	pPGI	1.76	18.06	0.11	-3.61	0.74	-11.13	1.18	-20.31
Nitrogen metabolism	NRVm	1.23	10.74	1.62	-22.54	3.04	-20.85	1.18	-15.24
	NRVs	1.66	8.16	0.92	-9.75	1.57	-8.76	1.08	-9.00

Supplemental table 4 continues.

	Gene ID	HVA22E		IQD22				UBP21	
	Locus	AT5G50720		AT4G23060				AT5G46740	
	marker	202544	202544	150426	150426	150428	150428	199255	199255
	Chromsome	5	5	4	4	4	4	5	5
	Position	20633709	20633709	12086182	12086182	12087455	12087455	18968187	18968187
Classification	Trait	$-\log_{10}(P)$	Effect	$-\log_{10}(P)$	Effect	$-\log_{10}(P)$	Effect	$-\log_{10}(P)$	Effect
Biomass	FW	3.96	0.04	0.51	0.01	2.32	0.03	3.18	-0.05
Structural components	ChlA	2.21	-0.03	0.93	-0.01	2.29	-0.03	0.95	0.02
	ChlB	0.38	0.00	0.14	0.00	0.69	0.00	0.17	0.00
	Prot	5.67	-0.54	0.71	-0.14	2.61	-0.40	3.04	0.55
AA	NO3	0.25	0.60	0.51	1.03	1.36	2.51	0.05	-0.22
	AA	4.10	-1.50	0.39	-0.30	2.18	-1.19	0.76	0.75
Sugars	Fru	0.32	-0.05	4.63	-0.30	2.52	-0.26	0.22	-0.06
	Glu	0.55	-0.15	2.20	-0.37	2.12	-0.45	0.38	0.17
	G6P	3.20	-11.77	0.20	-1.56	1.97	-10.26	0.87	7.54
	Suc	1.66	-0.09	0.31	-0.03	1.16	-0.09	1.56	0.14
	Starch	3.40	-5.19	0.05	-0.19	1.67	-3.89	0.95	3.37
Organic acids	Fum	1.17	0.65	0.02	-0.02	0.38	-0.33	0.10	-0.13
	Mal	2.89	-0.62	0.56	-0.19	1.89	-0.54	0.67	0.34
Sucrose breakdown	FK	1.23	-13.06	1.62	-15.32	2.35	-24.06	1.12	19.02
	GK	0.95	-15.29	0.13	-3.20	0.60	13.62	0.56	-17.11
	Ainv	1.98	-38.40	0.20	6.63	0.93	-26.61	0.13	6.97
	Ninv	1.78	-10.13	1.83	-9.49	4.14	-19.10	0.31	4.23
Glycolysis and Respiration	G6PDH	0.15	-8.04	0.00	0.24	0.88	-40.08	1.18	61.70
	FBP	0.13	2.14	0.17	2.66	0.17	-3.16	0.09	2.27
	FUMA	0.30	-69.77	0.29	-64.09	0.39	-99.64	2.23	414.78
Calvin-Benson Cycle	iRUB	0.74	-101.38	0.06	11.23	1.43	-185.07	1.90	279.75
	mRUB	1.09	-116.35	0.60	75.37	0.23	-45.07	3.66	382.32
Sucrose synthesis	UGP	3.11	-374.93	0.05	-13.88	1.68	-297.42	2.56	477.88
	PGM	1.83	-328.72	0.10	33.14	0.87	-236.51	3.71	731.08
	cPGI	1.72	-32.40	0.62	-14.86	3.77	-58.69	3.25	66.80
	tPGI	2.19	-50.03	0.21	-8.44	2.26	-58.37	4.22	104.15
Starch synthesis	AGP	2.55	-68.89	1.00	-35.32	2.30	-74.78	0.86	49.34
	pPGI	1.42	-21.08	0.22	4.82	0.02	0.51	1.78	34.56
Nitrogen metabolism	NRVm	0.83	-11.01	0.05	0.96	1.28	-16.99	3.67	40.22
	NRVs	0.94	-7.54	0.16	1.71	1.87	-13.57	4.23	27.25

Supplemental table 4 continues.

	Gene ID	LIP1		CIR1		CAC2		CRK30	
	Locus	AT5G64813		AT5G37260		AT5G35360		AT4G11460	
	marker	211982	211982	188153	188153	185972	185972	137560	137560
	Chromsome	5	5	5	5	5	5	4	4
	Position	25912160	25912160	14754311	14754311	13586045	13586045	6966308	6966308
Classification	Trait	$-\log_{10}(P)$	Effect	$-\log_{10}(P)$	Effect	$-\log_{10}(P)$	Effect	$-\log_{10}(P)$	Effect
Biomass	FW	1.29	-0.01	1.21	-0.01	0.66	-0.01	2.05	-0.04
Structural components	ChlA	0.65	0.01	0.12	0.00	1.15	-0.01	0.02	0.00
	ChlB	0.05	0.00	1.02	0.00	0.70	0.00	0.60	0.00
	Prot	1.50	0.19	1.77	0.23	0.63	0.11	2.13	0.46
AA	NO3	0.78	-1.10	0.57	-0.97	1.75	1.93	1.54	-3.51
	AA	0.78	0.40	0.11	0.09	0.40	0.25	2.15	1.54
Sugars	Fru	1.70	0.14	0.08	-0.01	3.37	-0.21	0.16	0.05
	Glu	0.24	0.06	0.19	0.06	4.64	-0.47	0.15	-0.08
	G6P	1.70	6.08	1.60	6.43	0.14	-0.95	2.74	16.26
	Suc	0.82	0.05	2.72	0.11	0.25	0.02	0.85	0.09
	Starch	0.87	1.65	1.05	2.05	0.07	-0.21	0.28	1.41
Organic acids	Fum	0.02	0.02	0.05	0.04	2.31	-0.76	0.72	-0.69
	Mal	2.10	0.38	1.78	0.37	0.33	0.11	1.27	0.55
Sucrose breakdown	FK	0.80	7.51	0.88	9.07	0.31	3.85	0.21	5.42
	GK	0.10	-1.85	0.06	1.42	0.50	7.78	0.04	-1.81
	Ainv	1.21	20.74	0.13	3.95	0.02	-0.59	0.87	33.23
	Ninv	1.78	7.58	1.02	5.77	0.10	0.89	1.88	15.66
Glycolysis and Respiration	G6PDH	1.44	34.43	0.32	13.39	0.08	3.57	2.46	99.87
	FBP	0.37	-3.91	1.15	9.87	0.41	-4.37	0.47	-9.47
	FUMA	2.30	220.09	0.96	138.47	0.07	-15.49	1.58	348.45
Calvin-Benson Cycle	iRUB	1.93	143.60	4.95	272.19	0.20	28.14	0.59	129.33
	mRUB	3.69	188.67	4.95	254.10	0.18	-23.77	0.21	52.81
Sucrose synthesis	UGP	2.36	238.35	0.87	137.65	0.02	-5.66	1.13	299.60
	PGM	4.84	438.11	4.09	438.95	0.24	-59.61	2.00	526.37
	cPGI	3.74	38.15	3.23	38.21	0.07	-2.12	5.93	98.30
	tPGI	4.09	53.67	3.99	57.79	0.02	-0.83	3.86	103.90
Starch synthesis	AGP	2.97	56.46	2.71	58.55	0.21	9.03	1.20	64.43
	pPGI	1.70	17.64	1.70	19.19	0.11	2.16	0.13	5.13
Nitrogen metabolism	NRVm	2.89	18.28	3.34	21.74	0.03	0.45	1.64	25.94
	NRVs	3.21	12.18	1.90	9.75	0.21	1.87	2.07	18.75

Supplemental table 4 continues.

	Gene ID	GA20OX1		LECRKA4.3		SRp34a		TAF4B	
	Locus	AT4G25420		AT5G01560		AT3G49430		AT1G27720	
	marker	152026	152026	161250	161250	115042	115042	16610	16610
	Chromsome	4	4	5	5	3	3	1	1
	Position	12995479	12995479	217923	217923	18332175	18332175	9643865	9643865
Classification	Trait	$-\log_{10}(P)$	Effect	$-\log_{10}(P)$	Effect	$-\log_{10}(P)$	Effect	$-\log_{10}(P)$	Effect
Biomass	FW	1.11	0.02	2.40	-0.02	0.03	0.00	1.04	-0.01
Structural components	ChlA	2.45	-0.02	0.01	0.00	1.04	-0.02	1.36	0.01
	ChlB	0.46	0.00	0.17	0.00	0.28	0.00	1.10	0.00
	Prot	0.46	-0.10	2.12	0.23	0.75	-0.17	1.43	0.17
AA	NO3	0.46	0.90	0.02	0.05	0.34	0.86	0.75	-1.04
	AA	1.85	-0.84	1.62	0.65	1.57	-0.92	0.75	0.37
Sugars	Fru	0.40	-0.06	1.31	-0.12	3.62	-0.31	0.72	0.07
	Glu	2.07	-0.34	1.21	-0.21	0.36	-0.12	0.46	0.10
	G6P	0.51	-3.21	1.60	5.90	1.72	-8.86	2.09	6.59
	Suc	0.58	-0.04	0.83	0.05	0.30	-0.03	0.06	0.01
	Starch	2.70	-4.05	0.45	-1.03	0.58	-1.78	0.07	0.18
Organic acids	Fum	3.11	-1.05	0.41	-0.23	0.85	0.57	0.03	-0.02
	Mal	0.90	-0.26	0.32	0.10	1.35	-0.42	2.65	0.41
Sucrose breakdown	FK	0.47	-6.26	0.78	7.64	0.54	-8.32	0.31	3.65
	GK	0.15	3.49	0.18	3.38	0.14	-3.84	0.51	7.61
	Ainv	0.42	11.69	0.01	-0.44	4.40	-65.65	3.99	40.45
	Ninv	0.72	-4.97	1.06	5.49	6.27	-22.61	5.81	14.25
Glycolysis and Respiration	G6PDH	0.84	-29.88	0.21	8.50	0.57	-27.32	0.33	12.14
	FBP	0.82	-8.46	1.09	8.72	0.46	-6.80	0.09	-1.12
	FUMA	0.97	-152.34	1.79	190.76	1.11	-200.81	1.13	134.05
Calvin-Benson Cycle	iRUB	1.19	-126.29	0.21	28.55	0.24	-46.07	1.18	100.26
	mRUB	1.40	-130.68	0.81	76.51	0.56	-83.38	0.55	56.04
Sucrose synthesis	UGP	1.60	-223.49	0.86	125.96	0.49	-120.49	1.47	169.13
	PGM	1.38	-248.71	0.86	153.13	0.86	-219.96	0.92	152.48
	cPGI	1.84	-29.92	1.12	18.33	2.34	-42.09	1.57	21.47
	tPGI	1.76	-38.94	1.04	23.35	1.70	-46.16	1.33	25.93
Starch synthesis	AGP	3.97	-79.50	0.61	20.29	1.64	-57.05	0.82	23.69
	pPGI	0.44	-8.21	0.17	3.14	0.17	-4.61	0.57	7.96
Nitrogen metabolism	NRVm	2.25	-18.77	4.16	22.68	0.56	-8.98	1.45	11.39
	NRVs	1.32	-8.43	5.60	16.71	0.14	-1.80	0.76	4.61

Supplemental table 5. Genes in LD interval of *XXT2*.
For each locus is given the gene identifier (ID) and description of the gene (TAIR). *XXT2* (At4G02500) is colored in grey.

Locus	Gene ID	Description
AT4G02350	*SEC15B*	exocyst complex component sec15B
AT4G02360		Protein of unknown function, DUF538
AT4G02370		Protein of unknown function, DUF538
AT4G02380	*SAG21*	senescence-associated gene 21
AT4G02390	*PP*	poly(ADP-ribose) polymerase
AT4G02400		U3 ribonucleoprotein (Utp) family protein
AT4G02410		Concanavalin A-like lectin protein kinase family protein
AT4G02420		Concanavalin A-like lectin protein kinase family protein
AT4G02430	*SR34b*	RNA-binding (RRM/RBD/RNP motifs) family protein
AT4G02440	*EID1*	F-box family protein
AT4G02450		HSP20-like chaperones superfamily protein
AT4G02460	*PMS1*	DNA mismatch repair protein, putative
AT4G02470		
AT4G02480		AAA-type ATPase family protein
AT4G02490		transposable element gene
AT4G02500	*XXT2*	UDP-xylosyltransferase 2
AT4G02510	*TOC86*	translocon at the outer envelope membrane of chloroplasts 159
AT4G02520	*GSTF2*	glutathione S-transferase PHI 2
AT4G02530		chloroplast thylakoid lumen protein
AT4G02540		Cysteine/Histidine-rich C1 domain family protein
AT4G02550		
AT4G02560	*LD*	Homeodomain-like superfamily protein
AT4G02570	*CUL1*	cullin 1
AT4G02580		NADH-ubiquinone oxidoreductase 24 kDa subunit, putative
AT4G02590	*UNE12*	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
AT4G02600	*MLO1*	Seven transmembrane MLO family protein
AT4G02610		Aldolase-type TIM barrel family protein
AT4G02620		vacuolar ATPase subunit F family protein
AT4G02630		Protein kinase superfamily protein
AT4G02640	*BZO2H1*	bZIP transcription factor family protein

Supplemental table 6. The effect of the xxt1, xxt2 and xxt1/xxt2 mutants on all traits.
For each trait is given the unit of measurement and the trait value for respectively wild-type Col-0 (WT), the *xxt1*, *xxt2* and *xxt1/xxt2* mutant. Significantly different trait values from WT are encircled with a thick box (Table S7). Significance scores (LSD-corrrected, P <0.05) for pair-wise comparisons of trait values between WT and *xxt1*, *xxt2* and *xxt1/xxt2* mutants.

		Trait value				P-value between WT and		
Trait	**Units**	**WT**	*xxt1*	*xxt2*	*xxt1xxt2*	*xxt1*	*xxt2*	*xxt1xxt2*
FW	g.plant-1	0.38	0.41	0.44	0.30	0.82	0.46	0.00
ChlA	mg.g-1 FW	0.99	1.04	1.02	0.98	0.31	0.43	0.93
ChlB	mg.g-1 FW	1.00	1.00	1.02	0.98	0.76	0.57	0.40
Prot	mg.g-1 FW	1.00	1.04	0.95	1.01	0.30	0.12	0.74
NO3-	µmol.g-1 FW	1.02	1.05	1.00	0.94	0.61	0.64	0.18
AA	µmol.g-1 FW	1.05	1.02	1.05	0.91	0.74	0.91	0.09
Fru	µmol.g-1 FW	0.76	1.15	1.18	1.07	0.02	0.01	0.05
Glu	µmol.g-1 FW	0.97	1.06	0.92	0.91	0.53	0.73	0.68
G6P	µmol.g-1 FW	1.14	0.95	1.00	1.05	0.26	0.39	0.60
Suc	µmol.g-1 FW	1.01	1.01	1.05	0.94	0.95	0.63	0.34
Starch	µmol.g-1 FW	1.05	1.05	1.00	0.85	0.90	0.49	0.01
Fum	µmol.g-1 FW	0.85	1.03	1.05	0.99	0.11	0.08	0.19
Mal	µmol.g-1 FW	0.86	1.05	1.06	1.00	0.14	0.13	0.29
FK	nmol.min-1.g-1 FW	225	155	210	156	0.01	0.53	0.01
GK	nmol.min-1.g-1 FW	85	89	101	75	0.84	0.40	0.60
Ainv	nmol.min-1.g-1 FW	643	539	596	564	0.01	0.21	0.04
Ninv	nmol.min-1.g-1 FW	153	128	142	131	0.06	0.39	0.10
G6PDH	nmol.min-1.g-1 FW	705	756	702	663	0.70	0.98	0.75
FBP	nmol.min-1.g-1 FW	95	92	120	112	0.89	0.11	0.26
FUMA	nmol.min-1.g-1 FW	4356	3786	4546	4059	0.31	0.73	0.59
iRub	nmol.min-1.g-1 FW	11551	11655	12702	10336	0.91	0.22	0.20
mRub	nmol.min-1.g-1 FW	14488	13045	14192	11774	0.07	0.69	0.00
UGP	nmol.min-1.g-1 FW	8960	7190	8564	7519	0.00	0.18	0.00
PGM	nmol.min-1.g-1 FW	10077	9091	9501	8610	0.02	0.14	0.00
SPS	nmol.min-1.g-1 FW	328	259	324	263	0.00	0.76	0.00
cPGI	nmol.min-1.g-1 FW	1165	1097	1159	1010	0.16	0.90	0.00
tPGI	nmol.min-1.g-1 FW	1481	1349	1432	1323	0.03	0.40	0.01
AGP	nmol.min-1.g-1 FW	1317	1050	1226	1057	0.22	0.81	0.25
pPGI	nmol.min-1.g-1 FW	317	252	274	313	0.22	0.26	0.94
NRVm	nmol.min-1.g-1 FW	884	789	842	817	0.16	0.52	0.32
NRVs	nmol.min-1.g-1 FW	455	419	428	385	0.35	0.48	0.08

Supplemental table 7. Spearman correlations (r_s) and adjusted *P*-values with Bonferroni correction between biomass formation (FW) and enzymes expressed on a protein basis. Trait abbreviations are listed in Supplemental table 1.

	FW	
Trait	**r_s**	***P*-value**
FK	-0.18	0.02
GK	-0.13	0.32
Ainv	-0.02	1.00
Ninv	-0.16	0.05
G6PDH	-0.15	0.09
FBP	0.01	1.00
FUMA	-0.04	1.00
iRub	0.00	1.00
mRub	0.12	0.53
UGP	-0.09	1.00
PGM	-0.03	1.00
cPGI	-0.16	0.07
tPGI	-0.17	0.04
AGP	-0.02	1.00
pPGI	-0.12	0.56
NRVm	-0.16	0.05
NRVs	-0.21	0.00

CHAPTER 5

Epigenetic basis of morphological variation and phenotypic plasticity in *Arabidopsis thaliana*

Authors:

Rik Kooke[1,2,3], Frank Johannes[4], René Wardenaar[4], Vincent Colot[5], Dick Vreugdenhil[1,3] and Joost J.B. Keurentjes[2,3]

[1] Laboratory of Plant Physiology, Wageningen University, the Netherlands
[2] Laboratory of Genetics, Wageningen University, the Netherlands
[3] Centre for Biosystems Genomics, Wageningen Campus, the Netherlands
[4] Groningen Bioinformatics Centre, University of Groningen, the Netherlands
[5] Ecole Normale Supérieure, Institut de Biologie (IBENS), Centre National de la Recherche Scientifique (CNRS) UMR8197, Institut National de la Santé et de la Recherche Médicale (INSERM) U1024, France

Submitted to Plant Cell

Abstract

Epigenetics is receiving growing attention in the plant science community and is likely to contribute to improvements in crop growth and productivity in the coming years. Epigenetic modifications are thought to play a particularly important role in fluctuating environments. It is hypothesized that epigenetics contributes considerably to plant phenotypic plasticity because epigenetic modifications, in contrast to DNA sequence changes, are more likely to be reversible. The population of *ddm1-2* derived epigenetic recombinant inbred lines (epiRILs) in *Arabidopsis thaliana* is well-suited for studying this hypothesis as DNA methylation differences are maximized and DNA sequence variation is minimized in this population. Here we report on the extensive variation in plant growth and morphology in neutral and saline conditions detected among the epiRILs. Variation in most traits was highly heritable and epigenomic regions could be associated with the differences in phenotypes. Plant performance, in terms of branching and leaf area, was both reduced and enhanced by different QTLs in the *ddm1-2* inherited epigenotype. The experimentally induced variation was found to affect salinity tolerance and phenotypic plasticity. The variation in plasticity associated significantly with certain genomic regions in which the *ddm1-2* inherited epigenotype caused an increased sensitivity to environmental changes. Many of the QTLs for morphology and plasticity overlapped suggesting major pleiotropic effects. Moreover, methylation differences in the promoter region of a salt tolerance gene, *HIGH-AFFINITY* K^+ *TRANSPORTER 1* (*HKT1*), associated significantly with leaf area in the epiRILs. These findings indicate that epigenetics contributes substantially to natural variation in plant growth and morphology, especially under stress conditions.

Introduction

Epigenetics is thought to be one of the reasons why genome-wide association studies fail to explain a substantial part of the heritable variation within species (Johannes et al., 2008; Bergelson and Roux, 2010; Korte and Farlow, 2013). DNA methylation, together with other chromatin modifications, is most often associated with silencing of transposable elements (TEs), and when present in *cis*-regulatory regions, with reduced gene expression. Although DNA methylation and de-methylation may occur spontaneously during development and in response to a changing environment, epigenetic patterns can be stably inherited through mitosis and meiosis, and could thus play a significant role in evolutionary processes (Rapp and Wendel, 2005; Richards, 2006; Baubec et al., 2010; Eichten et al., 2013). When genetic resources are exhausted or genetic diversity within species is low, epigenetic variation could become an important resource for optimizing plant yield (Hauben et al., 2009; Mirouze and Paszkowski, 2011; Springer, 2013).

Stressful environments can cause both hypomethylation and hypermethylation of DNA concurrent with up- or down-regulation of gene expression in different plant species (Steward et al., 2002; Hashida et al., 2006; Chinnusamy and Zhu, 2009; Grativol et al., 2012). In tobacco, oxidative stress, but not biotic stress, caused demethylation in promoter and coding regions of a glycerophosphodiesterase-like protein (GPDL) coinciding with induced *GPDL* expression (Choi and Sano, 2007). Moreover, there is a good correspondence between DNA methylation, histone modification and gene expression changes in both Arabidopsis and rice in response to salt stress (Bilichak et al., 2012; Karan et al., 2012). Through altered gene expression, DNA methylation may thus influence plant growth and development in stressful environments.

The genome-wide effects of epigenetic modifications on growth and development under stressful conditions have rarely been studied in detail. One of the main reasons is that the study of natural epigenetic variation is complicated due to the large contribution of DNA sequence variation to phenotypic variation within species. However, recently developed genome-wide bisulphite sequencing in natural and experimental populations of Arabidopsis and maize may open up new opportunities for studying epigenetic natural variation (Eichten et al., 2013; Schmitz et al., 2013b; Schmitz et al., 2013a). In addition, epigenetic recombinant inbred lines (epiRILs) provide an effective way to circumvent sequence variation. Two such epiRIL populations have been created in Arabidopsis by crossing wild-type Col-0 with the epigenetic DNA methylation mutants *ddm1-2* or *met1* in the same Col-0 background (Johannes et al., 2009; Reinders et al., 2009). Loss of *DDM1* results in a substantial reduction in DNA methylation and an increase in TE transcription and, although rare, transposition of TEs (Tsukahara et al., 2009). Loss of *MET1* results in almost complete loss of CG and non-CG methylation. The epiRIL populations consist of nearly isogenic lines (the *ddm1-2* and *met1* mutations have been eliminated by backcrossing and segregation in the F_2 progeny) with

stretches of DNA being differentially methylated that can be tested in multiple experiments and environments.

The *ddm1-2* derived epiRIL population has been analyzed for a number of growth-related morphological traits in both neutral and stressful conditions (Johannes et al., 2009; Reinders et al., 2009; Latzel et al., 2012; Zhang et al., 2013). The observed variation among the lines was found to be highly heritable, and recently, specific differentially methylated regions (DMRs) were shown to act as epigenetic quantitative trait loci accounting for most of the heritable variation in flowering time and root length (Cortijo et al., 2014). Besides phenotypic variation, phenotypic plasticity is an important property that can be induced or repressed through DNA methylation, as was recently demonstrated in epiRILs (Bossdorf et al., 2010; Mirouze and Paszkowski, 2011; Zhang et al., 2013). Phenotypic plasticity is defined as the ability of a genotype to express alternative phenotypes in different environments (Schlichting, 1986). Phenotypically plastic genotypes are able to display a variety of phenotypes, in both morphology and physiology, in response to changes in the environment and as such can have improved growth and reproduction (Lacaze et al., 2009). It has been proposed that this plasticity is hidden in wild-type plants through DNA methylation, and when unlocked, could be valuable for the improvement of crop production in unfavorable conditions (Mirouze and Paszkowski, 2011). Indeed, phenotypic plasticity in response to drought and nutrient stress is significantly increased in epiRILs and this increase is heritable, indicating that it could be subjected to selection (Zhang et al., 2013).

In the present study, a population of 99 *ddm1-2* derived epiRILs was grown under favorable and moderately saline conditions. The population was analyzed for a range of plant growth and morphology-related traits under both conditions. Ample variation between the epiRILs for all traits tested was observed and this variation was found to be highly heritable. We show here that experimentally-induced hypomethylation of chromosomes can render plants more sensitive to environmental variation and more plastic in their responses. DMR-based QTL mapping revealed many co-locating QTLs controlling growth, morphology and plasticity that coincided with previously published epigenetic QTLs for flowering time and root length, suggesting pleiotropic regulation via epigenetic mechanisms.

Results

Phenotypic characterization: Morphological traits

To assess the impact of DNA methylation on phenotypic variation in shoot growth and morphology, 99 epiRILs and their parents, Col-0 and *ddm1-2*, were analysed under neutral and moderately saline (25mM NaCl) conditions. Under saline conditions, plants were smaller, flowered later, produced fewer branches and had shorter internodes and inflorescence lengths (Figure 1A). The Col-0 parent was less affected by moderately saline conditions

than its *ddm1-2* counterpart in almost all traits (Figure 1A, Figure 3), which indicates that DNA hypomethylation, as is the case in *ddm1-2*, leads to higher sensitivity of plants to environmental perturbations. This conclusion is further strengthened by the observation that *ddm1-2* flowered earlier than Col-0 under favorable conditions, but later under saline conditions. Furthermore, for almost all traits Col-0 resembled the epiRIL population mean more than *ddm1-2*, which provides evidence for a stable heritable basis in the epiRILs as it agrees with the expected segregation from a back-cross scheme used for the population design (Johannes et al., 2009). Substantial variation between epiRILs was observed for each of the analysed traits although the range of variation was similar under optimal and saline conditions (Table 1). Projected leaf area varied by a factor five, whereas more than two weeks difference occurred between the earliest and latest flowering epiRIL. A two-fold difference in total plant height was observed and some lines were heavily branched, while others had almost no lateral branches (Figure 1A).

A Spearman rank correlation matrix was constructed to compare the growth and morphology-related traits across the two conditions. Leaf area correlated very well between neutral and saline conditions (r_{LA20} = 0.84), implying that fast growing epiRILs under control conditions also grew fast under saline conditions (Figure 2). Large plants produced more main stem

Table 1. Descriptive statistics for the morphological traits measured in the epiRIL population. Abbreviations used: AVG ± SEM is population average ± standard error of the population mean; V_G is among-genotype variance; V_E is residual variance; H^2 is broad-sense heritability calculated as $V_G/(V_G + V_E)$; CV_G is coefficient of genetic variation calculated as $\sqrt{(V_G)}/\bar{X}$*100% where $\bar{X}$ is the population mean; LA20, leaf area after 20 days; RGR, relative growth rate; FT, flowering time; AIL, average internode length; RB, rosette branching; MSB, main stem branching; PH1S, plant height 1st silique; TPH, total plant height.

Trait		nr of lines	AVG ± SEM	[V_G]	[V_E]	[H^2]	[CV_G]
LA20	C	96	118.84 (3.20)	981.2	1164.8	0.46	26
	S	97	101.13 (2.57)	638.3	1011.7	0.39	25
RGR	C	96	0.21 (0.00)	0.00027	0.00158	0.14	8
	S	97	0.22 (0.00)	0.00030	0.00170	0.15	8
FT	C	97	40.67 (0.32)	10.0	9.5	0.51	8
	S	93	43.61 (0.37)	12.6	16.5	0.43	8
RB	C	97	9.70 0.52)	26.5	54.7	0.33	53
	S	93	1.51 (0.21)	4.2	9.8	0.30	136
MSB	C	97	5.15 (0.12)	1.4	1.0	0.58	23
	S	93	5.35 (0.11)	1.2	1.8	0.40	20
AIL	C	97	2.48 (0.06)	0.3	0.4	0.46	22
	S	93	1.77 (0.04)	0.1	0.2	0.34	20
PH1S	C	97	12.11 (0.17)	2.9	5.1	0.37	14
	S	93	9.13 (0.17)	2.6	5.7	0.31	18
TPH	C	97	35.37 (0.33)	10.5	13.8	0.43	9
	S	93	30.19 (0.30)	8.3	23.4	0.26	10

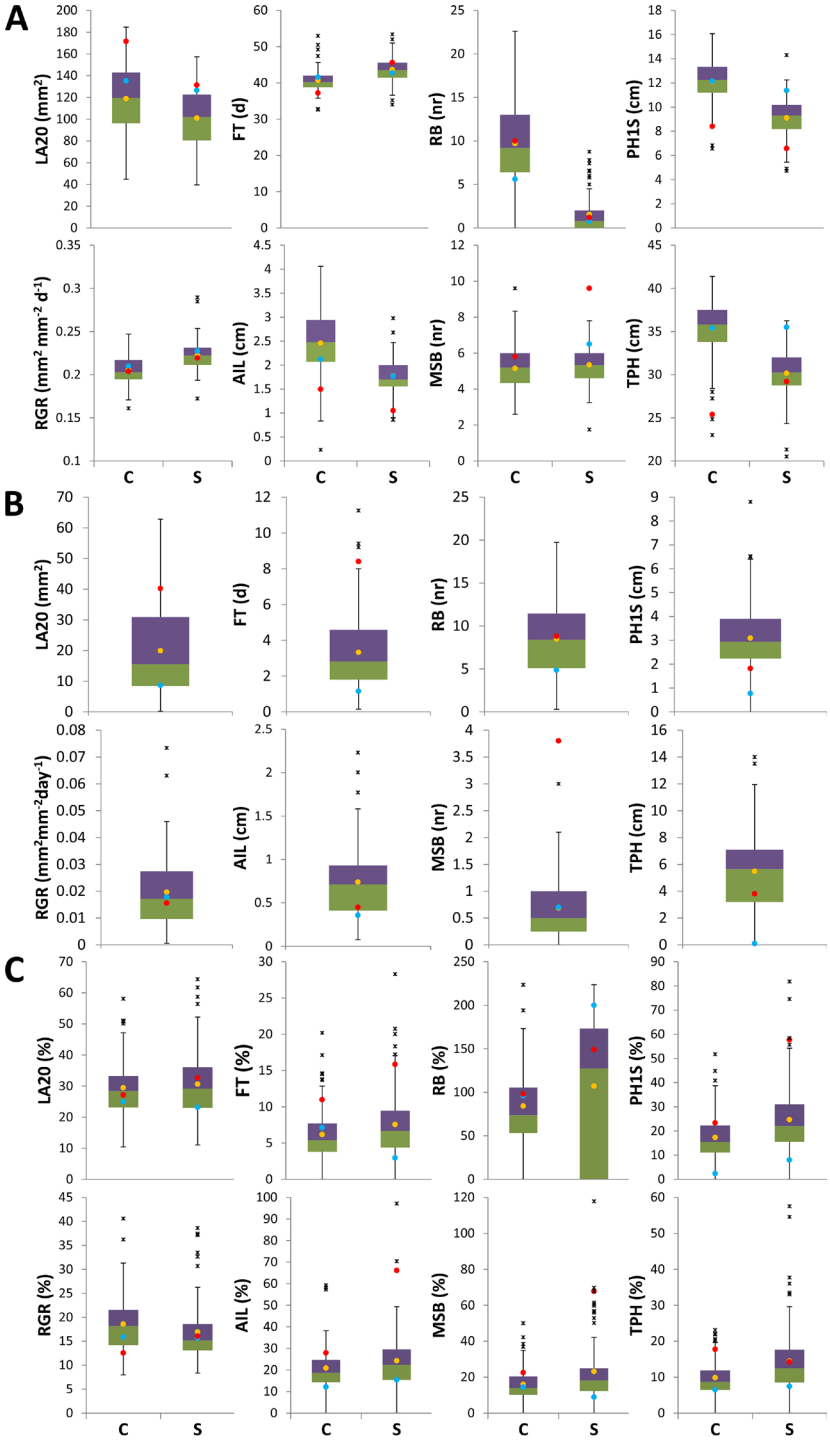
branches and acquired higher inflorescence heights, which suggests that these plants are also superior in terms of reproductive success (Clauss and Aarssen, 1994). Large plants, however, showed a much lower relative growth rate than small plants later in development, deduced from the highly negative correlation between leaf area and relative growth rate twenty days after germination ($r_{control}$ = -0.46, r_{salt} = -0.48) (Figure 2). All traits showed significant positive correlations between neutral and saline conditions. These correlations, ranging from 0.33 for total plant height to 0.74 for main stem branching, were, however, much lower than for leaf area, suggesting differential regulation of traits under control and saline conditions (Figure 2).

Phenotypic characterisation: Plasticity

Phenotypic plasticity was measured for all epiRILs as the absolute difference in average trait values between control and saline conditions. For all traits except relative growth rate, *ddm1-2* showed higher plasticity levels than Col-0, with values predominantly matching the highest and lowest quartile of the population range distribution, respectively (Figure 1B). These findings suggest that experimentally-induced DNA hypomethylation augments phenotypic plasticity. Moreover, some epiRILs showed increased trait values under saline conditions whereas others showed decreased trait values further indicating that epiRIL variation can alter the response to saline conditions (Figure 3).

The coefficient of variation (CV_G) was calculated for each epiRIL in both environments to quantify the within-line variation due to residual variation and developmental stability (Sangster et al., 2008). For most traits the average CV_G was higher under saline than under control conditions, indicating reduced stability in the saline environment (Figure 1C). CV_G values of the Col-0 parent again predominantly matched the lower quartiles of the population distribution further supporting the suggestion that DNA methylation buffers phenotypic plasticity (Figure 1C).

Figure 1. Boxplots of epiRIL variation and phenotypic plasticity.
(A) Boxplots showing variation within epiRIL population. (B) Boxplots showing distribution of phenotypic plasticity for all epiRILs. (C) Boxplots showing distribution of the coefficient of variation (CV_G) for the within-line variation for all epiRILs. The blue dot indicates the Col-0 parent value, the red dot indicates the *ddm1-2* parent value, the orange dot indicates the average value of all epiRIL lines and the black stars indicate suspected outliers (Tukey). Abbreviations used: LA20, leaf area after 20 days; RGR, relative growth rate; FT, flowering time; AIL, average internode length; RB, rosette branching; MSB, main stem branching; PH1S, plant height 1st silique; TPH, total plant height; C, control; S, saline.


A
LA20 (mm²)
FT (d)
RB (nr)
PH1S (cm)
RGR (mm² mm⁻² d⁻¹)
AIL (cm)
MSB (nr)
TPH (cm)
C
S
B
LA20 (mm²)
FT (d)
RB (nr)
PH1S (cm)
RGR (mm²mm⁻²day⁻¹)
AIL (cm)
MSB (nr)
TPH (cm)
C
S
C
LA20 (%)
FT (%)
RB (%)
PH1S (%)
RGR (%)
AIL (%)
MSB (%)
TPH (%)
C
S

		LA20		MSB		PH1S		TPH		FT		RB		AIL		RGR	
		C	S	C	S	C	S	C	S	C	S	C	S	C	S	C	S
LA20	C	*	0.8	0.4	0.5	0.4	0.3	0.3	0.3	-0.1	0.0	0.0	0.0	-0.3	-0.3	-0.5	-0.4
	S	0.00	*	0.5	0.5	0.4	0.3	0.4	0.3	-0.1	-0.1	0.0	0.0	-0.2	-0.3	-0.5	-0.5
MSB	C	0.00	0.00	*	0.7	0.4	0.4	0.1	0.2	0.4	0.4	-0.4	-0.3	-0.8	-0.5	-0.3	-0.3
	S	0.00	0.00	0.00	*	0.2	0.5	0.1	0.3	0.3	0.3	-0.3	-0.3	-0.6	-0.6	-0.2	-0.2
PH1S	C	0.00	0.00	0.00	0.02	*	0.4	0.7	0.4	0.1	0.1	0.2	0.0	0.2	0.1	-0.2	-0.1
	S	0.00	0.00	0.00	0.00	0.00	*	0.2	0.6	0.2	0.2	0.0	0.0	-0.1	0.2	-0.3	-0.2
TPH	C	0.00	0.00	0.30	0.48	0.00	0.09	*	0.3	0.0	-0.1	0.4	0.1	0.3	0.1	0.1	-0.1
	S	0.00	0.00	0.07	0.00	0.00	0.00	0.00	*	0.1	0.0	0.0	0.2	0.0	0.2	-0.1	-0.1
FT	C	0.17	0.38	0.00	0.01	0.47	0.13	0.71	0.19	*	0.6	-0.5	-0.3	-0.4	-0.1	-0.1	0.0
	S	0.96	0.28	0.00	0.01	0.37	0.05	0.34	0.86	0.00	*	-0.4	-0.4	-0.4	-0.1	-0.1	-0.1
RB	C	0.95	0.88	0.00	0.01	0.04	0.69	0.00	0.69	0.00	0.00	*	0.5	0.6	0.2	0.2	0.0
	S	0.88	0.86	0.00	0.01	0.70	0.68	0.23	0.12	0.00	0.00	0.00	*	0.4	0.3	0.0	0.1
AIL	C	0.01	0.01	0.00	0.00	0.06	0.39	0.00	0.97	0.00	0.00	0.00	0.00	*	0.6	0.2	0.2
	S	0.01	0.00	0.00	0.00	0.24	0.05	0.33	0.14	0.17	0.30	0.02	0.00	0.00	*	0.1	0.1
RGR	C	0.00	0.00	0.01	0.02	0.12	0.01	0.56	0.15	0.35	0.22	0.03	0.80	0.02	0.49	*	0.4
	S	0.00	0.00	0.01	0.03	0.19	0.06	0.51	0.63	0.93	0.47	0.67	0.27	0.11	0.22	0.00	*

Figure 2. Spearman's rho correlations and their respective P-values among morphological traits. The upper right panel shows the Spearman correlations, the lower left panel shows the significance values. Significant correlations in the upper right panel are in bold and encircled by a thick box. Abbreviations used: LA20, leaf area after 20 days; RGR, relative growth rate; FT, flowering time; AIL, average internode length; RB, rosette branching; MSB, main stem branching; PH1S, plant height 1st silique; TPH, total plant height.

QTL analysis for morphological traits

To quantify to what extent the phenotypic variation among the epiRILs was heritable, broad-sense heritability (H^2) was estimated. In general, moderate to high heritability values were observed, averaging at 0.37 and ranging from 0.14 to 0.58 for relative growth rate and main stem branching, respectively (Table 1). The H^2 for all traits, with the exception of relative growth rate, was higher under control than under saline conditions (across all traits, 0.41 and 0.33, respectively).

Previously, a genetic map was constructed for the epiRILs using differentially methylated regions (DMRs) as physical markers (Colome-Tatche et al., 2012). We employed this map to search for quantitative trait loci (QTL) that could account for the heritable variation in the morphological traits. For most traits, at least one QTL was detected. In total, 14 and 17 QTLs were detected under control and saline conditions, respectively (Figure 4, Table 2). The number and strength of QTLs varied between different traits and conditions with a maximum of five QTLs detected for leaf area under control conditions and main stem branching under

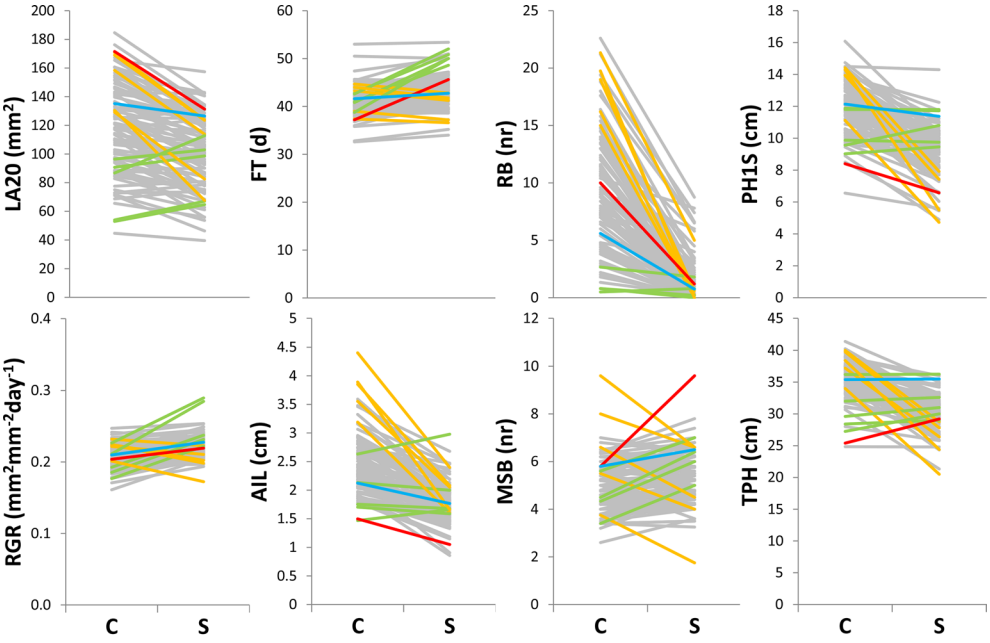


Figure 3. Reaction norm plots for all morphological traits tested in epiRIL population.
The blue line denotes Col-0, the red line denotes ddm1-2, the orange lines denote the highest negative effect lines and the green lines denote the highest positive effect lines (in some cases least negative). Abbreviations used: LA20, leaf area after 20 days; RGR, relative growth rate; FT, flowering time; AIL, average internode length; RB, rosette branching; MSB, main stem branching; PH1S, plant height 1st silique; TPH, total plant height.

saline conditions. Many QTLs were detected for multiple traits and under both conditions, indicating that these loci had pleiotropic effects independent of the growing conditions.

Twelve different QTL regions associating with one or more morphological traits could be assigned. Six of these regions were identified in both environments (QTL 2, 3, 4, 9, 10 and 14), while three QTLs were uniquely detected under control (QTL 6, 11 and 15) or saline (QTL 7, 8 and 12) conditions (Figure 4, Table 2). The similarities in QTL profiles of the neutral and saline conditions reflected the correlations between the two conditions. Traits with lower correlations showed higher numbers of unique QTLs (Figure 2 and 4). Significant LOD scores ranged from 2.6 for relative growth rate under saline conditions to 12.5 for average internode length under control conditions, explaining 12% or 46.4% of the variance, respectively. Large-effect QTLs, explaining more than 20% of the variance, were detected for leaf area, main stem branching and average internode length in neutral and saline conditions, representing two pleiotropic loci on chr 1 and 4. Although most QTLs displayed positive effects, 9 out of 31 QTLs displayed negative effects, indicating that *ddm1-2* inherited

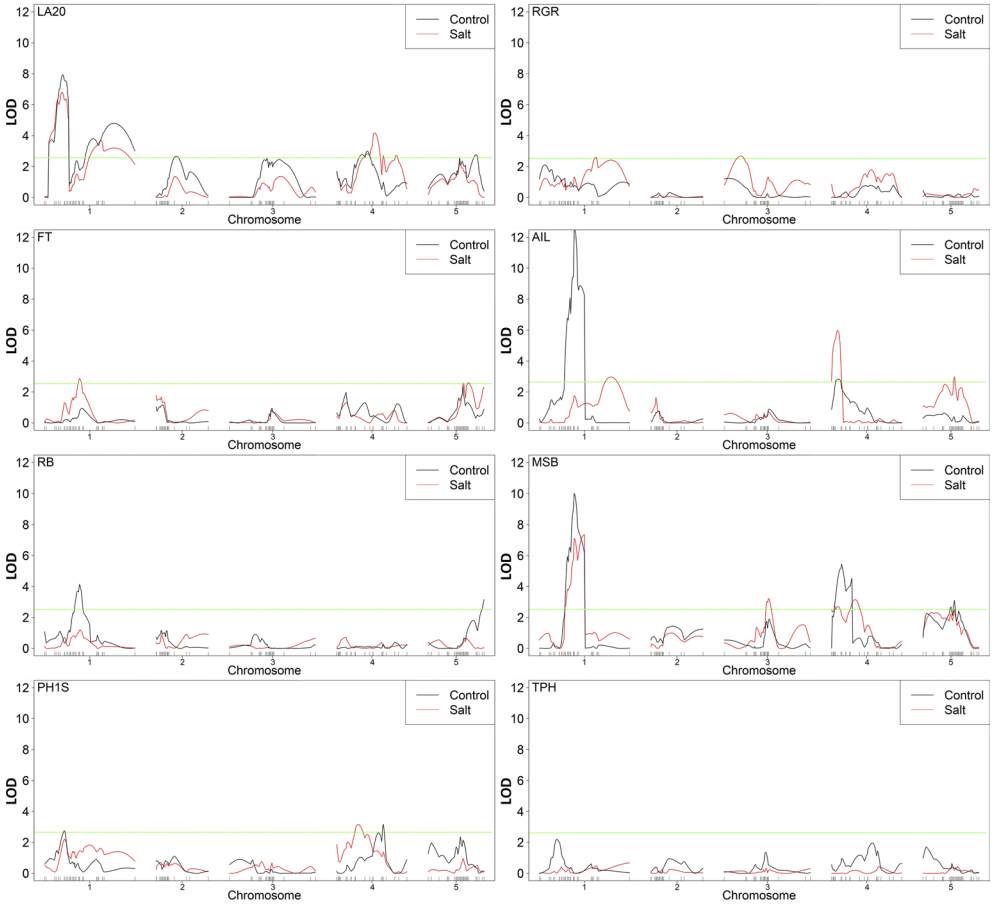


Figure 4. Epi-QTL plots for morphological traits tested in epiRIL population under control (black line) and saline (red line) conditions. LOD threshold was calculated using 1000 random permutations with α 0.05 as the genome-wide type I error level. The highest LOD threshold between the two conditions was used as LOD threshold in the figure and for determination of significance. Abbreviations used: LA20, leaf area after 20 days; RGR, relative growth rate; FT, flowering time; AIL, average internode length; RB, rosette branching; MSB, main stem branching; PH1S, plant height 1st silique; TPH, total plant height.

epigenotypes can both reduce and enhance plant morphological trait values. Positive-effect QTLs were detected for plant height, flowering time and main stem branching, while solely negative-effect QTLs were detected for rosette branching and average internode length. Opposite-effect QTLs were detected for leaf area and relative growth rate, indicating that effects are locus dependent rather than trait specific. For a number of pleiotropic QTLs (QTL 3, 4, 9 and 14) opposite effects were observed for different traits which was supported by the negative correlation between these traits (Figure 2, Table 2). The *ddm1-2* inherited

epigenotype in the QTL 3 region, for instance, was associated with decreased main stem branching but with increased rosette branching and internode length. Even though H^2 values were high, no significant QTLs were detected for total plant height under both controll and saline conditions. Similarly, no QTLs were detected for flowering time and relative growth rate under control conditions and rosette branching under saline conditions. The QTLs for flowering time did however resemble the highly significant QTL profiles from another epiRIL study (Cortijo et al., 2014), and might have gone undetected because of the smaller population size used here.

Table 2. QTLs detected in the epiRIL population for morphological traits.
The 1.5 LOD support interval is used. Co-factor indicates position of co-factors. The explained variance is calculated according to the following formula: EV (%) = $(1\text{-}10^{(-2*LOD/n)})*100\%$ where LOD is LOD score for the particular trait and n is number of epiRILs (R/QTL FAQ). The effect (%) is calculated as effect size (a) divided by mean (x 100%).Abbreviations used: M, morphology; CV, coefficient of variation; PP, phenotypic plasticity; C, control; S, saline; LA20, leaf area after 20 days; RGR, relative growth rate; FT, flowering time; AIL, average internode length; RB, rosette branching, MSB, main stem branching; PH1S, plant height 1st silique; TPH, total plant height.

nr	Trait	Morphology (M) or Plasticity (PP or CV)	Treatment	QTL	Chr	LOD	Support interval (cM)	Co-factor	Explained variance (%)	Effect (%)
1	LA20	CV	C	1	1	5.5	0-8	MM1 (1;0), MM707 (5;3.6)	24.1	-13.8
2	LA20	M	C	2	1	7.9	15-28	MM7 (1;16.4)	32.8	13.4
3	LA20	M	S		1	6.8	12-28	MM7 (1;16.4)	28.7	12.3
4	PH1S	M	C		1	2.7	12-28	-	12.9	4.1
5	RB	CV	S	3	1	2.7	14-59	-	12.5	17.1
6	LA20	CV	C		1	2.7	23.2-60	MM1 (1;0), MM707 (5;3.6)	12.8	-14.7
7	AIL	PP	-		1	4.1	37-46	MM5 (1;13.5), MM7 (1;16.4), MM123 (1;41.1)	18.7	-29.4
8	FT	M	S		1	2.9	31-51	-	13.4	3.1
9	RB	M	C		1	4.1	35-45	MM123 (1;41.1)	18.7	-22.7
10	RB	PP	-		1	4.2	35-44.8	MM123 (1;41.1)	18.9	-22.1
11	MSB	M	C		1	10.1	38-44.8	MM123 (1;41.1), MM661 (4;12.2)	39.6	12.3
12	MSB	M	S		1	7.6	37-54	MM123 (1;41.1)	31.7	9.8
13	AIL	M	C		1	12.5	40-44.8	MM123 (1;41.1)	46.4	-13.3
14	LA20	M	C	4	1	4.8	50-105	MM7 (1;16.4)	21.3	13.1
15	LA20	M	S		1	3.7	50-106	MM7 (1;16.4)	16.8	12.0
16	RGR	M	S		1	2.6	50-103	-	12.0	-2.4
17	AIL	M	S		1	3.0	70.1-101	MM551 (4;1.7)	13.8	-7.4
18	TPH	CV	C	5	2	2.8	0-13.8	-	13.1	-19.2
19	LA20	CV	S		2	4.0	3-9.9	MM330 (2;7)	18.1	-13.9
20	LA20	M	C	6	2	2.7	13.8-50	MM7 (1;16.4)	12.5	-8.3

Table 2. continues.

nr	Trait	Morphology (M) or Plasticity (PP or CV)	Treatment	QTL	Chr	LOD	Support interval (cM)	Co-factor	Explained variance (%)	Effect (%)
21	RGR	M	S	7	3	2.6	3-32	-	12.3	2.7
22	RGR	CV	C		3	2.8	3-31	-	13.0	-12.7
23	MSB	M	S	8	3	3.3	48.3-58	MM123 (1;41.1)	15.3	7.0
24	MSB	M	C	9	4	5.5	3.4-25	MM123 (1;41.1), MM661 (4;12.2)	24.0	11.4
25	MSB	M	S		4	2.8	0-17	MM123 (1;41.1)	13.2	8.6
26	AIL	M	C		4	2.8	1-21.5	MM123 (1;41.1)	13.2	-10.6
27	AIL	M	S		4	6.0	1-11	MM551 (4;1.7)	25.9	-9.7
28	RB	CV	S		4	2.8	0-21.5	-	12.9	18.2
29	LA20	M	C	10	4	2.7	18-50	MM7 (1;16.4)	12.7	11.0
30	LA20	M	S		4	4.2	33-52	MM7 (1;16.4)	18.8	12.4
31	MSB	M	S		4	3.5	18-37	MM123 (1;41.1)	16.2	8.2
32	PH1S	M	S		4	3.2	17-42	-	14.6	7.0
33	PH1S	M	C	11	4	3.2	43-58	-	14.7	5.5
34	LA20	M	S	12	4	2.7	53-80	MM7 (1;16.4)	12.7	8.6
35	LA20	CV	C	13	5	4.5	1-16	MM1 (1;0), MM707 (5;3.6)	20.0	-11.5
36	LA20	M	C	14	5	2.8	25-62	MM7 (1;16.4)	12.9	10.9
37	FT	M	S		5	2.6	37-56	-	12.2	3.1
38	MSB	M	C		5	3.1	25-47.4	MM123 (1;41.1), MM661 (4;12.2)	14.5	8.8
39	MSB	M	S		5	2.6	0-45	MM123 (1;41.1)	12.2	7.8
40	AIL	M	S		5	3.0	16-41	MM551 (4;1.7)	13.9	-6.7
41	RB	M	C	15	5	3.2	59-65.9	MM123 (1;41.1)	14.7	-30.0
42	RB	PP	-		5	2.7	58-65.9	MM123 (1;41.1)	12.8	-27.2

QTL analysis for phenotypic plasticity

Large differences were observed in the CV and phenotypic plasticity (PP) values between epiRILs and thus QTL mapping was subsequently performed on these traits. For PP, two QTLs were detected, one pleiotropic QTL on chr 1, explaining approximately 19% of the PP variation for internode length and number of branches from the rosette, and a second QTL on chr 5 explaining approximately 13% of the PP variation in rosette branching (Table 2). Both QTLs co-located with the identified QTLs for morphological trait variation, implying that the regulation of PP is governed by the same loci (Figure 2, Table 2). Eight QTLs were detected explaining the variation observed in CV, of which two coincided with the chr 1 QTL for PP. Two QTLs were pleiotropic and one QTL for relative growth rate was also detected for morphological trait variation (Table 2). Interestingly, a QTL was found for the CV of total plant height, for which no QTLs could be found for morphological variation, most likely due to the large within-line variation. The majority of CV and PP QTLs showed negative effect signs, illustrating that the *ddm1-2* inherted epigenotypes increase plant sensitivity to environmental variation.

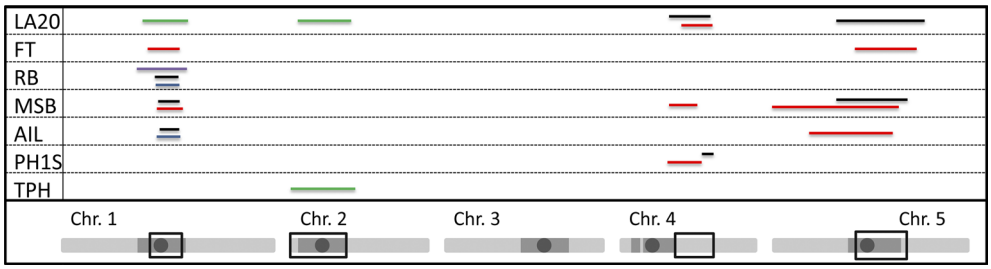


Figure 5. Overlap between QTLs found in our study and the validated epigenetic QTL for flowering time and root length in (Cortijo et al., 2014). Light grey bars indicate the five chromosomes of Arabidopsis, and the black box surrounding the chromosomes indicates the QTL intervals in (Cortijo et al., 2014). The colored lines indicate the QTL support intervals for the different traits. Different colors indicate in which conditions the QTLs were found: black, neutral conditions; red, saline conditions; green; coefficient of variation (CV) in neutral conditions; purple, CV in saline conditions; blue, phenotypic plasticity (PP). Abbreviations used: LA20, leaf area after 20 days; FT, flowering time; AIL, average internode length; RB, rosette branching; MSB, main stem branching; PH1S, plant height 1st silique; TPH, total plant height.

An epigenetic basis for pleiotropic QTLs

The QTL mapping results suggest that the variation for growth, morphology and plasticity is to a large extent due to DNA methylation differences in the epiRILs. Although, DNA sequence variation because of transposable element transposition cannot be ruled out on the basis of our results. However, the major pleiotropic QTLs found in our study coincide with six QTL intervals detected in an epigenetic QTL mapping study on flowering time and root length in which the QTL intervals for both traits were evaluated for transposable element insertions (Cortijo et al., 2014) (Figure 5). The few transposable element insertions that were found in the QTL intervals of the latter study were weakly associated, but not causal for the heritable variation, suggesting epigenetic regulation of flowering time and root length (Cortijo et al., 2014). Most importantly, it suggests that the overlapping QTLs found in our study are also epigenetically regulated, confirming the pleiotropic regulation by epigenetic mechanisms. It further gives strong supportive evidence for the epigenetic regulation of plasticity as six out of twenty-one overlapping QTLs were found for plasticity parameters. The remaining QTL intervals that did not coincide with the study of Cortijo et al., 2014 might be specific for the traits in our study and will have to be examined for transposable element insertions in future studies to confirm that these QTLs are regulated epigenetically.

The pleiotropic QTL interval on chr 4 includes the *HIGH-AFFINITY K⁺ TRANSPORTER 1*(*HKT1*) gene which is involved in Na^+ uptake by the root and Na^+ unloading from the xylem sap (Lin et al., 2004; Ren et al., 2005; Davenport et al., 2007; Baxter et al., 2010). Recently, DNA hypomethylation in the promoter region of *HKT1* was shown to induce *HKT1* expression and increase salt sensitivity (Baek et al., 2011). Therefore, the methylation levels in the

promoter region of the *HKT1* gene and the *HKT1* gene body were compared between the epiRILs in our study. Differentially methylated regions (DMRs) were observed at the end of the gene body and about 2.5 - 3 kb upstream of the gene (Figure 6A). Moreover, low gene expression variation was found between *ddm1-2* and Col-0, which was more pronounced around the DMR at the 3' end of the gene (Figure 6B). Interestingly, when the methylation profile of the two DMRs were compared with the morphological trait values, the first DMR in the promoter associated significantly with leaf area, and stronger in saline conditions ($P_{control} < 0.05$, $P_{salt} < 0.01$) (Figure 6C). These findings suggest that DNA methylation in the promoter region of *HKT1* regulates *HKT1* expression and plant growth, especially under saline conditions. To observe whether the epiRIL DMRs are also present in nature, the epiRIL DMRs were compared to DMRs found in natural populations (Cortijo et al., 2014). Although DMRs were found in natural accessions at the *HKT1* locus, they did not overlap with the DMRs in our study.

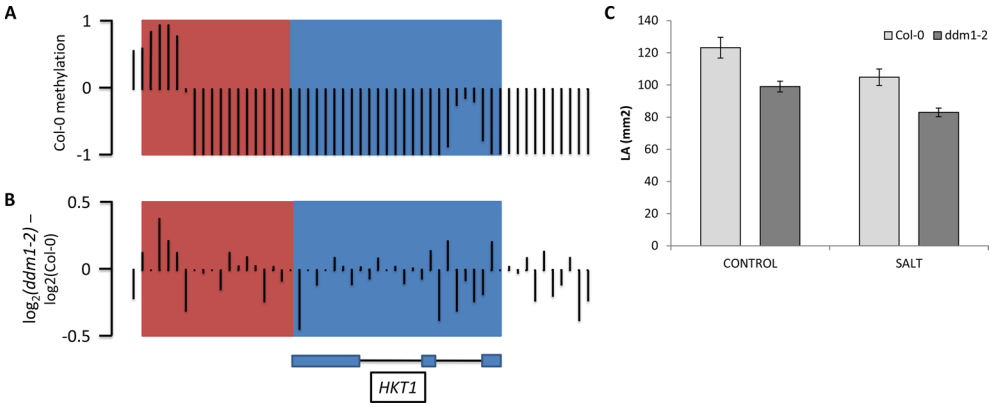


Figure 6. Methylation and expression variation at the *HKT1* locus.
(A) DNA methylation of Col-0 (bars) at 53 methylation probes in and around the *HKT1* gene. Three kb promoter region is shown in red; gene body is shown in blue; Arbitrary scaling reflects Col-0 methylation: values between -1 and -0.2 indicate unmethylated regions; values between -0.2 and 0.2 indicate intermediate methylation; value between 0.2 and 1 indicate methylated regions. *Ddm1-2* was completely unmethylated at all tested probes. (B) *HKT1* expression variation between Col-0 and *ddm1-2* (bars) at 53 probes in and around the *HKT1* gene. Three kb promoter region is shown in red; gene body is shown in blue. (C) Difference in leaf area (LA) in mm^2 in control and saline conditions between methylated (Col-0) and hypomethylated (< -0.2) (*ddm1-2*) epiRILs at the position of the DMR 2.6 kb before the *HKT1* gene.

DMR-based QTL analysis of epistasis

In the previous sections, it was outlined that morphological traits and phenotypic plasticity are to a large extent controlled by epigenetic loci. Because quantitative traits can be additively or epistatically regulated by different genetic factors (Kliebenstein et al., 2001)

Table 3. Testing for epistatic interactions by pairwise comparisons between loci in a two-dimensional two-QTL model (Broman & Sen, 2009). The full model, LOD_f, in the two-QTL model includes the main effects of the two loci and their interaction, the additive model, LOD_a, only includes the main effects of the two loci, and the epistatic model, LOD_i, tests specifically for interaction effects between the two loci ($LOD_i = LOD_f - LOD_a$). For full details of the models: see text, Material & Methods. LODs for the two-QTL model were found significant above an arbitrary threshold based on Broman and Sen, 2009: (LOD_f, LOD_{av1}, LOD_i) = (6.0, 5.0, 4.0).

Phenotype	Morphology (M) or Plasticity (PP or CV)	Treatment	Comparison between loci on chromosome:		Position of loci:		LOD score		
			1st Chr	2nd Chr	Pos_{1f}	Pos_{2f}	LOD_f	LOD_a	LOD_i
TPH	CV	C	1	3	23	101	6.5	2.1	4.3
LA20	M	S	4	5	41	38	10.6	5.9	4.7

this might also hold for epigenetic regulation. To test for epistatic interactions, pairwise comparisons were made among all loci and the interaction-effect (LOD_i) was estimated as the difference between the LOD-score of an additive model (LOD_a – not including interactions) and a full model (LOD_f) (Table 3) (Broman and Sen, 2009; Manichaikul et al., 2009). Significant epistatic interactions were found between the loci on chr 4 and chr 5 for leaf area under saline conditions and the loci on chr 1 and chr 3 for the phenotypic plasticity parameter, CV, for total plant height under control conditions (Table 3). Interestingly, the interacting loci on chr 4 and chr 5 for leaf area fall within the QTL-support intervals of the QTLs for flowering time and root length reported by (Cortijo et al., 2014), providing strong evidence for epigenetic regulation of the epistatic interaction.

Discussion

DNA methylation affects plant growth and productivity

In the present study, we show that *ddm1-2* induced DNA hypomethylation can give rise to a wide variety of highly heritable phenotypes with the exception of flowering time and total plant height for which no QTLs were detected under control conditions, most likely due to low levels of variation, which is in accordance with previous studies (Johannes et al., 2009; Zhang et al., 2013; Cortijo et al., 2014). Strong epigenetic QTLs were, however detected for flowering time in a previous epigenetic QTL study on 123 epiRILs and these QTLs resembled the QTL profiles in our study of 99 epiRILs (Figure 5) (Cortijo et al., 2014). When linkage tests of the flowering time QTL markers on chr 1,4 and 5 from the previous study (Cortijo et al., 2014) were applied to our flowering time data, a highly significant association on chr 1 ($P < 0.01$) in neutral conditions was detected. In saline conditions the QTLs on chr1 and 5 could be confirmed. These results indicate that the lack of genome-wide QTL detection for flowering time may simply be the result of reduced statistical power due to a lower number of epiRILs tested. For the other traits, high heritabilities were accompanied by strong

epigenetic variation resulting in the detection of multiple QTLs, more or less similar to genetic variation, heritability and number of QTLs found in conventional RIL populations (Ungerer et al., 2002; Bandaranayake et al., 2004; Keurentjes et al., 2007). For most traits, QTLs had positive additive effect signs, i.e. these loci increased trait values in the wild-type Col-0 background. However, negative effect-QTLs were also detected, e.g. for leaf area, rosette branching and average internode length, indicating that *ddm1-2* induced hypomethylation of genomic DNA can both reduce and enhance plant growth. Although the occurrence of trait variation due to transposable element remobilization after demethylation cannot be excluded, the strong overlap of most QTLs with the epigenetic QTLs found by Cortijo et al., 2014 suggests that these morphological traits are regulated by epigenetic mechanisms. This study thus significantly contributes to the increasing body of literature that suggests an important role for epigenetics in variation of plant growth and morphology.

DNA methylation can alter salinity tolerance

DNA hypomethylation could theoretically both increase and decrease salinity tolerance depending on genotype and site of methylation in the genome. In wheat seedlings, for example, chemical induction of DNA hypomethylation enhanced biomass and the activity of antioxidant enzymes under salt stress conditions in two different cultivars (Zhong et al., 2010). DNA hypomethylation was higher in one of the cultivars, concomitant with increased activity of the antioxidant enzymes. Genetic induction of hypomethylation through the *ddm1-2* mutation, however, reduced salinity tolerance to some extent in Arabidopsis seedlings (Yao et al., 2012). Opposite effects between genetic and chemical induction of hypomethylation were also observed for flowering time in Arabidopsis (Yaish et al., 2011). In our study, different lines of evidence illustrate that DNA hypomethylation decreased growth and reproductive success – in terms of branching and plant height (Clauss and Aarssen, 1994) - under saline conditions. For most traits, the wild-type Col-0 performed better than the population mean of the hypomethylated epiRILs under saline conditions. Furthermore, the QTL on chr 2, of which the hypomethylated allele increased leaf area under control conditions, was not observed under saline conditions. And, all leaf area QTLs detected under saline conditions had positive additive effect signs, reflecting higher trait values for Col-0 alleles.

That DNA methylation can enhance growth under saline conditions is further exemplified by our study on *HKT1*, a salt tolerance gene involved in Na^+ uptake by the root and Na^+ unloading from the xylem sap as revealed by QTL, GWAS and mutant studies (Lin et al., 2004; Ren et al., 2005; Davenport et al., 2007; Baxter et al., 2010). In our studies, a QTL on chr 4, relatively close to the *HKT1* gene, was found in both conditions but with a stronger effect in the saline environment. This could potentially mean that the tolerance towards saline conditions by *HKT1* is controlled by expression differences due to *cis*-regulatory epigenetic modifications. Methylation and low transcript variation were found at the *HKT1* gene (Figure 6 A and B),

and methylation differences in the promoter region of *HKT1* associated significantly with leaf area in the epiRILs, especially under saline conditions (Figure 6C). In another study, DNA hypomethylation in the promoter region of *HKT1* was shown to induce *HKT1* expression and increase salt sensitivity (Baek et al., 2011). Consistent with our results, a small putative RNA target region about 2.6 kb upstream of the *HKT1* gene is heavily methylated in Col-0 (Baek et al., 2011). Cytosine methylation in the small RNA region was removed in the *met1-3* mutant, coinciding with higher *HKT1* expression and increased salt sensitivity (Baek et al., 2011). These findings correspond with our results and suggest that *HKT1* expression is regulated by DNA methylation variation in the promoter region affecting plant growth, especially under saline conditions. Further studies are, however needed to prove that methylation variation in the promoter of *HKT1* is causal for the QTL in our studies (Richards, 2006). Comparison of DMRs between the epiRILs and 138 natural accessions of Arabidopsis revealed that approximately 30% of the epiRIL DMRs are present in nature (Cortijo et al., 2014). Although natural DMRs were found around the *HKT1* gene, they did not overlap with the epiRIL DMRs of our study.

An important observation in our study was the epistatic interaction between the two loci on chr 4 and chr 5 for leaf area, detected solely under saline conditions. Although interactions among epigenetic features have not been reported in earlier studies, our results clearly indicate that the regulation of complex traits may depend on the methylation status at multiple loci. DNA hypomethylation at one locus may for instance lead to the enhanced expression of a transcription factor whose functioning depends on the DNA methylation status of an unlinked target locus. The effect signs of both QTLs and their interaction are positive indicating that methylation at the two interacting loci increased growth under saline conditions.

DNA hypomethylation amplifies phenotypic plasticity

In many cases temporary adaptation to stressful conditions is beneficial for plants in fluctuating environments (Rando and Verstrepen, 2007). DNA mutations are irreversible and might thus be contra productive in such environments, whereas epigenetic modifications could be rapidly induced and reversed. Phenotypic plasticity, or the ability of a species to display different phenotypes according to variation in the environment, is, therefore, hypothesized to be (partly) controlled via epigenetic means (Schlichting, 1986; Mirouze and Paszkowski, 2011). This is supported by the differentially increased phenotypic plasticity after chemically induced DNA hypomethylation in several Arabidopsis accessions, suggesting that genotypes and epigenotypes may interact to define plasticity (Bossdorf et al., 2010). In the present study, phenotypic plasticity was greatly enhanced in the majority of epiRILs, implying that variation in DNA methylation profiles contributes substantially to plastic responses in changing conditions. We would like to emphasize here that the variation in plasticity is most likely due to stable heritable variation, and not due to plastic

de novo variation in the epiRILs (Richards et al., 2010). Similar results were obtained in drought and nutrient stress experiments, in which phenotypic plasticity was much higher in epiRILs than in the Col-0 wild-type and highly heritable (Zhang et al., 2013). In our study, large variation was detected in the plasticity response of the epiRILs to moderate salinity and three QTLs were mapped related to phenotypic plasticity (PP). Two out of three QTLs coincided with the QTL intervals of a previous study (Cortijo et al., 2014) indicating that part of the variation can be explained by epigenetic modifications. All PP QTLs coincided with QTLs explaining variation in the same morphological traits under control conditions. This indicates that the regulatory gene(s) underlying the QTLs are sensitive to variation in the environment and that modification of methylation profiles determines to some extent plasticity (Lacaze et al., 2009). In rice, a mutation in a gene leading to increased DNA methylation on repetitive sequences and decreased histone acetylation resulted in high expression variation in different environments, illustrating the regulation of phenotypic plasticity through epigenetic processes (Zhang et al., 2012). The regulation of phenotypic plasticity is thus most likely controlled through a complex regulatory network of epigenetic and genetic factors, depending on environment and development.

Besides environmental plasticity, within-line variation (Sangster et al., 2008) under both conditions was surveyed for epigenetic regulation. The level of within-variation among epiRILs was significantly associated with certain genomic regions. Most of the trait variation QTLs did not overlap with the trait value QTLs, which indicates that different loci explain the variation within and between lines. For relative growth rate and total plant height, no QTLs were detected under control conditions, but QTLs were detected explaining differences in the level of variation within lines. This suggests that the biological variation or developmental stability within lines was higher than the epigenetic variation between lines but that part of the within-line variation is controlled through epigenetics.

In conclusion, the majority of plasticity and stability QTLs showed negative effect signs suggesting that DNA hypomethylation increases environmental sensitivity. In many GWAS and QTL analyses, high variation is often observed between replicates of isogenic lines which could be due to subtle environmental differences. As outlined in this study the differences in within-line variation detected in such genetic resources might be due to epigenetic components that control the level of susceptibility of plants to small changes in the environment.

Material and methods

Plant growing conditions and trait descriptions

Seeds from 99 epiRILs and their parents, Col-0 and *ddm1-2*, all in the Arabidopsis thaliana Col-0 genetic background were sown on filter paper with demi water and stratified at 4˚C

in darkness for 5 d. Subsequently, seeds were transferred to a climate room (16 h light, 24˚C) to induce seed germination for 42 h. Seventeen replicates of each epiRIL and parental line were completely randomized transplanted to wet Rockwool blocks of 4 x 4 cm under both control and saline conditions (different flooding tables in same chamber) in a climate chamber (16 h light, 125 µmol m^{-2} s^{-1}, 70% RH, 20/18˚C day/night cycle). All plants were watered every morning for 5 min with 1/1000 Hyponex solution (Hyponex, Osaka, Japan) supplemented with (salt) or without (control) 25mM NaCl. Plants were photographed from above each hour for the entire growth period (until leaves started to overlap) to analyse leaf area after 20 days (LA20) and relative growth rate. Relative growth rate was calculated as RGR = ln(LA20)-ln(LA17)/d where LA20 is leaf area after 20 days, LA17 is leaf are after 17 days and d is the number of days between the two time points. At 28 days after germination, the first plants started to flower and flowering time (FT) was recorded for five pre-defined replicates out of the seventeen. Two weeks after flowering, main stem branching (MSB), rosette branching (RB), plant height at 1st silique (PH1S), total plant height (TPH) and average internode length (AIL) were measured for these five replicates.

Descriptive statistics

Spearman's rho correlation coefficient was determined using SPSS 21 using a two-tailed significance test. Boxplots were made using EXCEL 2010 based on the minimum (phenotypic value>first quartile-1.5*IQR), first quartile, median, third quartile and maximum (phenotypic value<third quartile+1.5*IQR). The interquartile range (IQR) is the difference between the upper (third quartile) and lower quartiles (first quartile). Suspected outliers were classified as phenotypic values above the minimum and maximum. Phenotypic plasticity was calculated as the absolute difference in means between the two conditions. Coefficient of variation (CV_G) was calculated as $\sqrt{(V_G)}/\overline{X}*100\%$; Broad sense heritability (H^2) was calculated as $V_G/(V_G+V_E)$ where V_G is genetic variation, V_E is environmental variation and $\overline{X}$ is the population average. Reaction norm plots were made using EXCEL 2010 based on the phenotypic values from neutral and saline conditions.

MQM mapping

QTL mapping was performed with multiple QTL mapping (MQM) implemented in the R/QTL software (Arends et al., 2010; Joosen et al., 2012). Co-factors were assigned to 42 out of the 126 markers based on the genetic map position and preliminary composite interval mapping (CIM) on the data. Backward elimination was used to remove cofactors that did not contribute to the fit of the model. MQM mapping was performed on each trait and each treatment separately and the results were compared to standard interval mapping, using Haley Knott regression (Haley and Knott, 1992). Thousand random permutations were generated for each phenotype to determine the LOD significance threshold with $\alpha = 0.05$ as the genome-wide type I error level. For Table 2 and Figure 4, the LOD threshold was determined for both neutral and saline conditions and the highest LOD threshold of both conditions was used for the significance determination. The explained variance per QTL

was calculated as EV (%) = $(1\text{-}10^{(-2*LOD/n)})*100$ where LOD is the LOD score for the particular phenotype and n is the number of epiRILs (R/QTL FAQ).

Calculation of methylation scores around *HKT1*

Probe-level methylation data was obtained from the MeDIP tiling arrays that were available for 123 epiRILs and the founder parents. The methylation calls were previously determined for each probe on these arrays using a Hidden Markov Model (Colome-Tatche et al., 2012). Based on these results, posterior probability for probe *i* to be unmethylated or methylated was calculated by $post(P_i = U)$ and $post(P_i = M)$, respectively. Using this, the methylation level of probe I was defines as $ML - post(P_i = U)*(-1) + post(P_i = M)*1$ (For further details, see Cortijo et al., 2014).

Calculation of expression differences between parental lines around *HKT1*

Whole-genome expression profiling was performed using a custom NimbleGen tiling array, as previously described (Pontier et al., 2012; Cortijo et al., 2014). DNA samples from the two parental lines were reverse-transcribed, differentially labelled and then co-hybridized on the tiling array in a dye-swap design. The resulting data were log transformed and averaged over the two experiments. Quantile normalization was applied to the averaged WT and *ddm1-2* data to bring both data sets to a common scale. Expression differences in the *HKT1* region were subsequently calculated as $log_2(ddm1\text{-}2)\text{-}log_2(WT)$. The log_2 expression difference of each annotation unit was defined as the probe of the unit with the maximum log_2 expression difference. In case of gene promoters the unit was given the maximum log_2 expression difference of the probes of the corresponding gene body (Cortijo et al., 2014).

Two-dimensional two-QTL genome scans

Two-QTL genome scans were performed using the scantwo function in the R/QTL software (Broman and Sen, 2009; Manichaikul et al., 2009; Arends et al., 2010). The output gives the results for five different QTL models: LOD_f, LOD_a, LOD_{fv1}, LOD_{av1} and LOD_i. The QTL positions used for the models can differ between the interaction models (LOD_f, LOD_{fv1} and LOD_i) and the additive models (LOD_a, LOD_{av1}). In the full model, LOD_f, main effects and interaction effects are included, whereas in LODa only the main effects are given. The interaction model, LOD_i, is the difference between LOD_f and LOD_a ($LOD_i = LOD_f - LOD_a$) testing for the significance of epistatic interactions. LOD_{fv1} compares the full model to the maximum single QTL model ($LOD_{fv1} = LOD_f - LOD_{QTLmax}$ in which LOD_{QTLmax} is the highest LOD of the two QTL (loci)) including interactions between the QTLs, whereas LOD_{av1} compares the additive model to the largest single-QTL model ($LOD_{av1} = LOD_a - LOD_{QTLmax}$) excluding interactions. LODs are found significant above an arbitrary threshold based on (Broman and Sen, 2009): (LOD_f, LODfv1, LOD_a, LOD_{av1}, LOD_i) = (6.0, 5.0, 5.0, 2.5, 4.0).

Acknowledgements

This research was supported by the Centre for Biosystem Genomics (CBSG), The Netherlands.

References

Arends D, Prins P, Jansen RC, Broman KW (2010) R/qtl: high-throughput multiple QTL mapping. Bioinformatics 26: 2990-2992

Baek D, Jiang J, Chung JS, Wang B, Chen J, Xin Z, Shi H (2011) Regulated AtHKT1 gene expression by a distal enhancer element and DNA methylation in the promoter plays an important role in salt tolerance. Plant Cell Physiol 52: 149-161

Bandaranayake CK, Koumproglou R, Wang XY, Wilkes T, et al (2004) QTL analysis of morphological and developmental traits in the Ler x Cvi population of Arabidopsis thaliana - QTL analysis in Arabidopsis. Euphytica 137: 361-371

Baubec T, Dinh HQ, Pecinka A, Rakic B, et al (2010) Cooperation of Multiple Chromatin Modifications Can Generate Unanticipated Stability of Epigenetic States in Arabidopsis. The Plant cell 22: 34-47

Baxter I, Brazelton JN, Yu D, Huang YS, et al (2010) A coastal cline in sodium accumulation in Arabidopsis thaliana is driven by natural variation of the sodium transporter AtHKT1;1. PLoS Genet 6: e1001193

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

Bilichak A, Ilnystkyy Y, Hollunder J, Kovalchuk I (2012) The Progeny of Arabidopsis thaliana Plants Exposed to Salt Exhibit Changes in DNA Methylation, Histone Modifications and Gene Expression. Plos One 7

Bossdorf O, Arcuri D, Richards CL, Pigliucci M (2010) Experimental alteration of DNA methylation affects the phenotypic plasticity of ecologically relevant traits in Arabidopsis thaliana. Evolutionary Ecology 24: 541-553

Broman KW, Sen S (2009) Two-dimensional, two-QTL scans. In A Guide to QTL Mapping with R/qtl, pp 213-239

Chinnusamy V, Zhu JK (2009) Epigenetic regulation of stress responses in plants. Curr Opin Plant Biol 12: 133-139

Choi CS, Sano H (2007) Abiotic-stress induces demethylation and transcriptional activation of a gene encoding a glycerophosphodiesterase-like protein in tobacco plants. Mol Genet Genomics 277: 589-600

Clauss MJ, Aarssen LW (1994) Phenotypic plasticity of size-fecundity relationships in Arabidopsis thaliana. Journal of Ecology 82: 447-455

Colome-Tatche M, Cortijo S, Wardenaar R, Morgado L, et al (2012) Features of the Arabidopsis recombination landscape resulting from the combined loss of sequence variation and DNA methylation. Proceedings of the National Academy of Sciences of the United States of America 109: 16240-16245

Cortijo S, Wardenaar R, Colomé-Tatché M, Gilly A, et al (2014) Mapping the Epigenetic Basis of Complex Traits. Science 343: 1145-1148

Davenport RJ, Munoz-Mayor A, Jha D, Essah PA, et al (2007) The Na+ transporter AtHKT1;1 controls retrieval of Na+ from the xylem in Arabidopsis. Plant Cell and Environment 30: 497-507

Eichten SR, Briskine R, Song J, Li Q, et al (2013) Epigenetic and Genetic Influences on DNA Methylation Variation in Maize Populations. The Plant cell 25: 2783-2797

Grativol C, Hemerly AS, Ferreira PC (2012) Genetic and epigenetic regulation of stress responses in natural plant populations. Biochim Biophys Acta 1819: 176-185

Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in crosses using flank marker. Heredity 69: 315-324

Hashida SN, Uchiyama T, Martin C, Kishima Y, et al (2006) The temperature-dependent change in methylation of the Antirrhinum transposon Tam3 is controlled by the activity of its transposase. Plant Cell 18: 104-118

Hauben M, Haesendonckx B, Standaert E, Van Der Kelen K, et al (2009) Energy use efficiency is characterized by an epigenetic component that can be directed through artificial selection to increase yield. Proceedings of the National Academy of Sciences of the United States of America 106: 20109-20114

Johannes F, Colot V, Jansen RC (2008) OPINION Epigenome dynamics: a quantitative genetics perspective. Nature Reviews Genetics 9: 883-890

Johannes F, Porcher E, Teixeira FK, Saliba-Colombani V, et al (2009) Assessing the Impact of Transgenerational Epigenetic Variation on Complex Traits. PLoS genetics 5

Joosen RVL, Arends D, Willems LAJ, Ligterink W, et al (2012) Visualizing the Genetic Landscape of Arabidopsis Seed Performance. Plant Physiology 158: 570-589

Karan R, DeLeon T, Biradar H, Subudhi PK (2012) Salt stress induced variation in DNA methylation pattern and its influence on gene expression in contrasting rice genotypes. PLoS One 7: e40203

Keurentjes JJB, Bentsink L, Alonso-Blanco C, Hanhart CJ, et al (2007) Development of a near-isogenic line population of Arabidopsis thaliana and comparison of mapping power with a recombinant inbred line population. Genetics 175: 891-905

Kliebenstein DJ, Gershenzon J, Mitchell-Olds T (2001) Comparative quantitative trait loci mapping of aliphatic, indolic and benzylic glucosinolate production in Arabidopsis thaliana leaves and seeds. Genetics 159: 359-370

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

Lacaze X, Hayes PM, Korol A (2009) Genetics of phenotypic plasticity: QTL analysis in barley, Hordeum vulgare. Heredity (Edinb) 102: 163-173

Latzel V, Zhang Y, Moritz KK, Fischer M, et al (2012) Epigenetic variation in plant responses to defence hormones. Annals of Botany 110: 1423-1428

Lin HX, Zhu MZ, Yano M, Gao JP, et al (2004) QTLs for Na+ and K+ uptake of the shoots and roots controlling rice salt tolerance. Theoretical and Applied Genetics 108: 253-260

Manichaikul A, Moon JY, Sen S, Yandell BS, et al (2009) A Model Selection Approach for the Identification of Quantitative Trait Loci in Experimental Crosses, Allowing Epistasis. Genetics 181: 1077-1086

Mirouze M, Paszkowski J (2011) Epigenetic contribution to stress adaptation in plants. Curr Opin Plant Biol 14: 267-274

Pontier D, Picart C, Roudier F, Garcia D, et al (2012) NERD, a Plant-Specific GW Protein, Defines an Additional RNAi-Dependent Chromatin-Based Pathway in Arabidopsis. Molecular cell 48: 121-132

Rando OJ, Verstrepen KJ (2007) Timescales of genetic and epigenetic inheritance. Cell 128: 655-668

Rapp RA, Wendel JF (2005) Epigenetics and plant evolution. New Phytologist 168: 81-91

Reinders J, Wulff BBH, Mirouze M, Mari-Ordonez A, et al(2009) Compromised stability of DNA methylation and transposon immobilization in mosaic Arabidopsis epigenomes. Genes & Development 23: 939-950

Ren ZH, Gao JP, Li LG, Cai XL, et al (2005) A rice quantitative trait locus for salt tolerance encodes a sodium transporter. Nature Genetics 37: 1141-1146

Richards CL, Bossdorf O, Verhoeven KJF (2010) Understanding natural epigenetic variation. New Phytologist 187: 562-564

Richards EJ (2006) Opinion - Inherited epigenetic variation - revisiting soft inheritance. Nature Reviews Genetics 7: 395-U392

Sangster TA, Salathia N, Undurraga S, Milo R, et al (2008) HSP90 affects the expression of genetic variation and developmental stability in quantitative traits. Proceedings of the National Academy of Sciences of the United States of America 105: 2963-2968

Schlichting CD (1986) The evolution of phenotypic plasticity. Annual Review of Ecology and Systematics 17: 667-693

Schmitz RJ, He Y, Valdes-Lopez O, Khan SM, Joshi T, et al (2013) Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population. Genome Res 23: 1663-1674

Schmitz RJ, Schultz MD, Urich MA, Nery JR, et al (2013) Patterns of population epigenomic diversity. Nature 495: 193-198

Springer NM (2013) Epigenetics and crop improvement. Trends Genet 29: 241-247

Steward N, Ito M, Yamaguchi Y, Koizumi N, et al (2002) Periodic DNA methylation in maize nucleosomes and demethylation by environmental stress. J Biol Chem 277: 37741-37746

Tsukahara S, Kobayashi A, Kawabe A, Mathieu O, et al (2009) Bursts of retrotransposition reproduced in Arabidopsis. Nature 461: 423-U125

Ungerer MC, Halldorsdottir SS, Modliszewski JL, Mackay TFC, et al (2002) Quantitative trait loci for inflorescence development in Arabidopsis thaliana. Genetics 160: 1133-1151

Yaish MW, Colasanti J, Rothstein SJ (2011) The role of epigenetic processes in controlling flowering time in plants exposed to stress. J Exp Bot 62: 3727-3735

Yao Y, Bilichak A, Golubov A, Kovalchuk I (2012) Ddm1 plants are sensitive to methyl methane sulfonate and NaCl stresses and are deficient in DNA repair. Plant Cell Rep 31: 1549-1561

Zhang CC, Yuan WY, Zhang QF (2012) RPL1, a gene involved in epigenetic processes regulates phenotypic plasticity in rice. Mol Plant 5: 482-493

Zhang YY, Fischer M, Colot V, Bossdorf O (2013) Epigenetic variation creates potential for evolution of plant phenotypic plasticity. New Phytol 197: 314-322

Zhong L, Xu YH, Wang JB (2010) The effect of 5-azacytidine on wheat seedlings responses to NaCl stress. Biologia Plantarum 54: 753-756

CHAPTER 6

The epigenetic regulation of secondary metabolism in different tissues of *Arabidopsis thaliana*

Authors:

Rik Kooke[1,2,3], Frank Johannes[4], Willem Kruijer[5], Ric C.H. de Vos[3,6,7], Dick Vreugdenhil[1,3], Joost J.B. Keurentjes[2,3]

[1] Laboratory of Plant Physiology, Wageningen University, the Netherlands
[2] Laboratory of Genetics, Wageningen University, the Netherlands
[3] Centre for Biosystems Genomics, Wageningen Campus, the Netherlands
[4] Biometris, Wageningen University, the Netherlands
[5] Plant Research International, Wageningen, the Netherlands
[6] Centre for Biosystems Genomics, Wageningen University, the Netherlands
[7] Netherlands Metabolomics Centre, Leiden University, the Netherlands

Abstract

The high diversity and flexibility of plant secondary metabolism allows plants to live in a constantly changing environment. Although the link between environment and epigenetic regulation of gene expression is well studied, only a few studies deal with the relationship between epigenetics and plant metabolism. In the present study, a subset of the Arabidopsis epiRIL population, consisting of differentially methylated lines derived from a cross between the hypomethylated *ddm1-2* mutant and Col-0 was analyzed for secondary metabolites in both leaves and flowers using an untargeted LC-MS-based metabolomics approach. There was abundant metabolite variation in both leaves and flowers across the lines, showing that epigenetics can add an extra dimension to the regulation of metabolism. Many metabolites showed strong correlations with each other and several QTLs were found for a small subset of the metabolites. The majority of these epigenetically regulated metabolites were identified as glucosinolates and flavonoids, and several candidate genes involved in their biosynthesis and metabolism were assigned. A number of morphological QTLs that were measured in the same population overlapped with metabolic QTLs, and a significant correlation existed between some morphological traits and specific metabolites. Furthermore, different sets of metabolites, including metabolites for which QTLs were found, could accurately predict morphological traits. These results suggest that the loci underlying the QTLs are involved in the regulation of both morphology and secondary metabolism.

Introduction

Plants possess a huge metabolic diversity to enhance survival and reproduction. Metabolites are essential for plant growth and development, for responses to stress and for signalling between organs. Moreover, plants deploy their metabolome to interact with all kinds of organisms for mineral exchange, to attract and repel insects and to compete for resources below -and aboveground with other plants (Allwood et al., 2008; Kegge and Pierik, 2010). This wide diversity of metabolites can be further explored due to recent and on-going innovations in plant metabolomics and can yield interesting insights for metabolic engineering, pharmaceutical research and plant breeding (De Vos et al., 2007; van der Hooft et al., 2013).

Plant metabolism is tightly fine-tuned towards changes in the environment, in order to support plant growth, defense, morphology and reproduction. The metabolome is the final outcome of cellular regulatory processes, and as such it incorporates all molecular, genetic, protein and feedback regulation that is involved in plant development (Kooke and Keurentjes, 2012). The metabolome is thus closest to the phenotype and can be used as a predictor or biomarker for plant growth, stress and development. Furthermore, unraveling the regulation of metabolic pathways might reveal regulators of plant growth and morphology. Most studies have focused on the link between primary metabolism and plant morphology as primary metabolites form most of the building blocks for plant growth (Schauer et al., 2006; Meyer et al., 2007; Sulpice et al., 2010; Sulpice et al., 2013). Plant secondary metabolites, on the other hand, are thought to be more diverse, more tissue -and development-specific and mostly involved with responses to changes in the biotic or abiotic environment (Keurentjes et al., 2006). Therefore, the metabolic profile needs to be highly plastic. Plants that invest more in secondary metabolism are thought to be better adapted to environmental changes, but as a consequence can invest less in growth and morphology (Herms and Mattson, 1992). For these reasons, plant secondary metabolism might have a comparable impact on growth and morphology-related traits as primary metabolism.

Epigenetic inheritance is the study of heritable changes in gene expression or cellular phenotype that are not caused by changes in the underlying DNA sequence, but are induced by variation in chromatin components including DNA methylation, histone modifications and small RNAs (Jablonka and Raz, 2009). The epigenetic chromatin marks are most often associated with the silencing of transposable elements (TEs) and repeat elements. However, when present in *cis*-regulatory regions, epigenetic chromatin marks could also affect gene expression. Loss of methylation is usually reset each generation by small RNAs to protect the genome against deleterious TE activity (Teixeira et al., 2009), but when the maintenance of silencing becomes independent of small RNAs, hypomethylation could lead to stable epi-alleles (Bond and Baulcombe, 2014). And such epi-alleles can be successfully passed on to future generations, and can thus be of evolutionary significance (Rapp and Wendel, 2005).

So far, research into epigenetic regulation of plant metabolism is limited. In the ice plant, *Mesembryanthemum crystallinum*, hypermethylation was observed after salt stress, which caused a shift from C3-photosynthesis to CAM metabolism (Dyachenko et al., 2006). Furthermore, carotenoid biosynthesis and lutein content in Arabidopsis were found to be under strict control of SDG8, a chromatin modifying methyltransferase enzyme (Cazzonelli et al., 2009). Other studies reported convincing evidence for epigenetic roles in proline, folate and myo-inositol biosynthesis (Zhang et al., 2012; Latrasse et al., 2013; Zhang et al., 2013). In view of the strong interaction of both epigenetics and secondary metabolism with the environment, a strong link between epigenetics and secondary metabolism may be expected. Although not yet established in plants, such a link between secondary metabolism and epigenetics has been reported in other organisms. In *Aspergillus nidulans*, mutants of histone de-acetylases showed increased gene expression and elevated levels of secondary metabolites (Shwab et al., 2007). Moreover, a strong connection was found between secondary metabolism and light-regulated development, both regulated via epigenetic mechanisms in the same organism (Bayram et al., 2008).

Recently, an epigenetic Recombinant Inbred Line (epiRIL) population of *Arabidopsis thaliana* was created through crossing of wild-type Col-0 with *ddm1-2*, a mutant impaired in the methylation of transposable elements and repeat elements (Johannes et al., 2009). Subsequent back-crossing to Col-0 and repeated self-fertilization created a population of genetically identical lines with differentially methylated chromosomes in the isogenic Col-0 background. This population can be studied in multiple experiments and environments, and could therefore be suitable for studying the epigenetic regulation of secondary metabolism.

In the present study, 96 epiRILs and their parents, wild-type Col-0 and *ddm1-2*, were compared for their secondary metabolite profiles using untargeted LC-QTOF-MS of both leaf and flower tissues. The variation in the relative abundance of all metabolites detected was subsequently associated, through QTL analyses, to epigenetic variation in DNA-methylation, and QTLs could be assigned to different chromosomal regions. The metabolic data were also compared to morphological data on growth, flowering, plant height and branching. Some of the metabolic QTLs co-located with the morphological QTLs, and different sets of metabolites could be used to predict the different morphological phenotypes.

Results

LC-MS analysis of leaf and flower tissue

To evaluate the effect of epigenetic variation on plant secondary metabolism, aqueous-methanol extracts of rosette leaves and flower heads from 96 epiRILs and the parents of the population, Col-0 and *ddm1-2*, were analyzed by an essentially untargeted LC-QTOF MS based metabolomics approach in negative electrospray ionization mode. This method

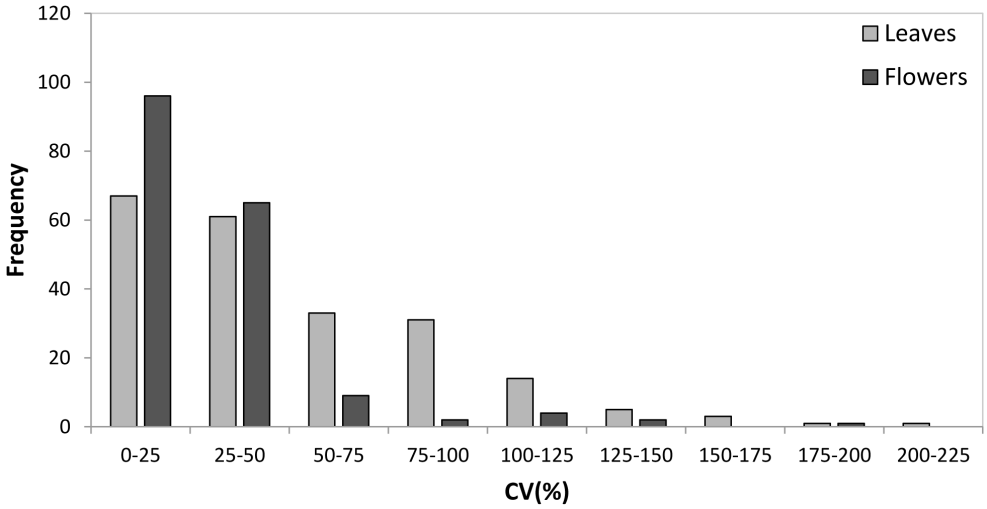


Figure 1. Frequency distribution of coefficient of variation (%) for all 216 leaf (light grey) and 179 flower (dark grey) metabolites detected in the Col-0 x *ddm1-2* epiRIL population using untargeted LC-QTOF-MS-based metabolomics.

is particularly suited for the analysis of semi-polar metabolites including glucosinolates, hydroxycinnamates, flavonoids and various other phytochemicals (De Vos et al., 2007; van der Hooft et al., 2012). In both tissues, qualitative and quantitative variation in metabolite accumulation could be observed among the epiRILs. In the leaves, 8955 reproducible mass signals corresponding to 216 reconstructed metabolites (mass clusters) were retrieved, using a Metalign- and MSClust-based untargeted data processing workflow (Tikunov et al., 2012). The average coefficient of variation of the leaf metabolites was 52%, ranging from 2 to 206% (Figure 1). Comparison of leaf metabolites between the parents showed that nine metabolites were only detected in Col-0, while sixteen were uniquely detected in *ddm1-2*. In the flowers, 6738 mass signals were extracted corresponding to 179 metabolites with an average coefficient of variation of 29% ranging from 2% to 191% (Figure 1). Fifteen metabolites were only detected in Col-0 flowers, while nine metabolites were only detected in *ddm1-2* flowers. The qualitative differences between the parental lines were in most cases also observed in the segregating epiRILs, both in leaves and in flowers.

Strong correlations between metabolites were detected in both tissues but much weaker correlations occurred between metabolites from the different tissues. Although the total number of correlated metabolites was quite similar between leaves and flowers, the proportion of negative correlations was much higher in leaves than in flowers (43% over 7%, respectively) (Figure 2), suggesting a stronger competition for resources in leaves than in flowers, possibly because of the dual role of leaves as both sink and source tissue. Flowers show a much more coordinated regulation of metabolite accumulation which might be

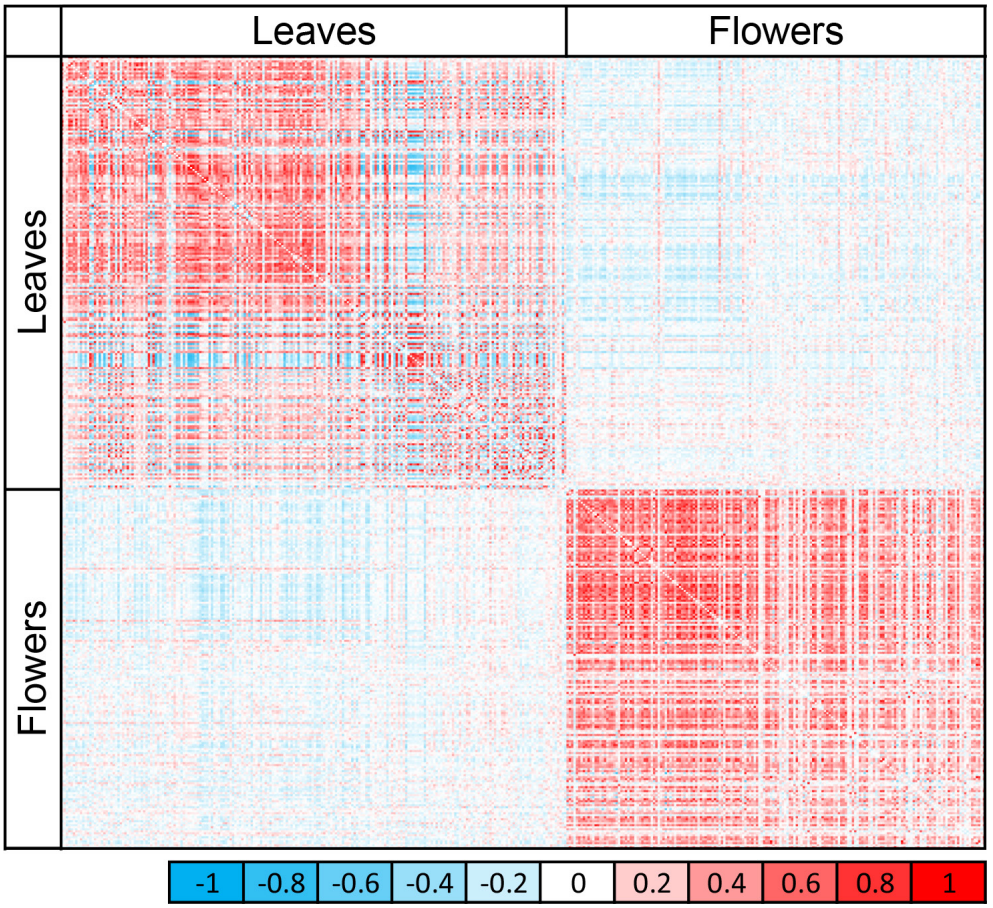


Figure 2. Correlation matrix of detected metabolites in epiRIL population. Pearson correlation between metabolites is indicated by color intensity from -1 (blue) to 1 (red).

caused by the tight developmental control and specific function of this tissue. Although the metabolic analyses for leaf and flower tissue were not performed on the same plants, there were some significant correlations between leaf and flower metabolites (9%, $r > 0.2$, $P < 0.05$) with the majority of them being negative (79%, $r < -0.2$, $P < 0.05$) (Figure 2). This illustrates the metabolic separation in tissue types and their functionally different roles, demanding distinct phytochemical profiles. The wide range of variation in metabolites between the wild type and demethylated parent of the population as well as between population individuals suggests that the methylation status is important for tissue-specific metabolic control.

QTL analysis for leaf and flower metabolites

To get a better understanding of the regulation of plant metabolism within the epiRIL

population, QTL analysis was performed on all metabolites using a genetic map based on differentially methylated regions (DMRs) as physical markers within the epiRIL population (Colome-Tatche et al., 2012). In total, 36 QTLs were identified of which ten were detected in the leaves and twenty-six in the flowers (Figure 3, Table 1). Many of the QTLs overlapped and thirteen different chromosomal regions could be assigned divided over the five chromosomes (Table 1). Four of these regions were shared among leaf and flower metabolites, while two were specific for the leaf metabolites and seven for the flower metabolites. A subset of these metabolites for which QTLs were detected could be annotated, based on matching accurate masses and retention times with an in-house developed Arabidopsis metabolite library. Most flower metabolites were either glucosinolates or flavonoids (Supplemental table 1 and 2). For most of the compounds, QTLs could only be detected in flowers although for quercetin-3-O-hexoside (a flavonoid), D-gluconic acid and a feruloyl malate-coniferyl alcohol conjugate, QTLs could only be detected in leaves. For dihydroxybenzoic acid xyloside III, QTLs were identified in both leaves and flowers (Table 1). For four flavonoids, two QTLs were detected on chr 1 and one on chr 4. The glucosinolate QTLs were detected at different positions in the genome and in several cases different glucosinolates were found to be associated with the same genomic region. Interestingly, the QTLs for dihydroxybenzoic acid xyloside III were associated to different genomic regions in leaves (chr 5) than in flowers (chr 3), which suggests that the accumulation of this metabolite is differently regulated between these two tissues. Significant LOD scores ranged from 2.49 for the unidentified compound number 1537 to 4.81 for 1-methoxy-3-indolmethyl glucosinolate, both found in flowers. Seventeen of the thirty-six QTLs had a negative effect sign, which indicates that *ddm1-2* induced hypomethylation of these QTLs had a positive effect on the metabolite content. In contrast, all glucosinolate QTLs had a positive effect sign, resulting in increased glucosinolate levels in lines with the Col-0 allele. These QTL analyses suggest that DNA methylation plays a prominent role in regulating tissue-specific accumulation of secondary metabolites.

Epigenetic regulation of metabolic pathway genes

The most likely candidates for differential epigenetic regulation of the metabolic pathways are the structural and regulatory genes, and genes encoding modifying enzymes involved in these pathways. All genes involved in metabolism, modification and biosynthesis underlying the QTLs were therefore analysed for differences in methylation status, gene expression and transposon insertion between Col-0 and *ddm1-2* using public databases (http://genomes.mcdb.ucla.edu/AthBSseq/) (Stroud et al., 2013). Candidate genes involved in corresponding metabolic pathways were found for all glucosinolates and flavonoids identified (Table 2). Identical QTLs were found for different glucosinolates, which suggests that the underlying gene regulates the accumulation of glucosinolates upstream in the pathway or through catalysing the same enzymatic side-chain modification. The C4 (and C8) aliphatic glucosinolates were found to be controlled by a QTL located at the end of chr 5, which might be caused by *MYB28* which is essential for the biosynthesis of aliphatic

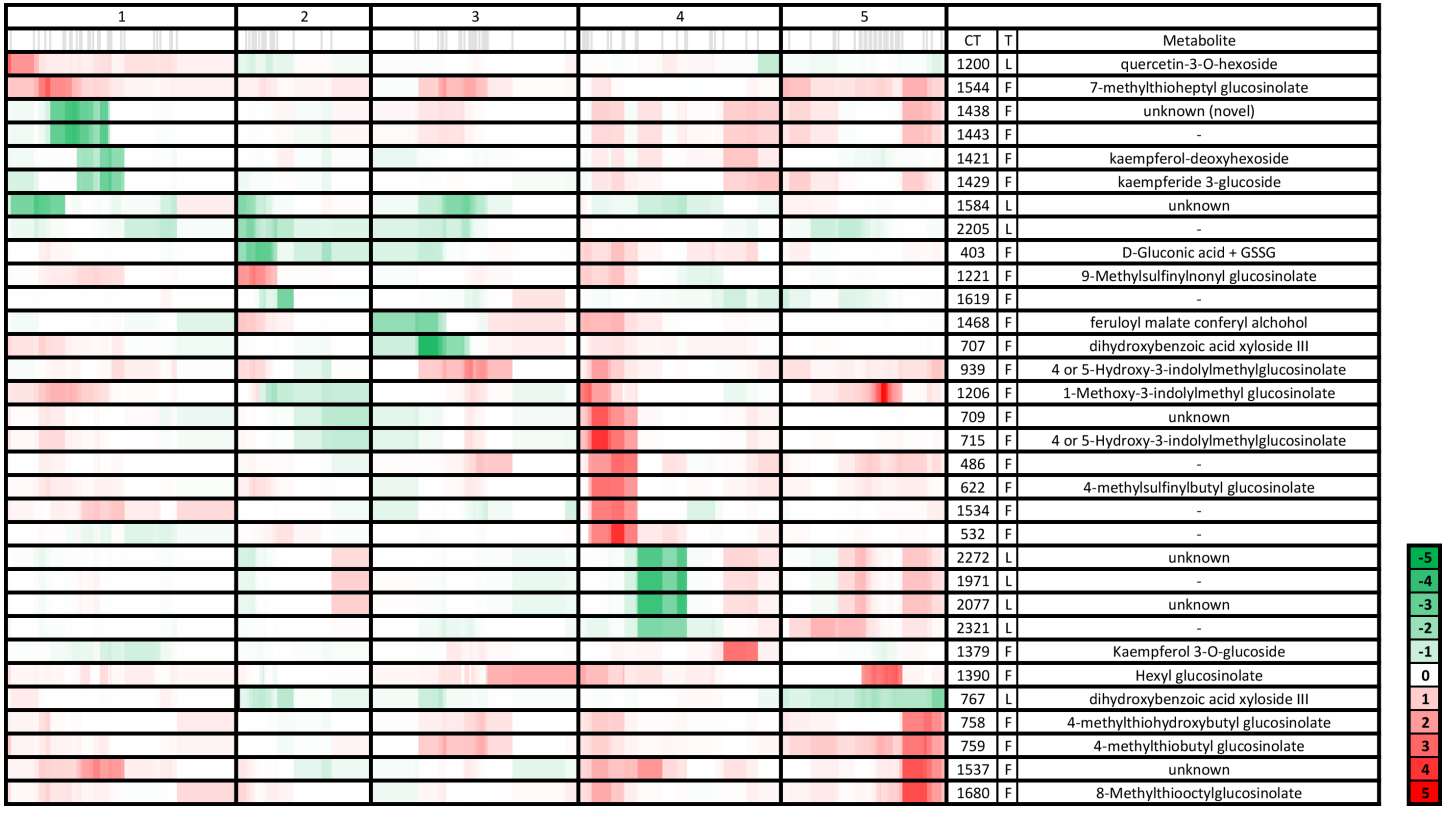


Figure 3. mQTL Heatmap showing the positions of the QTL divided over the 5 chromosomes.
The thin grey lines in the second row indicate the marker positions in cM. CT = centrotype nr, T =tissue, L = leaf, F = flower. The legend on the right indicates the QTL LOD score between -5 and 5. Nr, metabolite number; T, tissue.

Table 1. QTL overview for all metabolic QTLs in leaves (L) and flowers (F) of the epiRIL population. Tissue indicates leaf (L) or flower (F) tissue; QTL indicates the QTL nr; Chr, chromosome; LOD, QTL LOD score.

Metabolite nr	Tissue	QTL	Chr	LOD	Support interval (cM)	Explained variance (%)	Effect	Metabolite
1200	L	1	1	3.6	0-11.7	16.3	0.10	quercetin-3-O-hexoside
1584	L	2	1	3.6	6-26	16.7	-0.28	
1544	F		1	3.2	14-29	14.9	0.06	7-methylthioheptyl glucosinolate
1443	F		1	3.9	20-36	17.6	-0.13	
1438	F		1	3.8	20-36.5	17.2	-0.13	
1537	F	3	1	2.5	29-55	11.7	0.14	
1429	F		1	3.5	38-45.9	16.2	-0.10	kaempferide 3-glucoside
1421	F		1	3.1	37-54	14.5	-0.11	kaempferol-deoxyhexoside
1584	L	4	2	2.5	0-9	11.8	-0.29	
2205	L		2	2.7	0-10.4	12.6	-0.16	
1221	F		2	2.7	0-14	12.8	0.11	9-Methylsulfinylnonyl glucosinolate
403	F		2	3.3	2-15	15.0	-0.06	D-Gluconic acid + GSSG
1619	F	5	2	2.7	15-33	12.7	-0.04	
1468	F	6	3	3.0	22-36	14.0	-0.09	feruloyl malate conferyl alchohol
707	F		3	4.4	29-38	19.6	-0.10	dihydroxybenzoic acid xyloside III
1584	L	7	3	2.9	36-51	13.3	-0.31	
939	F		3	2.7	36-77	12.5	0.12	4 or 5-Hydroxy-3-indolylmethylglucosinolate
1206	F	8	4	3.1	0-14	14.5	0.05	1-Methoxy-3-indolylmethyl glucosinolate
709	F		4	3.4	5-17	15.5	0.09	
715	F		4	4.1	5-18	18.5	0.11	4 or 5-Hydroxy-3-indolylmethylglucosinolate
486	F	9	4	3.2	4-28	14.9	0.07	
622	F		4	2.8	4-29	13.3	0.09	4-methylsulfinylbutyl glucosinolate
1534	F		4	2.8	4.5-29	13.0	0.01	
532	F		4	4.1	12-21	18.6	0.06	
2272	L	10	4	3.6	23-45	16.3	-0.35	
1971	L		4	4.0	24-45	18.2	-0.39	
2077	L		4	4.0	24-45	18.2	-0.39	
1379	F	11	4	2.9	61-81	13.5	0.13	Kaempferol 3-O-glucoside
2321	L	12	5	2.7	24-50	12.8	-0.41	
767	L		5	2.5	31-65.9	11.8	-0.11	dihydroxybenzoic acid xyloside III
1206	F		5	4.8	40-43	21.4	0.06	1-Methoxy-3-indolylmethyl glucosinolate
1390	F		5	3.3	35-59.9	15.2	0.10	Hexyl glucosinolate
1537	F	13	5	3.4	47.4-65	15.7	0.15	
759	F		5	3.1	51-65	14.3	0.15	4-methylthiobutyl glucosinolate
1680	F		5	3.6	50-63.6	16.3	0.10	8-Methylthiooctylglucosinolate
758	F		5	2.9	51-65.9	13.6	0.07	4-methylthiohydroxybutyl glucosinolate

Table 2. List with metabolites and gene candidates within QTL support intervals (SI).
The colors within the locus ID and abbreviation columns indicate whether the gene is differentially methylated (light grey) or differentially expressed (grey) between Col-0 and *ddm1-2*, or whether transposons are located inside the Col-0 gene region (dark grey) (Stroud et al., 2013). Nr, metabolite number; T, tissue; L, leaves; F, flowers; Chr, chromosome.

Nr	T	Chr	SI (cM)	Metabolite	Locus ID	Abbreviation	Information, Details
1200	L	1	0-11.7	quercetin-3-O-hexoside	AT1G06000		Encodes a flavonol-7-O-rhamnosyltransferase
					AT1G13110	CYP71B7	Positive regulation of flovonoid biosynthesis
					AT1G17745	PGDH	Positive regulation of flovonoid biosynthesis, secondary metabolism
1544	F	1	14-29	7-methylthioheptyl glucosinolate	At1G12140	FMO GS-OX5	Glucosinolate biosynthetic process
1584	L	1	6-26		AT1G24100	UGT74B1	Involved in glucosylation for all glucosinolate pathways
1438	F	1	20-36.5		AT1G31230	AK-HSDH I	Aminoacid metabolism, glucosinolate biosynthetic process
1443	F	1	20-36		AT1G32640	MYC2	Indole glucosinolate biosynthetic process
1429	F	1	38-45.9	kaempferide 3-glucoside	AT1G27100		Positive regulation of flavonoid biosynthetic process
1421	F	1	37-54	kaempferol-deoxyhexoside	AT1G30530	UGT78D1	Quercetin 3-O-glucosyltransferase activity, kaempferol glucoside biosynthesis
1537	F	1	29-55		AT1G32640	MYC2	Indole glucosinolate biosynthetic process
					AT1G34790	TT1	Flavonoid biosynthetic process, transparent Testa 1
					AT1G43620	UGT80B1	Flavonoid biosynthetic process, sterol 3-beta-glucosyltransferase
					AT1G50740		Positive regulation of flavonoid biosynthetic process
					AT1G56650	MYB75	Anthocyanin biosynthetic process
1221	F	2	0-14	9-Methylsulfinylnonyl glucosinolate	AT2G04160	AIR3	Glucosinolate biosynthetic process, response to auxin stimulus
403	F	2	2-15	D-Gluconic acid + GSSG	AT2G05380	GRP3S	Glucosinolate biosynthetic process
1584	L	2	0-9		AT2G05520	GRP3	Glucosinolate biosynthetic process, tryptophan catabolic process
2205	L	2	0-10.4				
939	F	3	36-77	4 or 5-Hydroxy-3-indolylmethylglucosinolate	AT3G44320	NIT3	Nitrogen metabolism, glucosinolate catabolic process
1584	L	3	36-51		AT3G49860	BCAT3	Involved in glucosinolate biosynthesis
					AT3G56400	WRKY70	Plant-pathogen interaction, indole glucosinolate biosynthetic process

Table 2 continues.

Nr	T	Chr	SI (cM)	Metabolite	Locus ID	Abbreviation	Information, Details
1206	F	4	0-14	1-Methoxy-3-indolylmethyl glucosinolate	AT4G03050	AOP3	Glucosinolate biosynthesis from homomethionine
715	F	4	5-18	4 or 5-Hydroxy-3-indolylmethylglucosinolate	AT4G03060	AOP2	Glucosinolate biosynthesis from homomethionine
622	F	4	4-29	4-methylsulfinylbutyl glucosinolate	AT4G03070	AOP1.1	Glucosinolate biosynthetic process, Gibberellin inactivation
709	F	4	5-17		AT4G13430	IIL1	Co-expressed with AK-HSDH I, aliphatic glucosinolate biosynthesis
1534	F	4	4.5-29		AT4G13770	REF2	Co-expressed with MAM3, Glucosinolate biosynthesis in many glucosinolate pathways, but not from tryptophan
532	F	4	12-21				
486	F	4	4-28				
1379	F	4	61-81	Kaempferol 3-O-glucoside	AT4G33300	ADR1-L1	Positive regulation of flavonoid biosynthesis
					At4g34135	UGT73B2	flavonol 7-O-glucosyltransferase glucosylates flavonols (kaempferol and quercetin) at the 3-O-position, kaempferol glucoside biosynthesis
					AT4G36500		Positive regulation of flavonoid biosynthesis, Glutathione metabolism
1206	F	5	40-43	1-Methoxy-3-indolylmethyl glucosinolate	AT5G23010	MAM1	Aliphatic glucosinolate biosynthesis
1390	F	5	35-59.9	Hexyl glucosinolate	AT5G23020	MAM3	Aliphatic glucosinolate biosynthesis
767	L	5	31-65.9	dihydroxybenzoic acid xyloside III	AT5G26000	TGG1	The enzyme catalyzes the hydrolysis of glucosinolates into compounds that are toxic to various microbes and herbivores, glucosinolate catabolic process
2321	L	5	24-50		AT5G42650	DDE2	JA biosynthesis, Alpha-linolenic acid biosynthesis, glucosinolate biosynthetic process
					AT5G44070	PCS1	Indole glucosinolate biosynthetic process
759	F	5	51-65	4-methylthiobutyl glucosinolate	AT5G42650	DDE2	JA biosynthesis, Alpha-linolenic acid biosynthesis, glucosinolate biosynthetic process
1680	F	5	50-63.6	8-Methylthiooctylglucosinolate	AT5G44070	PCS1	Indole glucosinolate biosynthetic process
758	F	5	51-65.9	4-methylthiohydroxybutyl glucosinolate	AT5G61420	MYB28	Involved in positive regulation of aliphatic glucosinolate biosynthesis
1537	F	5	47.4-65				

glucosinolates (Beekwilder et al., 2008). Furthermore, *MYB28* was detected previously as candidate gene in a QTL study for aliphatic glucosinolate content and transcript level of genes involved in aliphatic glucosinolate biosynthesis in the Bay x Sha RIL population (Sonderby et al., 2007; Wentzell et al., 2007). There was a small difference in the methylation profile between Col-0 and *ddm1-2* which may affect *MYB28* expression and thus aliphatic glucosinolate content (http://genomes.mcdb.ucla.edu/AthBSseq/) (Stroud et al., 2013). The major QTL for the accumulation of indole glucosinolates was detected on chr 4 and might be representing CYP83A1 monooxygenase, *REF2*, which is involved in aliphatic glucosinolate biosynthesis but when mutated significantly increases indole glucosinolate levels (Hemm, 2002). The indole glucosinolate content was higher in the Col-0 background of the epiRIL population, which would be in agreement with reduced expression of *REF2*. The gene is indeed hypomethylated in *ddm1-2* (http://genomes.mcdb.ucla.edu/AthBSseq/) which could result in increased expression and a reduction in indole glucosinolate content. Moreover, *REF2* also has other functions in secondary metabolism and might thus also be involved in regulation of the unknown metabolites for which the QTLs co-located on chr 4 (Hemm, 2002). Another interesting candidate gene is *UGT74B1* (AT1G24100) which is located in the QTL interval for 7-methylthioheptyl glucosinolate on chr 1 and is involved in glycosylation of glucosinolates (Table 2). This gene was found to have small methylation differences between Col-0 and *ddm1-2* (http://genomes.mcdb.ucla.edu/AthBSseq/). For the flavonoids, an interesting candidate gene was found for kaempferide-3-glucoside and kaempferol-deoxyhexoside, *UGT78D1* (AT1G30530), which is involved in flavonol aglycone biosynthesis (Jones et al., 2003). The 5'end of the gene was found to be differently methylated between Col-0 and *ddm1-2* (http://genomes.mcdb.ucla.edu/AthBSseq/). These findings suggest that a large part of the variation in metabolite levels, detected within the epiRIL population, can be explained by differences in the methylations status. This is particularly true for the highly abundant chemical families of glucosinolates and flavonoids for which likely candidate genes could be assigned.

Correlation between metabolism and morphology

Besides leaf and flower metabolism, a series of morphological traits were analysed for all epiRILs and the parents on different plants grown in the same experimental set up. It was subsequently investigated to what extent the metabolite levels correlate with the morphological traits and whether they may be used as trait markers. To get more insight into these relationships, pairwise correlations were calculated between each metabolite and the morphological traits using Spearman's rank correlation. Two hundred eighty eight significant correlations ($r_s > 0.2$, $P < 0.05$) were detectedd between leaf metabolites and morphological traits (Figure 4). Of these correlations, 155 were negative and 127 were positive. Significant positive correlations were detected between leaf metabolites showing a QTL and flowering time (i.e. time to flower), and mostly negative correlations were detected between the leaf metabolites showing a QTL and leaf area and relative growth rate (Table 5).

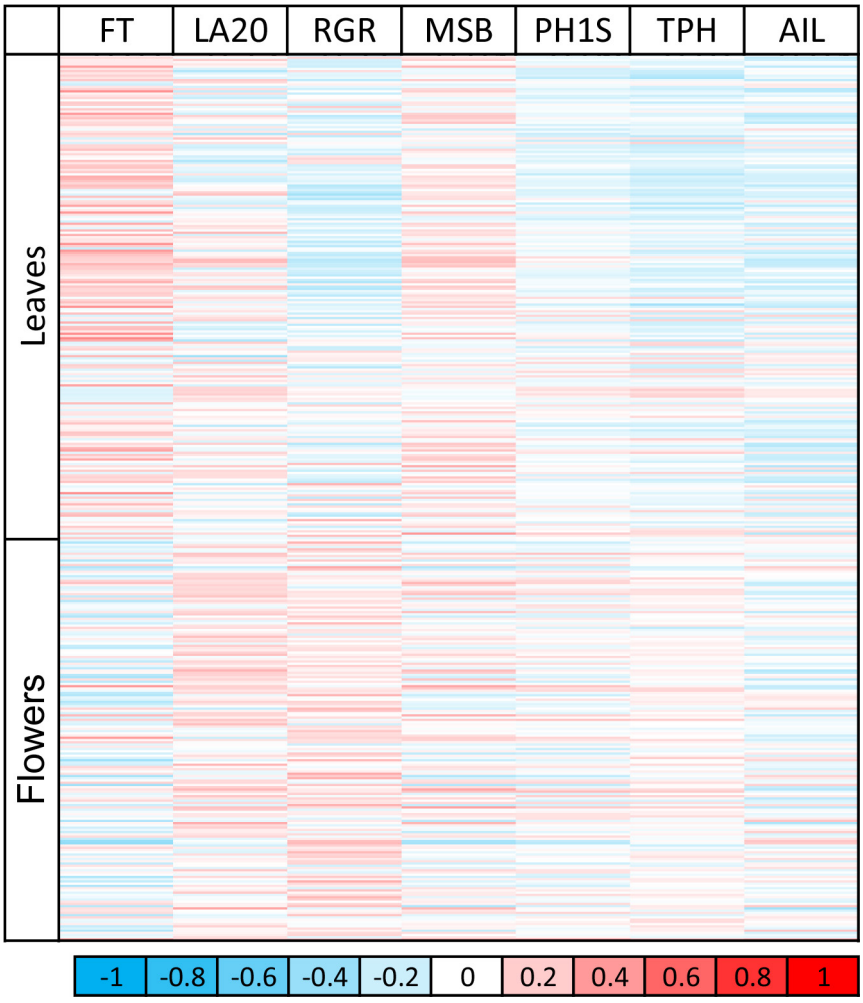


Figure 4. Correlation matrix between morphological traits and leaf and flower metabolites. Pearson correlation between metabolites and morphological traits is indicated by color intensity from -1 (blue) to 1 (red).

When we compared the percentage of metabolites with a QTL that significantly correlated with a morphological trait to the percentage of significant correlations among all metabolites with that morphological trait, we found a significant over-representation of the metabolic QTLs that correlate with leaf area ($P(\chi^2) < 0.05$, Table 3). This suggests that there is a strong relationship between these metabolites and leaf area, and that they are both affected by the underlying epigenetic regulation.

For the flower metabolites, a total of 202 correlations were found of which 115 were

Table 3. Spearman correlation between leaf (L) and flower (F) metabolites and morphological traits. Only metabolites for which a significant QTL was detected were taken into account. *P*(χ^2) gives the chi square *P*-value for the chance that the number of observed correlations does not deviate from the expected number of significant correlations among all metabolites.

Metabolite nr	Tissue	FT	LA20	RGR	MSB	PH1S	TPH	AIL
767	L	-0.10	0.03	-0.16	-0.01	-0.16	-0.21	-0.04
1200	L	0.21	0.10	-0.21	0.17	-0.04	-0.19	-0.17
1584	L	-0.03	-0.24	0.09	-0.13	-0.05	-0.10	0.09
1971	L	-0.04	-0.08	0.20	-0.06	-0.04	-0.01	0.05
2077	L	0.01	-0.25	0.30	0.02	0.00	-0.03	-0.01
2205	L	0.06	-0.08	0.14	0.09	0.12	-0.01	0.06
2272	L	-0.03	-0.24	0.31	-0.01	0.00	-0.02	0.02
2321	L	-0.08	-0.07	0.25	-0.06	0.08	0.09	0.13
P (χ2)	L	0.30	0.01	0.27	*	*	*	*

Metabolite nr	Tissue	FT	LA20	RGR	MSB	PH1S	TPH	AIL
403	F	-0.14	0.08	-0.01	0.00	0.12	0.05	0.09
486	F	0.13	0.17	0.09	0.24	0.17	0.12	-0.18
532	F	0.01	0.07	0.13	0.07	0.08	0.07	-0.07
622	F	0.09	0.16	0.11	0.25	0.12	0.14	-0.25
707	F	0.06	-0.17	0.18	0.05	-0.02	-0.19	-0.06
709	F	0.24	0.13	0.05	0.29	0.17	0.08	-0.23
715	F	0.27	0.16	0.02	0.31	0.17	0.07	-0.25
758	F	0.14	0.22	-0.02	0.20	0.24	0.18	-0.09
759	F	0.18	0.22	-0.05	0.21	0.23	0.20	-0.08
939	F	0.06	0.18	-0.05	0.17	0.03	0.04	-0.21
1206	F	0.40	0.19	-0.05	0.35	0.24	0.04	-0.23
1221	F	0.07	0.11	0.17	0.16	-0.05	0.03	-0.27
1379	F	-0.31	0.15	0.21	-0.18	-0.17	0.01	0.08
1390	F	0.10	0.28	0.08	0.35	0.14	0.15	-0.28
1421	F	-0.43	-0.07	0.27	-0.35	-0.28	-0.04	0.20
1429	F	-0.40	-0.01	0.25	-0.33	-0.23	0.03	0.22
1438	F	-0.30	-0.07	0.30	-0.21	-0.21	0.03	0.07
1443	F	-0.31	-0.07	0.29	-0.22	-0.20	0.03	0.08
1468	F	0.11	-0.08	0.33	0.05	-0.02	-0.05	-0.10
1534	F	-0.15	0.00	0.04	0.01	-0.06	0.01	-0.10
1537	F	0.12	0.36	-0.10	0.29	0.02	0.07	-0.36
1544	F	-0.06	0.22	-0.04	0.03	0.19	0.17	0.08
1619	F	-0.16	0.19	-0.11	0.01	-0.04	-0.02	-0.03
1680	F	-0.09	0.19	0.06	0.06	0.10	0.17	-0.02
P (χ2)	F	0.26	0.10	0.11	0.00	0.00	0.33	0.00

positive and 87 were negative. Interestingly, flowering time showed a high number of positive correlations with leaf metabolites but a high number of negative correlations with flower metabolites. Although solely based on the correlations, this could suggest that early flowering plants grow fast (small negative correlation between flowering time and leaf area after 20 days) and invest more in flower metabolism, while late flowering plants, grow slower and invest more in leaf metabolism. This point is strengthened by the high number of negative correlations between relative growth rate and leaf metabolites, and the high number of positive correlations between relative growth rate and leaf area with flower metabolites.

Both positive and negative correlations were observed between flower metabolites for which a QTL was detected and morphological traits. The highest negative correlation was found between flowering time and kaempferol-deoxyhexoside with a linear correlation of -0.43, explaining roughly 18% of the total variation for flowering time (Table 3). The highest positive correlation (r = 0.4) was found between flowering time and 1-methoxy-3-indolmethyl glucosinolate, explaining roughly 16% of the total variation in flowering time (Table 3). When we compared the percentage of metabolites with a QTL that significantly

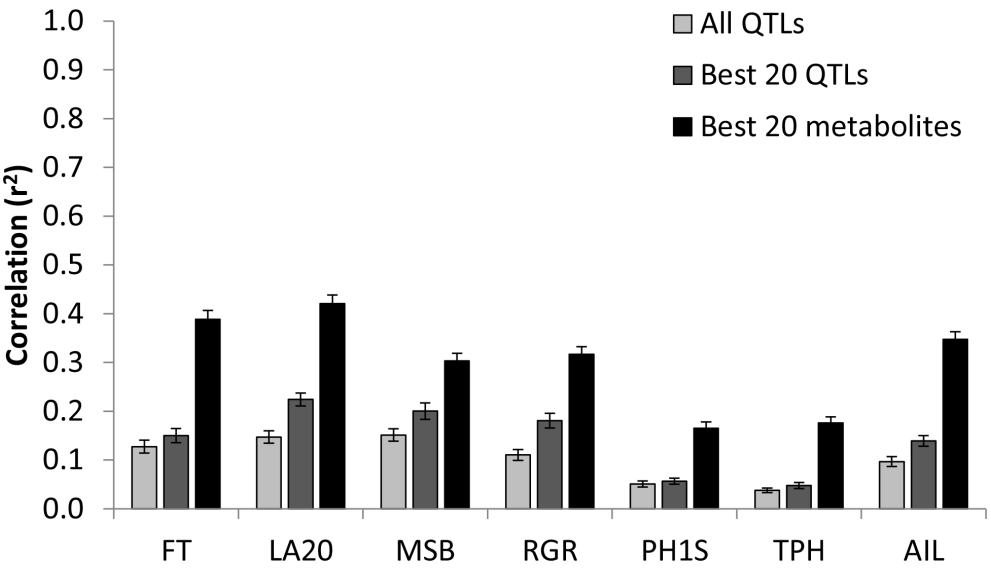


Figure 5. Metabolic prediction with Random Forest for morphological traits showing the correlation (r^2) between the predicted and observed values in the test set (20% of all epiRILs) based on the training set (80% of all epiRILs) The light grey bar represents the metabolic predictions using all metabolites with a significant QTL, the dark grey bar represents the metabolic prediction for the 20 best predictive metabolites with a significant QTL and the black bar represents the metabolic prediction for the 20 best predictive metabolites from all measured metabolites.

correlated with a morphological trait to the percentage of significant correlations among all metabolites with that morphological trait, we found a significant over-representation of the metabolic QTLs that correlate with main stem branching, plant height at 1st silique and average internode length ($P(\chi^2) < 0.05$). This suggests that these metabolites are strongly connected to the morphological traits and that they might be regulated by the same epigenetic mechanisms.

Table 4. The importance (higher value corresponds to higher importance) of the metabolic variables in Random Forest estimated in the training set and used to estimate the correlation in the validation set for all metabolites with a QTL.

Metabolite nr	Tissue	Metabolite	FT	LA20	RGR	MSB	PH1S	TPH	AIL
767	L	dihydroxybenzoic acid xyloside III	227	191	396	146	146	249	131
1200	L	quercetin-3-O-hexoside	176	146	225	200	116	150	138
1584	L		121	277	129	158	114	152	148
1971	L		126	130	121	138	140	169	144
2077	L		128	403	204	166	294	183	173
2205	L		116	198	158	211	165	207	304
2272	L		110	248	218	169	349	255	142
2321	L		83	134	216	71	105	135	219
403	F	D-Gluconic acid + GSSG	235	124	102	217	226	417	188
486	F		122	135	137	152	379	307	162
532	F		126	119	152	92	140	136	118
622	F	4-methylsulfinylbutyl glucosinolate	85	123	100	187	160	162	213
707	F	dihydroxybenzoic acid xyloside III	115	255	94	109	243	443	137
709	F		222	139	174	245	179	169	214
715	F	4 or 5-Hydroxy-3-indolylmethylglucosinolate	226	149	155	306	220	182	224
758	F	4-methylthiohydroxybutyl glucosinolate	89	136	159	113	185	183	134
759	F	4-methylthiobutyl glucosinolate	128	211	156	149	299	333	135
939	F	4 or 5-Hydroxy-3-indolylmethylglucosinolate	152	201	172	172	146	213	177
1206	F	1-Methoxy-3-indolylmethyl glucosinolate	427	199	139	435	214	230	436
1221	F	9-Methylsulfinylnonyl glucosinolate	91	109	149	96	171	127	391
1379	F	Kaempferol 3-O-glucoside	266	194	146	253	267	136	158
1390	F	Hexyl glucosinolate	241	271	125	462	220	318	201
1421	F	kaempferol-deoxyhexoside	750	293	582	481	360	119	295
1429	F	kaempferide 3-glucoside	415	168	156	281	170	79	430
1438	F		273	159	273	257	256	129	146
1443	F		399	175	197	308	205	92	132
1468	F	feruloyl malate conferyl alchohol	84	256	869	77	170	149	97
1534	F		283	286	87	147	144	209	153
1537	F		231	465	218	293	174	210	457
1544	F	7-methylthioheptyl glucosinolate	120	266	104	98	230	222	212
1619	F		128	249	143	106	139	191	89
1680	F	8-Methylthiooctylglucosinolate	75	123	121	92	122	151	185

Predictive power of metabolites for morphology

To further investigate the relationship between secondary metabolism and morphology, the predictive power of epigenetically regulated metabolite content on the morphological traits was analysed with Random Forest, a multivariate regression method that can be used for metabolic prediction (Carreno-Quintero et al., 2012). First, we only used the metabolites for which a QTL was detected in Random Forest. On average, the metabolites for which a QTL was detected could explain 11% of the variance of the morphological traits, with a minimum of 4% for total plant height and a maximum of 15% for main stem branching (Figure 5). These

Table 5. The importance (higher value corresponds to higher importance) of the metabolic variables in Random Forest estimated in the training set and used to estimate the correlation in the validation set for the 20 best predictive metabolites with a QTL.

Metabolite nr	Tissue	Metabolite	FT	LA20	RGR	MSB	PH1S	TPH	AIL
767	L	dihydroxybenzoic acid xyloside III	279	303	468			331	
1200	L	quercetin-3-O-hexoside	240		306	271			
1584	L			363		211			204
1971	L		181					242	
2077	L		185	468	271	228	383	248	258
2205	L			263	229	276		273	384
2272	L			331	292	228	435	323	
2321	L				287				311
403	F	D-Gluconic acid + GSSG	322			298	296	508	264
486	F					218	472	404	245
532	F		173		210				
622	F	4-methylsulfinylbutyl glucosinolate				269		250	282
707	F	dihydroxybenzoic acid xyloside III		324			328	537	
709	F		280		244	312	262	238	282
715	F	4 or 5-Hydroxy-3-indolylmethylglucosinolate	290		239	383	314	265	292
758	F	4-methylthiohydroxybutyl glucosinolate			246		247	259	
759	F	4-methylthiobutyl glucosinolate	198	265	238	241	368	415	
939	F	4 or 5-Hydroxy-3-indolylmethylglucosinolate	205	250	244	235		286	252
1206	F	1-Methoxy-3-indolylmethyl glucosinolate	484	293		480	278	292	505
1221	F	9-Methylsulfinylnonyl glucosinolate			198		264		483
1379	F	Kaempferol 3-O-glucoside	308	230	191	290	357		226
1390	F	Hexyl glucosinolate	322	344		556	307	421	272
1421	F	kaempferol-deoxyhexoside	782	369	624	572	431		386
1429	F	kaempferide 3-glucoside	490	232	208	346	234		515
1438	F		363	215	347	314	322		
1443	F		496	272	255	378	276		
1468	F	feruloyl malate conferyl alchohol		390	936		251		
1534	F		313	367				283	219
1537	F		286	559	315	386	238	274	512
1544	F	7-methylthioheptyl glucosinolate		334			299	299	298
1619	F		176	325				265	
1680	F	8-Methylthiooctylglucosinolate							264

values correspond to an average correlation of 0.28 with a minimum correlation of 0.16 and a maximum correlation of 0.35. The variable importance, which signifies the importance of a variable for classifying the data, showed that the metabolites with the highest correlation with a certain morphological trait were also the most important variables in the Random Forest selection (Table 4). As many of the metabolites only contributed marginally to the prediction, it was tested whether the twenty most important metabolites for which a QTL was detected could predict a higher percentage of the variance for the morphological traits. The average explained variance increased to 15% (r = 0.39) with a minimum of 5% (r = 0.17) for total plant height and a maximum of 22% (r = 0.45) for leaf area (Figure 5). The individual importance of each of the metabolites increased (Table 5). Especially, the variable importance of feruloyl malate coniferyl alcohol for relative growth rate (VIP=936) was exceptionally high. Finally, it was investigated to what extent the morphological traits can be predicted by Random Forest by taking the 20 metabolites with the highest importance values out of all metabolites (Table 6). The correlation increased significantly to an average explained variance of 33% (r = 0.58) with a minimum value of 17% (r = 0.36) for plant height at 1st silique and a maximum value of 46% (r = 0.67) for leaf area (Figure 5). The finding that these metabolites explain more of the variance for morphological traits than the twenty metabolites, for which a QTL was detected, might seem counter-intuitive. This is most likely due to the high correlations between the metabolites with a QTL and, therefore, their individual importance in classification is reduced. Interestingly, feruloyl malate coniferyl alchohol was also the best predictor for relative growth rate among all metabolites (Table 6). These analyses show that different sets of metabolites are important for different morphological traits. Unraveling the epigenetic regulation of these metabolites might assist in understanding the epigenetic regulation of morphology.

Overlap between mQTLs and morphological QTLs

To study whether the same genomic regions can explain part of the variation for metabolism and morphology, the QTL profiles of all metabolites and morphological traits were aligned (Figure 6). All morphological QTLs overlapped with different mQTLs and most of the metabolites with overlapping QTLs were also highly correlated with the morphological traits (Figure 6, Tables 3 and 7). For example, leaf area correlated with both the unknown metabolite number 1584 and 7-methylthioheptyl glucosinolate for which a QTL was detected

Table 6. The importance (higher value corresponds to higher importance) of the metabolic variables in Random Forest estimated in the training set and used to estimate the correlation in the validation set for the 20 best metabolites. Given next to the RF importance is the spearman correlation between metabolite and morphological trait. Metabolites for which QTLs were detected in leaf tissue are highlighted in green, in flower tissue in purple. Metabolite numbers are given under morphological trait.

FT	T	RF	Sp
846	L	298	0.40
850	L	323	-0.07
903	L	347	0.06
906	L	274	-0.10
955	L	430	0.39
965	L	488	0.44
1005	L	367	0.42
1025	L	226	-0.20
1158	L	360	0.14
1510	L	238	0.45
1650	L	269	0.36
1852	L	284	0.23
1920	L	258	0.30
2036	L	272	0.14
682	F	456	-0.31
1184	F	427	-0.31
1206	F	339	0.40
1421	F	301	-0.43
1522	F	292	-0.31
1550	F	247	-0.22

LA20	T	RF	Sp
591	L	289	-0.31
611	L	287	-0.30
779	L	586	0.24
983	L	215	0.04
1078	L	330	0.27
1090	L	308	0.26
1132	L	393	0.16
1253	L	275	0.24
1579	L	397	-0.36
1584	L	267	-0.24
2077	L	392	-0.25
698	F	425	0.32
733	F	251	0.30
1335	F	407	-0.22
1400	F	265	-0.06
1421	F	322	-0.07
1537	F	383	0.36
1600	F	290	0.15
1715	F	222	0.28
1806	F	239	0.07

RGR	T	RF	Sp
757	L	387	-0.28
773	L	358	-0.27
779	L	471	-0.35
823	L	287	-0.26
1078	L	294	-0.28
1279	L	312	-0.30
1455	L	245	-0.27
2163	L	284	0.19
2272	L	252	0.31
68	F	259	0.19
99	F	268	0.02
130	F	355	0.00
217	F	285	-0.01
1335	F	295	0.34
1374	F	392	0.30
1421	F	330	0.27
1468	F	687	0.33
1590	F	240	-0.19
1744	F	260	-0.01
1817	F	265	0.07

AIL	T	RF	Sp
801	L	286	0.11
1265	L	273	-0.16
1437	L	483	-0.26
2005	L	298	-0.29
2013	L	255	0.12
2163	L	368	0.13
2300	L	283	0.14
2306	L	365	-0.29
660	F	369	-0.16
682	F	302	0.24
801	F	348	0.19
1206	F	348	-0.23
1221	F	315	-0.27
1335	F	347	0.20
1382	F	372	0.21
1384	F	265	0.01
1386	F	284	0.18
1429	F	305	0.22
1448	F	265	-0.04
1537	F	392	-0.36

MSB	T	RF	Sp
481	L	384	0.16
1078	L	289	0.25
1132	L	352	0.24
1437	L	284	0.21
1455	L	288	0.16
1510	L	269	0.24
2005	L	332	0.33
2306	L	369	0.40
698	F	454	0.29
709	F	221	0.29
715	F	269	0.31
733	F	314	0.26
1206	F	337	0.35
1214	F	335	-0.30
1382	F	333	-0.30
1386	F	332	-0.30
1390	F	339	0.35
1421	F	345	-0.35
1528	F	407	0.37
1537	F	302	0.29

PH1S	T	RF	Sp
102	L	219	-0.18
301	L	265	-0.18
526	L	260	-0.14
591	L	476	-0.15
611	L	384	-0.17
788	L	302	0.05
1090	L	312	0.08
1451	L	320	0.16
1516	L	313	0.04
1630	L	357	0.14
1939	L	390	0.12
2272	L	354	0.00
698	F	268	0.07
720	F	266	-0.16
759	F	298	0.23
1421	F	394	-0.28
1545	F	277	-0.22
1552	F	353	-0.06
1671	F	312	-0.05
1724	F	347	0.11

TPH	T	RF	Sp
301	L	272	-0.12
326	L	260	-0.32
526	L	355	-0.13
591	L	388	-0.14
611	L	480	-0.21
779	L	275	-0.23
788	L	291	-0.09
838	L	323	-0.28
1068	L	233	-0.07
1143	L	240	-0.19
1516	L	244	-0.15
1565	L	324	-0.01
1630	L	238	0.16
384	F	487	-0.08
403	F	357	0.05
707	F	318	-0.19
759	F	300	0.20
1390	F	328	0.15
1454	F	357	-0.25
1550	F	372	0.11

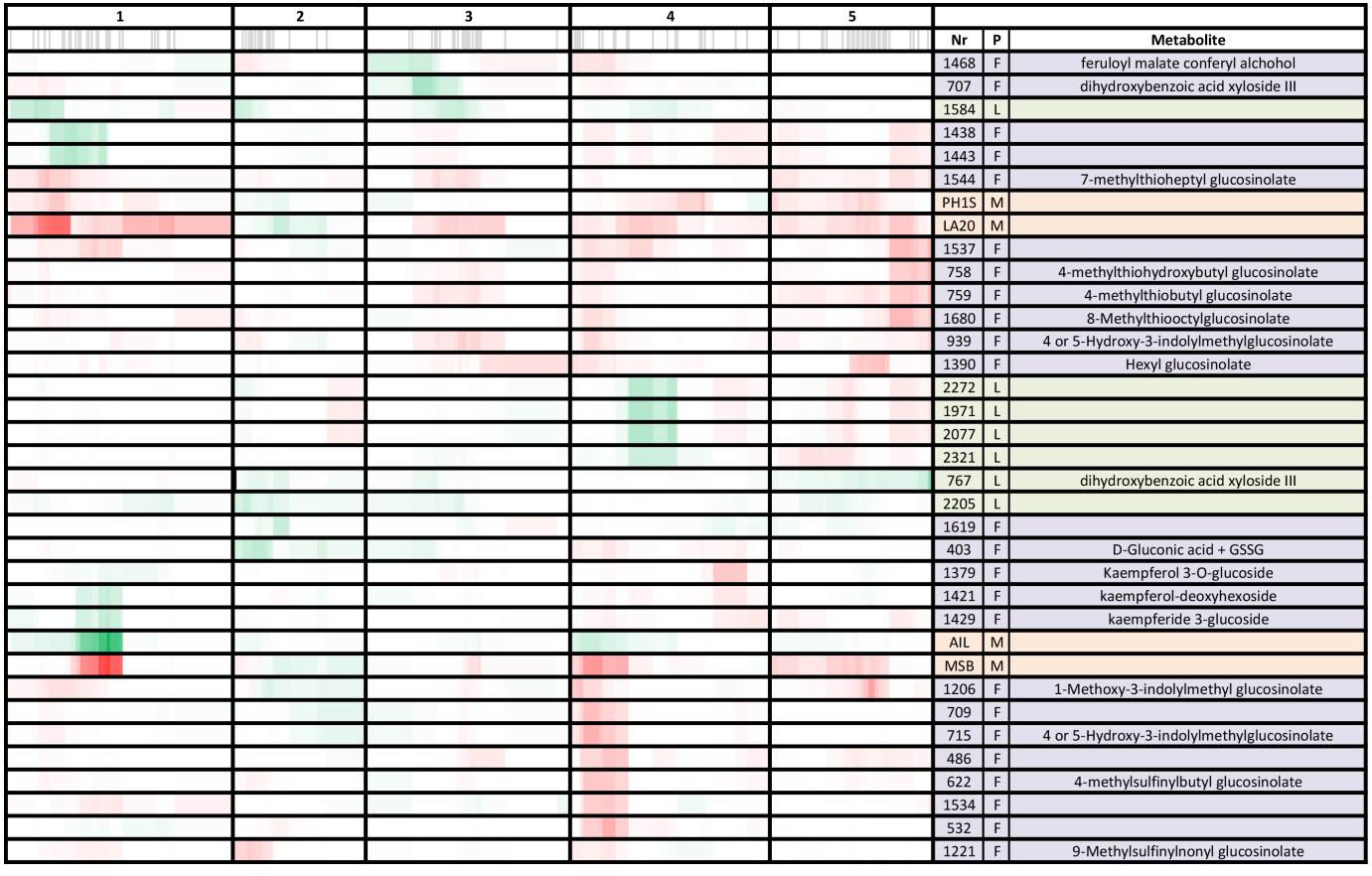


Figure 6. QTL overlap between mQTLs and morphological QTLs divided over the 5 chromosomes. The thin grey lines in the second row indicate the marker positions in cM. Trait is centrotype nr or morphological phenotype, P = phenotype, F = flower metabolite, L = leaf metabolite, M = morphology. The legend on the right indicates the LOD score between -15 and 15.

Table 10. Overlap between mQTL and morphological QTLs among the 5 chromosomes. Abreviations: L, leaf; F, flower; M, morphology.

Trait	Phenotype	Chr	Support interval (cM)	LOD	Explained variance (%)	Effect	Metabolite
1584	L	1	6-26	3.6	16.7	-0.28	
1544	F	1	14-29	3.2	14.9	0.06	7-methylthioheptyl glucosinolate
1443	F	1	20-36	3.9	17.6	-0.13	
1438	F	1	20-36.5	3.8	17.2	-0.13	
LA20	M	1	15-28	7.9	32.8	13.4	
PH1S	M	1	12-28	2.7	12.9	4.1	
1537	F	1	29-55	2.5	11.7	0.14	
1429	F	1	38-45.9	3.5	16.2	-0.10	kaempferide 3-glucoside
1421	F	1	37-54	3.1	14.5	-0.11	kaempferol-deoxyhexoside
MSB	M	1	38-44.8	10.1	39.6	12.3	
AIL	M	1	40-44.8	12.5	46.4	-13.3	
1619	F	2	15-33	2.7	12.7	-0.04	
LA20	M	2	13.8-50	2.7	12.5	-8.3	
1206	F	4	0-14	3.1	14.5	0.05	1-Methoxy-3-indolylmethyl glucosinolate
709	F	4	5-17	3.4	15.5	0.09	
715	F	4	5-18	4.1	18.5	0.11	4 or 5-Hydroxy-3-indolylmethylglucosinolate
486	F	4	4-28	3.2	14.9	0.07	
622	F	4	4-29	2.8	13.3	0.09	4-methylsulfinylbutyl glucosinolate
1534	F	4	4.5-29	2.8	13.0	0.01	
532	F	4	12-21	4.1	18.6	0.06	
MSB	M	4	3.4-25	5.5	24.0	11.4	
AIL	M	4	1-21.5	2.8	13.2	-10.6	
2272	L	4	23-45	3.6	16.3	-0.35	
1971	L	4	24-45	4.0	18.2	-0.39	
2077	L	4	24-45	4.0	18.2	-0.39	
LA20	M	4	18-50	2.7	12.7	11.0	
2321	L	5	24-50	2.7	12.8	-0.41	
767	L	5	31-65.9	2.5	11.8	-0.11	dihydroxybenzoic acid xyloside III
1206	F	5	40-43	4.8	21.4	0.06	1-Methoxy-3-indolylmethyl glucosinolate
1390	F	5	35-59.9	3.3	15.2	0.10	Hexyl glucosinolate
LA20	M	5	25-62	2.8	12.9	10.9	
MSB	M	5	25-47.4	3.1	14.5	8.8	

on chr 1 (Tables 3 and 7). Furthermore, plant height at 1st silique correlated significantly with the two other unknown metabolites, number 1443 and 1438 (Tables 3 and 7). For the second QTL on chr 1, a similar observation was made as two flavonoids and the unknown metabolite number 1537 correlated significantly with both main stem branching and average internode length (Tables 3 and 7). These results suggest that the genes underlying the QTLs are involved in the regulation of both the metabolic and morphological traits.

Discussion

Epigenetic variation can have a profound impact on plant secondary metabolism

This is the first study in plants to focus on the role of DNA methylation in the regulation of secondary metabolism. The metabolic variation due to epigenetic variation observed in the present study was moderate to large, and this suggests that epigenetics could be an important factor in determining plant phenotypic variation, also within natural populations (Figure 1). In flowers, glucosinolates were the most variable metabolites showing substantial variation and differential epigenetic regulation revealed by the different QTLs. It is well-known that the accumulation of glucosinolates is tissue-specific and under tight control by genetic factors, development and environment (Wentzell and Kliebenstein, 2008). In the present study, an extra dimension of epigenetic regulation was revealed in the complex regulation of glucosinolate content in flowers. Both aliphatic and indole glucosinolates, derived from methionine and tryptophan respectively, were at least partly under epigenetic control (Table 2).

One recent study also points towards a role for epigenetics in the regulation of glucosinolate metabolism (Rasmann et al., 2012). The authors showed that mutants dysfunctional in small RNA biogenesis with no other obvious phenotypic effects had significantly reduced amounts of glucosinolates in their leaves and these glucosinolate levels were not increased upon caterpillar feeding, in contrast to those in wild-type Col-0 (Rasmann et al., 2012). The authors further suggested that altered gene expression through siRNA signaling can be maintained in future generations via DNA methylation.

Other major metabolic QTLs in the flowers were related to flavonoid content, specifically to flavonol glycosides. Flavonoid glycosides have been shown to be implicated in several different processes such as virulence, UV-protection, biotic stress resistance, flowering, pigmentation, nodulation and the regulation of auxin transport (Shirley, 1996; Graham, 1998). Flavonol glycosides consist of an aglycone, such as quercetin and kaempferol, and sugars, such as hexoses, pentoses and deoxyhexoses. Depending upon their complexity, flavonol glycosides are found throughout the plant or in specific plant tissues (Saito et al., 2013). Here, one QTL was found for quercetin-3-O-hexoside accumulation in the leaves and three QTLs were found for different flavonoids in the flowers (Table 2). A few genome-wide expression profiling studies have suggested epigenetic regulation of flavonoid biosynthesis. In methylation mutants as well as F_1 hybrids in Arabidopsis, respective up –and down-regulation of flavonoid biosynthesis genes was found together with altered methylation states (Kurihara et al., 2008; Shen et al., 2012). Furthermore, in maize, tissue-specific pigmentation by the anthocyanin-type of flavonoids was found to be controlled through DNA methylation (Cocciolone et al., 2001).

Trade-off between metabolism, growth and reproduction

The long-standing growth/defence hypothesis states that investments in secondary metabolism and defense go at the expense of growth and reproduction (Herms and Mattson, 1992). Recently, it was found that the production of glucosinolates in leaves is negatively correlated with relative growth rate and knock-out mutants in the glucosinolate pathway were found to outperform Col-0 wild-type, providing evidence to support this hypothesis (Paul-Victor et al., 2010; Zust et al., 2011). In the present study, similar observations were made as the majority of significant correlations (91%) between relative growth rate and leaf metabolite levels were negative, while 98% of the significant correlations between flowering time and leaf metabolite levels were positive. Early-flowering, fast growing plants thus seem to invest less in leaf metabolism. Furthermore, all leaf metabolite QTLs were found to have negative effects signs, which indicates that *ddm1-2* induced hypomethylation increases the leaf metabolite content. If we assume that the higher accumulation of leaf metabolites is associated with increased defense to pathogens and herbivores, these results suggest that hypomethylation may induce the production of defense metabolites at the expense of growth.

Interestingly, early-flowering, fast growing plants were found to invest much more in flower metabolism. It has been noted that defense strategies in plants are expected to change before and after flowering (Briggs and Schultz, 1990; Herms and Mattson, 1992). The chemical defense theory suggests that the allocation of metabolites depends on the fitness costs associated with loss of the organ and the likelihood that the organ will be attacked (McKey, 1974). Flowers and seeds are the most important plant tissues in terms of fitness and should thus be well protected (McCall and Irwin, 2006). Indeed, glucosinolates in Arabidopsis are found in the highest concentrations in the reproductive organs, i.e. inflorescence, siliques and seeds (Brown et al., 2003). In the present study, it was found that all glucosinolate QTLs had positive effect signs, indicating higher glucosinolate levels in Col-0 than in *ddm1-2*, supporting the above studies (Table 2). It would be interesting to test whether these investments in flower defense metabolism go at the expense of seed set, seed size or silique size. Although this was not within the scope of this study, plant height and main stem branching correlated positively with leaf area, suggesting that larger plants also had a higher productivity. But this needs to be assessed more rigorously.

Metabolic prediction of plant morphology

To our knowledge, this is the first study in Arabidopsis to report on the relation between secondary metabolites and morphological traits. All morphological traits could be fairly well predicted using a set of the 20 most predictive metabolites. Different sets of metabolites were found to be important for different traits and metabolites with a QTL were found in each set. In a similar experiment using a RIL population between Col-0 and C24, a correlation of 0.58 between primary metabolites and biomass was reported (Meyer et al., 2007). In a

study on 97 natural accessions of Arabidopsis primary metabolites correlated with biomass under different conditions in the range of 0.21 - 0.58 (Sulpice et al., 2013). This is slightly lower than the correlation obtained in the present study between leaf area and the 20 best predictive metabolites (LA20 = 0.63). It must be noted here, however, that the other studies used partial least square regression and looked at true biomass, while the present study used random forest and leaf area. Our study indicates that secondary metabolites, which in contrast to primary metabolites are not directly associated with plant growth, can be good predictors - or biomarkers - for morphological and growth-related traits, and that prediction is at least comparable with prediction using primary metabolites.

The metabolome can be seen as the outcome of both genetic and epigenetic regulation, and as such it could be a good predictor for plant phenotypes, as shown in this study. It must be noted, however, that the metabolome is extremely flexible over time and the metabolic profile at any point in time cannot give a complete picture of a plant's morphological phenotype. Moreover, the complete metabolome of a plant cannot be measured using one or even multiple metabolomics platform(s). Nevertheless, in hybrid maize the metabolic prediction accuracy (using 130 metabolites) was only slightly lower than genomic prediction accuracy (using thousands of SNPs) (Riedelsheimer et al., 2012).

The present study revealed by QTL analyses and multivariate correlation analyses that the tissue-specific accumulation of secondary metabolites is tightly linked to plant growth and morphology, and might be partly regulated by epigenetic mechanisms. Epi-allelic variation in the epiRIL population resulted in wide variation in secondary metabolites in both leaves and flowers, and the variation associated significantly with certain genomic regions that also explained epigenetic variation for growth and morphology. Pleiotropic epigenetic loci are therefore expected to underlie the QTLs and to be involved in the regulation of both secondary metabolism and morphology.

Material and Methods

Plant growth conditions and phenotyping

Seeds from 99 epiRILs and the parents, Col-0 and *ddm1-2* (in the Col-0 background) were sown on filter paper with demi water and stratified at 4˚C in darkness for 5 d. Subsequently, seeds were transferred to a culture room (16 h light, 24˚C) to induce seed germination for 42 h. Seventeen replicates per epiRIL and parent were completely randomly transplanted to wet Rockwool blocks of 4 x 4 cm in a climate chamber (16 h light, 125 µmol m^{-2} s^{-1}, 70% RH, 20/18˚C day/night cycle). All plants were watered every morning for 5 min at 9am with 1/1000 Hyponex solution (Hyponex, Osaka, Japan). Plants were photographed from above each hour for the entire growth period (until leaves started to overlap) to analyse LA and RGR. At 21 days after germination (DAG), six randomly selected replicates were harvested for

leaf tissue. At the time of flowering, the flower head was harvested for six other randomly selected replicates. For the other five replicates, (FT) was noted at the opening of the first flower. Two weeks after flowering, main stem branching (MSB), plant height at 1st silique (PH1S), total plant height (TPH) and average internode length (AIL) were measured for these five replicates (Chapter 5).

LC-QTOF-MS analysis of leaf and flower tissue

For both leaves and flowers, three replicates were pooled to make one representative sample. For the leaves, 0.05 g tissue was grinded and extracted using 200 µL aqueous methanol (methanol (94%), formic acid (0.125%) and demi-water) in 1.5 ml Eppendorf tubes. For the flower tissue, between 0.015 and 0.06 g tissue, depending on the sample, was grinded and extracted using tissue/methanol (methanol (94%), formic acid (0.125%) and demi-water) 0.01 g/100 µL, proportionally in 1.5 ml Eppendorf tubes. After addition of methanol, the eppendorf tubes were immediately vortexed. Subsequently, all samples were sonicated for 15 min and then centrifuged for 10 min. Supernatant was vacuum filtrated using 96-well protein filtration plates (Captiva 0.45mm, Ansys Technologies) and collected in 700ml glass cuvets in 96-well autosampler plates (Waters) using a Genesis worksystem (Tecan Systems).

Metabolic profiles of the prepared methanol extracts were obtained using reverse phase liquid chomatography combined with a quadrupole time of flight high-resolution mass spectrometer (LC-QTOF-MS) (De Vos et al., 2007). In short, 5 µl of extract was injected in an Alliance 2796 HPLC system equipped with a Luna C18 (150 x 2.0 mm, 3 µm) column (Phenomenex). Separation was performed using a linear gradient of 5% acetonitrile in ultrapure water (both acidified with 0.1% formic acid) to 35% acetonitrile in 45 min at a flow rate of 0.19 ml min^{-1}. After detection of the compounds eluting from the column with a photodiode array detector (PDA; 200-700nm), negative electrospray ionization was applied and masses in the range of m/z 80 to 1500 were detected in a QTOF Ultima MS (Waters).
Metabolite profiles obtained were processed using Metalign software (www.metalign.nl) for baseline correction, noise estimation and ion wise mass spectral alignment. This resulted in 8955 mass signals for leaf samples and 6738 mass signals for flower samples. All masses with amplitudes above 100 in at least 10 samples were kept, resulting in 2334 masses for leaf samples and 1818 masses for flower samples. MSClust software (Tikunov et al., 2012) was used for clustering masses that originate from the same parent ion based on their corresponding retention times and intensity patterns over samples. In the end, 216 mass clusters representing reconstructed metabolites were obtained for leaf samples and 179 for flower samples.

Identification of metabolites was based on matching the retention time and accurate masses of parent ions and their (in-source) fragment s with an in-house experiment-based database of previously reported Arabidopsis metabolites detected under the same chromatographic

conditions (van der Hooft et al., 2012). Masses not present in the in-house database were matched with masses present in other databases such as: the Dictionary of Natural Products (http://dnp.chemnetbase.com) and Metabolomics Japan (http://metabolomics.jp).

QTL mapping with R/QTL

QTL mapping was performed with multiple QTL mapping (MQM) implemented in the R/QTL software (Arends et al., 2010; Joosen et al., 2012). Co-factors were assigned to 42 out of the 126 markers based on their physical cM position and preliminary composite interval mapping (CIM) on the data. Backward elimination was used to remove cofactors that did not contribute to the fit of the model. MQM mapping was performed on each trait and each treatment separately and the results were compared to standard interval mapping, using Haley Knott regression (Haley and Knott, 1992). Thousand random permutations were generated for each phenotype to determine the LOD significance threshold with 0.05 as the genome-wide type I error level.

Metabolic prediction of morphology

Random Forest (RF) was used to estimate the prediction of the morphological traits by the metabolite levels using the "randomForest" package in R (Breiman, 2001). Random Forest generates many decision trees, which are built using a deterministic algorithm. The trees are different owing to two factors. First, at each node, a best split is chosen from a random set of metabolites. Second, every tree is built using a bootstrap sample of the trait values. The out-of bag data (20% of the trait values) are then used to estimate the prediction accuracy. The overall prediction is then calculated by averaging over all the trees (n_{tree}=1000). The r^2 in Random Forest may be interpreted as a measure for predictive quality.

The variable importance is measured by the mean decrease in accuracy of a variable (in our case a metabolite). The more the accuracy of the random forest decreased due to addition of a single variable, the more important the variable is in classification. The variables with a large decrease in accuracy are more important for classifying the data.

References

Allwood JW, Ellis DI, Goodacre R (2008) Metabolomic technologies and their application to the study of plants and plant-host interactions. Physiologia plantarum 132: 117-135

Arends D, Prins P, Jansen RC, Broman KW (2010) R/qtl: high-throughput multiple QTL mapping. Bioinformatics 26: 2990-2992

Bayram O, Krappmann S, Ni M, Bok JW, et al (2008) VelB/VeA/LaeA complex coordinates light signal with fungal development and secondary metabolism. Science 320: 1504-1506

Beekwilder J, van Leeuwen W, van Dam NM, Bertossi M, et al (2008) The Impact of the Absence of Aliphatic Glucosinolates on Insect Herbivory in Arabidopsis. Plos One 3: 12

Bond DM, Baulcombe DC (2014) Small RNAs and heritable epigenetic variation in plants. Trends in cell biology 24: 100-107

Breiman L (2001) Random forests. Machine Learning 45: 5-32

Briggs MA, Schultz JC (1990) Chemical defense production in Lotus corniculatus: 2. Trade-offs among growth, reproduction and defense. Oecologia 83: 32-37

Brown PD, Tokuhisa JG, Reichelt M, Gershenzon J (2003) Variation of glucosinolate accumulation among different organs and developmental stages of Arabidopsis thaliana. Phytochemistry 62: 471-481

Carreno-Quintero N, Acharjee A, Maliepaard C, Bachem CW, et al (2012) Untargeted metabolic quantitative trait loci analyzes reveal a relationship between primary metabolism and potato tuber quality. Plant Physiol 158: 1306-1318

Cazzonelli CI, Cuttriss AJ, Cossetto SB, Pye W, et al (2009) Regulation of carotenoid composition and shoot branching in Arabidopsis by a chromatin modifying histone methyltransferase, SDG8. The Plant cell 21: 39-53

Cocciolone SM, Chopra S, Flint-Garcia SA, McMullen MD, et al (2001) Tissue-specific patterns of a maize Myb transcription factor are epigenetically regulated. Plant Journal 27: 467-478

Colome-Tatche M, Cortijo S, Wardenaar R, Morgado L, et al (2012) Features of the Arabidopsis recombination landscape resulting from the combined loss of sequence variation and DNA methylation. Proceedings of the National Academy of Sciences of the United States of America 109: 16240-16245

De Vos RCH, Moco S, Lommen A, Keurentjes JJB, et al (2007) Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. Nature Protocols 2: 778-791

Dyachenko OV, Zakharchenko NS, Shevchuk TV, Bohnert HJ, et al (2006) Effect of hypermethylation of CCWGG sequences in DNA of Mesembryanthemum crystallinum plants on their adaptation to salt stress. Biochemistry (Moscow) 71: 461-465

Graham TL (1998) Flavonoid and flavonol glycoside metabolism in Arabidopsis. Plant Physiology and Biochemistry 36: 135-144

Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in crosses using flank marker. Heredity 69: 315-324

Hemm MR (2002) The Arabidopsis ref2 Mutant Is Defective in the Gene Encoding CYP83A1 and Shows Both Phenylpropanoid and Glucosinolate Phenotypes. The Plant Cell Online 15: 179-194

Herms DA, Mattson WJ (1992) The dilemma of plants - To grow or defend. Quarterly Review of Biology 67: 283-335

Jablonka E, Raz G (2009) Transgenerational epigenetic inheritance: Prevalence, mechanisms, and implications for the study of heredity and evolution. Quarterly Review of Biology 84: 131-176

Johannes F, Porcher E, Teixeira FK, Saliba-Colombani V, et al (2009) Assessing the Impact of Transgenerational Epigenetic Variation on Complex Traits. PLoS genetics 5

Jones P, Messner B, Nakajima JI, Schaffner AR, et al (2003) UGT73C6 and UGT78D1, glycosyltransferases involved in flavonol glycoside biosynthesis in Arabidopsis thaliana. Journal of Biological Chemistry 278: 43910-43918

Joosen RVL, Arends D, Willems LAJ, Ligterink W, et al (2012) Visualizing the Genetic Landscape of Arabidopsis Seed Performance. Plant Physiology 158: 570-589

Kegge W, Pierik R (2010) Biogenic volatile organic compounds and plant competition. Trends in plant science 15: 126-132

Keurentjes JJ, Fu J, de Vos CH, Lommen A, Hall RD, et al (2006) The genetics of plant metabolism. Nature genetics 38: 842-849

Kooke R, Keurentjes JJB (2012) Multi-dimensional regulation of metabolic networks shaping plant development and performance. Journal of experimental botany 63: 3353-3365

Kurihara Y, Matsui A, Kawashima M, Kaminuma E, et al (2008) Identification of the candidate genes regulated by RNA-directed DNA methylation in Arabidopsis. Biochem Biophys Res Commun 376: 553-557

Latrasse D, Jegu T, Meng PH, Mazubert C, et al (2013) Dual function of MIPS1 as a metabolic enzyme and transcriptional regulator. Nucleic Acids Res 41: 2907-2917

McCall AC, Irwin RE (2006) Florivory: the intersection of pollination and herbivory. Ecology Letters 9: 1351-1365

McKey D (1974) Adaptive patterns in alkaloid physiology. American Naturalist 108: 305-320

Meyer RC, Steinfath M, Lisec J, Becher M, et al (2007) The metabolic signature related to high plant growth rate in Arabidopsis thaliana. Proceedings of the National Academy of Sciences of the United States of America 104: 4759-4764

Paul-Victor C, Zust T, Rees M, Kliebenstein DJ, et al (2010) A new method for measuring relative growth rate can uncover the costs of defensive compounds in Arabidopsis thaliana. New Phytol 187: 1102-1111

Rapp RA, Wendel JF (2005) Epigenetics and plant evolution. New Phytologist 168: 81-91

Rasmann S, De Vos M, Casteel CL, Tian D, et al (2012) Herbivory in the previous generation primes plants for enhanced insect resistance. Plant Physiol 158: 854-863

Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, et al (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nat Genet 44: 217-220

Saito K, Yonekura-Sakakibara K, Nakabayashi R, Higashi Y, et al (2013) The flavonoid biosynthetic pathway in Arabidopsis: structural and genetic diversity. Plant Physiol Biochem 72: 21-34

Schauer N, Semel Y, Roessner U, Gur A, et al (2006) Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nature biotechnology 24: 447-454

Shen H, He H, Li J, Chen W, et al (2012) Genome-wide analyzis of DNA methylation and gene expression changes in two Arabidopsis ecotypes and their reciprocal hybrids. Plant Cell 24: 875-892

Shirley BW (1996) Flavonoid biosynthesis: 'New' functions for an 'old' pathway. Trends in Plant Science 1: 377-382

Shwab EK, Bok JW, Tribus M, Galehr J, et al (2007) Histone deacetylase activity regulates chemical diversity in Aspergillus. Eukaryotic Cell 6: 1656-1664

Sonderby IE, Hansen BG, Bjarnholt N, Ticconi C, et al (2007) A Systems Biology Approach Identifies a R2R3 MYB Gene Subfamily with Distinct and Overlapping Functions in Regulation of Aliphatic Glucosinolates. Plos One 2

Stroud H, Greenberg MVC, Feng SH, Bernatavichute YV, et al (2013) Comprehensive Analyzis of Silencing Mutants Reveals Complex Regulation of the Arabidopsis Methylome. Cell 152: 352-364

Sulpice R, Nikoloski Z, Tschoep H, Antonio C, et al (2013) Impact of the carbon and nitrogen supply on relationships and connectivity between metabolism and biomass in a broad panel of Arabidopsis accessions. Plant Physiol 162: 347-363

Sulpice R, Trenkamp S, Steinfath M, Usadel B, et al (2010) Network analyzis of enzyme activities and metabolite levels and their relationship to biomass in a large panel of Arabidopsis accessions. The Plant cell 22: 2872-2893

Teixeira FK, Heredia F, Sarazin A, Roudier F, et al (2009) A Role for RNAi in the Selective Correction of DNA Methylation Defects. Science 323: 1600-1604

Tikunov Y, Laptenok S, Hall R, Bovy A, et al (2012) MSClust: a tool for unsupervised mass spectra extraction of chromatography-mass spectrometry ion-wise aligned data. Metabolomics 8: 714-718

Tikunov YM, Laptenok S, Hall RD, Bovy A, et al (2012) MSClust: A tool for unsupervised mass spectra extraction of chromatography-mass spectrometry ion-wise aligned data. Metabolomics 8: 714-718

van der Hooft JJ, Vervoort J, Bino RJ, de Vos RC (2012) Spectral trees as a robust annotation tool in LC–MS based metabolomics. Metabolomics 8: 691-703

van der Hooft JJJ, de Vos RCH, Ridder L, Vervoort J, et al (2013) Structural elucidation of low abundant metabolites in complex sample matrices. Metabolomics 9: 1009-1018

van der Hooft JJJ, Vervoort J, Bino RJ, de Vos RCH (2012) Spectral trees as a robust annotation tool in LC-MS based metabolomics. Metabolomics 8: 691-703

Wentzell AM, Kliebenstein DJ (2008) Genotype, age, tissue, and environment regulate the structural outcome of glucosinolate activation. Plant physiology 147: 415-428

Wentzell AM, Rowe HC, Hansen BG, Ticconi C, et al (2007) Linking metabolic QTLs with network and cis-eQTLs controlling biosynthetic pathways. PLoS genetics 3: 1687-1701

Zhang CY, Wang NN, Zhang YH, Feng QZ, et al (2013) DNA methylation involved in proline accumulation in response to osmotic stress in rice (Oryza sativa). Genet Mol Res 12: 1269-1277

Zhang H, Deng X, Miki D, Cutler S, et al (2012) Sulfamethazine suppresses epigenetic silencing in Arabidopsis by impairing folate synthesis. Plant Cell 24: 1230-1241

Zust T, Joseph B, Shimizu KK, Kliebenstein DJ, et al (2011) Using knockout mutants to reveal the growth costs of defensive traits. Proc Biol Sci 278: 2598-2603

Supplemental table 1. Metabolic annotation, empirical formula and name for glucosinolates found in the epiRIL population.

Metabolite nr	EF	Name	Other name
622	$C_{12}H_{22}NO_{10}S_3$	glucoraphanin	4-methylsulfinylbutyl glucosinolate
715	$C_{16}H_{20}N_2O_{10}S_2$	hydroxyglucobrassicin I	4 or 5-hydroxy-3-indolylmethylglucosinolate
758	$C_{12}H_{23}NO_{10}S_3$	4-methylthiohydroxybutyl glucosinolate	4-methylthiohydroxybutyl glucosinolate
759	$C_{12}H_{23}NO_9S_3$	glucoerucin	4-methylthiobutyl glucosinolate
939	$C_{16}H_{20}N_2O_{10}S_2$	hydroxyglucobrassicin II	4 or 5-hydroxy-3-indolylmethylglucosinolate
1206	$C_{17}H_{22}N_2O_{10}S_2$	1-methoxy-3-indolylmethyl glucosinolate	1-methoxy-3-indolylmethyl glucosinolate
1221	$C_{17}H_{33}NO_{10}S_3$	glucoarabin	9-methylsulfinylnonyl glucosinolate
1390	$C_{13}H_{25}NO_9S_2$	hexyl glucosinolate	aliphatic glucosinolate
1419	$C_{17}H_{22}N_2O_{10}S_2$	1 or 4-methoxy glucobrassicin	4-methoxy-3-indolylmethyl glucosinolate
1544	$C_{15}H_{29}NO_{10}S_3$	7-methylthioheptyl glucosinolate; S-Oxide derivative	7-methylthioheptyl glucosinolate
1680	$C_{16}H_{31}NO_9S_3$	glucoarabishirsuin	8-methylthiooctylglucosinolate

Supplemental table 2. Metabolic annotation, empirical formula and name for flavonoids found in the epiRIL population.

Metabolite nr	EF	Name
1200	$C_{21}H_{20}O_{12}$	quercetin-3-O-hexoside
1421	$C_{21}H_{20}O_{10}$	kaempferol-deoxyhexoside
1429	$C_{22}H_{22}O_{11}$	kaempferide 3-glucoside
1379	$C_{21}H_{20}O_{11}$	kaempferol 3-O-glucoside

CHAPTER 7

Biotic and abiotic stress-induced transgenerational inheritance of transcript, metabolite and morphological variation in Arabidopsis

Authors:

Rik Kooke[1,2,3], Henriëtte van Eekelen[3,4], Elio Schijlen[3,4], Paul Keizer[6], Fred van Eeuwijk[6], Ric C.H. de Vos[3,4,5], Dick Vreugdenhil[1,3] and Joost J.B. Keurentjes[2,3]

[1] Laboratory of Plant Physiology, Wageningen University, the Netherlands
[2] Laboratory of Genetics, Wageningen University, the Netherlands
[3] Centre for Biosystems Genomics, Wageningen Campus, the Netherlands
[4] Plant Research International, Wageningen, the Netherlands
[5] Netherlands Metabolomics Centre, Leiden University, the Netherlands
[6] Biometris, Wageningen University, the Netherlands

Abstract

When plants germinate in similar environments as their parents, environmentally induced transgenerational inheritance of epigenetic modifications might provide the progeny with an evolutionary benefit. The evidence, however, for transgenerational inheritance remains elusive. Pure epialleles, epigenetically regulated alleles that are initiated and maintained independent of genetic variation, are rare and stress-induced transgenerational inheritance of epigenetic changes over more than one non-stressed generation has not yet been proven. Here, we describe a study using sixty different lines derived from the same *Arabidopsis thaliana* ecotype Col-0 line by treatment of different plants for four generations under control, salt, mock or methyljasmonate (MeJA) conditions. Significant variation between the lines was detected at the transcript, metabolite and morphological trait level, and convincing evidence was found for both parental and transgenerational inheritance. When parents or grandparents were stressed with salt or methyljasmonate, stress-specific responses in gene expression, metabolite abundance and morphological and growth related traits were elicited in their progeny. For several traits, the stress-induced epigenetic changes were transmitted over more than one non-stressed generation, and they appear to be adaptive. Because all lines are derived from one common Arabidopsis founder line, epigenetic variation is most likely causal for the stress-induced transgenerational inheritance of trait variation.

Introduction

Ever since land colonization, plants have become increasingly adapted to a wide range of environments though random genetic mutations, recombination, chromosome rearrangements and genome duplications (Koornneef et al., 2004; De Bodt et al., 2005; Schranz et al., 2006). Beneficial mutations in terms of fitness are naturally selected for throughout evolution and superior genotypes dominate the specific environment in which they thrive. Over shorter or longer periods of time, the environment may change and differently adapted genotypes can then invade such environments. If environmental changes are slow over long periods on an evolutionary timescale, genetic mutation is an efficient mechanism for adaptation. If, however, the environment is fluctuating or unstable in time, epigenetic modifications, which are heritable and reversible, might give plants an evolutionary adaptive advantage (Rando and Verstrepen, 2007).

Epigenetic inheritance is defined as cellular or organismal transgenerational inheritance of phenotypic variation that has a different origin than DNA sequence variation (Jablonka and Raz, 2009). This heritable phenotypic variation is produced through variation in DNA methylation, histone modification, chromatin remodelling and small RNAs. In plants, several meiotically stable epialleles have been found (Cubas et al., 1999; Stokes and Richards, 2002; Manning et al., 2006; Rangwala et al., 2006; Martin et al., 2009). In fact, the first natural morphological mutant functionally characterized in plants was an epigenetic mutant. A mutant in *Linaria vulgaris*, originally described by Linnaeus, displays radial instead of bilateral symmetric flowers due to hypermethylation of the C*YCLOIDEA-like Lcyc* gene, causing transcriptional silence (Cubas et al., 1999). Another natural, stable epigenetic mutant in melon determines the sexual fate of flowers through DNA methylation of *CmWIP1*. The epiallele depends on the insertion of a transposon that initiates and maintains DNA methylation (Martin et al., 2009). The latter example illustrates the genetic initiation and regulation of a meiotically stable epiallele and because most epialleles discovered so far are dependent on genetic initiation, it is debated whether there are in fact true, exclusively epigenetically regulated epialleles (Paszkowski and Grossniklaus, 2011; Becker and Weigel, 2012; Pecinka and Scheid, 2012).

Besides genetic regulation of epialleles, the environment can also play an important role in the initiation of epigenetic modifications. The best studied example is probably the vernalisation-induced histone methylation of the flowering locus *FLC*. Upon cold exposure, *FLC* expression is progressively reduced and transcriptional repression is maintained after plants are returned to warmer temperatures (Michaels and Amasino, 1999; Sheldon et al., 1999). The initial down-regulation of *FLC* is mediated by non-coding RNAs that catalyse the enrichment of tri-methylated histone H3 Lys (27) chromatin at the *FLC* locus upon exposure to cold temperatures (Heo and Sung, 2011). The enrichment of histone methylation at the *FLC* locus is strongly associated with transcriptional repression, and once plants are fully

vernalised, exposure to warmer temperatures initiates flowering. The epigenetic state of *FLC* is reset during reproductive development, and is thus not inherited to subsequent generations to ensure that progeny also undergo the vernalisation process (Choi et al., 2009). However, a number of recent studies point to environmentally induced epigenetic modifications that are inherited to future generations (Molinier et al., 2006; Whittle et al., 2009; Boyko et al., 2010; Kathiria et al., 2010; Lang-Mladek et al., 2010; Verhoeven et al., 2010; Bilichak et al., 2012; Rasmann et al., 2012; Suter and Widmer, 2013). As seed dispersal is mostly local and a plant's offspring thus experience a similar environments as its parents, the transfer of epigenetic modifications that enhance the offspring's ability to survive under these environmental conditions seems plausible. The first study reporting on environmentally induced transgenerational epigenetic inheritance indicated that homologous recombination frequency (HRF) was increased in at least four generations of untreated progeny after exposure of parental plants to UV-C (Molinier et al., 2006). However, numerous studies, including an exact replication of the UV-C experiment, yielded contradicting results (Pecinka et al., 2009), indicating the difficulty of unequivocally establishing transgenerational epigenetic relationships.

Recently, three studies were reported in which biotic factors primed plants for disease resistance in multiple subsequent generations (Luna et al., 2012; Rasmann et al., 2012; Slaughter et al., 2012). One study concluded that untreated progeny of primed parents display an enhanced capacity in terms of defense related gene expression and disease resistance to infection with different pathogens (Slaughter et al., 2012). The primed state of the progeny did, however, not proceed to subsequent generations, suggesting a parental effect of priming rather than transgenerational inheritance of epigenetic modifications. In two other studies, the signal inherited over one non-stressed generation (Luna et al., 2012; Rasmann et al., 2012). Using mutants in small RNA biogenesis, these studies further showed that progeny of mutants grown in the presence of the caterpillar *Pieris rapae* did not show enhanced resistance to caterpillar feeding, in contrast to Col-0 wild type progeny. This provides further evidence for an epigenetic signal that is possibly maintained in future generations via DNA methylation (Rasmann et al., 2012). Moreover, it was found that the chromatin structure of the promoter region of biotic stress responsive genes was changed in non-stressed progeny of stressed plants, and that hypomethylated mutants also showed increased resistance (Luna et al., 2012). It must be noted that the follow-up work was performed only on the direct progeny of stressed plants, which cannot rule out the effect of the maternal environment. It might be that the stress-induced epigenetic signals are either experienced by the reproductive organs themselves or are transmitted onwards from other cells, and are thus found back in the progeny (Kumar et al., 2013). Another issue in transgenerational inheritance is the high stochasticity in observed effects, illustrated by the high variation in transgenerational priming effects in wild radish, and often insufficient replication in other experiments (Agrawal, 2002; Pecinka et al., 2009; Pecinka and Scheid, 2012). Furthermore,

in several lines generated for 30 generations from the same Arabidopsis Col-0 founder line, stochastic methylome variation arose between the lines under neutral conditions, indicating that epimutations can also be gained spontaneously (Becker et al., 2011; Schmitz et al., 2011). Because of these reasons and because most epialleles discovered so far are initiated by genetic mutations, the inheritance of epigenetic signals over more than one generation is still widely debated, and firm evidence for transgenerational inheritance of an exclusively and stable epigenetic mark is still missing (Paszkowski and Grossniklaus, 2011; Becker and Weigel, 2012; Pecinka and Scheid, 2012).

In the present study, Col-0 wild type plants were exposed to either stress or control treatments for three generations. Seeds from each generation and treatment were collected, stored and then grown simultaneously, again exposed to control or stress treatments. All plants were phenotyped for morphological characteristics, such as flowering time and plant height, and were analysed for secondary metabolites using UPLC-Orbitrap-FTMS. A selected number of lines were analysed using RNA sequencing to determine parental and transgenerational effects on gene expression.

Results

Experimental design

Col-0 plants derived from seeds of a single plant were either stressed with 50mM NaCl or 100μM methyljasmonate (MeJA), or control or mock treated for four generations, in all possible sequential combinations (Figure 1). Saline conditions were considered representative for abiotic stress whereas MeJA treatment mimics biotic stress. Methyljasmonate is a volatile, organic plant hormone that is involved in defence against various pathogens, and can be used as a chemical elicitor of defence responses. Seed batches were coded according to their treatment history over multiple generations. For instance, plants grown from seeds of a line that was control treated (C) in the first generation and were themselves salt-stressed (S) in the second generation are coded CS, while plants grown from seeds of a line that was mock treated (M) in the first generation and MeJA-stressed (J) for the following two generation and its own generation, are coded MJJJ (Figure 1). Four different populations were derived based on the final treatment of each line. These populations are the C population, the S population, the M population and the J population corresponding to final treatment under control, saline, mock or MeJA conditions, respectively (Supplemental table 1). Classification depends solely on the treatment in the final experiment, independent of the treatments in the earlier generations. The CS line thus belongs to the S population, while the MJJJ line belongs to the J population. In total, each population consists of 15 lines with different treatment histories and 60 different lines are considered altogether (Table 1). In each generation, seeds were dried after harvesting and stored at -80°C until use in the final experiment. In the final experiment, the 30 lines from the first, second and third generation

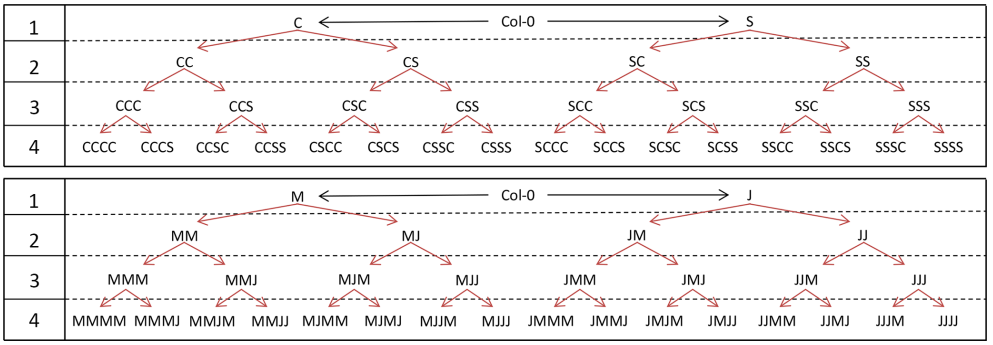


Figure 1. Experimental set-up.
Col-0 plants were either stressed with salt or methyljasmonate, or control or mock treated in three subsequent generations: control (C), salt (S), mock (M) or methyljasmonate (J). Seeds of different lines from generation 1 to 3 were grown in a single final experiment again under either salt and control or mock and methyljasmonate conditions, simultaneously generating the fourth generation.

were sown in a completely randomized design. All 30 lines were grown either in C and S or in M and J conditions, creating 60 different lines in total (Figure 1).

Each line was analysed in the final experiment for variation in growth and morphology related traits and in secondary metabolites using untargeted UPLC-Orbitrap FTMS. A representative subset of the lines (CCCC, CSSC, SSCC, MMMM, MJJM and JJMM) was also analysed for gene expression with RNA sequencing. Different comparisons were made among the lines that are studied separately in the four populations (the C, S, M and J population). For all line comparisons, we thus only consider 15 lines per population (Table 1). In a first step, for both the morphological traits and the metabolic traits, the effect of different treatment histories is studied by comparing the average trait values of the 15 lines in each population (Between-lines: Table 1). If there are significant differences between the lines, this suggests that the different treatment histories affect the trait values. However, these differences do not provide information regarding parental or transgenerational inheritance. Second, we compare the between-generation effects (Between-generations: Table 1). Although the lines were grown for four consecutive generations and in the final experiment in the same climate chamber and under the same conditions, the maternal environment might have been slightly different within each generation, and that might have caused morphological trait variation between lines of different generations, independent of previous stress treatment. Furthermore, growing plants for successive generations in the same climate chamber might habituate the plants to the conditions, either through maternal effects or stochastic epi-mutation, and this could affect the morphological trait values. Third, we compare the within-generation effects and differentiate parental, transgenerational and great transgenerational effects (Within-generations: Table 1). Parental effects can be studied by comparing the progeny of control

Table 1. Description of the possible comparisons among the 15 lines in each population (C, S, M and J population). The **between-lines** comparison tests for significant differences between the 15 lines of each population. The **between-generation** comparison tests for significant differences between generations possibly due to maternal effects of the climate chamber conditions that might have differed within the different generations. It compares for significant differences between generation 1, 2, 3 and 4. The **within-generation** comparisons test for parental, transgenerational and great transgenerational effects. The between-line comparisons are used to look for significant differences between the 1's and 2's indicated with the same color. So, for parental effects, we analyse 10 specific comparisons for significance (e.g. CC vs SC in the C population and JMJ vs JJJ in the J population). For the transgenerational comparisons, we also study the between-line differences for significance between the 1's and 2's of the same color (e.g. CCC vs SCC in the C population and MMMM vs JJMM in the M population). For the great transgenerational comparison, we compare the between-line difference for significance between the orange 1 and 2 (e.g. CCCS vs SCCS in the S population and MMMM vs JMMM in the M population). The general comparisons are tested for parental effects by comparing the average of the light grey 1's with the average of the grey 2's, and for transgenerational effects by comparing the average of the light grey 1's with the dark grey 2's. Significance level for all tests: $P < 0.05$. Abbreviations: P, parental; T, transgenerational; G, great transgenerational.

							Within-generations						**General**	
Generation	Population				**Between-lines**	**Between-generations**	**P**			**T**		**G**	**P**	**T**
1	C	S	M	J	1	1								
2	CC	CS	MM	MJ	2	2	1						1	
	SC	SS	JM	JJ	3		2						2	
3	CCC	CCS	MMM	MMJ	4	3	1	1		1			1	1
	CSC	CSS	MJM	MJJ	5		2						2	
	SCC	SCS	JMM	JMJ	6		1			2			1	2
	SSC	SSS	JJM	JJJ	7		2	2					2	
4	CCCC	CCCS	MMMM	MMMJ	8	4	1	1	1	1	1	1	1	1
	CCSC	CCSS	MMJM	MMJJ	9		2						2	
	CSCC	CSCS	MJMM	MJMJ	10		1			2			1	2
	CSSC	CSSS	MJJM	MJJJ	11		2	2					2	
	SCCC	SCCS	JMMM	JMMJ	12		1					2	1	
	SCSC	SCSS	JMJM	JMJJ	13		2						2	
	SSCC	SSCS	JJMM	JJMJ	14		1				2		1	2
	SSSC	SSSS	JJJM	JJJJ	15		2		2				2	

treated and salt stressed parents, or progeny of mock treated and MeJA treated parents (e.g. CC vs SC or MMMM vs MMJM). Ten different comparisons can be made within each population by analysing the pair-wise comparisons between two lines (Parental effects: Table 1, each color represents a different pair-wise comparison between a certain line 1 and a certain line 2 that are derived from a control (1) or a stressed (2) parent, respectively). Transgenerational effects can be investigated by comparing progeny of control treated and salt stressed grandparents, or progeny of mock treated and MeJA treated grandparents

(e.g. CCS vs SCS or MMMM vs JJMM) . Three different pair-wise comparisons can be made within each population (Transgenerational effects: Table 1, each color represents a different pair-wise comparison between a certain line 1 and a certain line 2 that are derived from a control (1) or a stressed (2) grandparent, respectively). Great transgenerational effects can be studied by comparing progeny of control treated and salt stressed great grandparents, and progeny of mock treated and MeJA treated great grandparents (e.g. CCCC vs SCCC or MMMJ vs JMMJ; great transgenerational effect in Table 1). And finally, general parental effects can be investigated by comparing the average trait values of all lines derived from control parents with the average trait values of all lines derived from salt stressed parents, and similarly for mock and methyljasmonate (Table 1: General Parental effect). General transgenerational effects can be studied by comparing the average trait values of all lines derived from control grandparents with the average trait values of all lines derived from salt stressed grandparents, and similarly for mock and methyljasmonate (Table 1: General Transgenerational effects).

Morphological traits

Phenotypic characterisation

To investigate the effect of environmental stress on morphological traits, the average trait values of the four populations were compared. Exposure to saline conditions delayed flowering and reduced biomass, leaf length, plant height and total seed weight compared with control conditions (Figure 2). Exposure to methyl jasmonate also reduced biomass, leaf length and plant height and delayed flowering, but total seed weight was slightly increased (Figure 2). The morphological traits measured on the same plants showed strong correlations in all four populations. In the C, M and J populations, flowering time strongly positively correlated with plant height, leaf length and total seed weight, indicating that late flowering is advantageous in relation to plant size and reproductive success (Figure 3). Within the S population, however, flowering time strongly negatively correlated with total plant height, leaf length and total seed weight, indicating that early flowering in lines grown under saline conditions is beneficial in terms of plant size and reproductive success (Figure 3).

Between-lines effects

For all morphological traits in each population it was investigated whether the different treatment histories of the 15 lines resulted in significantly different trait values. Indeed, within all four populations significant effects of the treatment history were found for some of the morphological traits (Supplemental table 1). There were significant differences in leaf length between the lines in all populations, and in total seed weight between the lines in the C, M and J populations. Furthermore, total plant height in the C population, flowering time in the M population, and biomass and plant height 1st silique in the J population showed significant treatment effects between the lines. These initial analyses show that different treatment histories affect plant growth and morphology, but do not provide information

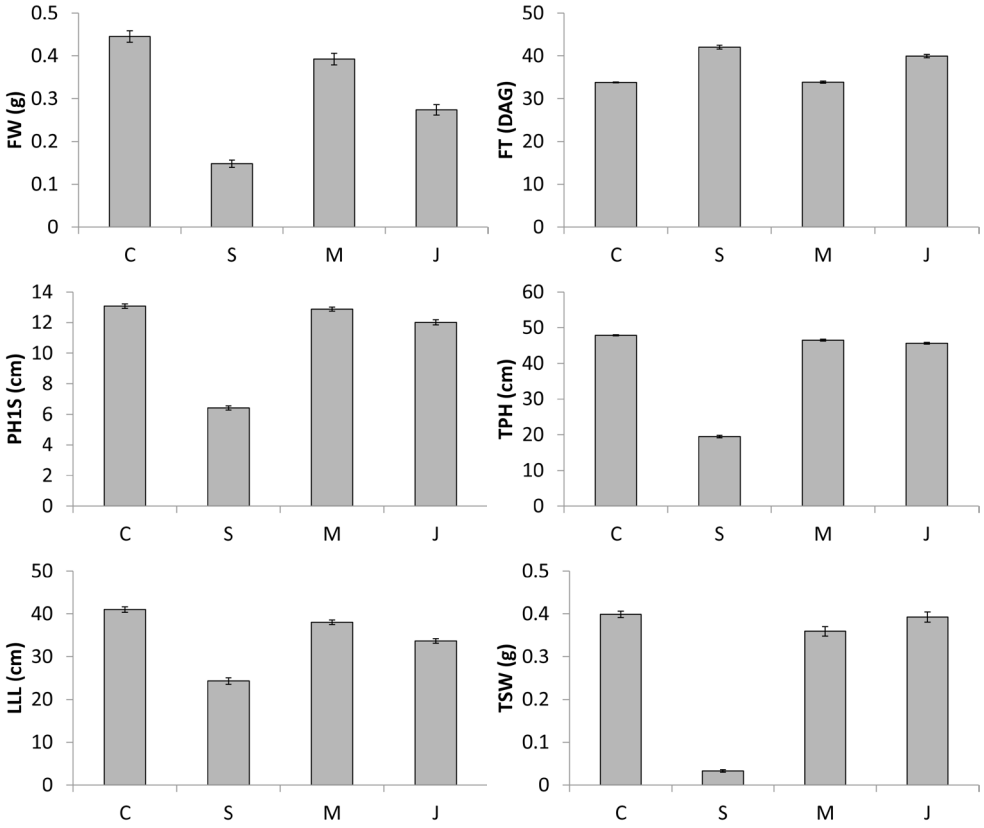


Figure 2. The average trait value of all plants from all generations in their respective environments: control (C), salt (S), mock (M), methyl jasmonate (J). FW, biomass; FT, flowering time (days after germination); PH1S, plant height at 1st silique; TPH, total plant height; LLL, largest leaf length; TSW, total seed weight.

regarding parental or transgenerational inherited effects.

Between-generation effects

Significant between-generation effects were rare but were found between all generations and for all traits, except for total plant height (Supplemental table 2). In the C, S and J populations, there were only two significant between-generation effects, but in the M population, ten significant between-generation effects were detected ($P < 0.05$). The large majority of between-generation effects (69%) were found between lines grown in the first generation and lines grown in all other generations, where the smallest plants in size and height, and the lowest total seed weight were always found in the first generation, which indeed suggests that plants become habituated to the environmental conditions of the

C	FT	PH1S	TPH	LLL	TSW
FT	*	0.38	0.32	0.51	0.32
PH1S	.000	*	0.58	0.49	0.50
TPH	.000	.000	*	0.60	0.56
LLL	.000	.000	.000	*	0.66
TSW	.000	.000	.000	.000	*

S	FT	PH1S	TPH	LLL	TSW
FT	*	-0.02	-0.51	-0.44	-0.52
PH1S	.824	*	0.67	0.07	0.44
TPH	.000	.000	*	0.44	0.84
LLL	.000	.472	.000	*	0.42
TSW	.000	.000	.000	.000	*

M	FT	PH1S	TPH	LLL	TSW
FT	*	0.16	0.15	0.51	0.54
PH1S	.044	*	0.40	0.44	0.46
TPH	.060	.000	*	0.42	0.50
LLL	.000	.000	.000	*	0.65
TSW	.000	.000	.000	.000	*

J	FT	PH1S	TPH	LLL	TSW
FT	*	0.18	0.19	0.46	0.23
PH1S	.032	*	0.51	0.38	0.36
TPH	.019	.000	*	0.49	0.53
LLL	.000	.000	.000	*	0.65
TSW	.005	.000	.000	.000	*

Figure 3. Spearman correlation matrices between morphological traits measured in a specific environment: control (C), salt (S), mock (M), methyl jasmondate (J). Spearman correlations are presented top-right of each figure, P-values are presented top-left of each figure (C, S, M and J). FT, flowering time; PH1S, plant height at 1st silique; TPH, total plant height; LLL, largest leaf length; TSW, total seed weight.

climate chamber. This is very useful information, because the parental and transgenerational comparisons only consider lines from the second, third and fourth generation. The differences between plants grown in the second, third and fourth generation do not show a general trend. Morphological trait values are sometimes higher in the previous generation, sometimes in the next generation.

Within-generation effects

In all populations, significant parental and transgenerational effects were detected for different traits within different generations (Supplemental table 3, Figure 4), of which the most striking ones will be shortly discussed. Plant height at 1st silique, total plant height, leaf length and total seed weight all showed significant parental effects in the third generation in the C population (Figure 4 A and B, and Supplemental table 3; light grey vs grey). Progeny of stressed parents had reduced leaf length, shorter height and a lower total seed weight compared with progeny of unstressed parents. Furthermore, a significant transgenerational effect was found in the same population for total plant height between CCCC and SSCC where total plant height was reduced in the SSCC line (Figure 4B and Supplemental table 3; light grey vs dark grey). In the J population, significant parental effects were detected in three comparisons in the fourth generation where the progeny of stressed parents had reduced plant height at the 1st silique (Figure 4C and Supplemental table 3). For some of the comparisons, the effects were also significant for total plant height, leaf length and total seed

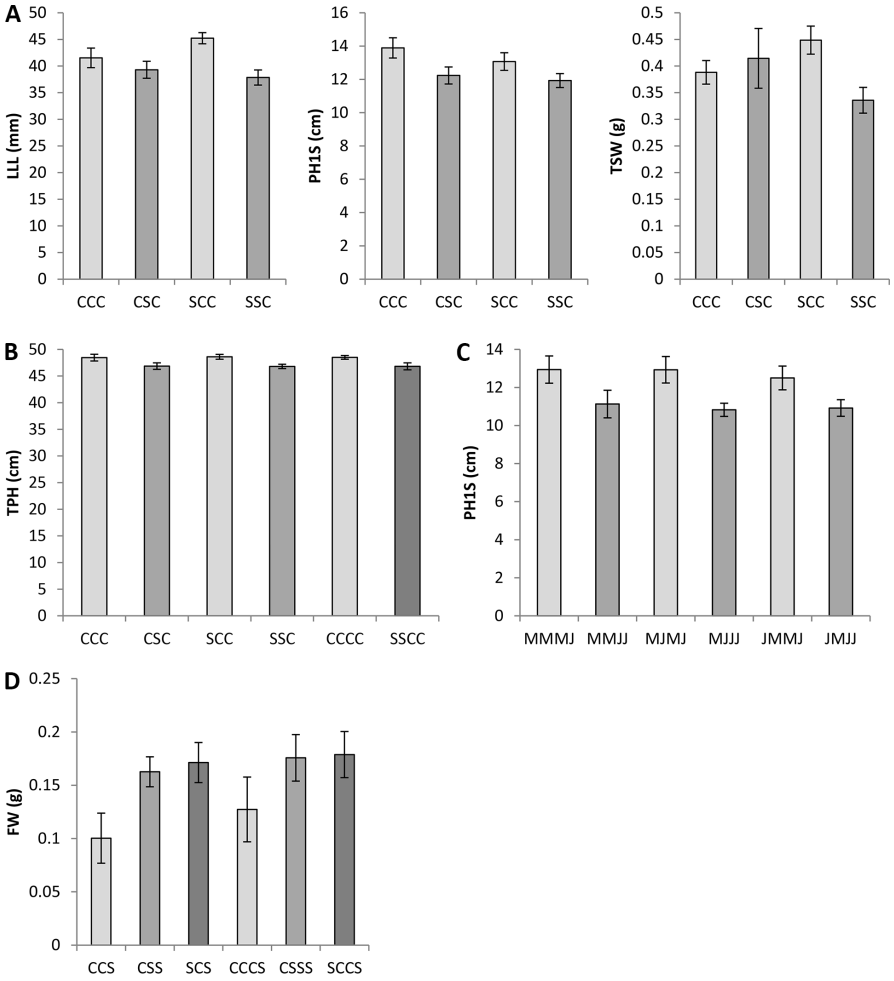


Figure 4. Detection of significant ($P < 0.05$) differences between progeny of stressed and non-stressed parents or grandparents in the third and fourth generation for morphological traits in different population. Light grey bar indicates progeny of unstressed parents, grey bar indicates progeny of stressed parents and dark grey bar indicates progeny of stressed (great) grandparents.
(A) Parental effects for largest leaf length (LLL), plant height at 1st silique (PH1S) and total seed weight (TSW) in the third generation in the C population. C, control; S, salt.
(B) Parental effects in the third generation and transgenerational effects in the fourth generation for total plant height (TPH) in the C population. C, control; S, salt.
(C) Parental effects in the fourth generation for PH1S in the J population. M, mock; J, methyljasmonate.
(D) Parental and transgenerational effects in the third and fourth generation for biomass formation (FW) in the S population. C, control; S, salt.

weight (Supplemental table 3). In the S population, parental and transgenerational effects were detected for biomass in the third and fourth generation, where parental and (great) grand-parental stress treatment increased biomass in the stressed progeny (CCS vs CSS and SCS, and CCCS vs CSSS and SCCS) (Figure 4D and Supplemental table 3). These analyses show that parental and transgenerational effects can be observed, and they suggest that environmentally induced phenotypic effects can be inherited to future generations.

General effects

To test for general parental effects in each population, the mean value was calculated for all progeny derived from salt or MeJA stressed parents and compared with the mean value of all progeny derived from control or mock treated parents (Table 1). Three significant general parental effects were found, one in the S population and two in the J population (Supplemental table 4). Progeny in the S population flowered earlier when the parental plants were stressed, and as indicated above, early flowering seems to be beneficial for plants under saline conditions (Supplemental table 3 and 4). In the J population, the progeny of stressed parents flowered earlier and had reduced height compared to progeny of unstressed parents (Supplemental table 3 and 4). There were no significant general transgenerational effects.

The analyses above show that the environment can induce morphological trait variation that is passed on to future generations. Within the generations, convincing evidence was detected for parental and transgenerational effects, although limited to a number of traits, generations and populations. Transgenerational inheritance does not appear to be a general response, which might be caused by the different generation histories between lines that impede general comparisons.

Secondary metabolism

Biochemical characterisation

To investigate to what extent previous generation histories determine the biochemical composition of Arabidopsis, all lines were analyzed for secondary metabolites using untargeted UPLC-Orbitrap FT-MS. In total, 18,352 mass signals corresponding to 209 metabolites were retrieved using Metalign and MSClust based untargeted data processing. Thirty-nine metabolites were partly or fully annotated within different secondary metabolic classes. Fifteen metabolites were identified as glucosinolates, thirteen as flavonoids, ten

Table 2. Average metabolite levels (multiple ion count (MIC)) in each population: *P*-value comparing metabolite levels of C with S population and M with J population. *P*, adjusted *P*-value using Bonferroni correction; C, control; S, salt; M, mock; J, methyljasmonate. Orange color indicates flavonoids, blue color indicates glucosinolates, green color indicates isoprenoid and purple color indicates phenylpropanoids.

Nr	Candidate molecule	C	S	*P*	M	J	*P*
84	quercetin 7-O-rhamnoside 3-O-rhamnosyl-glucoside	1243995	644760	0.000	754604	611465	0.000
95	kaempferol-3-O-(2-rhamnosylgluco-side)-7-O-rhamnoside	6407702	5398578	0.000	5795455	5440110	0.001
102	quercetin x-O-rhamnoside y-O-rhamnoside II	99545	51201	0.000	103122	83223	0.000
112	quercetin x-O-glucoside y-O-rhamnoside	1542161	516278	0.000	1007767	476597	0.000
116	kaempferol-3-O-(glucosyl-1,6-glucoside)-7-O-rhamnoside	25270	23273	1	22134	6067	0.000
124	quercetin 3-O-rhamnoside 7-O-rhamnoside	13552	189	0.000	6160	2145	0.000
128	kaempferol-3-O-gentiobioside-7-O-rhamno-side + FA	70081	38466	0.000	50469	18023	0.000
142	isorhamnetin 3-O-glucoside 7-O-rhamnoside	73526	19196	0.000	79144	32101	0.000
147	quercetin 3-O-hexoside	62027	14547	0.000	44304	33047	0.000
168	methyl-quercetin-dideoxyhexoside	40277	5090	0.000	43434	15919	0.000
169	kaempferitrin	67856	72067	1	57665	86181	0.000
175	kaempferol 3-O-glucoside	49203	15561	0.000	39914	24178	0.000
222	kaempferol rhamnoside II	53421	26777	0.304	32013	7161	0.000
8	2-propenyl glucosinolate	407887	38262	0.000	353383	157612	0.000
16	3-butenylglucosinolate	0	373	1.000	0	666	0.047
19	2-propenyl glucosinolate	0	0	1	230	0	0.156
33	6-methylsulfinylhexyl glucosinolate	144200	131730	1	162173	201990	0.000
41	7-methylsulfinylheptyl glucosinolate	96397	98600	1	140498	299934	0.000
48	7-methylsulfinylheptyl glucosinolate	1037089	941850	1	1235232	1987467	0.000
49	4-methylthiobutyl glucosinolate	1045370	466095	0.000	1231307	818070	0.000
69	8-methylsulfinyloctyl glucosinolate	7647286	8525582	1	10022630	16273148	0.000
72	2-phenylethyl glucosinolate	217248	175260	0.004	199154	276432	0.000
77	4-methoxy-3-indolylmethyl-glucosinolate	2682990	2238215	0.001	3721396	4374128	0.000
97	9-ethylthiononyl glucosinolate	37020	101417	0.000	66158	133281	0.000
117	6-methylthiohexyl glucosinolate	47483	10655	0.000	88929	46367	0.000
163	heptyl glycosinolate	185598	200292	1	291248	335193	0.000
179	7-methylthioheptyl glucosinolate	375752	98170	0.000	657940	663073	0.726
210	8-methylthiooctyl glucosinolate	721922	290026	0.005	1501934	2189631	0.036
34	2-C-methyl-D-erythritol 1-O-beta-D-fructo-furanoside	274499	193781	0.000	248648	262941	0.003
43	dihydroxybenzoic acid glucoside I	9342	13981	0.050	17181	18808	0.452
52	dihydroxybenzoic acid xyloside II	373361	474076	0.000	493916	605082	0.000
56	hydroxyferulic acid glucoside I	114127	31834	0.000	71497	13832	0.000
67	hydroxyferulic acid glucoside II	46204	30989	0.000	48042	28339	0.000
140	sinapoyl quinic acid conjugate	14668	4060	0.000	15734	42668	0.000
150	sinapoyl malate I	12492359	11711910	0.180	12377492	11428927	0.000
152	trans-dihydrodehydrodiconiferyl alco-hol-9-O-beta-D-glucoside	541600	421390	0.000	587008	764350	0.000
158	feruloyl malate coupled to coniferyl alcohol IV	496575	394623	0.000	393106	347090	0.000
174	feruloyl malate coupled to coniferyl alcohol II	21244	17688	1.000	2556	6281	0.002
180	dehydrodiconiferyl alcohol glucoside	207874	119836	0.000	170836	198701	0.000

as phenylpropanoids and one as isoprenoid (Supplemental table 5). Glucosinolates are classified into aliphatic, aromatic or indole glucosinolates based on their biosynthesis from methionine, phenylalanine or tryptophan, respectively. Twelve glucosinolates were aliphatic glucosinolates, one was an aromatic glucosinolate and one was an indole glucosinolate (Supplemental table 6).

Next, the treatment effects on the metabolic traits were investigated. Under saline conditions, 33% of the metabolites had significantly increased levels and 32% had significantly decreased levels compared to control conditions ($P < 0.05$). Upon methyljasmonate treatment, 35% of the metabolites had significantly increased levels and 24% had significantly decreased levels compared to mock treatment (Supplemental table 7). Interestingly, all flavonoid levels, except kaempferitrin decreased in both stress conditions, and the majority highly significantly ($P < 0.05$); Table 2). Seven glucosinolates showed significant differences between control and saline conditions, and their levels all decreased under saline conditions (Table 2). However, upon methyljasmonate treatment, the levels of ten glucosinolates increased significantly, which fits with their function in biotic stress resistance (Table 2). The isoprenoid, 2-C-methyl-D-erythritol 1-O-beta-D-fructofuranoside, decreased under saline conditions, but increased under methyljasmonate conditions (Table 2). The majority of phenylpropanoids decreased significantly upon salt treatment, but not upon methyljasmonate treatment (Table 2). These analyses show that biotic and abiotic stress elicit very different responses in phenylpropanoid and glucosinolate levels, but very similar responses in the flavonoid levels. Furthermore, the metabolites within each class respond rather similar to the different stresses.

Between-lines effects

For all metabolites, it was assessed whether the different generation histories among the 15 lines in each population result in significantly different metabolite levels. Thirty-eight metabolites were significantly altered by the treatment history in the different populations. Twelve metabolites had significantly different levels between lines in the C population, three in the S population, eight in the M population and sixteen in the J population (Supplemental table 8). This indicates that the treatment history has a significant effect on metabolite accumulation, but it does not differentiate between-generation, parental and transgenerational effects. Therefore, in the following sections on between-generation, within-generation and general effects, we will elaborate on the effects of the treatment histories.

Between-generation effects

In total, 355 significant between-generation effects were detected for the metabolic traits of which about 50% were found between generation 1 and the other three generations. These effects are less important because the parental and transgenerational effects are only investigated in the second, third and fourth generation. The large majority (208) of

between-generation effects were detected in the C population ($P < 0.05$) (Supplemental table 9). This indicates that the maternal environment (climate chamber conditions) was most likely slightly different in the consecutive generations, and had a strong impact on the biochemical composition of plants grown in control conditions in the final experiment. The between-generation effects are lower in the S, M and J populations, suggesting that stress treatment, but also mock treatment in the M population, has in most cases a larger effect on the metabolite accumulation than the variation in maternal environments (climate chamber conditions).

Within-generation effects

In total, 484 significant parental and transgenerational effects were observed (Supplemental table 10). Most significant effects were parental effects observed in the fourth generation in all three populations, probably due to the higher number of comparisons. We did not increase the significance threshold for multiple testing, because stress-induced parental and transgenerational effects are not expected on all metabolite levels, but rather on a small subset, and these effects are actually expected to be small. Furthermore, a few very striking cases were detected where the same metabolite or a number of metabolites from the same metabolic class showed altered levels due to parental or transgenerational effects, sometimes in different generations. Some of these convincing observations will be shortly discussed here. In the C population, the levels of three aliphatic glucosinolates were significantly decreased in the fourth generation in progeny of stressed parents and grandparents, showing both parental and transgenerational inheritance (Figure 5A). In the M population, on the other hand, the levels of three flavonoids were significantly increased in the fourth generation in progeny of stressed parents, grandparents and great grandparents, again providing evidence for both parental and transgenerational inheritance (Figure 5B). Similarly, the levels of the unidentified metabolite number 146 were significantly increased in both the third and fourth generation in progeny of differently stressed ancestors (Figure 5C). In the J population, two phenylpropanoids showed significantly different levels in the second, third and fourth generation in progeny of stressed parents, grandparents and great grandparents (Figure 5D). Furthermore, the accumulation of two unknown metabolites (nr 130 and 132) in both the third and fourth generation was affected by different grandparental treatment (Figure 5E). These results show that in all different metabolic classes, both parental and transgenerational effects of previous generation treatments can be detected, both in the third and fourth generation for the same metabolites. This provides convincing evidence that some level of transgenerational inheritance also exists at the metabolic trait level.

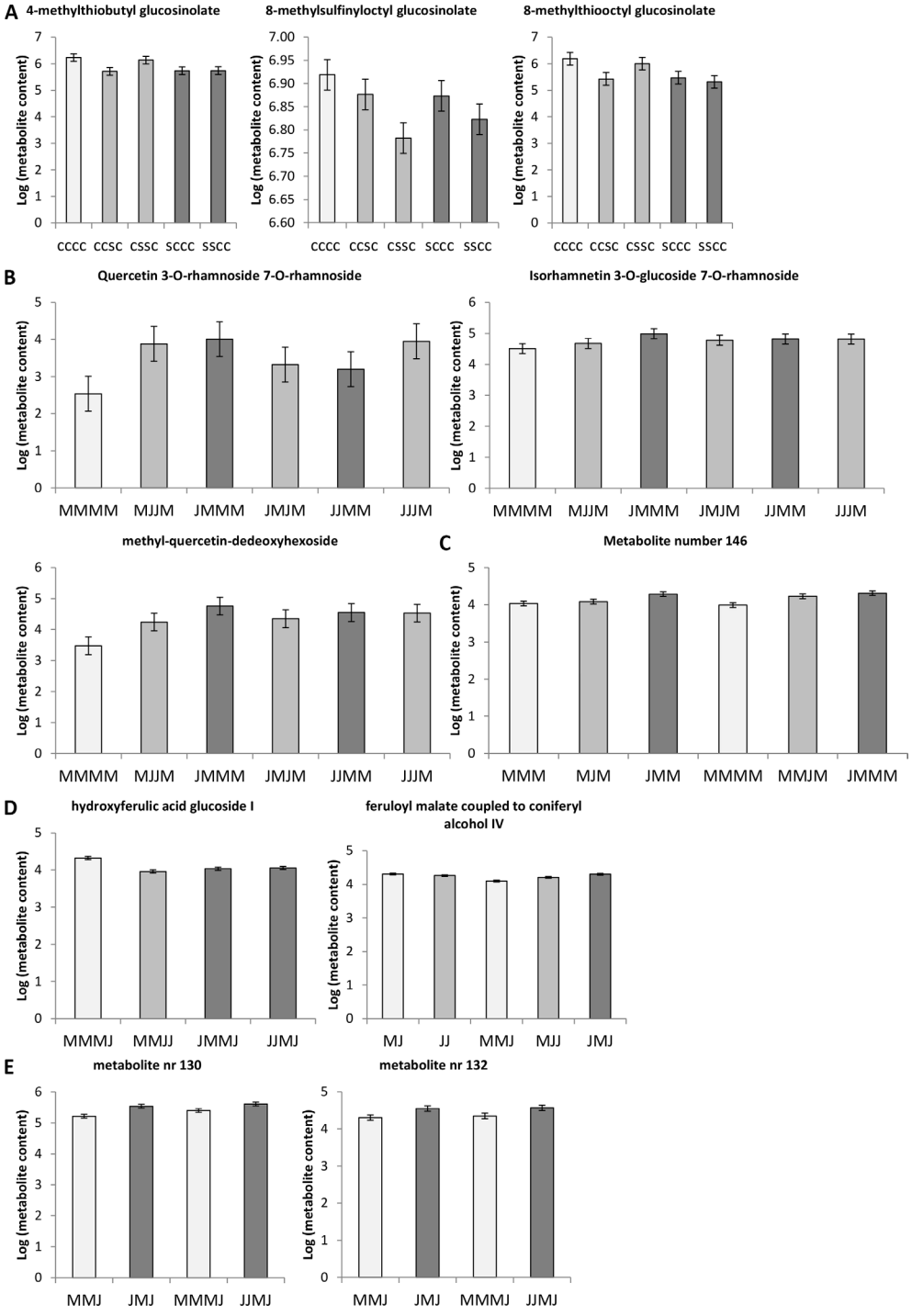
General effects

To observe general parental and transgenerational effects among all lines, the mean value was calculated for all progeny derived from salt or MeJA-stressed parents or grandparents

and compared to the mean value of all progeny derived from control or mock-treated parents or grandparents (Table 1). There were more significant transgenerational effects (44) than parental effect (17). Twelve significant parental effects were found in the M population, three in the C population and two in the J population (Supplemental table 11). One of the metabolites in the M population was a phenylpropanoid, hydroxyferulic acid glucoside I. Twenty-four significant transgenerational effects were detected in the M population, fourteen in the J population, four in the S population and two in the C population (Supplemental table 12). Among the metabolites with transgenerational effects in the M population were four aliphatic glucosinolates and one phenylpropanoid. The glucosinolate content in the mock-treated progeny increased when the grandparents had been stressed with methyljasmonate (Figure 6), concomitant with the role that glucosinolates play in biotic stress resistance. Measured over all lines, however, two of these glucosinolates showed greatly reduced levels after MeJA treatment compared with mock treatment, while one glucosinolate had greatly increased levels and the other glucosinolate did not show different levels between the two treatments (Supplemental table 8). Interestingly, two of the four metabolites were also found to have maternal and transgenerational effects in the fourth generation of the C population where previous generational salt treatment reduced the glucosinolate levels, concomitant with the general effect of salt on these glucosinolates (Figure 5A). For metabolite number 3 and 234, both general parental and transgenerational effects were observed, although for metabolite number 3, the parental and transgenerational effects were opposite in effect (Supplemental table 11 and 12).

This large-scale untargeted metabolomics approach demonstrates that parental and transgenerational effects, although small, can be observed in the progeny of stressed and unstressed parents and grandparents. Because all the lines were derived from the same Col-0 parental line, epigenetic variation, and not DNA sequence variation is most likely responsible for the metabolic variation detected.

Figure 5. Detection of significant ($P < 0.05$) differences between progeny of stressed and non-stressed parents or grandparents in the third and fourth generation for different metabolites in different population. Light grey bar indicates progeny of unstressed parents, grey bar indicates progeny of stressed parents and dark grey bar indicates progeny of stressed (great) grandparents.
(A) Parental (grey) and transgenerational (dark grey) effects on the levels of three glucosinolates in the fourth generation of the C population. (B) Parental (grey) and transgenerational (dark grey) effects on the levels of three flavonoids in the fourth generation of the M population. (C) Parental (grey) and transgenerational (dark grey) effects on the metabolite levels of unknown metabolite nr 146 in both the third and fourth generation in the M population. (D) Parental (grey) and transgenerational (dark grey) effects on the levels of two phenylpropanoids in the second, third or fourth generation of the J population. (E) Transgenerational effects (dark grey) on the accumulation of two unknown metabolites with nr 130 and 132 in the third and fourth generation in the J population.


A
4-methylthiobutyl glucosinolate
8-methylsulfinyloctyl glucosinolate
8-methylthiooctyl glucosinolate
Log (metabolite content)
CCCC CCSC CSSC SCCC SSCC
B
Quercetin 3-O-rhamnoside 7-O-rhamnoside
Isorhamnetin 3-O-glucoside 7-O-rhamnoside
MMMM MJJM JMMM JMJM JJMM JJJM
methyl-quercetin-dedeoxyhexoside
C
Metabolite number 146
MMM MJM JMM MMMM MMJM JMMM
D
hydroxyferulic acid glucoside I
feruloyl malate coupled to coniferyl alcohol IV
MMMJ MMJJ JMMJ JJMJ
MJ JJ MMJ MJJ JMJ
E
metabolite nr 130
metabolite nr 132
MMJ JMJ MMMJ JJMJ

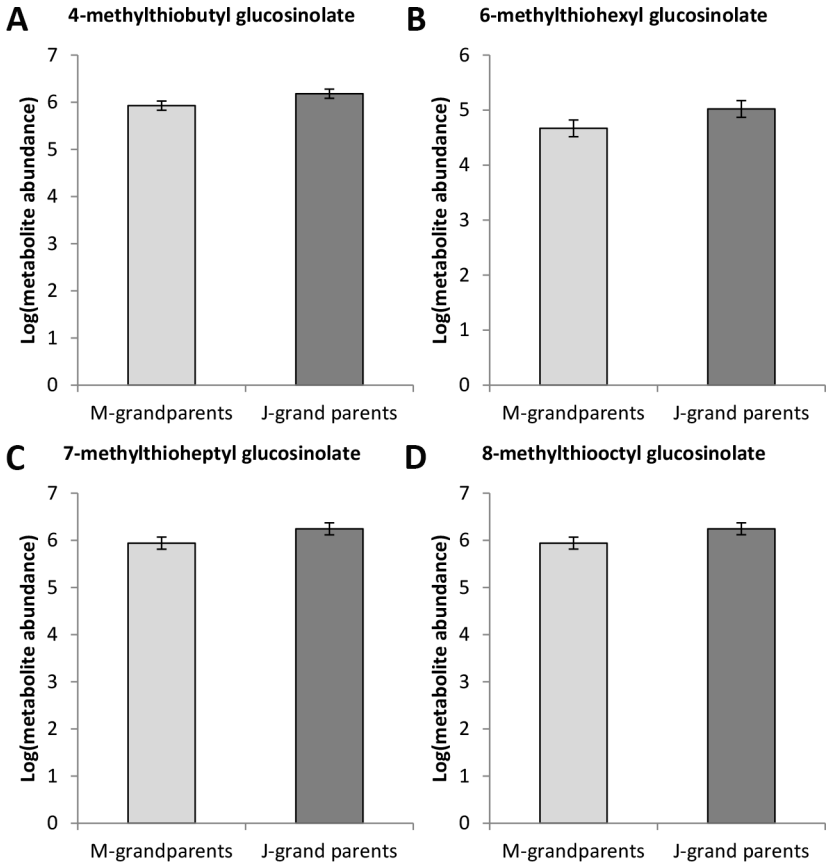


Figure 6. The general effect of methyljasmonate (J) treatment or control treatment in the grandparental generation on the accumulation of four aliphatic glucosinolates in the unstressed progeny in the M population ($P < 0.05$). (A-D) Light grey bar represent unstressed progeny from unstressed (M) grandparents, dark grey bars represent unstressed progeny from stressed (J) grandparents.

RNA sequencing

To investigate whether the observed parental and transgenerational effects at the morphological and metabolic level can be explained by transcript variation, RNA sequencing was performed on three selected lines from the fourth generation of both the C and M population: CCCC, CSSC and SSCC, and MMMM, MJJM and JJMM. The SSCC and JJMM were selected to look at the transgenerational effect of salt and MeJA on the unstressed progeny and in both lines plants were stressed only during the first two generations. The CSSC and MJJM were selected to look into the parental effects of salt and MeJA on the unstressed progeny, and stress was applied for the two consecutive generations two and three.

Salt experiment

First, the parental effects on transcript variation in the lines were studied. Seven genes were differentially expressed between the lines CCCC and CSSC with a significant FDR-adjusted *P*-value (*FDR-P* < 0.05) (Supplemental table 13). One of these genes, *DELTA1-PYRROLINE-5-CARBOXYLATE SYNTHASE 1* (*P5CS1*), is involved in salt stress responses (Strizhov et al., 1997; Abrahám et al., 2003). Some of the other genes are involved in disease resistance or cell wall maintenance. If we allow a more loose significance level (*FDR-P* < 0.1), ten genes were found to be significantly differently expressed between the two lines. One more salt-responsive gene, *EARLY RESPONSE TO DEHYDRATION 6* (*ERD6*) (Yamada et al., 2010) and a cell wall maintenance gene were now found. Eight genes were up-regulated in the CSSC line, while two genes, among which the *P5CS1* gene were down-regulated. The strongest down-regulation was found for *P5CS1*, 1.6 ($\log_2 = 0.7$), while the strongest up-regulation was found for an *EXS family protein* gene, 2.3 ($\log_2 = 1.2$) (Supplemental table 13).

The transgenerational effect on transcript variation was quite similar to the parental effect, but much stronger. Fifty-two genes were significantly differently expressed between the lines SSCC and CCCC (*FDR-P* < 0.05) (Table 3). Fourty-nine genes were up-regulated in the SSCC line with a maximum up-regulation of 3.2 fold ($\log_2 = 1.7$) for *ERD5*, and three genes were down-regulated with a maximum down-regulation of 1.5 ($\log_2 = 0.6$). Interestingly, of the twenty genes that showed the highest expression differences between the parental lines, twelve genes were present in the top-30 genes of the transgenerationally differentially expressed genes (Table 3). This indicates that salt stress in previous generations both parentally and transgenerationally affects the expression of specific genes in the unstressed progeny. This further suggests a salt stress-induced epigenetic mechanism that controls the expression of certain genes or transcription factors.

Six of the significantly differentially expressed genes are annotated to be involved in salt stress responses and eight in responses to water deprivation. Genes involved in stress, abiotic stress and water deprivation were enriched in the top-25 differentially expressed genes (Table 4, *P* < 0.01). This suggests that the environmentally induced epigenetic mechanism is adaptive, and that these plants may perform better under saline conditions. Indeed, transgenerational effects are present in the third generation and maternal effects in the third and fourth generation for biomass in the S population (Figure 4D). Biomass of plants grown under stress conditions significantly increased when parents or grandparents were also stressed.

Interestingly, no significantly differently expressed genes were found between the CSSC and SSCC lines, further demonstrating that parental and transgenerational salt stress effects on gene expression are very similar.

Table 3. Trans-generational effects salt: Top-30 differentially expressed genes
Top-30 RNA sequencing results between CCCC and SSCC. Results are sorted on most significant *P*-value between CCCC and SSCC. Red-coloured gene descriptions are involved in responses to salt stress (TAIR). Blue-coloured genes are also found in CCCC-CSSC top-20 genes. *FDR-P*, FDR-adjusted *P*-value; Abbr., gene abbreviation.

	BaseMean	BaseMean	foldChange				
Gene id	**CCCC**	**SSCC**	**SSCC/CCCC**	***P*-value**	***FDR-P***	**Abbr.**	**Description**
AT4G30270	450.53	894.74	1.99	0	0	XTH24	MERI5B (meristem-5)
AT1G35350	161.85	478.87	2.96	0	1.00E-06		
AT1G56510	665.2	1150.09	1.73	0	0.00026	WRR4	WHITE RUST RESISTANCE 4
AT4G19530	1489.61	2637.43	1.77	0	0.002445		disease resistance protein
AT3G04210	2700.67	4044.17	1.5	1.00E-06	0.002832		disease resistance protein
AT4G19520	743.89	1170.17	1.57	1.00E-06	0.002832		disease resistance protein
AT5G65730	972.06	1529.55	1.57	1.00E-06	0.002832	XTH6	xyloglucan:xyloglucosyl transferase
AT4G20260	2554.42	4261.95	1.67	1.00E-06	0.003291	PCAP1	DREPP plasma membrane
AT1G63860	133.94	258.83	1.93	1.00E-06	0.003704		ATP binding
AT2G01190	220.19	372.58	1.69	3.00E-06	0.006044	PDE331	octicosapeptide
AT5G03120	208.84	466.19	2.23	3.00E-06	0.006044		unknown protein
AT2G48030	88.36	200.47	2.27	3.00E-06	0.006049		endonuclease
AT4G08930	180.46	328.78	1.82	3.00E-06	0.006049	ATAPRL6	ATAPRL6 (APR-like 6)
AT1G52290	440.02	687.9	1.56	4.00E-06	0.00622	PERK15	protein kinase family protein
AT5G65470	693.6	1188.33	1.71	4.00E-06	0.006674		unknown protein
AT5G35750	327.45	545.29	1.67	5.00E-06	0.006875	HK2	ARABIDOPSIS HISTIDINE KINASE 2
AT1G08930	1466.57	3046.44	2.08	5.00E-06	0.006972	ERD6	EARLY RESPONSE TO DEHYDRATION 6)
AT4G00970	191.65	348.75	1.82	7.00E-06	0.008588	CRK41	protein kinase family protein
AT3G59310	367.25	610.88	1.66	1.70E-05	0.020284		unknown protein
AT1G53430	490.69	984.06	2.01	1.80E-05	0.020375		leucine-rich repeat family
AT1G29660	3272.06	4777.55	1.46	3.40E-05	0.033774		GDSL-motif lipase
AT5G06530	1669.1	3314.61	1.99	3.40E-05	0.033774	AtABCG22	ABC transporter family protein
AT1G12110	1654.11	2495.1	1.51	3.60E-05	0.033774	NRT1.1	nitrate transmembrane transporter
AT1G23480	271.67	454.65	1.67	3.70E-05	0.033774	CSLA3	CELLULOSE SYNTHASE-LIKE A3
AT2G23600	1425.87	2096.8	1.47	3.80E-05	0.033774	MES2	ACETONE-CYANOHYDRIN LYASE
AT5G58670	486.71	806.66	1.66	3.80E-05	0.033774	PLC1	PHOSPHOLIPASE C1
AT3G30775	182.91	581.3	3.18	4.20E-05	0.035777	PRODH	EARLY RESPONSIVE TO DEHYDRATION 5
AT1G66940	561.77	824.06	1.47	4.80E-05	0.038467		protein kinase-related
AT1G17990	295.33	508.46	1.72	5.10E-05	0.038467		12-oxophytodienoate reductase
AT3G60320	1170.98	2002.2	1.71	5.20E-05	0.038467		DNA binding

Table 4. Over-representation of genes in certain biological processes.
AmiGO v1.8 was used to calculate GO enrichment. Sample frequency is number of genes involved in the process / total number of entered genes. TAIR gives the total number of genes in process/ total number of genes in Arabidopsis thaliana (from TAIR10).

Biological process	*P*-value	Sample frequency	TAIR
Response to water deprivation	0.0000	8/30 (26.7%)	421/30320 (1.4%)
Response to water	0.0000	8/30 (26.7%)	428/30320 (1.4%)
Response to abiotic stimulus	0.0055	12/30 (40.0%)	2833/30320 (9.3%)
Response to stress	0.0068	14/30 (46.7%)	4030/30320 (13.3%)
Response to acid	0.0120	9/30 (30.0%)	1604/30320 (5.3%)
Response to inorganic substance	0.0377	8/30 (26.7%)	1417/30320 (4.7%)
Response to oxygen-containing compound	0.0580	10/30 (33.3%)	2459/30320 (8.1%)
Response to salt stress	0.0723	6/30 (20.0%)	780/30320 (2.6%)

MeJA experiment

Upon correction for multiple testing, no significantly differentially expressed genes were detected between the MJJM and MMMM lines, but fourteen genes were differentially expressed between JJMM and MMMM, which suggests that the transgenerational effect on transcript variation is larger than the parental effect (*FDR-P* < 0.05) (Supplemental table 14 and Table 5). Similar to salt stress, seven genes from the top-25 genes, based on the uncorrected *P*-value, between MJJM and MMMM were also present in the top-25 genes between JJMM and MMMM. This indicates that MeJA application to parents and grandparents elicits a similar pattern of gene expression variation in the unstressed progeny. This further suggests that MeJA-induced epigenetic changes control the expression of specific genes or transcription factors. There was no overlap whatsoever in the differentially expressed genes of the unstressed progeny of salt stressed and MeJA treated grandparents, which suggests that Me JA and salt stress elicit different, specific, epigenetic changes.

Of the top-25 genes that were significantly differentially expressed between JJMM and MMMM, seven genes were detected that are involved in organ development (Table 5). Interestingly, also *FLAVONOL SYNTHASE1* (*FLS1*), which catalyzes the formation of flavonols from dihydroflavonols, was detected (Owens et al., 2008). It was in the top-5 of most significantly differentially expressed gene and was 1.51 ($\log_2$ = 0.59) times stronger expressed in JJMM than in MMMM (Table 5). In the metabolic study, three flavonoids were found that showed transgenerational effects (Figure 5B) and they all had higher levels in the transgenerationally stressed lines. This suggests that a transgenerational stress signal caused higher gene expression of *FLS1* resulting in increased flavonoid content in the unstressed progeny. The strongest differentially expressed gene is *QUA-QUINE STARCH* (*QQS*), a gene involved in starch metabolism that has previously been linked to natural epigenetic regulation (Silveira et al., 2013) (Table 5).

Table 5. Trans-generational effects JA: Top-25 differentially expressed genes
Top-25 RNA sequencing results of differentially expressed genes between MMMM and JJMM. Results are sorted on most significant P-value between MMMM and JJMM. Red-colored gene descriptions are involved in responses to JA stress (TAIR). Blue-colored gene descriptions are involved in organ development. Purple-colored genes are also found in top-25 of MMMM and MJJM. *FDR-P*, FDR-adjusted P-value; Abbr., gene abbreviation.

	BaseMean	BaseMean	foldChange				
Gene id	**MMMM**	**JJMM**	**JJMM/ MMMM**	***P*-value**	***FDR-P***	**Abbr.**	**Description**
AT3G30720	125.22	37.64	0.3	1.00E-06	0.013749	QQS	QUA-QUINE STARCH
AT2G27310	186.54	103.87	0.56	3.00E-06	0.013749		F-box family protein
AT3G54820	105.37	189.76	1.8	3.00E-06	0.013749	PIP2D	PLASMA MEMBRANE INTRINSIC PROTEIN 2;5
AT4G08150	47.32	120.7	2.55	3.00E-06	0.013749	KNAT1	KNOTTED-LIKE
AT5G08640	622.16	936.8	1.51	3.00E-06	0.013749	FLS1	FLAVONOL SYNTHASE
AT3G16460	311.67	175.17	0.56	5.00E-06	0.016376	JAL34	jacalin lectin family protein
AT3G63200	198.78	321.83	1.62	5.00E-06	0.016376	PLP9	PATATIN-LIKE PROTEIN 9
AT3G18000	455.29	692.7	1.52	7.00E-06	0.01902	XPL1	XPL1 (XIPOTL 1)
AT1G78000	178.5	93.02	0.52	1.00E-05	0.023971	SULTR1;2	SULFATE TRANSPORTER1;2
AT5G16250	237.03	366.08	1.54	1.00E-05	0.023971		unknown protein
AT1G60730	417.06	245.58	0.59	1.50E-05	0.030001		aldo/keto reductase family
AT1G13710	23.85	66.7	2.8	1.60E-05	0.030001	KLU	CYP78A5
AT1G18370	135.15	216.96	1.61	2.10E-05	0.038049	NACK1	HINKEL
AT3G17998	451.51	685.24	1.52	2.40E-05	0.039529	CPuORF30	Conserved peptide upstream open reading frame 30
AT3G15500	17.92	2.1	0.12	3.30E-05	0.050664	NAC3	NAC DOMAIN CONTAINING PROTEIN 55
AT2G45170	282.62	174.5	0.62	4.40E-05	0.057123	ATG8E	AtATG8e
AT1G22590	305.84	147.17	0.48	4.50E-05	0.057123	AGL87	MADS-box family protein
AT4G13540	56.21	118.86	2.11	4.50E-05	0.057123		unknown protein
AT3G13960	25.2	57.51	2.28	4.70E-05	0.057123	GRF5	GROWTH-REGULATING FACTOR 5
AT5G15780	326.95	504.62	1.54	5.60E-05	0.062597		pollen Ole e 1 allergen and extensin family protein
AT5G23940	497.96	776.77	1.56	5.70E-05	0.062597	PEL3	embryo defective 3009
AT1G08920	216.21	136.43	0.63	6.30E-05	0.066792	ESL1	sugar transporter
AT4G23800	378.64	544.56	1.44	7.90E-05	0.079278	3xHMG-box2	high mobility group
AT2G24940	1502.03	1096.54	0.73	8.30E-05	0.080559	MAPR2	membrane-associated progesterone binding protein 2
AT5G39610	39.39	12.49	0.32	9.00E-05	0.082372	ORE1	NAC DOMAIN CONTAINING PROTEIN 6

Discussion

Convincing evidence for transgenerational inheritance was produced at the transcript, biochemical and morphological plant level. When parents or grandparents were stressed with salt or methyljasmonate, specific responses in gene expression, metabolite abundance and morphological and growth-related traits were elicited in their progeny. Because all lines were derived from the same homozygous founder plant, the trait variation is most likely caused by epigenetic variation, and not due to DNA sequence variation. The results strongly indicate that the epigenetic signal is not stochastically produced, but is induced by the applied stresses. This is supported by a number of observations. First, seeds from different but identically treated plants were randomly sown in the next generation. The variation between lines is thus not the result of epigenetic modifications that arose in one, specific plant, but more likely, similar epigenetic signals that arose in multiple plants independently. It must be noted, however, that some of the plants died and others produced no or only a few seeds in the salt treatment, which obviously biased the random selection towards epigenetic modifications in high productive plants. Second, both parental and transgenerational effects of the same sign were found at all phenotypic levels. The clearest example was demonstrated by RNA sequencing where a strong overlap was detected in the effects of parental and grand-parental stress treatment on the differential expression of genes in the non-stressed progeny. Both parental and grandparental stress treatment induced transgenerational responses that increased or decreased the expression of the same genes in the non-stressed progeny. Third, at the metabolic level, transgenerational effects were detected in both the third and fourth generation for a number of metabolites, and general trans-generational effects measured over multiple lines were observed for glucosinolates and a phenylpropanoid, among others. Finally, the transgenerational signal induced by salt or methyljasmonate was specific for both stresses. In all four populations (C, S, M and J), salt and methyljasmonate increased the accumulation of metabolites from different chemical classes. In addition, different genes were significantly up-regulated in the progeny when grandparents had been stressed with salt or with methyljasmonate. There was no overlap between the two stresses. These results strongly suggest that abiotic and biotic stress elicit stress-specific epigenetic changes that are inherited to future generations. However, more studies are needed to prove that the transgenerational signal is an epigenetic signal, including analysing DNA methylation, chromatin and small RNAs and the molecular mechanisms that govern such responses. The most obvious candidate for the transgenerational epigenetic signal is DNA methylation, which can be induced by stress and can be stably inherited to future generations (Verhoeven et al., 2010). Therefore, the logical next step in proving stress-induced transgenerational inheritance will be whole-genome bisulphite sequencing on selected lines. This will also allow differentiating between epigenetic and genetic polymorphisms.

If a transgenerational epigenetic mechanism exists, it may be adaptive in certain

environments. Plants and especially annual plants are quite likely to grow in the same environment as their parents and when a certain biotic or abiotic stress continues for several generations, transgenerational inheritance can be adaptive (Karban et al., 1999; Rando and Verstrepen, 2007). However, if the stress persists for much longer periods of time, stable, genetic polymorphisms that increase adaptation to the environment are most likely beneficial. Thus, in environments where conditions fluctuate, epigenetic modifications might be particularly adaptive, because, in theory, they can be reversed when the environment changes.

Research on transgenerational responses to salt stress showed that salt stress in previous generations can increase salt stress tolerance of progeny (Boyko et al., 2010; Suter and Widmer, 2013). The increased tolerance of the progeny is most likely associated with epigenetic modifications because exposure to salt also increased DNA methylation and the methylome variation was maintained in non-stressed progeny concomitant with changes in gene expression (Boyko et al., 2010). However, the epigenetic responses were not transferred to further consecutive generations and the authors hypothesized that only repeated exposure to the same stress would maintain the transgenerational response over multiple generations (Boyko et al., 2010; Boyko and Kovalchuk, 2011). The non-stressed lines analysed with RNA sequencing in our study were derived from ancestors that had been exposed to stress for two consecutive generations, and we indeed detected differentially expressed genes that are involved in the response to salt. Most genes, however, were responsive to water deprivation, which indicates that the increase in external osmotic pressure due to salt stress probably had a larger effect on the phenotypes than the toxicity of the ions. Besides, many genes involved in cell wall maintenance, such as *XTH24*, *XTH6* and *CELLULOSE SYNTHASE LIKE A3* (*CSLA3*) were significantly up-regulated in SSCC compared to CCCC, indicating that the grand-parental stress treatment also had a significant effect on growth. We indeed observed trans-generational effects for biomass in the third generation and maternal effects in the third and fourth generation in the S population (Figure 3). Biomass of stressed plants was significantly increased when parents or grand-parents were also stressed, providing supportive evidence for this hypothesis. However, biomass formation and other morphological traits did not improve further when plants were stressed for more consecutive generations. This probably indicates that some level of stochasticity is present and that the epigenetic modifications may be reversed, even when stress is maintained.

Three independent studies on biotic stress and priming showed that progeny of stressed parents were better protected against pathogens and herbivores than progeny of non-stressed parents (Luna et al., 2012; Rasmann et al., 2012; Slaughter et al., 2012). Caterpillar mass was significantly reduced on offspring of parents or grandparents that had been exposed to caterpillar feeding (Rasmann et al., 2012). Furthermore, repeated inoculations with virulent *Pseudomonas syringae* pv *tomato* DC3000 revealed transgenerational resistance

that was sustained over one non-stressed generation (Luna et al., 2012). These studies provide convincing evidence that the biotic stress-induced transgenerational responses can be inherited over one non-stressed generation, similar to our study. Herbivory in previous generations increased gene expression of two JA-responsive gene, *LIPOXYGENASE2* (*LOX2*) and *ALLINE OXIDE SYNTHASE* (*AOS*) (Rasmann et al., 2012). In our study, expression of these genes was not affected. Nevertheless, a number of other genes were differentially expressed between MMMM and JJMM, among which *FLAVONOL SYNTHASE1* (*FLS1*), which catalyses the formation of flavonols from dihydroflavonols (Owens et al., 2008). Interestingly, three flavonoids were detected that had higher metabolite levels in non-stressed progeny of stressed as compared to non-stressed progeny of non-stressed grandparents. Flavonoids are beneficial for plants as stress protecting agents and for attraction of pollinators. Moreover, several insects are sensitive to flavonoids that may hence function as feeding deterrents (Treutter, 2006). Therefore, the increased accumulation of flavonoids could be beneficial in MeJA elicited defense.

The strongest differentially expressed gene between JJMM and MMMM was *QUA-QUINE STARCH* (*QQS*), a gene involved in starch metabolism in leaves. RNAi-directed silencing of this gene slightly increased starch levels at the end of the day (Li et al., 2009). *QQS* is a de novo, recently originated gene that is present in various Arabidopsis accessions, but not in any other sequenced species (Li et al., 2009; Silveira et al., 2013). Hypomethylation in different epigenetic mutants caused increased *QQS* expression (Jordan et al., 2007; Lister et al., 2008). Interestingly, *QQS* expression varies widely in natural accessions as well as in wild populations, and the level of expression correlates negatively with DNA methylation (Silveira et al., 2013). The DNA methylation and gene expression patterns can be stably inherited for several generations. Because we detected MeJA induced epigenetic changes in the expression of this gene and because MeJA mimics herbivore attack, herbivore pressure might affect the DNA methylation and expression of this gene. *QQS* was down-regulated in JJMM compared to MMM and might thus cause higher sugar levels. Increased investment in primary metabolism could be beneficial for plants in stressful environments where primary metabolites can be used as resources for defense responses (Bolton, 2009) (Chapter 4).

Stress-induced transgenerational inheritance was observed at the transcript, metabolite and morphological trait level. For several traits, the stress-induced changes were transmitted over two non-stressed generations, clearly demonstrating transgenerational inheritance. Furthermore, specific salt-responsive genes were up-regulated in non-stressed progeny of salt stressed parents and grandparents. The salt-elicited response clearly differed from the MeJA elicited response suggesting that the transgenerational signals are stress-specific. Because all lines were derived from a common founder line, epigenetic variation, and not DNA sequence variation is most likely causal for the transgenerational effects.

Material and methods

Experimental Set-up

Seeds of Arabidopsis thaliana ecotype Col-0 derived from a single plant were sown on filter paper with demi water and stratified at 4˚C in darkness for 5 d. Thereafter, seeds were transferred to a culture room (16 h LD, 24˚C) to induce seed germination for 42 h. 80 replicate seedlings were transplanted to separate Rockwool blocks of 4 x 4 cm in a climate chamber (16 h light, 125 µmol m^{-2} s^{-1}, 70% RH, 20/18˚C day/night cycle). The replicates were divided over four different treatment groups: control (C), salt (S), Mock (M) and methyl-jasmonate (J), each group consisting of 20 replicates. The C, M and J plants were watered every morning with 1/1000 Hyponex solution (Hyponex, Osaka, Japan). The S plants were watered each morning with 1/1000 Hyponex solution supplemented with 50mM NaCl. The C and S plants were watered in trays where the solution was removed after 5 min, the M and J plants were watered by a flooding system for 5 min. The J plants were sprayed bi-weekly with 100µM MeJA (5 µl MeJA in 218 µl 96% ethanol, and supplemented with 7.5 µl Silwet-77, added to a total volume of 50 ml with demi water), while the M plants were sprayed with the same solution lacking MeJA. Two weeks after flowering, the MeJA treatment was stopped. Seeds were harvested in batch per treatment at the end of the growth period and then left on the bench for 1 month. Subsequently, seeds were dried for 4 d at a relative humidity of 20% and then either stored for further experiments in -80 ˚C or sown on filter paper for the experiment in the next generation. The C and S seeds were grown in both C and S treatment (CC, CS, SC and SS), while the M and J seeds were grown in both M and J treatment (MM, MJ, JM, JJ) (Figure 1). After seed-drying the exact same experiment was repeated again with the CC, CS, SC and SS seeds growing on C and S, and the MM, MJ, JM and JJ seeds growing on M and J (Figure 1). For the final experiment, all seeds from all treatment histories were taken from the -80°C and sown in the exact same circumstances in the same climate chamber, as previously mentioned in a completely randomized design per treatment group with 20 replicates per line (Figure 1).

Morphological phenotyping

Nine out of twenty replicates were harvested twenty-five days after germination (25 DAG) at the end of the day for metabolomics and sequencing, and biomass (FW) was noted for each individually harvested plant. For the other eleven replicates, flowering time (FT) was noted at the time the first flower opened. Three weeks after flowering for each individual plant, plant height to the 1st siliques (PH1S), total plant height (TPH) and largest leaf length (LLL) were measured. At the end of the experiment, all seeds were harvested per individual plant and total seed weight (TSW) was measured.

Untargeted metabolomics using UPLC-Orbitrap FTMS

Three individual replicates were pooled and grinded to make one replicate, and for each line there were in total three replicates. Metabolites were extracted in 75% methanol containing 0.1% formic acid and 100 mg fresh weight of frozen powdered leaf material, as described in

Chapter 6. After sonication and filtration, the extracts were subjected to untargeted LCMS-based metabolomics, using an AQUITY UPLC (Waters) system coupled to a PDA detector (Waters) and an LTQ-Orbitrap FTMS hybrid mass spectrometer (Thermo). UPLC conditions and MS settings were as described (van Duynhoven et al., 2014). In short, 5 µl of extract was injected and compounds were separated using a 1.7µm AQUITY UPLC BEH C18 column (2.1*150 mm; Waters), held at 40˚C, and a linear 20 min-gradient from 5 to 35% acetonitrile (acidified with 0.1% FA) at a flow rate of 400 µl/min. An additional 15 min was used to wash and equilibrate the column before next injection. The MS analysis was carried out in ESI-negative ionization mode at a source voltage of 4.5 kV. The spectra were collected at the mass range m/z 92-1200 at a resolution of 60,000 (FWHM at m/z 400) in centroid mode.

Untargeted processing of the raw data files was performed using Metalign software (Lommen, 2009). Sub-ppm mass accuracy enhancement was applied using both multiple internal lock masses, for scan by scan correction, and multiple external masses (PEG) for sample by sample correction (Lommen et al., 2011). The data file obtained was subsequently filtered for mass signals being present (i.e. signal to noise ratio > 3) in at least 3 samples and having an intensity of more than 1000 ions/scan in at least 1 sample, resulting in 18,352 reproducible mass features. Finally, MSClust software (Tikunov et al., 2012) was used for clustering masses that originate from the same molecular ion based on their corresponding retention time and intensity patterns over samples, resulting in 209 metabolites.

RNA sequencing analysis

Total RNA was extracted from the same tissue and for the same three replicates per line as the metabolomics samples, but only for the lines CCCC, CSSC and SSCC, and MMMM, MJJM and JJMM using Qiagen's RNeasy kit after a Qiagen DNase treatment. We used 50-75 mg rosette tissue per sample and RNA concentration was checked with the Qubit 2.0 fluorometer. Per sample 1 µg total RNA was used for RNA library preparation suitable for illumina HiSeq paired end sequencing according to TruSeq RNA™ Sample preparation LT protocol (Illumina Inc, San Diego CA, USA). In short, poly-adenylated RNA was captured using oligo dT beads followed by chemo-thermal fragmentation. Fragmented RNA was subsequently used for end repair, adaptor ligation, first and second strand cDNA synthesis and final library amplification following manufacturer's protocol (Illumina). Final libraries were eluted in 30 µl elution buffer. Library quality was analysed using a Bioanalyzer 2100 DNA1000 chip (Agilent Technologies) and quantified on a Qubit quantitation platform using Quant-iT PicoGreen (Invitrogen, Life Technologies).

Indexed libraries were equimolar pooled and diluted to 6 pM for TruSeq Paired End v2 DNA clustering on two flow cell lanes. Final sequencing was done on a HiSeq2000 platform using 101, 7, 101 flow cycles for sequencing paired end reads plus indexes reads. All steps for clustering and subsequent sequencing were carried out according to manufacturer's protocol. Reads were split per sample by corresponding index demultiplexing using CASAVA 1.8 software.

The resulting sequence reads were checked for quality using FastQC 0.10.1 (Andrews, 2010), and then quality-trimmed using CLC bio 6.0 (http://www.clcbio.com), and again checked for quality using FastQC (Andrews, 2010). BWA was used for mapping the reads to the Phi-X174 genome and the mapped reads were removed (Li and Durbin, 2009). The remaining reads were mapped to the Arabidopsis Col-0 (TAIR10) genome using CLC bio Genomics Server 5.0.1 (http://www.clcbio.com). Differential expression was finally determined using the Bioconductor DESeq R package v. 1.6.1 (Anders and Huber, 2010). The applications mentioned here are all integrated in an analysis pipeline developed by Applied Bioinformatics (Plant Research International, Wageningen). Gene ontology (GO) enrichment was analysed with AmiGO v. 1.8 (http://amigo1.geneontology.org/).

Statistics

Biomass (FW) values for the lines were transformed using the natural logarithm, and flowering time (FT) and total seed weight (TSW) values were transformed using the square root. All metabolite values for the lines were $\log_{10}$-transformed. Data were analysed in a generalized linear mixed model using restricted maximum likelihood estimation. Trait values were fixed effects; tray position in climate chamber was taken up in the model as random variable for the C and S population; individual plant positions (row and column) in the climate chamber were taken up as random variables in the M and J population. Significance ($P < 0.05$) was assessed for the fixed (trait value) effect between lines, and for the pair-wise comparisons between individual lines. Significance of between-generation effects was assessed by comparing the average trait values between two generations. Significance of general parental effects was studied by comparing the average trait value of all progeny derived from stressed parents with the average trait value of all progeny derived from non-stressed parents. Significance of transgenerational effects was assessed by comparing the average trait values of CCC and CCCC with the average trait values of SCC, CSCC and SSCC.

References

Abrahám E, Rigó G, Székely G, Nagy R, et al (2003) Light-dependent induction of proline biosynthesis by abscisic acid and salt stress is inhibited by brassinosteroid in Arabidopsis. Plant molecular biology 51: 363-372

Agrawal AA (2002) Herbivory and maternal effects: mechanisms and consequences of transgenerational induced plant resistance. Ecology 83: 3408-3415

Anders S, Huber W (2010) Differential expression analyzis for sequence count data. Genome biology 11: R106

Andrews S (2010) FastQC: A quality control tool for high throughput sequence data. Reference Source

Becker C, Hagmann J, Muller J, Koenig D, et al (2011) Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. Nature 480: 245-249

Becker C, Weigel D (2012) Epigenetic variation: origin and transgenerational inheritance. Curr Opin Plant Biol 15: 562-567

Bilichak A, Ilnystkyy Y, Hollunder J, Kovalchuk I (2012) The Progeny of Arabidopsis thaliana Plants Exposed to Salt Exhibit Changes in DNA Methylation, Histone Modifications and Gene Expression. Plos One 7

Bolton MD (2009) Primary Metabolism and Plant Defense-Fuel for the Fire. Molecular Plant-Microbe Interactions 22: 487-497

Boyko A, Blevins T, Yao Y, Golubov A, et al (2010) Transgenerational adaptation of Arabidopsis to stress requires DNA methylation and the function of Dicer-like proteins. PloS one 5: e9514

Boyko A, Kovalchuk I (2011) Genome instability and epigenetic modification--heritable responses to environmental stress? Curr Opin Plant Biol 14: 260-266

Choi J, Hyun Y, Kang MJ, Yun H, et al (2009) Resetting and regulation of FLOWERING LOCUS C expression during Arabidopsis reproductive development. Plant Journal 57: 918-931

Cubas P, Vincent C, Coen E (1999) An epigenetic mutation responsible for natural variation in floral symmetry. Nature 401: 157-161

De Bodt S, Maere S, Van de Peer Y (2005) Genome duplication and the origin of angiosperms. Trends Ecol Evol 20: 591-597

Heo JB, Sung S (2011) Vernalization-Mediated Epigenetic Silencing by a Long Intronic Noncoding RNA. Science 331: 76-79

Jablonka E, Raz G (2009) Transgenerational epigenetic inheritance: Prevalence, mechanisms, and implications for the study of heredity and evolution. Quarterly Review of Biology 84: 131-176

Jordan ND, West JP, Bottley A, Sheikh M, et al (2007) Transcript profiling of the hypomethylated hog1 mutant of Arabidopsis. Plant molecular biology 65: 571-586

Karban R, Agrawal AA, Thaler JS, Adler LS (1999) Induced plant responses and information content about risk of herbivory. Trends in Ecology & Evolution 14: 443-447

Kathiria P, Sidler C, Golubov A, Kalischuk M, et al (2010) Tobacco mosaic virus infection results in an increase in recombination frequency and resistance to viral, bacterial, and fungal pathogens in the progeny of infected tobacco plants. Plant Physiol 153: 1859-1870

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

Kumar S, Kumari R, Sharma V, Sharma V (2013) Roles, and establishment, maintenance and erasing of the epigenetic cytosine methylation marks in plants. Journal of Genetics 92: 629-666

Lang-Mladek C, Popova O, Kiok K, Berlinger M, et al (2010) Transgenerational inheritance and resetting of stress-induced loss of epigenetic gene silencing in Arabidopsis. Mol Plant 3: 594-602

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25: 1754-1760

Li L, Foster CM, Gan Q, Nettleton D, et al (2009) Identification of the novel protein QQS as a component of the starch metabolic network in Arabidopsis leaves. The Plant Journal 58: 485-498

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, et al (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133: 523-536

Lommen A (2009) MetAlign: interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. Analytical chemistry 81: 3079-3086

Lommen A, Gerssen A, Oosterink JE, Kools HJ, et al (2011) Ultra-fast searching assists in evaluating sub-ppm mass accuracy enhancement in U-HPLC/Orbitrap MS data. Metabolomics 7: 15-24

Luna E, Bruce TJA, Roberts MR, Flors V, et al (2012) Next-Generation Systemic Acquired Resistance. Plant Physiology 158: 844-853

Manning K, Tor M, Poole M, Hong Y, et al (2006) A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. Nature Genetics 38: 948-952

Martin A, Troadec C, Boualem A, Rajab M, et al (2009) A transposon-induced epigenetic change leads to sex determination in melon. Nature 461: 1135-U1237

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

Molinier J, Ries G, Zipfel C, Hohn B (2006) Transgeneration memory of stress in plants. Nature 442: 1046-1049

Owens DK, Alerding AB, Crosby KC, Bandara AB, et al (2008) Functional analyzis of a predicted flavonol synthase gene family in Arabidopsis. Plant physiology 147: 1046-1061

Paszkowski J, Grossniklaus U (2011) Selected aspects of transgenerational epigenetic inheritance and resetting in plants. Curr Opin Plant Biol 14: 195-203

Pecinka A, Rosa M, Schikora A, Berlinger M, et al (2009) Transgenerational stress memory is not a general response in Arabidopsis. PLoS One 4: e5202

Pecinka A, Scheid OM (2012) Stress-Induced Chromatin Changes: A Critical View on Their Heritability. Plant and Cell Physiology 53: 801-808

Rando OJ, Verstrepen KJ (2007) Timescales of genetic and epigenetic inheritance. Cell 128: 655-668

Rangwala SH, Elumalai R, Vanier C, Ozkan H, et al (2006) Meiotically stable natural epialleles of Sadhu, a novel Arabidopsis retroposon. PLoS Genet 2: e36

Rasmann S, De Vos M, Casteel CL, Tian D, et al (2012) Herbivory in the previous generation primes plants for enhanced insect resistance. Plant Physiol 158: 854-863

Schmitz RJ, Schultz MD, Lewsey MG, O'Malley RC, et al (2011) Transgenerational Epigenetic Instability Is a Source of Novel Methylation Variants. Science 334: 369-373

Schranz ME, Lysak MA, Mitchell-Olds T (2006) The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. Trends Plant Sci 11: 535-542

Sheldon CC, Burn JE, Perez PP, Metzger J, et al (1999) The FLF MADS box gene: A repressor of flowering in Arabidopsis regulated by vernalization and methylation. Plant Cell 11: 445-458

Silveira AB, Trontin C, Cortijo S, Barau J, et al (2013) Extensive natural epigenetic variation at a de novo originated gene. PLoS Genet 9: e1003437

Slaughter A, Daniel X, Flors V, Luna E, et al (2012) Descendants of primed Arabidopsis plants exhibit resistance to biotic stress. Plant Physiol 158: 835-843

Stokes TL, Richards EJ (2002) Induced instability of two Arabidopsis constitutive pathogen-response alleles. Proc Natl Acad Sci U S A 99: 7792-7796

Strizhov N, Abraham E, Ökrész L, Blickling S, et al (1997) Differential expression of two P5CS genes controlling proline accumulation during salt-stress requires ABA and is regulated by ABA1, ABI1 and AXR2 in Arabidopsis. The Plant Journal 12: 557-569

Suter L, Widmer A (2013) Environmental heat and salt stress induce transgenerational phenotypic changes in Arabidopsis thaliana. PloS one 8: e60364

Suter L, Widmer A (2013) Phenotypic effects of salt and heat stress over three generations in Arabidopsis thaliana. PLoS One 8: e80819

Tikunov YM, Laptenok S, Hall RD, Bovy A, et al (2012) MSClust: A tool for unsupervised mass spectra extraction of chromatography-mass spectrometry ion-wise aligned data. Metabolomics 8: 714-718

Treutter D (2006) Significance of flavonoids in plant resistance: a review. Environmental Chemistry Letters 4: 147-157

van Duynhoven J, van der Hooft JJJ, van Dorsten FA, Peters S, et al (2014) Rapid and sustained systemic circulation of conjugated gut microbial catabolites after single-dose black tea extract consumption. Journal of proteome research

Verhoeven KJ, Jansen JJ, van Dijk PJ, Biere A (2010) Stress-induced DNA methylation changes and their heritability in asexual dandelions. New Phytol 185: 1108-1118

Whittle CA, Otto SP, Johnston MO, Krochko JE (2009) Adaptive epigenetic memory of ancestral temperature regime inArabidopsis thalianaThis paper is one of a selection of papers published in a Special Issue from the National Research Council of Canada – Plant Biotechnology Institute. Botany 87: 650-657

Yamada K, Osakabe Y, Mizoi J, Nakashima K, et al (2010) Functional analyzis of an Arabidopsis thaliana abiotic stress-inducible facilitated diffusion transporter for monosaccharides. Journal of biological chemistry 285: 1138-1146

Supplemental table 1. Significance of between-lines effects.
P-value describing whether there are significant differences in morphological trait values between the lines in all four populations. FW, biomass; FT, flowering time; PH1S, plant height at 1st silique; TPH, total plant height; LLL, largest leaf length; TSW, total seed weight; n.s. not significant.

ANOVA	**C**	**S**	**M**	**J**
Trait	*P*-value	*P*-value	*P*-value	*P*-value
FW	n.s.	n.s.	n.s.	0.042
FT	n.s.	n.s.	0.017	n.s.
PH1S	n.s.	n.s.	n.s.	0.041
TPH	0.040	n.s.	n.s.	n.s.
LLL	0.021	0.008	0.007	0.024
TSW	0.006	n.s.	0.038	0.026

Supplemental table 2. The significance of between-generation effects in all populations.
P-value describing whether there are significant between-generation effects in morphological trait values between specific generations in all four populations (C, S, M and J). FW, biomass; FT, flowering time; PH1S, plant height at 1st silique; TPH, total plant height; LLL, largest leaf length; TSW, total seed weight; n.s. not significant.

ANOVA		**C**	**S**	**M**	**J**
Trait	Generation	*P*-value	*P*-value	*P*-value	*P*-value
FW	1 vs 2	n.s.	0.015	n.s.	n.s.
FW	1 vs 3	n.s.	n.s.	n.s.	n.s.
FW	1 vs 4	n.s.	n.s.	0.023	n.s.
FW	2 vs 3	n.s.	n.s.	n.s.	n.s.
FW	2 vs 4	n.s.	n.s.	0.005	n.s.
FW	3 vs 4	n.s.	n.s.	n.s.	n.s.
FT	1 vs 2	n.s.	n.s.	n.s.	0.024
FT	1 vs 3	n.s.	n.s.	n.s.	0.002
FT	1 vs 4	n.s.	n.s.	n.s.	n.s.
FT	2 vs 3	n.s.	n.s.	n.s.	n.s.
FT	2 vs 4	n.s.	n.s.	n.s.	n.s.
FT	3 vs 4	0.036	n.s.	0.046	n.s.
PH1S	1 vs 2	n.s.	n.s.	n.s.	n.s.
PH1S	1 vs 3	n.s.	n.s.	0.015	n.s.
PH1S	1 vs 4	n.s.	n.s.	n.s.	n.s.
PH1S	2 vs 3	0.044	n.s.	n.s.	n.s.
PH1S	2 vs 4	n.s.	n.s.	n.s.	n.s.
PH1S	3 vs 4	n.s.	n.s.	n.s.	n.s.
TPH	1 vs 2	n.s.	n.s.	n.s.	n.s.
TPH	1 vs 3	n.s.	n.s.	n.s.	n.s.
TPH	1 vs 4	n.s.	n.s.	n.s.	n.s.
TPH	2 vs 3	n.s.	n.s.	n.s.	n.s.
TPH	2 vs 4	n.s.	n.s.	n.s.	n.s.
TPH	3 vs 4	n.s.	n.s.	n.s.	n.s.
LLL	1 vs 2	n.s.	n.s.	0.001	n.s.
LLL	1 vs 3	n.s.	n.s.	0.000	n.s.
LLL	1 vs 4	n.s.	0.003	n.s.	n.s.
LLL	2 vs 3	n.s.	n.s.	0.012	n.s.
LLL	2 vs 4	n.s.	n.s.	n.s.	n.s.
LLL	3 vs 4	n.s.	n.s.	n.s.	n.s.
TSW	1 vs 2	n.s.	n.s.	0.043	n.s.
TSW	1 vs 3	n.s.	n.s.	0.003	n.s.
TSW	1 vs 4	n.s.	n.s.	0.033	n.s.
TSW	2 vs 3	n.s.	n.s.	n.s.	n.s.
TSW	2 vs 4	n.s.	n.s.	n.s.	n.s.
TSW	3 vs 4	n.s.	n.s.	n.s.	n.s.

Supplemental table 3. Within-generation effects between lines in all populations.
Square root transformed average trait values (FT and TSW), natural log-transformed average trait values (FW) and untransformed average trait values (PH1S, TPH, and LLL) are given for different lines with a significance letter ($P < 0.05$). Different letters mean statistically different trait values. Parental effects (grey) and trans-generational effects (dark grey) are shown relative to the lowest trait value (light grey). FW, biomass; FT, flowering time; PH1S, plant height at 1st silique; TPH, total plant height; LLL, largest leaf length; TSW, total seed weight.

CONTROL	Treatment	FW AVG	signficance (α = 0.05)	FT AVG	signficance (α = 0.05)	PH1S AVG	signficance (α = 0.05)
1	C	0.28	a	2.54	ab	12.77	abcd
2	CC	0.47	a	2.61	ab	13.71	...d
	SC	0.21	a	2.59	ab	13.46	..cd
3	CCC	0.06	a	2.54	ab	13.85	...d
	CSC	0.23	a	2.43	a.	12.02	ab..
	SCC	0.14	a	2.58	ab	13.05	abcd
	SSC	0.34	a	2.55	ab	11.83	a...
4	CCCC	0.26	a	2.61	ab	13.38	.bcd
	CCSC	0.50	a	2.64	ab	12.99	abcd
	CSCC	0.06	a	2.67	.b	13.64	...d
	CSSC	0.16	a	2.65	.b	13.11	abcd
	SCCC	0.17	a	2.68	.b	13.37	.bcd
	SCSC	0.31	a	2.60	ab	13.26	abcd
	SSCC	0.23	a	2.58	ab	12.16	abc.
	SSSC	0.33	a	2.56	ab	13.04	abcd
CONTROL	**Treatment**	**TPH AVG**	**signficance (α = 0.05)**	**LLL AVG**	**signficance (α = 0.05)**	**TSW AVG**	**signficance (α = 0.05)**
1	C	47.38	abcd	38.27	ab...	0.59	abc..
2	CC	47.97	.bcd	41.55	abcde	0.62	abcde
	SC	47.86	abcd	41.98	.bcde	0.64	.bcde
3	CCC	48.43	..cd	40.89	abcde	0.62	abcde
	CSC	46.71	ab..	38.66	abc..	0.58	ab...
	SCC	48.57	..cd	44.76	....e	0.67	....e
	SSC	46.54	ab..	38.26	ab...	0.56	a....
4	CCCC	48.66	..cd	41.16	abcde	0.62	abcde
	CCSC	47.78	abcd	42.59	.bcde	0.61	abcd.
	CSCC	48.83	..cd	43.13	..cde	0.67	...de
	CSSC	47.92	abcd	41.75	.bcde	0.65	..cde
	SCCC	47.95	abcd	39.66	abcd.	0.61	abcde
	SCSC	48.43	..cd	44.07	...de	0.64	.bcde
	SSCC	46.37	a...	37.14	a....	0.56	a....
	SSSC	47.46	abcd	39.12	abc..	0.64	.bcde

Supplemental table 3 continues.

SALT	Treatment	FW AVG	signficance (α = 0.05)	FT AVG	signficance (α = 0.05)	PH1S AVG	signficance (α = 0.05)
1	S	-1.40	abc..	3.75	ab...	6.18	ab
2	CS	-0.72	.bcde	4.06	ab...	6.63	ab
	SS	-0.72	.bcde	3.76	ab...	6.08	ab
3	CCS	-1.60	a....	4.13	.b...	6.76	ab
	CSS	-0.72	.bcde	4.02	ab...	5.29	a.
	SCS	-0.69	.bcd.	3.84	ab...	6.31	ab
	SSS	-1.62	a....	3.64	ab...	7.22	.b
4	CCCS	-1.42	ab...	4.17	.b...	6.42	ab
	CCSS	-1.00	abcd.	3.73	ab...	5.95	ab
	CSCS	-0.80	abcd.	3.66	ab...	6.10	ab
	CSSS	-0.66	..cde	3.59	a....	6.78	ab
	SCCS	-0.63	...d.	3.80	ab...	6.62	ab
	SCSS	-0.80	abcd.	3.59	a....	6.02	ab
	SSCS	-1.03	abcde	3.99	ab...	6.05	ab
	SSSS	-1.00	abcde	3.89	ab...	6.80	ab
SALT	**Treatment**	**TPH AVG**	**signficance (α = 0.05)**	**LLL AVG**	**signficance (α = 0.05)**	**TSW AVG**	**signficance (α = 0.05)**
1	S	18.68	ab...	18.96	a....	0.11	a.
2	CS	19.27	ab...	21.50	ab..	0.14	ab
	SS	17.95	a....	23.89	abcd	0.14	ab
3	CCS	19.55	ab...	20.81	ab...	0.15	ab
	CSS	18.55	ab...	22.14	abc.	0.16	ab
	SCS	19.97	ab...	22.68	abc.	0.16	ab
	SSS	22.58	.b...	28.49	...d	0.21	.b
4	CCCS	20.53	ab...	26.43	.bcd	0.16	ab
	CCSS	19.71	ab...	27.19	..cd	0.18	ab
	CSCS	18.49	ab...	27.95	..cd	0.15	ab
	CSSS	22.60	.b...	27.10	..cd	0.19	ab
	SCCS	19.19	ab...	28.01	..cd	0.17	ab
	SCSS	19.45	ab...	23.06	abcd	0.20	ab
	SSCS	16.96	a....	23.02	abcd	0.12	ab
	SSSS	18.85	ab...	22.07	abc.	0.17	ab

Supplemental table 3 continues.

MOCK	Treatment	FW AVG	signficance (α = 0.05)	FT AVG	signficance (α = 0.05)	PH1S AVG	signficance (α = 0.05)
1	M	-0.10	ab...	2.43	abc	11.96	a...
2	MM	-0.19	a....	2.62	.bc	13.09	abc
	JM	0.03	abcde	2.80	..c	12.66	ab..
3	MMM	0.22	.bcd.	2.56	.bc	13.39	.bc
	MJM	0.04	abcd.	2.63	..c	13.16	abc
	JMM	-0.03	abc..	2.64	..c	13.44	.bc
	JJM	0.10	abcde	2.80	..c	12.81	ab..
4	MMMM	0.29	..cde	2.67	..c	12.64	ab..
	MMJM	0.18	.bcde	2.64	..c	14.14	..c
	MJMM	0.33	...de	2.52	.bc	12.47	ab..
	MJJM	0.28	..cde	2.23	ab.	12.61	ab..
	JMMM	0.11	abcd.	2.67	..c	12.78	ab..
	JMJM	0.26	.bcde	2.09	a....	12.56	ab..
	JJMM	-0.04	abc..	2.44	abc	12.92	abc
	JJJM	0.22	.bcde	2.69	..c	12.56	ab..
MOCK	**Treatment**	**TPH AVG**	**signficance (α = 0.05)**	**LLL AVG**	**signficance (α = 0.05)**	**TSW AVG**	**signficance (α = 0.05)**
1	M	45.14	ab.	34.84	a....	0.52	a....
2	MM	46.68	abc	38.91	..cde	0.61	.bcd
	JM	47.21	.bc	41.49	...de	0.59	abc..
3	MMM	47.24	.bc	38.77	..cde	0.61	.bc..
	MJM	46.91	abc	39.07	..cde	0.61	.bcd
	JMM	46.99	.bc	37.09	abc..	0.60	.bc..
	JJM	45.63	ab.	37.83	abc..	0.59	abc..
4	MMMM	44.83	a..	35.08	ab...	0.54	ab...
	MMJM	48.53	..c	41.51	....e	0.68	..d.
	MJMM	45.79	ab.	38.06	abcd.	0.57	abc..
	MJJM	46.40	ab.	36.24	abc..	0.57	abc..
	JMMM	46.25	ab.	37.12	abc..	0.59	abc..
	JMJM	46.51	ab.	37.60	abc..	0.56	abc..
	JJMM	47.25	.bc	38.57	.bcde	0.61	.bcd
	JJJM	46.11	ab.	37.96	abcd.	0.63	..cd

Supplemental table 3 continues.

MEJA	Treatment	FW AVG	signficance (α = 0.05)	FT AVG	signficance (α = 0.05)	PH1S AVG	signficance (α = 0.05)
1	J	-0.18	.bcd.	3.24	a....	12.53	.bc..
2	MJ	-0.40	abcd.	3.58	abc..	11.97	abc..
	JJ	-0.21	.bcd.	3.53	ab...	11.32	ab...
3	MMJ	-0.79	a....	3.68	abc..	12.32	abc..
	MJJ	-0.10	..c..	3.46	ab...	12.39	abc..
	JMJ	-0.39	abcde	4.06	..c..	11.76	abc..
	JJJ	-0.35	.bcde	3.59	abc..	12.98	.bc..
4	MMMJ	-0.03	..cde	3.44	ab...	12.94	.bc..
	MMJJ	-0.59	ab...	3.48	ab...	10.58	a....
	MJMJ	-0.13	..cde	3.81	.bc..	13.18	..c..
	MJJJ	-0.15	..cde	3.44	ab...	10.62	a....
	JMMJ	-0.04	..c..	3.64	abc..	12.91	.bc..
	JMJJ	-0.36	.bcde	3.25	a....	10.70	a....
	JJMJ	-0.19	.bcd.	3.55	ab...	11.93	abc..
	JJJJ	-0.16	..c..	3.50	ab...	12.31	abc..
MEJA	**Treatment**	**TPH AVG**	**signficance (α = 0.05)**	**LLL AVG**	**signficance (α = 0.05)**	**TSW AVG**	**signficance (α = 0.05)**
1	J	45.57	abc..	32.09	abc..	0.61	abc..
2	MJ	45.28	a....	33.04	abcd.	0.62	abc..
	JJ	44.94	a....	32.35	abc..	0.63	abcd.
3	MMJ	45.69	abc..	31.15	ab...	0.58	ab...
	MJJ	45.89	abc..	32.78	abcd.	0.63	.bcd.
	JMJ	45.08	a....	36.41	..cde	0.62	abc..
	JJJ	47.89	.bc..	38.11	....e	0.70	...d.
4	MMMJ	45.38	a....	34.49	abcde	0.62	abc..
	MMJJ	44.67	a....	31.45	ab...	0.60	abc..
	MJMJ	45.82	abc..	32.40	abc..	0.57	ab...
	MJJJ	45.74	abc..	32.20	abc..	0.64	.bcd.
	JMMJ	48.03	..c..	37.09	...de	0.67	..cd.
	JMJJ	43.69	a....	31.04	a....	0.55	a....
	JJMJ	45.61	abc..	34.67	abcde	0.63	abcd.
	JJJJ	45.45	ab...	35.49	.bcde	0.63	.bcd.

Supplemental table 4. Significance of general parental effects.
P-value describing whether there are significant parental effects in morphological trait values between the lines in all four populations. FW, biomass; FT, flowering time; PH1S, plant height at 1st silique; TPH, total plant height; LLL, largest leaf length; TSW, total seed weight; n.s. not significant.

ANOVA	C	S	M	J
Trait	*P*-value	*P*-value	*P*-value	*P*-value
FW	n.s.	n.s.	n.s.	n.s.
FT	n.s.	0.035	n.s.	0.022
PH1S	n.s.	n.s.	n.s.	0.029
TPL	n.s.	n.s.	n.s.	n.s.
LLL	n.s.	n.s.	n.s.	n.s.
TSW	n.s.	n.s.	n.s.	n.s.

Supplemental table 5. Overview of metabolic compounds.
Question mark indicated by (*) mean that the metabolite belongs to the metabolic class, but it might be a different metabolite. The metabolite that is shown is the most likely metabolite based on membership, retention time and molecular mass.

Nr	Candidate molecule	Metabolic class	Chemical formula
84	quercetin 7-O-rhamnoside 3-O-rhamnosylglucoside	Flavonoid	C33H40O20
95	kaempferol-3-O-(2-rhamnosylglucoside)-7-O-rhamnoside	Flavonoid	C33H40O19
102	quercetin x-O-rhamnoside y-O-rhamnoside II	Flavonoid	C27H30O15
112	quercetin x-O-glucoside y-O-rhamnoside	Flavonoid	C27H30O16
116	kaempferol-3-O-(glucosyl-1,6-glucoside)-7-O-rhamnoside	Flavonoid	C33H40O20
124	quercetin 3-O-rhamnoside 7-O-rhamnoside	Flavonoid	C27H30O15
128	kaempferol-3-O-gentiobioside-7-O-rhamnoside + FA	Flavonoid	C33H40O20
142	isorhamnetin 3-O-glucoside 7-O-rhamnoside	Flavonoid	C28H32O16
147	quercetin 3-O-hexoside	Flavonoid	C21H20O12
168	methyl-quercetin-dideoxyhexoside	Flavonoid	C28H32O15
169	kaempferitrin	Flavonoid	C27H30O14
175	kaempferol 3-O-glucoside	Flavonoid	C21H20O11
222	kaempferol rhamnoside II	Flavonoid	C21H20O10
8	2-propenyl glucosinolate	Glucosinolate	C10H17O9S2N1
16	3-butenylglucosinolate	Glucosinolate	C11H19O9S2N1
19	2-propenyl glucosinolate	Glucosinolate	C10H17O9S2N1
33	6-methylsulfinylhexyl glucosinolate	Glucosinolate	C14H27O10S3N1
41	7-methylsulfinylheptyl glucosinolate	Glucosinolate	C15H29O10S3N1
48	7-methylsulfinylheptyl glucosinolate	Glucosinolate	C15H29O10S3N1
49	4-methylthiobutyl glucosinolate	Glucosinolate	C12H23O9S3N1
69	8-methylsulfinyloctyl glucosinolate	Glucosinolate	C16H31O10S3N1
72	2-phenylethyl glucosinolate	Glucosinolate	C15H21O9S2N1
77	4-methoxy-3-indolylmethyl-glucosinolate	Glucosinolate	C17H22O10S2N2
97	9-ethylthiononyl glucosinolate	Glucosinolate	C17H33O10S3N1
117	6-methylthiohexyl glucosinolate	Glucosinolate	C14H27O9S3N1
163	heptyl glycosinolate	Glucosinolate	C14H27O9S2N1
179	7-methylthioheptyl glucosinolate	Glucosinolate	C15H29O9S3N1

Supplemental table 5 continues.

Nr	Candidate molecule	Metabolic class	Chemical formula
210	8-methylthiooctyl glucosinolate	Glucosinolate	C16H31O9S3N1
43	dihydroxybenzoic acid glucoside I	Phenylpropanoid	C13H16O9
52	dihydroxybenzoic acid xyloside II	Phenylpropanoid	C12H14O8
56	hydroxyferulic acid glucoside I	Phenylpropanoid	C16H20O10
67	hydroxyferulic acid glucoside II	Phenylpropanoid	C16H20O10
140	sinapoyl quinic acid conjugate	Phenylpropanoid	C37H46O21
150	sinapoyl malate I	Phenylpropanoid	C15H16O9
152	trans-dihydrodehydrodiconiferyl alcohol-9-O-beta-D-glucoside	Phenylpropanoid	C26H34O11
158	feruloyl malate coupled to coniferyl alcohol IV	Phenylpropanoid	C24H26O12
174	feruloyl malate coupled to coniferyl alcohol II	Phenylpropanoid	C24H26O12
180	dehydrodiconiferyl alcohol glucoside	Phenylpropanoid	C26H32O11
34	2-C-methyl-D-erythritol 1-O-beta-D-fructofuranoside	Isoprenoid	C11H22O9

Supplemental table 6. Overview of glucosinolates.

Nr	Metabolite name	Other name	Derived from
8	2-propenyl glucosinolate	sinigrin	Methionine
16	3-butenylglucosinolate	gluconapin	Methionine
19	2-propenyl glucosinolate	sinigrin	Methionine
33	6-methylsulfinylhexyl glucosinolate	glucohesperin	Methionine
41	7-methylsulfinylheptyl glucosinolate	glucoibarin	Methionine
48	7-methylsulfinylheptyl glucosinolate	glucoibarin	Methionine
49	4-methylthiobutyl glucosinolate	glucoerucin	Methionine
69	8-methylsulfinyloctyl glucosinolate	glucohirsutin	Methionine
72	2-phenylethyl glucosinolate	gluconasturtiin	Phenylalanine
77	4-methoxy-3-indolylmethyl-glucosinolate	1 or 4-methoxyglucobrassicin	Tryptophan
97	9-ethylthiononyl glucosinolate	9-ethylthiononyl glucosinolate	Methionine
117	6-methylthiohexyl glucosinolate	6-methylthiohexyl glucosinolate	Methionine
163	heptyl glycosinolate	heptyl glycosinolate	Methionine
179	7-methylthioheptyl glucosinolate	7-methylthioheptyl glucosinolate	Methionine
210	8-methylthiooctyl glucosinolate	8-methylthiooctyl glucosinolate	Methionine

Supplemental table 7. Number of metabolites present in each population and number of metabolites significantly increased in one population compared to the other. *P*-values are adjusted *P*-values using Bonferroni correction. C, control; S, saline; M, mock; J, methyljasmonate.

	Population	Number of metabolites	
	Total	209	
Present in	C + S	192	
	C	195	
	S	202	
	M + J	199	
	M	203	
	J	202	Percentage
Significantly ($P < 0.05$) higher in	C vs S	62	32%
	S vs C	64	33%
	M vs J	48	24%
	J vs M	69	35%

Supplemental table 8. Significance of between-lines effects.
P-value describing whether there are significant differences in metabolic trait values between the lines in all four populations. C, control population; S, salt population; M, mock population; J, methyljasmonate population; n.s., not significant.

Trait	**Metabolite**	**C**	**S**	**M**	**J**
		P-value	*P*-value	*P*-value	*P*-value
14		0.031	n.s.	0.024	n.s.
38		n.s.	n.s.	n.s.	0.033
40		n.s.	0.047	n.s.	n.s.
46		0.024	n.s.	n.s.	n.s.
52	dihydroxybenzoic acid xyloside II	n.s.	n.s.	n.s.	0.005
54		0.023	n.s.	n.s.	n.s.
56	hydroxyferulic acid glucoside I	n.s.	n.s.	0.047	n.s.
57		n.s.	0.028	n.s.	n.s.
64		n.s.	n.s.	n.s.	0.041
70		n.s.	n.s.	0.039	n.s.
99		0.033	n.s.	n.s.	n.s.
102	Quercetin x-O-rhamnoside y-O-rhamnoside II	n.s.	n.s.	n.s.	0.043
103		0.015	n.s.	n.s.	n.s.
109		n.s.	n.s.	n.s.	0.003
130		n.s.	n.s.	n.s.	0.008
144		n.s.	n.s.	n.s.	0.034
146		n.s.	n.s.	0.004	n.s.
149		n.s.	n.s.	n.s.	0.003
156		n.s.	n.s.	0.046	n.s.
158	feruloyl malate coupled to 4-O-8 coniferyl alcohol IV	n.s.	n.s.	n.s.	0.008
160		n.s.	n.s.	0.043	n.s.
163	heptylglycosinolate	n.s.	n.s.	n.s.	0.014
168	methyl-quercetin-dideoxyhexoside	n.s.	0.048	n.s.	n.s.

Supplemental table 8 continues.

171		0.043	n.s.	n.s.	n.s.
178		n.s.	n.s.	n.s.	0.008
180	Dehydrodiconiferyl alcohol glucoside	0.036	n.s.	n.s.	n.s.
182		n.s.	n.s.	n.s.	0.022
184		n.s.	n.s.	n.s.	0.004
188		n.s.	n.s.	n.s.	0.009
192		n.s.	n.s.	0.047	n.s.
193		n.s.	n.s.	n.s.	0.031
198		0.035	n.s.	n.s.	n.s.
223		n.s.	n.s.	0.000	n.s.
239		0.041	n.s.	n.s.	n.s.
241		n.s.	n.s.	n.s.	0.028
243		0.033	n.s.	n.s.	n.s.
244		0.026	n.s.	n.s.	n.s.
245		0.045	n.s.	n.s.	n.s.

Supplemental table 9. Number of significant between-generation effects in all populations. Between-generation effects were separately tested between each two generations ($P < 0.05$). C, control population; S, salt population; M, mock population; J, methyljasmonate population.

Generation	C	S	M	J	Total
1 vs 2	28	35	5	3	71
1 vs 3	24	2	10	16	52
1 vs 4	22	5	15	4	46
2 vs 3	46	0	3	15	64
2 vs 4	12	0	7	4	23
3 vs 4	76	0	13	10	99
Total	208	42	53	52	355

Supplemental table 10. Number of significant parental, trans-generational and great grand-parental effects in different generations. Within-generation effects were separated into parental, trans-generational and great grand-parental effects depending on the generation in which stress was applied, and tested for significance ($P < 0.05$). Given are the numbers of significant effects. Great transgenerational effects are CCCC vs SCCC, CCCS vs SCCS, MMMM vs JMMM and MMMJ vs JMMJ. C, control; S, salt; M, mock; J, methyljasmonate.

		Population				
Effect	**Generation**	**C**	**S**	**M**	**J**	**Total**
Parental	2	1	0	2	8	11
	3	8	125	9	30	172
	4	49	6	37	50	142
Trans-generational	3	6	1	8	36	51
	4	19	1	35	14	69
Great Trans-generational	4	18	0	19	2	39
Total	all	101	133	110	140	484

Supplemental table 11. General parental effects on the metabolite levels in the C, M and J populations. No effects were detected in the S-population. Significance of the parental affects (*P* < 0.05) and the log(metabolite abundance) average of the progeny between unstressed parents (C or M parents) and stressed parents (S or J parents). Nr corresponds with metabolite number. SEM=standard error of the mean. n.s., not significant., blue nr 56=hydroxyferulic acid glucoside I.

	C-population				M-population				J-population			
Nr	**P-value**	**C-parents**	**S-parents**	**SEM**	**P-value**	**M-parents**	**J-parents**	**SEM**	**P-value**	**M-parents**	**J-parents**	**SEM**
3					0.044	5.96	5.93	0.017				
27					0.003	4.85	4.92	0.023				
56					0.014	4.81	4.93	0.045				
57					0.019	4.34	4.68	0.140				
80					0.043	5.92	6.01	0.043				
105	0.039	1.81	2.26	0.212								
129					0.017	3.13	2.22	0.363				
136	0.021	4.35	4.43	0.031								
153	0.032	4.49	4.63	0.062								
159					0.001	3.26	2.17	0.313				
223					0.018	2.48	1.88	0.242				
233					0.032	4.36	4.18	0.083				
234									1.4	4.51	4.40	0.055
237									1.4	4.63	4.53	0.048
239					0.042	5.44	5.31	0.065				
243					0.035	5.14	4.98	0.071				
251					0.014	4.62	4.40	0.087				

Supplemental table 12. General trans-generational effects on the metabolite levels in the C, S, M and J populations. Significance of trans-generational effects (P-value < 0.05) and the log(metabolite abundance) average of both progeny between unstressed grandparents (C or M-grandparents) and stressed grand-parents (S or J- grandparents). Nr corresponds with metabolite number. SEM = standard error of the mean. C-GP, C-grandparents; S-GP, S-grandparents; M-GP, M-grandparents; J-GP, J-grandparents; C, control; S, salt; M, mock; J, methyljasmonate; n.s., not significant.

	C-population				S-population				M-population				J-population			
Nr	***P***	**C-GP**	**S-GP**	**SEM**	***P***	**C-GP**	**S-GP**	**SEM**	***P***	**M-GP**	**J-GP**	**SEM**	***P***	**M-GP**	**J-GP**	**SEM**
3	n.s.				n.s.				0.049	5.93	5.99	0.03	n.s.			
12	n.s.				n.s.				0.036	4.78	4.92	0.06	n.s.			
14	n.s.				n.s.				0.003	5.88	5.98	0.03	n.s.			
33	n.s.				0.045	5.03	5.17	0.04					n.s.			
34	n.s.				0.016	5.32	5.21	0.03	n.s.				n.s.			
40	n.s.				n.s.				0.047	4.17	3.09	0.49	n.s.			
44	n.s.				0.047	5.07	4.74	0.10	n.s.				n.s.			
45	n.s.				n.s.				0.009	5.34	5.39	0.02	n.s.			
49	n.s.				n.s.				0.021	5.93	6.18	0.10	n.s.			
51	n.s.				n.s.				n.s.				0.050	5.00	5.20	0.06

Supplemental table 12 continues.

	C-population				S-population				M-population				J-population			
Nr	***P***	**C-GP**	**S-GP**	**SEM**	***P***	**C-GP**	**S-GP**	**SEM**	***P***	**M-GP**	**J-GP**	**SEM**	***P***	**M-GP**	**J-GP**	**SEM**
52	n.s.				n.s.				0.028	5.71	5.61	0.04	n.s.			
71	n.s.				0.04	4.84	4.77	0.02	n.s.				n.s.			
74	n.s.				n.s.				0.039	4.93	5.24	0.14	n.s.			
91	n.s.				n.s.				0.019	4.63	4.78	0.06	n.s.			
96	n.s.				n.s.				0.005	5.40	5.55	0.04	n.s.			
103	0.044	4.91	4.82	0.04	n.s.				n.s.				n.s.			
104	n.s.				n.s.				0.005	5.13	4.68	0.13	n.s.			
106	n.s.				n.s.				n.s.				0.021	4.16	4.27	0.03
109	n.s.				n.s.				n.s.				0.030	5.28	5.53	0.07
117	n.s.				n.s.				0.038	4.67	5.02	0.15	n.s.			
130	n.s.				n.s.				n.s.				0.014	5.32	5.52	0.05
131	n.s.				n.s.				0.043	4.83	4.68	0.07				
132	n.s.				n.s.				n.s.				0.026	4.33	4.50	0.05
135	n.s.				n.s.				0.008	5.47	5.57	0.03				
144	n.s.				n.s.				n.s.				0.019	6.53	6.62	0.02
145	n.s.				n.s.				n.s.				0.036	4.88	5.01	0.04
146	n.s.				n.s.				0.039	4.01	4.16	0.06				
148	n.s.				n.s.				n.s.				0.044	4.79	5.04	0.08
156	n.s.				n.s.				0.004	4.94	4.64	0.09	0.011	4.89	5.08	0.04
160	n.s.				n.s.				0.003	5.53	5.67	0.04				
166	n.s.				0.031	4.49	4.34	0.04	n.s.							
172	n.s.				n.s.				n.s.				0.012	4.73	5.02	0.07
179	n.s.				n.s.				0.040	5.59	5.88	0.13				
183	n.s.				n.s.				0.033	3.32	4.27	0.40				
188	n.s.				n.s.				n.s.				0.004	4.92	5.02	0.02
207	n.s.				n.s.				0.016	4.84	4.38	0.17	0.038	5.01	5.27	0.08
209	n.s.				n.s.				n.s.				0.033	4.83	4.98	0.04
210	n.s.				n.s.				0.031	5.94	6.24	0.13				
219	n.s.				n.s.				0.018	4.16	2.81	0.50				
232	n.s.				n.s.				n.s.				0.013	4.26	4.61	0.09
234	n.s.				n.s.				0.035	4.64	4.31	0.14				
244	n.s.				n.s.				0.027	4.51	4.60	0.04				
245	0.045	4.91	4.82	0.04	n.s.				n.s.							

Supplemental table 13. Parental effects salt: Top-25 differentially expressed genes.
Top-25 RNA sequencing results between CCCC and CSSC. Results are sorted on most significant P-value between CCCC and CSSC. Grey-coloured gene descriptions are involved in responses to salt stress (TAIR). *FDR-P*, FDR-adjusted *P*-value; Abbr., gene abbreviation.

	Base-Mean	**Base-Mean**	**fold-Change**				
Gene id	**CCCC**	**CSSC**	**CSSC/ CCCC**	***P*-value**	***FDR-P***	**Abbr.**	**Description**
AT1G35350	161.85	371	2.29	0	0.000486		
AT1G56510	665.2	1096.98	1.65	0	0.000486	WRR4	WHITE RUST RESISTANCE 4
AT3G04210	2700.67	3929.95	1.46	0	0.002345		disease resistance protein
AT4G30270	450.53	782.95	1.74	0	3.90E-05	XTH24	meristem-5
AT4G19520	743.89	1102.19	1.48	8.00E-06	0.033208		disease resistance protein
AT2G36885	431.65	273.36	0.63	1.00E-05	0.033208		unknown protein
AT2G39800	2065.05	1304.74	0.63	1.00E-05	0.033208	P5CS1	DELTA1-PYRROLINE-5-CARBOXY-LATE SYNTHASE 1
AT1G23480	271.67	440.95	1.62	2.20E-05	0.061178	CSLA3	CELLULOSE SYNTHASE-LIKE A3
AT1G08930	1466.57	2635.38	1.8	2.70E-05	0.067385	ERD6	EARLY RESPONSE TO DEHYDRA-TION 6
AT4G08930	180.46	296	1.64	3.20E-05	0.072477	ATAP-RL6	APR-like 6
AT4G13250	2671.71	1949.97	0.73	7.50E-05	0.154031	NYC1	short-chain dehydrogenase
AT2G26560	382.93	639.03	1.67	9.50E-05	0.180556	PLP2	PHOSPHOLIPASE A 2A
AT1G66940	561.77	761.7	1.36	0.000135	0.235011		protein kinase-related
AT3G59310	367.25	543.52	1.48	0.000187	0.282446		unknown protein
AT4G20260	2554.42	3606.7	1.41	0.000187	0.282446	PCAP1	DREPP plasma membrane polypep-tide family protein
AT1G63860	133.94	210.26	1.57	0.000207	0.290811		ATP binding
AT2G40300	424.78	263.92	0.62	0.000218	0.290811	FER4	ferritin 4
AT1G69840	651.47	887.81	1.36	0.000299	0.357721		band 7 family protein
AT3G05320	84.83	142.71	1.68	0.000299	0.357721		unknown protein
AT1G72180	830.49	1152.09	1.39	0.000318	0.36135		leucine-rich repeat transmembrane protein kinase
AT1G07890	9041.36	6838.35	0.76	0.000348	0.366488	MEE6	ascorbate peroxidase 1
AT1G64980	999.54	1337.35	1.34	0.000355	0.366488		unknown protein
AT4G18010	226.7	421.61	1.86	0.000376	0.366488	IP5PII	MYO-INOSITOL POLYPHOSPHATE 5-PHOSPHATASE 2
AT5G13630	7921.61	6253.77	0.79	0.00039	0.366488	GUN5	GENOMES UNCOUPLED 5
AT1G33415	55.19	99.81	1.81	0.000417	0.366488		other RNA

Supplemental table 14. Parental effects JA: Top-25 differentially expressed genes.
Top-25 RNA sequencing results between MMMM and MJJM. Results are sorted on most significant P-value. Grey-coloured genes are involved in organ development (TAIR). *FDR-P*, FDR-adjusted *P*-value; Abbr., gene abbreviation.

	BaseMean	BaseMean	fold-Change				
Gene id	**MMMM**	**MJJM**	**MJJM/ MMMM**	**P-value**	**FDR-P**	**Abbr.**	**Description**
AT4G22485	42.4	171.03	4.03	1.50E-05	0.255865		Encodes a Protease inhibitor
AT4G08150	47.32	109.52	2.31	2.20E-05	0.255865	KNAT1	KNOTTED-LIKE 1
AT3G54820	105.37	180.34	1.71	5.30E-05	0.413149	PIP2D	PLASMA MEMBRANE INTRINSIC PROTEIN 2;5
AT5G08000	212.75	333.53	1.57	0.00015	0.621461	PDCB2	GLUCAN ENDO-1,3-BETA-GLUCOSIDASE-LIKE PROTEIN 3
AT3G02640	154.04	229.79	1.49	0.00021	0.621461		unknown protein
AT3G51740	211.82	315.08	1.49	0.000223	0.621461	IMK2	INFLORESCENCE MERISTEM RECEPTOR-LIKE KINASE 2
AT1G77120	220.06	392.34	1.78	0.000231	0.621461	ATADH1	ALCOHOL DEHYDROGENASE 1
AT3G11520	34.77	63.64	1.83	0.000241	0.621461	CYCB1;3	CYCLIN B1;3
AT3G18000	455.29	629.26	1.38	0.000252	0.621461	XPL1	XIPOTL 1
AT1G80760	50.36	106.57	2.12	0.000266	0.621461	NLM7	NOD26-LIKE INTRINSIC PROTEIN 6;1
AT3G02120	73.22	115.93	1.58	0.000556	1		hydroxyproline-rich glycoprotein family protein
AT4G33790	41.21	76.19	1.85	0.000589	1	G7	ECERIFERUM 4
AT5G61480	279.93	394.96	1.41	0.000695	1	TDR	leucine-rich repeat transmembrane protein kinase
AT1G49320	10.81	29.9	2.77	0.000786	1	USPL1	BURP domain-containing protein
AT4G23800	378.64	516.61	1.36	0.000811	1	3xHMG-box2	high mobility group (HMG1/2) family protein
AT4G22505	24.92	87.74	3.52	0.000886	1		
AT2G19920	22.9	46.14	2.01	0.000945	1		RNA-dependent RNA polymerase family protein
AT4G10270	53.3	91.68	1.72	0.001103	1		wound-responsive family protein
AT5G23940	497.96	705.89	1.42	0.001221	1	PEL3	embryo defective 3009
AT3G17998	451.51	604.75	1.34	0.001225	1	CPuORF30	Conserved peptide upstream open reading frame 30
AT3G12145	264.66	411.51	1.55	0.001393	1	FLR1	enzyme inhibitor
AT5G66400	77.12	40.81	0.53	0.001429	1	RAB18	RESPONSIVE TO ABA 18
AT1G18370	135.15	196.15	1.45	0.001504	1	NACK1	HINKEL
AT3G15550	65.16	99.49	1.53	0.001782	1		unknown protein
AT3G07320	249.95	354.81	1.42	0.001835	1		glycosyl hydrolase family 17 protein

Chapter 8

General discussion

Every living organism, from the smallest unicellular bacterium to the largest giant sequoia, arises from the same genetic material, DNA. It is the organization of DNA into genes and the translation of genes into proteins that generate different cell types and the endless diversity of nature. However, the phenotype is not simply a product of DNA and proteins, but it is the complexity in which DNA and proteins are expressed in response to environmental and developmental cues that generates the phenotype. Natural diversity is not static due to the evolutionary forces of mutation and recombination that allow the genetic material to change with each generation. Due to recent advances in next generation sequencing methods, the genetic variation can be fully surveyed down to the nucleotide level. Hundreds of genotypes within different species have been resequenced, which enables the association of natural phenotypic variation within species with the smallest possible genetic variation, single nucleotide polymorphisms (SNPs) (Schneeberger and Weigel, 2011).

Genome-wide association studies (GWAS) associate phenotypic variation in a large population of individuals to genome-wide SNP variation. Since the first GWA study in plants by Atwell et al., 2010, several studies have followed and they have identified numerous SNPs that associate with hundreds of different phenotypes (Korte and Farlow, 2013). Although previously validated genes involved in flowering time, such as *FLOWERING LOCUS C* (*FLC*) and *FRIGIDA* (*FRI*), and simple Mendelian inherited traits, such as disease resistance, have been identified, a large part of the genetic variation for most quantitative traits remains unidentified in GWAS (Atwell et al., 2010; Brachi et al., 2010; Ingvarsson and Street, 2011). Because many important agronomic traits are highly quantitative, GWAS may fail to identify the causal variants for those traits (Ingvarsson and Street, 2011; Korte and Farlow, 2013). Furthermore, the verification of novel causal variants is complicated due to the high number of false-positives and false-negatives in GWAS.

Missing heritability

In human GWAS, the fraction of heritability that cannot be explained by genetic variants is known as the missing heritability (Manolio et al., 2009) and several reasons have been given to explain this missing heritability that could also be of importance for GWAS in plants (Gibson, 2011). Heritability estimates may be over-estimated due to interactions between loci that cannot be determined genome-wide with conventional GWA methods (Zuk et al., 2012). Furthermore, genotype-by-environment interactions may account for a substantial part of the heritability (Thomas, 2010; van Ijzendoorn et al., 2011) and epigenetic modifications, independent of genetic variation, may impact phenotypic variation (Eichler et al., 2010). It might be, however, that a large part of the missing heritability is simply hidden in the genetic architecture of quantitative traits with many loci of small effect or few, large-effect, rare alleles that cause phenotypic variation (Gibson, 2011). Genetic variants that determine disease risks, for instance, are likely to be deleterious and are thus kept at low frequency in the population, supporting the latter argument (McClellan et al., 2007; Zuk et al., 2014).

A study on human height, on the other hand, revealed that the estimated additive effects of all SNPs explain a large percentage of the heritability, supporting the former argument (Yang et al., 2011). This last study suggests that most of the genetic variation that causes the phenotypic variation is hidden below the thresholds used for GWA studies (Gibson, 2010). Both studies further indicate that the reason for missing heritability is most likely highly dependent on the investigated trait and its genetic architecture.

Chapter 2 of this thesis describes a GWA study on morphological traits in which the fraction of the phenotypic variation explained by the genetic loci was small in relation to the heritability estimates, suggesting missing heritability similar to human GWAS. This hypothesis was investigated with genomic selection models that used the 100,000 most informative SNPs for prediction of the morphological traits. The genomic selection analyses showed high correlations between observed and predicted phenotypic values with very small effect sizes for most traits. These results comply with the hypothesis that the heritable phenotypic variation is regulated by the additive and interaction effects of the individual loci, but that the allelic effect sizes are too small to be identified with GWAS. Quantitative traits in plants, such as the morphological traits described here, are indeed thought to be extremely complex due to their polygenic inheritance and strong dependence on the environment (Ungerer et al., 2002; Ungerer et al., 2003; Alonso-Blanco et al., 2005; Keurentjes et al., 2007; Ingvarsson and Street, 2011).

Because the genetic loci that cause the phenotypic variation have very small effect sizes, they are most likely hidden below the significance threshold (Gibson, 2010). When the significance threshold in our study was lowered to $-\log_{10}(P) = 4$, numerous plausible candidate genes that had been linked to the phenotypic trait either directly or indirectly in previous studies were detected. Most of the candidate genes experience purifying selection at their locus and they contain several non-synonymous SNPs (nsSNPs) or SNPs in the promoter region in linkage disequilibrium (LD) with the SNPs from the GWA study. These results suggest that quantitative morphological traits in plants are determined by many genes of small effect and that the heritability is hidden rather than missing.

Verification of candidate genes

Although plausible candidate genes were identified for the morphological traits by lowering the significance threshold, they need to be verified. Some of the genes could be false-positive associations due to non-causal SNPs that co-segregate with the trait values. Verification of candidate genes is most often performed through a number of different steps. Initially, next-generation sequencing data from the accessions can be used to calculate the extent of linkage disequilibrium (LD) surrounding the most significant SNPs. The genes within the LD interval may then be scanned for non-synonymous SNPs, frame shifts, insertions and deletions that are in strong LD with the SNPs from the GWA study (chapter 2 and

4). Moreover, haplotypes can be built on the basis of the resequence data and it can be evaluated whether they co-segregate with the trait values (chapter 2). T-DNA insertion lines are usually the next step in candidate gene confirmation as many different homozygous knock-outs (KOs) in the Col-0 background are available from the stock centers. Many mutant phenotypes, however, depend on the genetic background and might thus be concealed in KOs in the Col-0 background (Huang et al., 2012). Apart from that, gene redundancy can mask the mutant phenotype, as was the case in chapter 2 and 4, in which only double or multiple mutants of a gene family showed a mutant phenotype. Alternatively, mutant phenotypes can be studied in different genetic backgrounds using artificial microRNAs that target and silence the gene transcript (Ossowski et al., 2008). The artificial microRNAs can be transformed into different accessions to study the effect of the genetic background on the mutant phenotype. Moreover, the recent discoveries of zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) enable targeted gene modification and might further facilitate gene confirmation (Joung and Sander, 2013). Furthermore, analysis of variation in gene expression among the accessions can aid in the validation of candidate genes (chapter 2). Such gene expression analyses may also provide information on the effect of different genetic backgrounds on the phenotypic trait. Note, however, that regulatory variation can also affect protein function and stability, and in some cases regulatory variation does thus not lead to transcript variation. However, changes in gene expression through regulatory variation in the promoter are less likely to be negatively selected for than non-synonymous polymorphisms, because they are less likely to be deleterious, and they might thus be more common (Ingvarsson and Street, 2011). For further validation, the SNPs from GWAS may be compared to QTL intervals of the same traits in bi-parental QTL mapping populations (chapter 4) (Brachi et al., 2010). For Arabidopsis, more than sixty of such populations are available from the stock centers, and they have been phenotyped for numerous developmental and biochemical traits. Alternatively, the recent discovery of the *cenh3* mutant which enables the production of doubled haploids (DHs) within three generations can accelerate the creation of new mapping populations (Ravi & Chan, 2010). Using data generated by GWAS, divergent accessions can be selected based on the SNPs and the phenotypic data to create such populations. Because the onset of flowering initiates numerous developmental changes and many developmental traits correlate strongly with flowering time (chapter 2 and 5), accessions with similar flowering time are preferred provided that flowering time is not the trait of interest. Late flowering accessions that require vernalisation have rarely been studied by linkage analysis because the generation of such populations would take many years. The discovery of the *cenh3* mutant, however, greatly accelerates population development and a late flowering population might reveal as yet undiscovered genetic variation. If the interest of the GWA study is local adaptation, two accessions that differ for the trait of interest, but are not the extremes of the population, might be preferred. If the extremes of a population are used, QTL mapping could result in the detection of recently established, large-effect rare alleles that are not important in an

evolutionary context (Rockman, 2012). Rather two accessions should be chosen that are not too far away from each other in the population distribution, but differ for the SNPs of interest. Segregation within the population will result in phenotypic variation. However, because the SNPs of interest might be false-positives, the selection of accessions for follow-up studies remains elusive.

Recently, different approaches have been undertaken to combine the advantages of linkage and association mapping, using one or multiple founder lines that are crossed to several other accessions, such as the Multi-parent Advanced Generation Inter-Cross (MAGIC) lines and the Arabidopsis Multi-Parent RILs (AMPRILs) (Huang et al., 2011; Kover and Mott, 2012). Resolution and allelic diversity in such populations are improved compared to RIL populations, while mapping power and the detection of epistasis are increased compared to GWAS (Huang et al., 2011; Huang et al., 2012; Huang et al., 2013). With the advancements made in the creation of DHs, the combination of GWAS and QTL analysis can be further enhanced by creating numerous DH populations in intelligent designs in which genetic diversity, resolution and mapping power are maximized.

Although the above-mentioned analyses will provide more confidence on the association between candidate gene and phenotype, complementation is needed for complete confirmation. Quantitative or transgenic complementation in which different alleles at a certain locus are examined in a knock-out or other isogenic background may be used for confirmation (Mackay, 2001; Weigel, 2012). In quantitative complementation, the different alleles are analysed in the hemizygous state in which the accessions, NILs or RILs are crossed to a loss-of-function mutant. If the different alleles induce different phenotypes in the mutant crosses, they are most likely causal for the trait variation (Mackay, 2001). In transgenic complementation, the genomic sequence of different alleles including its own promoter, or otherwise fused to a constitutive or inducible promoter, is transferred into the KO or any other isogenic background to study the specific allelic effects. The genomic sequence of one allele may also be fused to the promoter of the other allele, and vice versa, and other forms of transformation are possible to confirm causality between allelic variation and trait variation.

A system-wide genetical genomics approach

Because the morphological traits analysed in chapter 2 are extremely polygenic, lowering the threshold will not result in the identification of all genetic variants that are involved in the regulation of the traits. Furthermore, some of the identified SNPs are probably false-positive associations and confirmation of all candidate genes is highly laborious and thus not feasible. Moreover, regulation of morphology occurs at multiple intermediate levels that can be studied individually in so-called genetical genomics. The integration of transcript, protein and/ or metabolite data with morphological traits may partition the phenotypic variation

in its underlying components and thus assist in candidate gene identification and selection (Jansen and Nap, 2001; Keurentjes et al., 2006; Keurentjes et al., 2007; Fu et al., 2009).

Chapter 4 describes a GWA study in which several structural components, amino acids, sugars, organic acids and enzymes of carbon and nitrogen metabolism were analysed together with plant biomass. This allows the dissection of heritability for plant biomass into some of its underlying components. Indeed, many pleiotropic regulators were identified with mostly opposite effects on biomass and primary metabolism. A good example is the genome-wide association of *ACCELERATED CELL DEATH 6* (*ACD6*) with biomass and several components of primary metabolism, providing strong support that natural variation at this locus contributes to developmental variation. Higher activity of the *ACD6* allele increases resistance to microbial infection and herbivory, but it induces necrosis and reduces biomass formation (Todesco et al., 2010). The higher enzyme activities and higher levels of protein and amino acids in accessions with higher *ACD6* activity suggest higher metabolic rates to support constitutive defenses at the expense of growth (chapter 4).

Comparison of the heritabilities between chapter 2 and 4 reveals that the heritabilities for morphological traits and biomass are much higher than for metabolic traits. This is in accordance with increased robustness at higher phenotypic levels, which was observed previously (Fu et al. 2009). The robustness at the morphological level corresponds well with an infinitesimal model with many loci of small effect that together ensure the progress of development, despite small changes in the environment. The decreased robustness at lower phenotypical levels should allow us to narrow down the number of candidate genes. Although the morphological traits were analysed under long day conditions (chapter 2) and the metabolic traits under short-day conditions (chapter 4), significant Bonferroni-corrected correlations were found between the two experiments. Flowering time correlated positively with total amino acids and UDP-glucose pyrophosphorylase (UGP) activity, but negatively with biomass. Furthermore, the petiole to leaf length ratio correlated positively with UGP activity and relative growth rate correlated negatively with total amino acids and cytosolic phosphoglucose isomerase. Both rosette branching and relative growth rate correlated positively with biomass. Although these were the only significant Bonferroni corrected correlations, there are many negative correlations between relative growth rate and primary metabolism and positive correlations between flowering time and primary metabolism. When the SNPs with $-\log_{10}(P) > 3$ were compared between morphological and metabolic traits, 313 candidate genes could be assigned in both datasets, among which obvious ones such as *FLC*, *PHYTOCHROME B* (*PHYB*), *DELAY OF GERMINATION 1* (*DOG1*), *PHOSPHOGLUCOSE ISOMERASE 1* (*PGI1*) and *ACS11*. The majority of these genes have been implicated to play a role in flowering, further emphasizing the large effect of flowering time in studies on natural variation (Huang et al., 2013). Interestingly, variation in protein, glucokinase activity and the ratio of petiole to leaf length (chapter 2) associated significantly

with *ACS11* variation, providing additional evidence that natural variation at this locus plays a role in developmental processes. Although system-wide genetical genomics reveals many additional candidate genes, the overlap between metabolism and morphology can assist in narrowing down the number of candidate genes found with GWAS.

A further partitioning of the heritability can be revealed by RNA sequencing and bisulphite sequencing (Schmitz et al., 2013). Variation in the transcriptome and methylome might identify novel regulators of phenotypic traits or aid in the confirmation of GWA candidate genes. Epigenetic polymorphisms remain unidentified with conventional GWA methods, but they might explain a considerable amount of the broad-sense heritability when the epigenetic polymorphisms are shared between individuals of the same genotype (Weigel, 2012; Korte and Farlow, 2013). If such changes are not shared, but they are formed through stochastic processes such as bet hedging (Simons, 2011; Herman et al., 2013), they might explain part of the residual variation.

As indicated in the previous sections on missing heritability, GWAS have several limitations for the identification of genes for quantitative traits. Two of the main drawbacks of GWAS are the inability to identify rare, large-effect variants and many variants of small effect. The latter problem is alleviated by lowering the significance threshold at the expense of an increase in false-positives (chapter 2 and 4) and by dividing the heritability of robust phenotypes into the less complex underlying components of the primary metabolic network (chapter 4). If rare variants cause the phenotypic variation, they most likely remain unidentified in GWAS. One possible solution to identify rare variants is the use of recombinant inbred lines in which the power to identify such variants is greatly enhanced because the applied crossing scheme ensures that the alleles are at equal frequency in the population. Alternatively, genes of small effect or rare variants might be identified by increasing the effective population size and/ or sampling in local populations with reduced population structure and heterogeneity. Such strategies might increase the power to find the causal variants, and especially variants important for local adaptation (Korte and Farlow, 2013). Moreover, resolution of GWA mapping might be improved in the near future when the full genome sequences of more than 1000 Arabidopsis accessions become available (www.1001genomes.org).

Hard versus soft inheritance

The Lamarckian idea of soft inheritance, i.e. that an organism can pass on environmentally acquired characteristics to subsequent generations, was refuted by Neo-Darwinians after the rediscovery of Mendel's laws and the unsuccessful attempts to prove soft inheritance (Mayr, 1982). Hard inheritance is defined as the constant, stable inheritance of hereditary material to subsequent generations, modified only through stochastic and random mutation, and was believed to be the sole, heritable cause of phenotypic variation for the major part of the 20th century (Richards, 2006). Recent discoveries in the field of epigenetics,

however, suggest that epigenetic modifications can be induced by the environment and that they can be stably transmitted from one generation to the next, reviving the idea of soft inheritance (Verhoeven et al., 2010; Mirouze and Paszkowski, 2011; Becker and Weigel, 2012; Sahu et al., 2013). It is, however, fiercely debated whether such epigenetic modifications are pure epialleles, independent of genetic variation, whether they are stochastically or environmentally induced and whether they are adaptive (Richards, 2006; Pecinka and Scheid, 2012; Turck and Coupland, 2013; Verhoeven and Preite, 2013). The main difficulty when studying the effect of epigenetic variation on phenotypic variation is the presence of DNA sequence variation. Recent developments together with reduced costs of sequencing technologies have enabled new opportunities to disentangle epigenetic variants from sequence variants. Whole-genome bisulphite sequencing, for example, allows the comparison between DNA sequence and methylome variation. In several lines generated for 30 generations by single seed descent from a common Arabidopsis founder line, the single methylation polymorphism rate was found to be about four times higher than the genetic mutation rate (Becker et al., 2011; Schmitz et al., 2011). The stochastic methylome variation among the lines most likely arises through the imperfect resetting of epigenetic marks through cell division and meiosis (Schmitz et al., 2011). More recently, genome, methylome and transcriptome variation were analysed in 150 different Arabidopsis accessions, and widespread natural variation in DNA methylation was found (Schmitz et al., 2013). The methylation variation was in many cases linked to genetic variation, which indicates that a genetic variant is most likely causal for the variation in DNA methylation. An alternative strategy to circumvent sequence variation is the use of asexual species in which there is less confounding from genetic variation (Verhoeven and Preite, 2013). As asexual species do not undergo meiosis, epigenetic resetting is less likely and epigenetic variation might be the main source of phenotypic variation as there is no recombination and segregation of alleles in such species (Verhoeven and Preite, 2013). Alternatively, epigenetic recombinant inbred lines (epiRILs) consisting of homozygous mosaic lines with stable, differentially methylated regions originating from either a wild-type or a hypomethylated mutant parent, such as *ddm1*, could be used for studying epigenetic variation. The mutant *ddm1* allele was out-crossed by a back-cross to Col-0 and subsequent selection to ensure that methylation variation was the main source of variation in this population (Johannes et al., 2009).

Chapter 5 and 6 describe the epigenetic basis of morphology and phenotypic plasticity in different environments (chapter 5) and secondary metabolism in different tissues (chapter 6) using the *ddm1-2*-derived epiRIL population (Johannes et al., 2009). Consistent with previous analyses of the epiRIL populations (Johannes et al., 2009; Reinders et al., 2009; Roux et al., 2011; Latzel et al., 2012; Zhang et al., 2013) there was abundant variation for morphological traits under different conditions. Chapter 6 also describes for the first time the strong variation for secondary metabolites of leaves and flowers in the epiRILs. The variation

was highly heritable and comparable to heritabilities detected in conventional genetic RIL populations. This suggests that epigenetic variation can be an important source for heritable, phenotypic variation. Differentially methylated regions (DMRs) within the epiRIL population were used as physical markers to generate an epigenetic map (Colome-Tatche et al., 2012). The DMRs of the epigenetic map associated significantly with the phenotypic trait values and different epigenomic regions could be assigned to various phenotypes. Many of the QTLs for morphology, plasticity and metabolism overlapped, suggesting major pleiotropic epigenetic regulation. The co-location between metabolic and morphological QTLs indicates that unraveling the epigenetic regulation of metabolism can provide insight into the epigenetic regulation of morphology, provided that they are controlled by the same underlying epigenetic factors. Given the correlations between the metabolites and the morphological phenotypes, pleiotropic epigenetic regulation is expected. Moreover, QTLs are usually quite accurate despite the rather large QTL intervals (Price, 2006).

Although the epiRILs were designed to differentiate between epigenetic and genetic variation, genetic variation through transposable element (TE) insertion cannot be ruled out. Because the major pleiotropic QTLs found in our study co-locate with the confirmed epigenetic loci (QTLepi) detected in another epigenetic QTL study on flowering time and root length, these QTLs are most likely caused by epigenetic factors (Cortijo et al., 2014). Furthermore, opposite-effect QTLs were found for leaf area, which indicates that DNA hypomethylation can both increase and reduce leaf area depending on the site of methylation. These findings illustrate that adaptive phenotypic variation can be induced through epigenetic variation that is stable for at least eight generations, and can thus contribute to (short-term) evolutionary adaptation. Moreover, comparison of DMRs between the epiRILs and 138 natural accessions of Arabidopsis revealed that approximately 30% of the epiRIL DMRs are also present in nature (Cortijo et al., 2014). This suggests that these experimentally induced DMRs are also significant in natural settings and that they might contribute to local adaptation.

DNA methylation might be relevant in stressful environments. It is well-known that stressful environments can induce hypo- and hypermethylation of DNA in different plant species, concomitant with variation in gene expression (Choi and Sano, 2007; Chinnusamy and Zhu, 2009; Grativol et al., 2012). However, thus far it is not known whether DNA methylation induced by stress is adaptive and whether it is regulated or stochastic. In our analyses in chapter 5, a pleiotropic QTL was found on chr 4 that includes the salt tolerance gene *HIGH AFFINITY* K^+ *TRANSPORTER 1* (*HKT1*) (Davenport et al., 2007). Both methylation and small expression differences for this gene were found between *ddm1-2* and Col-0 and methylation variation in the promoter region of the gene associated with leaf area in the epiRILs where reduced methylation decreased leaf area, especially under saline conditions. The DMR in the promoter region coincides with a small RNA target region that is heavily methylated in

Col-0 (Baek et al., 2011). Removal of methylation at this region increases *HKT1* expression and salt sensitivity, consistent with our results (Baek et al., 2011). These findings illustrate that stable DNA methylation might be adaptive, especially under stressful environments.

Chapter 7 elaborates on these findings and describes the transgenerational inheritance of stress-induced epigenetic modifications that cause phenotypic variation at the transcript, metabolite and morphological trait level. Convincing evidence was produced for the specific environmental induction of epigenetic variation at multiple phenotypic levels. Furthermore, the stress-induced epigenetic changes were transmitted over more than one stress-free generation, and they appear to be adaptive. Biotic, methyljasmonate stress and abiotic, salt stress induced stress-specific transgenerational inheritance of variation in gene expression, metabolite abundance and morphology in unstressed progeny. When parents or grandparents had been grown under saline conditions, salt-stress responsive genes and cell wall maintenance genes were significantly up or down regulated in the progeny. Furthermore, transgenerational effects were observed that increased biomass in the stressed progeny of stressed grandparents, suggesting that the transgenerational inherited phenotypes are adaptive. It needs to be proven with further analyses, however, that these transgenerational inherited phenotypes are strictly determined by epigenetic modifications, and not initiated or maintained by genetic variation.

Phenotypic plasticity

It is thought that environmentally-induced epigenetic changes are of special importance in fluctuating environments, because epigenetic variants, in contrast to DNA sequence variants, can be reversed (Rando and Verstrepen, 2007). Epigenetic modifications might unlock phenotypic plasticity, and could thus enhance adaptation in such environments. However, if the environment changes very rapidly, between one generation and the next, the environmental sensitivity of the epigenetic loci might result in immediate reversion and the epigenetic chromatin marks are not inherited. It is reasoned that transgenerational inheritance of epigenetic regulation may evolve when the environment fluctuates over somewhat longer timescales, so that epigenetic marks can be stably transmitted (Furrow and Feldman, 2013). If the environment, however, changes over much longer timescales, in the order of thousands of generations, genetic mutations are most likely preferred.

In chapter 5, the phenotypic plasticity was calculated as the absolute difference in average trait values between control and saline conditions in the epiRILs. The majority of epiRILs showed stronger plasticity under changing conditions than wild-type Col-0. Furthermore two pleiotropic QTLs for phenotypic plasticity that co-located with the morphological trait QTLs were found. This indicates that DNA hypomethylation can unlock phenotypic plasticity, and that the plasticity is most likely regulated by the same loci that regulate the morphological traits. It is reasonable to assume that by removing silencing from TEs and

genes, the *ddm1-2* induced hypomethylation increases the environmental sensitivity of such genes, and therefore increases plasticity. It must be noted that the variation observed here is due to stable, heritable variation, and not due to plastic de novo variation induced by the stress (Richards et al., 2010). Phenotypic plasticity was also enhanced in the epiRILs in response to drought and nutrient stress, which suggests that phenotypic plasticity is a general response in the epiRILs (Zhang et al., 2013). Phenotypic plasticity is also observed in conventional RIL populations and most often the plasticity QTLs overlap with the trait value QTLs (Lacaze et al., 2009; Tetard-Jones et al., 2011; El-Soda et al., 2014). It is difficult to compare the epigenetic with the genetic contribution to phenotypic plasticity as epigenetic variation might contribute to phenotypic variation in conventional RIL populations when the epigenetic variation associates with the genetic markers (Schmitz et al., 2013). Recently, a RIL population in soybean was analysed for genome, methylome and transcriptome variation and indeed, the majority of the DMRs co-segregated with the genetic background, and for 90% of the DMRs, genetic QTLs explaining the methylation variation were identified (Schmitz et al., 2013). Similar results were obtained for natural accessions of Arabidopsis (Schmitz et al., 2013). Although this suggests that most epigenetic variants are dependent on genetic variation, rare examples of DMRs not linked to genetic variation were identified, and such DMRs could be pure epialleles (Schmitz et al., 2013; Schmitz et al., 2013). In our study, the pleiotropic QTL on chr. 1 included two plasticity QTLs that overlapped with the epigenetic QTL interval from Cortijo et al., 2014 and suggests that pure epigenetic alleles determine phenotypic plasticity and morphological trait variation at this QTL. These findings illustrate that stable, epigenetic modifications can increase phenotypic plasticity. It needs, however, additional experiments to prove that the environment can induce such changes, that they are adaptive and that such environmentally induced changes are inherited to future generations. In violets, variation in DNA methylation was associated with different levels of herbivory in natural settings, suggesting local adaptation via epigenetic phenotypic plasticity (Herrera and Bazaga, 2010). Moreover in a study on nectar-associated yeast, sugar composition and concentration initiated different non-random DNA methylation states and when DNA methylation was chemically inhibited, proliferation was repressed on media containing sugar. DNA methylation thus allowed successful growth in extreme sugar environments and enhanced phenotypic plasticity (Herrera et al., 2012). These studies indicate that DNA methylation variation can be adaptive in certain environments, but do not provide conclusive evidence for environmentally induced soft inheritance.

Besides phenotypic plasticity, chapter 5 showed that developmental instability was generally increased in the epiRILs and that several genomic regions associated with developmental instability, indicating epigenetic regulation. Increased developmental instability conforms to an adaptive strategy called bet hedging. Bet hedging is described as the ability of a genotype to produce phenotypically diverse offspring where variation in fitness is reduced at the cost of overall fitness (Simons, 2011). This strategy could be adaptive in fluctuating environments,

because at least some individuals within the population will thrive (Herman et al., 2013). Since bet hedging is found in adverse environments in nature (Evans et al., 2007; Venable, 2007), and because epigenetic variation can be induced by different environmental stresses (Chinnusamy and Zhu, 2009; Verhoeven et al., 2010; Grativol et al., 2012) and because DNA hypomethylation increases developmental instability for many morphological traits (chapter 5), epigenetic mechanisms might determine bet hedging strategies.

To claim soft inheritance, it must be proven that the epigenetic polymorphisms are independent of genetic variation and that they are induced directly by the environment in a non-random process. Most epi-alleles discovered so far are not pure epi-alleles, but rather depend on some sort of genetic variation (Becker and Weigel, 2012; Pecinka and Scheid, 2012; Turck and Coupland, 2013) or are induced stochastically (Becker et al., 2011; Schmitz et al., 2011). The analyses of chapter 5, 6 and 7 and the study of Cortijo et al., 2014 provide strong evidence that pure epi-alleles can regulate both morphological and metabolic trait variation (chapter 5 and 6), that they can be specifically induced by the environment (chapter 7), that they can be found in nature (Cortijo et al., 2014), and that they can be adaptive (chapter 5 and 7). These analyses together strongly suggest that soft inheritance can play an important role in determining phenotypic variation between individuals of the same species. However, it must be emphasized that numerous DNA sequence variants have been identified in crop and wild species that cause naturally occurring phenotypes (Weigel, 2012; Alonso-Blanco and Méndez-Vigo, 2014). Although epigenetics is an interesting field of research and a better understanding of its occurrence and mechanisms might contribute to improved agricultural crop growth and productivity, only a small number of epialleles have been identified so far, and the majority is dependent on genetic initiation.

To grow fast or to grow strong

Chapter 3 describes the metabolic trade-off between investment of resources for growth and reproduction, or for defense and survival. When pathogen pressure is low or absent, plants are expected to evolve increased competitive ability, while high pathogen pressure would favor the evolution of resistance and tolerance (Hare, 2012). This balance between growth and defense may be seen in the light of primary and secondary metabolites, because primary metabolites are mostly involved with growth, development and reproduction, while secondary metabolites are mostly involved with responses to changes in the environment. Moreover, primary metabolites are often used as substrate, cofactor or ligand in enzymatic reactions for the biosynthesis of secondary metabolites, and these primary metabolic resources are, in such cases, thus diverted away from growth related processes (Logemann et al., 2000; Bolton, 2009). Furthermore, energy required for the activation of the defense pathways is merely generated via primary metabolic pathways (Scheideler et al., 2002). Interestingly, chapter 4 describes negative correlations between primary metabolites and growth related traits (RGR and FW), suggesting a trade-off between investment in primary

metabolism and growth, quite contrary to expectation (chapter 3 and 4). Moreover, the positive associations between primary metabolism and stress-related genes, such as *ACD6* and *HVA22E* further illustrate that investment in primary metabolism contrasts with fast growth. The stress related associations, however, also suggest that primary metabolites are needed in harsh environments, most likely as resources for stress tolerance and pathogen resistance. Consistent with this hypothesis, transcripts involved in primary metabolic regulation, such as TCA cycle and biosynthesis of amino acids increased upon exposure to various abiotic and biotic stresses, which suggests indeed that primary metabolites provide the energy for defense responses (Less et al., 2011; Etalo et al., 2013). Furthermore carbohydrates and amino acids positively regulate the expression of defense-related genes and may trigger defense responses (Rojas et al., 2014). Therefore, plants that experience high pathogen and herbivore pressure might increase their primary metabolic status to enhance defense responses in the case of attack (Bolton, 2009).

Flowering time on the other hand correlated positively with the primary metabolites, which suggests that fast growing accessions use their carbon resources very efficiently to enable fast growth, early flowering and fast reproduction. In stressful environments, these fast growing accessions are most likely outcompeted by their defense-oriented neighbors, while in less stressful environments, they grow faster and outcompete their slow growing neighbors. This indicates that environmental conditions act as selective pressures that can shape the natural variation in plant metabolism, growth and morphology, and drive local adaptation (Baxter et al., 2010; Hare, 2012; Juenger, 2013; Keith and Mitchell-Olds, 2013). Further studies that analyse natural variation in both primary and secondary metabolites, possibly on different analytical platforms and in different environments, integrated with plant growth analyses, are needed to present a more comprehensive overview of the relationships between growth, metabolism and adaptation. Such studies would also contribute to a better understanding of the possible trade-off between investment in primary and secondary metabolism.

Ddm1-induced hypomethylation caused both qualitative and quantitative variation in secondary metabolism of leaves and flowers (chapter 6). For a number of metabolites, several QTLs were found in leaves and flowers that coincided with the morphological QTLs and the epigenetic QTLs published earlier (Cortijo et al., 2014) and suggest pleiotropic, tissue-specific regulation of secondary metabolites by epigenetic marks. Interestingly, the glucosinolates in the flowers correlated positively with leaf area and most other morphological traits and had the same QTL-effect sign, while all leaf metabolites had opposite QTL-effect signs and correlated negatively with most morphological traits. In general, investment in secondary leaf metabolism was associated with reduced growth rate and late flowering, but investment in secondary flower metabolism was associated with increased growth rate and early flowering (chapter 6), suggesting that early flowering plants grow fast and invest more in flower metabolism, while late flowering plants, grow slower and invest more in

leaf metabolism. The chemical defense theory suggests that the allocation of metabolites depends on the costs associated with loss of the organ and the likelihood that the organ will be attacked (McKey, 1974). Because the floral organs have greater fitness value, it was suggested that plants allocate more defensive compounds to these organs (McCall and Irwin, 2006). Indeed, defense compounds are most concentrated in the reproductive organs of Arabidopsis (Brown et al., 2003) and floral herbivory is common in natural settings and affects reproductive success (McCall and Irwin, 2006). Nevertheless, a recent meta-analysis found that young leaves had more defense compounds than older leaves, but flowers were not better protected than leaves (McCall and Fordyce, 2010). The flowers of *Brassica nigra*, however, had five times higher concentrations of different glucosinolates than leaves, but nevertheless the growth rate of caterpillars was much higher on flowers than on leaves (Smallegange et al., 2007). These results indicate that the allocation of defense compounds to different tissues is not straightforward. It is conceivable that these different strategies evolve under different herbivore pressures, where high herbivore pressures move selection towards high vegetative defense investment and slow growth, while low herbivore pressures moves selection towards fast growing plants that focus on their reproductive outcome to improve their competitive ability.

Epigenetic regulation might provide an evolutionary advantage in fluctuating environments when different defense strategies optimize reproductive success between generations, or when a sudden transition in the environment is likely to occur (Furrow and Feldman, 2013; Herman et al., 2013). Epigenetic modifications could be beneficial in such environments, as they are relatively inexpensive, and they allow resetting when the environment changes.

Conclusions and future perspectives

The regulation of the flexible plant phenotype by the interaction between the stable genetic loci and the ever-changing environment is becoming clearer through recent breakthroughs in sequencing and 'omics' technologies (De Vos et al., 2007; Schneeberger and Weigel, 2011; Schmitz et al., 2013). These technological advancements allow the association between transcriptome, metabolome and phenome with single nucleotide polymorphisms (SNPs) in genome-wide association studies (GWAS) (Nordborg and Weigel, 2008). It is shown in the above sections that GWA studies have several limitations to identify the causal variants for quantitative traits in natural populations, but they do pave the way for further analyses and therefore greatly facilitate causal variant identification. By lowering the stringent significance threshold of GWAS (chapter 2 and 4) and by partitioning the heritability into the underlying components, such as enzymes, proteins and metabolites (chapter 4), GWAS identified numerous plausible candidate genes for quantitative traits. Intelligent crossing designs that combine the strengths of GWAS and QTL analyses by maximizing genetic diversity, mapping power and resolution will most likely increase the likelihood of identifying such variants. The recent discovery of the *cenh3* mutant in Arabidopsis opens up new exciting opportunities

for the accelerated creation of doubled haploids (DHs) that can be used in such designs (Ravi and Chan, 2010; Wijnen and Keurentjes, 2014). If such studies are optimized for crop species, plant breeding will most likely see a revolution in the identification of candidate genes for many agronomic traits. Furthermore, epistatic interactions that are extremely difficult to identify in GWAS can be studied in recently developed chromosome substitution lines, in which chromosomes from different genotypes can be studied in an isogenic background (Kooke et al., 2012; Wijnker et al., 2012). And finally, the environment has an unsurpassed impact on quantitative traits which should not be underestimated. Therefore, future studies need to be performed in natural environments and under conditions that mimic local environments to discover genes that are important for local adaptation (Anderson et al., 2014).

The suspected hard line between the stable genotype and the plastic phenotype is softened due to recent advances in the field of epigenetics that suggest the stable, adaptive inheritance of environment-induced epigenetic modifications to subsequent generations (Richards, 2006). It is however widely disputed whether such epigenetic polymorphisms act completely independent of genetic variation, whether they are adaptive, whether they are present in nature and whether they are directly induced by the environment. In chapter 5, 6 and 7 we find highly convincing evidence that certain epialleles are stable for several generations (chapter 5 and 6), induced by the environment (chapter 7), adaptive for some morphological traits (chapter 5 and 7), and that they can increase phenotypic plasticity (chapter 5). The presence of 30% of the epiRIL DMRs in nature further suggests that epialleles could also be important in natural settings (Cortijo et al., 2014).

In conclusion, the genetic and epigenetic analyses in this thesis help to understand the quantitative architecture of metabolic and morphological plant traits that are regulated by multiple, additive genetic loci, epigenetic variation, the environment and the interactions among them.

References

Alonso-Blanco C, Méndez-Vigo B (2014) Genetic architecture of naturally occurring quantitative traits in plants: an updated synthesis. Current Opinion in Plant Biology 18: 37-43

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

Anderson JT, Wagner MR, Rushworth CA, Prasad K, et al (2014) The evolution of quantitative traits in complex environments. Heredity 112: 4-12

Atwell S, Huang YS, Vilhjalmsson BJ, Willems G, et al (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature 465: 627-631

Baek D, Jiang J, Chung JS, Wang B, et al (2011) Regulated AtHKT1 gene expression by a distal enhancer element and DNA methylation in the promoter plays an important role in salt tolerance. Plant Cell Physiol 52: 149-161

Baxter I, Brazelton JN, Yu D, Huang YS, et al (2010) A coastal cline in sodium accumulation in Arabidopsis thaliana is driven by natural variation of the sodium transporter AtHKT1;1. PLoS Genet 6: e1001193

Becker C, Hagmann J, Muller J, Koenig D, et al (2011) Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. Nature 480: 245-249

Becker C, Weigel D (2012) Epigenetic variation: origin and transgenerational inheritance. Curr Opin Plant Biol 15: 562-567

Bolton MD (2009) Primary Metabolism and Plant Defense-Fuel for the Fire. Molecular Plant-Microbe Interactions 22: 487-497

Brachi B, Faure N, Horton M, Flahauw E, et al(2010) Linkage and Association Mapping of Arabidopsis thaliana Flowering Time in Nature. PLoS genetics 6

Brown PD, Tokuhisa JG, Reichelt M, Gershenzon J (2003) Variation of glucosinolate accumulation among different organs and developmental stages of Arabidopsis thaliana. Phytochemistry 62: 471-481

Chinnusamy V, Zhu JK (2009) Epigenetic regulation of stress responses in plants. Curr Opin Plant Biol 12: 133-139

Choi CS, Sano H (2007) Abiotic-stress induces demethylation and transcriptional activation of a gene encoding a glycerophosphodiesterase-like protein in tobacco plants. Mol Genet Genomics 277: 589-600

Colome-Tatche M, Cortijo S, Wardenaar R, Morgado L, et al (2012) Features of the Arabidopsis recombination landscape resulting from the combined loss of sequence variation and DNA methylation. Proceedings of the National Academy of Sciences of the United States of America 109: 16240-16245

Cortijo S, Wardenaar R, Colomé-Tatché M, Gilly A, et al (2014) Mapping the Epigenetic Basis of Complex Traits. Science 343: 1145-1148

Davenport RJ, Munoz-Mayor A, Jha D, Essah PA, et al (2007) The Na+ transporter AtHKT1;1 controls retrieval of Na+ from the xylem in Arabidopsis. Plant Cell and Environment 30: 497-507

De Vos RCH, Moco S, Lommen A, Keurentjes JJB, et al (2007) Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. Nature Protocols 2: 778-791

Eichler EE, Flint J, Gibson G, Kong A, et al (2010) VIEWPOINT Missing heritability and strategies for finding the underlying causes of complex disease. Nature Reviews Genetics 11: 446-450

El-Soda M, Boer MP, Bagheri H, Hanhart CJ, et al (2014) Genotype-environment interactions affecting preflowering physiological and morphological traits of Brassica rapa grown in two watering regimes. Journal of Experimental Botany 65: 697-708

Etalo DW, Stulemeijer IJ, van Esse HP, de Vos RC, et al (2013) System-Wide Hypersensitive Response-Associated Transcriptome and Metabolome Reprogramming in Tomato. Plant physiology 162: 1599-1617

Evans MEK, Ferriere R, Kane MJ, Venable DL (2007) Bet hedging via seed banking in desert evening primroses (Oenothera, Onagraceae): Demographic evidence from natural populations. American Naturalist 169: 184-194

Fu J, Keurentjes JJ, Bouwmeester H, America T, et al (2009) System-wide molecular evidence for phenotypic buffering in Arabidopsis. Nature genetics 41: 166-167

Furrow RE, Feldman MW (2013) Genetic variation and the evolution of epigenetic regulation. Evolution

Gibson G (2010) Hints of hidden heritability in GWAS. Nat Genet 42: 558-560

Gibson G (2011) Rare and common variants: twenty arguments. Nat Rev Genet 13: 135-145

Grativol C, Hemerly AS, Ferreira PC (2012) Genetic and epigenetic regulation of stress responses in natural plant populations. Biochim Biophys Acta 1819: 176-185

Hare JD (2012) How Insect Herbivores Drive the Evolution of Plants. Science 338: 50-51

Herman JJ, Spencer HG, Donohue K, Sultan SE (2013) How stable 'should' epigenetic modifications be? Insights from adaptive plasticity and bet hedging. Evolution

Herrera CM, Bazaga P (2010) Epigenetic differentiation and relationship to adaptive genetic divergence in discrete populations of the violet Viola cazorlensis. New Phytologist 187: 867-876

Herrera CM, Pozo MI, Bazaga P (2012) Jack of all nectars, master of most: DNA methylation and the epigenetic basis of niche width in a flower-living yeast. Molecular ecology 21: 2602-2616

Huang XQ, Ding J, Effgen S, Turck F, et al (2013) Multiple loci and genetic interactions involving flowering time genes regulate stem branching among natural variants of Arabidopsis. New Phytologist 199: 843-857

Huang XQ, Effgen S, Meyer RC, Theres K, et al (2012) Epistatic Natural Allelic Variation Reveals a Function of AGAMOUS-LIKE6 in Axillary Bud Formation in Arabidopsis. Plant Cell 24: 2364-2379

Huang XQ, Paulo MJ, Boer M, Effgen S, et al (2011) Analysis of natural allelic variation in Arabidopsis using a multiparent recombinant inbred line population. Proceedings of the National Academy of Sciences of the United States of America 108: 4488-4493

Ingvarsson PK, Street NR (2011) Association genetics of complex traits in plants. New Phytol 189: 909-922

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

Johannes F, Porcher E, Teixeira FK, Saliba-Colombani V, et al (2009) Assessing the Impact of Transgenerational Epigenetic Variation on Complex Traits. PLoS genetics 5

Joung JK, Sander JD (2013) TALENs: a widely applicable technology for targeted genome editing. Nat Rev Mol Cell Biol 14: 49-55

Juenger TE (2013) Natural variation and genetic constraints on drought tolerance. Current Opinion in Plant Biology 16: 274-281

Keith R, Mitchell-Olds T (2013) Genetic variation for resistance to herbivores and plant pathogens: hypotheses, mechanisms and evolutionary implications. Plant Pathology 62: 122-132

Keurentjes JJ, Fu J, de Vos CH, Lommen A, et al (2006) The genetics of plant metabolism. Nature genetics 38: 842-849

Keurentjes JJ, Fu J, Terpstra IR, Garcia JM, et al (2007) Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. Proc Natl Acad Sci U S A 104: 1708-1713

Keurentjes JJB, Bentsink L, Alonso-Blanco C, Hanhart CJ, et al (2007) Development of a near-isogenic line population of Arabidopsis thaliana and comparison of mapping power with a recombinant inbred line population. Genetics 175: 891-905

Kooke R, Wijnker E, Keurentjes JB (2012) Backcross Populations and Near Isogenic Lines. In SA Rifkin, ed, Quantitative Trait Loci (QTL), Vol 871. Humana Press, pp 3-16

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

Kover PX, Mott R (2012) Mapping the genetic basis of ecologically and evolutionarily relevant traits in Arabidopsis thaliana. Curr Opin Plant Biol 15: 212-217

Lacaze X, Hayes PM, Korol A (2009) Genetics of phenotypic plasticity: QTL analyzis in barley, Hordeum vulgare. Heredity (Edinb) 102: 163-173

Latzel V, Zhang Y, Moritz KK, Fischer M, et al (2012) Epigenetic variation in plant responses to defence hormones. Annals of Botany 110: 1423-1428

Less H, Angelovici R, Tzin V, Galili G (2011) Coordinated Gene Networks Regulating Arabidopsis Plant Metabolism in Response to Various Stresses and Nutritional Cues. Plant Cell 23: 1264-1271

Logemann E, Tavernaro A, Schulz WG, Somssich IE, et al (2000) UV light selectively coinduces supply pathways from primary metabolism and flavonoid secondary product formation in parsley. Proceedings of the National Academy of Sciences of the United States of America 97: 1903-1907

Mackay TFC (2001) The genetic architecture of quantitative traits. Annual Review of Genetics 35: 303-339

Manolio TA, Collins FS, Cox NJ, Goldstein DB, et al (2009) Finding the missing heritability of complex diseases. Nature 461: 747-753

Mayr E (1982) The growth of biological thought: Diversity, evolution, and inheritance. Harvard University Press

McCall AC, Fordyce JA (2010) Can optimal defence theory be used to predict the distribution of plant chemical defences? Journal of Ecology 98: 985-992

McCall AC, Irwin RE (2006) Florivory: the intersection of pollination and herbivory. Ecology Letters 9: 1351-1365

McClellan JM, Susser E, King MC (2007) Schizophrenia: a common disease caused by multiple rare alleles. British Journal of Psychiatry 190: 194-199

McKey D (1974) Adaptive patterns in alkaloid physiology. American Naturalist 108: 305-320

Mirouze M, Paszkowski J (2011) Epigenetic contribution to stress adaptation in plants. Curr Opin Plant Biol 14: 267-274

Nordborg M, Weigel D (2008) Next-generation genetics in plants. Nature 456: 720-723

Ossowski S, Schwab R, Weigel D (2008) Gene silencing in plants using artificial microRNAs and other small RNAs. Plant Journal 53: 674-690

Pecinka A, Scheid OM (2012) Stress-Induced Chromatin Changes: A Critical View on Their Heritability. Plant and Cell Physiology 53: 801-808

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

Rando OJ, Verstrepen KJ (2007) Timescales of genetic and epigenetic inheritance. Cell 128: 655-668

Ravi M, Chan SW (2010) Haploid plants produced by centromere-mediated genome elimination. Nature 464: 615-618

Reinders J, Wulff BBH, Mirouze M, Mari-Ordonez A, et al (2009) Compromised stability of DNA methylation and transposon immobilization in mosaic Arabidopsis epigenomes. Genes & Development 23: 939-950

Richards CL, Bossdorf O, Verhoeven KJF (2010) Understanding natural epigenetic variation. New Phytologist 187: 562-564

Richards EJ (2006) Opinion - Inherited epigenetic variation - revisiting soft inheritance. Nature Reviews Genetics 7: 395-U392

Rojas CM, Senthil-Kumar M, Tzin V, Mysore KS (2014) Regulation of primary plant metabolism during plant-pathogen interactions and its contribution to plant defense. Frontiers in Plant Science 5: 12

Roux F, Colome-Tatche M, Edelist C, Wardenaar R, et al (2011) Genome-Wide Epigenetic Perturbation Jump-Starts Patterns of Heritable Variation Found in Nature. Genetics 188: 1015-U1402

Sahu PP, Pandey G, Sharma N, Puranik S, et al (2013) Epigenetic mechanisms of plant stress responses and adaptation. Plant Cell Rep 32: 1151-1159

Scheideler M, Schlaich NL, Fellenberg K, Beissbarth T, et al (2002) Monitoring the switch from housekeeping to pathogen defense metabolism in Arabidopsis thaliana using cDNA arrays. Journal of Biological Chemistry 277: 10555-10561

Schmitz RJ, He Y, Valdes-Lopez O, Khan SM, et al (2013) Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population. Genome Res 23: 1663-1674

Schmitz RJ, Schultz MD, Lewsey MG, O'Malley RC, et al (2011) Transgenerational Epigenetic Instability Is a Source of Novel Methylation Variants. Science 334: 369-373

Schmitz RJ, Schultz MD, Urich MA, Nery JR, et al (2013) Patterns of population epigenomic diversity. Nature 495: 193-198

Schneeberger K, Weigel D (2011) Fast-forward genetics enabled by new sequencing technologies. Trends in Plant Science 16: 282-288

Simons AM (2011) Modes of response to environmental change and the elusive empirical evidence for bet hedging. Proceedings of the Royal Society B: Biological Sciences 278: 1601-1609

Smallegange RC, van Loon JJA, Blatt SE, Harvey JA, et al (2007) Flower vs. leaf feeding by Pieris brassicae: Glucosinolate-rich flower tissues are preferred and sustain higher growth rate. Journal of Chemical Ecology 33: 1831-1844

Tetard-Jones C, Kertesz MA, Preziosi RF (2011) Quantitative trait loci mapping of phenotypic plasticity and genotype-environment interactions in plant and insect performance. Philos Trans R Soc Lond B Biol Sci 366: 1368-1379

Thomas D (2010) Gene-environment-wide association studies: emerging approaches. Nature Reviews Genetics 11: 259-272

Todesco M, Balasubramanian S, Hu TT, Traw MB, et al (2010) Natural allelic variation underlying a major fitness trade-off in Arabidopsis thaliana. Nature 465: 632-U129

Turck F, Coupland G (2013) Natural variation in epigenetic gene regulation and its effects on plant developmental traits. Evolution

Ungerer MC, Halldorsdottir SS, Modliszewski JL, Mackay TFC, et al (2002) Quantitative trait loci for inflorescence development in Arabidopsis thaliana. Genetics 160: 1133-1151

Ungerer MC, Halldorsdottir SS, Purugganan MD, Mackay TFC (2003) Genotype-Environment Interactions at Quantitative Trait Loci Affecting Inflorescence Development in Arabidopsis thaliana. Genetics 165: 353-365

van Ijzendoorn MH, Bakermans-Kranenburg MJ, Belsky J, Beach S, Bet al (2011) Gene-by-environment experiments: a new approach to finding the missing heritability. Nature Reviews Genetics 12: 2

Venable DL (2007) Bet hedging in a guild of desert annuals. Ecology 88: 1086-1090

Verhoeven KJ, Jansen JJ, van Dijk PJ, Biere A (2010) Stress-induced DNA methylation changes and their heritability in asexual dandelions. New Phytol 185: 1108-1118

Verhoeven KJ, Preite V (2013) Epigenetic variation in asexually reproducing organisms. Evolution

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

Wijnen CL, Keurentjes JJ (2014) Genetic resources for quantitative trait analyzis: novelty and efficiency in design from an Arabidopsis perspective. Current opinion in plant biology 18: 103-109

Wijnker E, van Dun K, de Snoo CB, Lelivelt CLC, et al (2012) Reverse breeding in Arabidopsis thaliana generates homozygous parental lines from a heterozygous plant. Nat Genet 44: 467-470

Yang J, Manolio TA, Pasquale LR, Boerwinkle E, et al (2011) Genome partitioning of genetic variation for complex traits using common SNPs. Nat Genet 43: 519-525

Zhang YY, Fischer M, Colot V, Bossdorf O (2013) Epigenetic variation creates potential for evolution of plant phenotypic plasticity. New Phytol 197: 314-322

Zuk O, Hechter E, Sunyaev SR, Lander ES (2012) The mystery of missing heritability: Genetic interactions create phantom heritability. Proceedings of the National Academy of Sciences of the United States of America 109: 1193-1198

Zuk O, Schaffner SF, Samocha K, Do R, et al (2014) Searching for missing heritability: Designing rare variant association studies. Proceedings of the National Academy of Sciences of the United States of America 111: E455-E464

Summary

The plant phenotype is shaped by complex interactions between its genotype and the environment. Although the genotype is stable and determined by the genomic sequence, plants are able to respond flexibly to changes in environmental conditions by orchestrated signal transduction pathways. The genomic sequence may change with each generation through chromosome rearrangements, meiotic recombination and spontaneous mutations. Through natural selection on these randomly induced changes, genotypes become adapted to their local environment. Because different genotypes adapt to different environments, natural variation within species expands in time and gives rise to a wide variety of genotypes and phenotypes. The genetic architecture that specifies the phenotype can be investigated by analysing different genotypes in the same environment and associate the phenotypic variation with molecular markers that discriminate the genotypes. Recent advances in next-generation sequencing technology enabled the fast sequencing of entire genomes, and in *Arabidopsis thaliana* alone, more than 1000 different genotypes have been fully resequenced. The sequencing allows the association of phenotypic variation with large numbers of single nucleotide polymorphisms (SNPs) that greatly enhance resolution in genome-wide association studies (GWAS).

GWAS on human diseases suffer from missing heritability that is most likely caused by the genetic architecture of the disease traits. Many variants of small effect or rare variants most likely determine a large part of the genetic variation and these variants are difficult to identify in GWAS due to lack of statistical power. In plants, several GWAS have been performed and they have identified previously validated genes and genes involved in monogenic disease resistance, but elucidating quantitative traits such as many agronomic important traits might be problematic in plants as well. Chapter 2 describes a GWA study in which quantitative morphological traits, such as leaf area, flowering time and branching were examined in 350 accessions of Arabidopsis for association with about 200,000 SNPs. The morphological traits showed extensive variation and were highly heritable, but GWA mapping could not identify the genetic variants that explain the heritability. Therefore, missing heritability was addressed using genomic selection models and these models confirmed the quantitative complex architecture of the morphological traits. Based upon these results, the heritability was assumed to be hidden below the significance threshold, and indeed lowering the significance threshold enabled the identification of many candidate genes that have been implicated to play a role in the phenotype directly or indirectly, in previous studies. One candidate gene was studied in more detail; natural variants of *ACS11*, an ethylene biosynthesis gene, associated significantly with the petiole to leaf length ratio. *ACS11* is indeed expressed in petioles and ectopically supplied ethylene abolished the difference in the phenotype of natural variants at this locus, strongly suggesting that *ACS11* is involved in the regulation of petiole growth.

However, lowering the significance threshold also increases the number of false-positive associations, non-causal alleles that co-segregate with the trait values. Because regulation of the morphological traits occurs at multiple intermediate levels, increased certainty on the associations can be obtained by performing GWA mapping on the intermediate levels from genotype to phenotype such as gene expression, and protein and metabolite content. Chapter 3 describes a literature survey into the multi-dimensional regulation of metabolic networks that are regulated by inputs from the clock, the communication between cells and between source and sink tissues, and the environment. The metabolic status of the plant can be seen as the final product of the interaction with the environment, and as such, it can serve as a blueprint for growth and development. Chapter 4 describes the abundant variation in enzyme activities and metabolites involved in primary carbon and nitrogen metabolism. The metabolite and enzyme activity data were analysed together with plant biomass data, and many pleiotropic regulators were identified with opposite effects on primary metabolism and biomass formation. Natural variants in two stress-responsive genes were oppositely associated with biomass and many enzymes and metabolites involved in primary metabolism, suggesting that higher enzyme activities and higher levels of sugars and proteins might be needed to support plant resistance to stress at the expense of growth.

Some studies indicated that epigenetic variation, independent of the genetic SNPs, may contribute to missing heritability. Epigenetic inheritance is defined as the inheritance of phenotypic variation to future generations without changes in DNA sequence. Epigenetic variation is caused by variation in chromatin marks such as DNA methylation, histone modifications and small RNAs. Recently, a recombinant inbred line (RIL) population was developed in Arabidopsis where the chromosomes are differentially methylated in lines with an otherwise isogenic background by crossing wild-type Col-0 with a hypomethylated *ddm1-2* mutant. Chapter 5 describes the epigenetic regulation of morphology and phenotypic plasticity by studying morphological variation in 99 epiRILs under control and saline conditions. The morphology and plasticity trait values were associated with differentially methylated regions (DMRs) that were used as molecular markers in QTL mapping. Many QTLs for various morphological traits and phenotypic plasticity parameters co-located, suggesting pleiotropic epigenetic regulation of growth, morphology and plasticity. Furthermore, methylation variation in the promoter of a salt-tolerance gene, *HIGH-AFFINITY K^+ TRANSPORTER1* (*HKT1*) associated significantly with leaf area, especially under saline conditions.

To gain more insight into the epigenetic regulation of plant growth and morphology, chapter 6 describes the epigenetic regulation of secondary metabolite levels in leaves and flowers and studies the relationship with the morphological traits determined in chapter 5. Many of the QTLs that were found for growth and morphology overlapped with the QTLs for metabolic traits, and suggest pleiotropic regulation. Furthermore, subsets of the metabolites

correlated well with the morphological traits and might thus be regulated by the same loci. The majority of metabolite QTLs were detected for glucosinolates and flavonoids in the flowers, and methylation variation was observed for some of the biosynthetic pathway genes of these compounds when comparing Col-0 and *ddm1-2*, which indicates a role for epigenetic regulation of these biosynthesis pathways.

Although stable, natural epialleles have been found in plant species and the environment can induce hypo- and hypermethylation of DNA, it remains elusive whether environmentally induced epigenetic changes can be inherited to subsequent generations, independent of genetic variation. Chapter 7 describes the transgenerational inheritance of phenotypic variation in progeny derived from a common Arabidopsis founder line. The progeny of stressed parents and grandparents showed variation in morphological traits, metabolite accumulation and gene expression. For example, many salt-responsive genes were up-regulated in progeny of salt-stressed grandparents. The responses to biotic (methyljasmonate) and abiotic (salt) stress differed strongly and this suggests that different environments can cause different transgenerational responses. Because all lines are derived from a single ancestor, epigenetic variation and not DNA variation is most likely causal for the phenotypic variation. Further studies are, however, needed to provide conclusive evidence for transgenerational inheritance.

Chapter 8 provides a synthesis of the work and discusses the GWA studies in the light of missing heritability, genetic architecture and the verification of candidate genes. The work on epigenetic regulation of phenotypic plasticity, morphology and metabolism is discussed in relation to Lamarckian soft inheritance that gained new enthusiasm after some recent discoveries in the field of epigenetics. And finally, the metabolomics work is discussed in the light of the growth-defense hypothesis that states that investments in defense occur at the expense of growth.

Samenvatting

Het fenotype van een plant wordt bepaald door de complexe interacties tussen zijn genotype en de omgeving waarin hij opgroeit. Hoewel het genotype onveranderlijk is en gedefinieerd wordt door de genomische DNA sequentie zijn planten in staat om flexibel te reageren op veranderingen in de omgeving door sterk georganiseerde signaal transductie routes. De genomische sequentie kan elke generatie worden aangepast door chromosoom reorganisatie, meiotische recombinatie en spontane mutaties. Door het proces van natuurlijke selectie op deze willekeurig aangebrachte veranderingen passen genotypes zich in de tijd aan naar hun lokale omgeving. Omdat verschillende genotypes zich verschillend aanpassen aan (verschillende) omstandigheden breidt de natuurlijke variatie zich uit met de tijd en ontstaat er een grote verscheidenheid aan genotypes en fenotypes. De genetische opmaak van de plant die het fenotype specificeert kan onderzocht worden door verschillende genotypes van dezelfde plant op te groeien onder dezelfde omstandigheden en een associatie te maken tussen de uiterlijke variatie en moleculaire merkers die de genotypes onderscheiden. De onderzoeken naar de associatie tussen vele moleculaire merkers en de fenotypes worden aangeduid als genoomwijde associatiestudies (GWAS). Recente vorderingen in genotyperingstechnologie hebben het mogelijk gemaakt om de volledige DNA sequentie van een organisme te ontrafelen, en in *Arabidopsis thaliana* (de zandraket) alleen zijn nu meer dan 1000 verschillende genotypes volledig genetisch in kaart gebracht.

GWAS wordt ook toegepast in humane genetica om de genetische oorzaak van ziekten te achterhalen, maar het vinden van de associaties tussen ziekte en moleculaire merkers wordt sterk vermoeilijkt door de ontbrekende erfelijkheid (erfelijkheid die niet kan worden aangetoond met de moleculaire merkers). Dit wordt zeer waarschijnlijk veroorzaakt door de genetische opmaak van ziekteverschijnselen. Vele genetische varianten van klein effect of zeldzame genvarianten bepalen waarschijnlijk voor een groot gedeelte de genetische variatie. Deze varianten zijn echter moeilijk te achterhalen met GWAS door het ontbreken van statistische kracht. In planten zijn verschillende GWAS uitgevoerd en zij hebben verscheidene, eerder bevestigde genen gevonden en genen betrokken bij monogenetische ziekteresistentie. Echter, het vinden van genen voor kwantitatieve eigenschappen zoals belangrijke eigenschappen voor de landbouw lijkt ook in planten problematisch. Hoofdstuk 2 beschrijft een GWA studie naar kwantitatieve morfologische eigenschappen, zoals bladoppervlak, bloeitijd en stengel vertakkingen in 350 verschillende accessies van Arabidopsis gegenotypeerd met ongeveer 200,000 merkers. We vinden veel variatie voor de morfologische eigenschappen met een hoge erfelijkheid, maar GWAS leidde niet tot het in kaart brengen van de genetische varianten die de erfelijkheid verklaren. Daarom gebruikten wij genomische selectie modellen om het probleem van de ontbrekende erfelijkheid verder te onderzoeken en deze modellen bevestigden de complexe, kwantitatieve opmaak van de

morfologische eigenschappen. Gebaseerd op deze resultaten werd aangenomen dat de erfelijkheid niet ontbreekt, maar verborgen is onder de significantie drempel. Het verlagen van de significantie drempel zorgde inderdaad voor de identificatie van vele genen waarvoor eerder direct of indirect bewijs was geleverd in andere studies dat zij te maken hebben met de eigenschappen die wij onderzoeken. Eén gen werd in onze studies in meer detail bestudeerd: natuurlijke varianten van *ACC synthase 11*, *ACS11*, een ethyleen biosynthese gen, associeerden significant met de ratio tussen petiole -en bladlengte. *ACS11* komt tot expressie in de petiole en het blootstellen van de planten aan ethyleen deed het verschil in de ratio tussen petiole- en bladlengte tussen de verschillende genotypes teniet. Deze vindingen duiden er sterk op dat *ACS11* betrokken is bij de regulatie van petiole groei.

Echter, het verlagen van de significantie drempel leidt ook tot een verhoging van het aantal vals positieve associaties, niet causale allelen die cosegregeren met de fenotypische waarden. Omdat regulatie van de morfologische eigenschappen plaatsvindt op verschillende, intermediaire niveaus kan meer zekerheid over de juistheid van de associaties verkregen worden door GWAS toe te passen op deze intermediaire niveaus tussen genotype en fenotype, zoals op genexpressie, eiwit en metaboliet niveau. Hoofdstuk 3 beschrijft een literatuuronderzoek naar de multidimensionale regulatie van metabolische netwerken, die geregeld worden door signalen van de klok, de communicatie tussen cellen en tussen weefsels, en de omgeving. De metaboliet huishouding van de plant kan worden gezien als het uiteindelijke product van de interactie tussen genotype en omgeving en op deze manier kan het dienen als een blauwdruk voor groei en ontwikkeling. Hoofdstuk 4 beschrijft de veelvuldige variatie in enzymactiviteiten en metabolieten betrokken bij primair carbon en stikstof metabolisme. De metaboliet huishouding en enzymactiviteit data werden geanalyseerd ten opzichte van de biomassa van de plant, en vele pleiotropische genen werden geïdentificeerd met tegenovergestelde effecten op primair metabolisme en biomassa vorming. Natuurlijke variatie in twee stress responsieve genen bijvoorbeeld suggereert dat hoge enzymactiviteit en hogere suiker en eiwit niveaus nodig zijn om de resistentie van planten te verhogen ten koste van snelle groei.

Verscheidene studies wijzen erop dat epigenetische variatie, onafhankelijk van genetische variatie, ook zou kunnen bijdragen aan de ontbrekende erfelijkheid. Epigenetische erfelijkheid wordt gedefinieerd als de overerving van fenotypische variatie aan volgende generaties, onafhankelijk van veranderingen in de DNA sequentie. Epigenetische variatie wordt veroorzaakt door veranderingen in het chromatine zoals variatie in DNA methylatie, histon modificaties en kleine RNAs. Recentelijk is een recombinante inteelt lijn populatie ontwikkeld (RIL) in Arabidopsis waar de chromosomen differentieel gemethyleerd zijn in een anderzijds isogene achtergrond door het kruisen van wildtype Col-0 met een hypogemethyleerde mutant, *ddm1-2*. Deze populatie wordt de epigenetische recombinante inteelt lijn (epiRIL) populatie genoemd. Hoofdstuk 5 beschrijft de epigenetische regulatie van

morfologie en fenotypische plasticiteit aan de hand van een studie naar de morfologische variatie in 99 epiRILs opgegroeid onder ofwel controle ofwel zoute omstandigheden. De morfologische en plasticiteits-gerelateerde eigenschappen werden geassocieerd met de differentieel gemethyleerde gebieden, die gebruikt worden als moleculaire merkers voor kwantitatieve kenmerk analyses (QTL, quantitative trait locus). Verscheidene QTLs werden gevonden voor zowel de morfologische eigenschappen als de fenotypische plasticiteit parameters op dezelfde positie in het genoom, wat er op duidt dat er sprake is van pleiotropische, epigenetische regulatie van groei, morfologie en plasticiteit. Bovendien, variatie in DNA methylatie in de promotor van een gen betrokken bij zout tolerantie, *HKT1*, associeerde significant met bladoppervlak, vooral onder zoute omstandigheden.

Om dieper in te gaan op de epigenetische regulatie van plantengroei en morfologie beschrijft hoofdstuk 6 de epigenetische regulering van secundaire metaboliet niveaus in bladeren en bloemen en wordt de relatie tussen het metabolisme en de morfologische eigenschappen onderzocht. Velen van de QTLs die werden gevonden voor groei en morfologie overlapten met de QTLs die werden gevonden voor het secundaire metabolisme, en dit wijst opnieuw op pleiotropische regulatie. Bovendien correleren bepaalde metabolieten zeer goed met de morfologische eigenschappen en zouden zij dus beiden geregeld kunnen worden door epigenetische invloeden. De meeste metabolische QTLs werden gevonden voor glucosinolaten en flavonoïden in de bloemen. Bovendien werd variatie in DNA methylatie waargenomen tussen de ouders van de populatie Col-0 en *ddm1-2* voor enkele biosynthese genen van deze moleculen. Dit duidt sterk op epigenetische regulatie van de biosynthese routes.

Hoewel stabiele, natuurlijke epiallelen zijn gevonden in verschillende plantensoorten en hoewel omgevingsfactoren hypo –en hypermethylatie van het DNA kunnen veroorzaken blijft het onzeker of omgeving-geinduceerde epigenetische veranderingen kunnen worden doorgegeven aan het nageslacht, onafhankelijk van genetische variatie. Hoofdstuk 7 beschrijft de transgenerationele overerving van fenotypische variatie in nakomelingen van eenzelfde Arabidopsis ouder. De nakomelingen van gestreste ouders en voorouders tonen variatie in morfologische eigenschappen, metaboliet accumulatie en genexpressie ten opzichte van nakomelingen van niet gestreste ouders en voorouders. Verscheidene zout responsieve genen bijvoorbeeld werden verhoogd afgeschreven in de nakomelingen van grootouders die opgroeiden onder zoute omstandigheden in vergelijking met nakomelingen van grootouders die opgroeiden onder optimale omstandigheden. De respons van nakomelingen na biotische (methyljasmonaat) en abiotische (zout) stress verschilde sterk en dit suggereert dat verschillende omgevingen verschillende transgenerationele responsen veroorzaken. Omdat alle lijnen van dezelfde homozygote ouder afkomstig zijn via zelfbestuiving is epigenetische variatie waarschijnlijk verantwoordelijk voor de fenotypische variatie, en niet genetische variatie. Vervolgstudies zijn echter nodig om volledig aan te

tonen dat het hier gaat om epigenetishce transgenerationele overerving.
Hoofdstuk 8 voegt het werk samen tot één geheel en bediscussieert GWAS in het licht van ontbrekende erfelijkheid, genetische opmaak en de verificatie van kandidaat genen. Het werk over de epigenetische regulatie van fenotypische plasticiteit, morfologie en metabolisme wordt bediscussieerd met betrekking tot Lamarckiaanse erfelijkheid die recentelijk hernieuwde aandacht kreeg door ontdekkingen in het epigenetische onderzoek. En als laatste wordt het metabolieten werk bediscussieerd in het licht van de groei-verdedigings hypothese die aanduidt dat investeringen in verdediging gebeuren ten koste van groei.

Acknowledgements

Thinking back of my first day in Wageningen I recall that when Dick was introducing me to some of my fellow PhDs I did not believe at all that we really had to sit in the open space, because I was certain that PhDs would sit in offices. I told myself that they must have an office somewhere else, and I would be sure to get one. And when Joost introduced me to the Arabidopsis plants in the greenhouse, I thought I must have misunderstood him when he talked about natural genetic variation and that all these different plants were from the same species. You can see from these two small anecdotes on my first day that I sure had to learn and have learned a lot during my PhD.

Joost, you are top priority in my list of gratitude. I had received a golden member card to enter your office or call your mobile at any time I wanted, no matter how simple or stupendous the questions. And whenever I needed help in the laboratory or the greenhouse, you were eager to help me out. With my way of planning, usually a few days before a new experiment started, I cannot be less grateful. Furthermore, there was and still is never a shortage of ideas in your laboratory to work on, and numerous times when I walked casually into your office you had come up with new ideas for genetically sound plant populations. And you brought me straight into contact with many of the people I collaborated with for my thesis. Joost, your enthusiasm and passion for science is admirable and your knowledge on genetics and genomics is beyond dispute. Besides the great scientific help that you were, you also helped me with difficult choices that I faced during my PhD. You respected my honesty and listened to me well, and although I seriously doubted about continuing with my PhD, you really wanted me to make the best decisions for me as a person leaving yourself and others out of the equation.

Dick, your help and support during my thesis was really valuable, putting things into perspective for me, and asking and challenging me about ideas and experiments. You also knew very well when to ask about my well-being and give me compliments when I needed them the most. During the last few months, you have been a great support, giving me good comments within a short time, and knowing how to motivate me to keep writing.

Harro, I really wonder how you maintain an athletic body under so much pressure and such a workload. I heard about your skills from other PhDs, but only witnessed it when you were giving me valuable comments on my chapters late at night for many days in a row, and looked perfectly fine in the morning. Furthermore, you always showed interest in my projects and in me as a person. I would also like to thank Sander, Henk, Richard, Léonie and Wilco for their encouragement, chats and advice.

I would also like to thank the professors from genetics, starting with Maarten who was always keen to talk to me about my project and about Arabidopsis and genetics in general. I admire your passion for research, illustrated for me by your walk every Monday to the greenhouse to look for interesting specimens. When publications were to my interest, you would always indicate them to me, either personally or through e-mail. I greatly appreciate

your words and advice over the years. Mark, I always cherished the small moments in the coffee corner, in the hallways or after a presentation where we talked about my project or the pitfalls of GWAS. Hans, you have always been very kind to me, perhaps knowing that I was not a natural-born scientist, we talked about photography, microscopy and travels. Your enthusiastic lectures in the GATC course that I followed in the first months of my PhD have been of great value to me. Bas, I remember sitting in the train with you one day travelling east, mostly talking about our favourite photographers. Whenever I gave a talk in GSS, you always asked me important questions and mostly came to me afterwards to give me advice.

Frank, we have spent weeks, maybe months together in a climate chamber measuring, photographing and harvesting plants, listening to sets and albums of different DJ's and bands. I especially remember a set of Tiga in Paradiso that we could play at 9am in the morning while we started our daily routine. It was hard work, but it did create a good bond and occasionally we met at some abstract techno party making ugly faces. You have been such a great help to me and without you this thesis would have been a few chapters shorter. You always showed great spirit and motivation, and you have helped me numerous times throughout my PhD project, either in the climate chambers or in the molecular lab.

Laurens, although the time you spent in our lab was much shorter, I could count on you as much as I could count on Frank. You have given tremendous support in setting up the doubled haploid populations, helping me with the harvests of my experiments and the molecular lab work anytime I needed assistance, and you always did this with a big smile. It was great having you around the lab.

Johanna, when you joined the laboratory you brought some new élan, opening up discussions about our projects between the people in the laboratory, enabling each one of us to learn more from each other's work. You have also been an outstanding help for me, especially when we were planting and harvesting experiments together. I also liked it that you enjoyed listening to the music that I played in the climate chambers and that we could talk on subjects within and outside science.

There were a lot of others that helped me inside or outside the laboratory. Ralph, we have spent some hours per day for a couple of weeks measuring petioles and leaves, until space became the only noise. Pingping, Léa, Yanxia and Imran, you are my molecular lab heroes, when I had almost lost all hope you guys were around to support me and get me back on track. Pádraic, Nihal, Myriam, Cris and many others, thank you for helping me with the planting and harvests of my experiments. Zeshan, Desalegn and Emilie, thank you for working with me on hydroponics, metabolomics and the *ACS11* gene. Diaan, Lidia, Carin and Corrie, thank you for not kicking my ass too much in your laboratories. And Mr. Juventus, Giovanni, I appreciate your kindness for letting me steal your alcohol. I would also like to thank all the other members of the Diaan, Carin and Lidia laboratory: Mahdere, Beatriz, Nasr, Sankseok, Xi, Junwei, among others.

The next cheers is for the lovely girls that surrounded me in the open space: Lidia, I loved to talk to you about Russia, Tolstoy, Dostoevsky, Chekov, movies, art, and many other things and I will definitely bring you a visit in the near future. Cecilia, how nice was it when you became my neighbor singing the songs to me that you woke up to in the morning. Your cheerfulness was a delight. Phuong, it was always nice to smile with you sitting at the other side of the desk. Yanxia, it was nice to share smiles with you too and I really admire your working spirit. Caroline, it was always good to see you and have a little chat. And for the boys that had been sitting in our little office space for a while: Wouter, you always knew something to make me laugh and Jamil, I appreciate the kind advice that you gave me.

Erik, my Arabidopsis collection team is still waiting for the Maarten Koornneef trophy and the Arabidopsis salad. However, it was always a pleasure to talk with you and oh boy, you have some brilliant ideas. Ronny, it was always good to discuss the GWAS with you, but besides that we could also share a good laugh. Wilco, thanks for your patience with me those times that I came by your desk to ask something relatively futile about Excel or GeneMaths, and further I really appreciate your warmhearted character. Rina and Wytske, thank you for your help when I came by with trivial questions about trips, conferences and others, and Rina, I really appreciate the times that you drove me to the station after PhD celebrations.

And now, I would like to thank all the others from the PPH and GEN laboratories that helped me in my project or enjoyed emptying a glass of pure fluid gold with me. Manickam, your presence is amazing and I really enjoyed our conversations together with Ralph. Julio and Pádraic, I really enjoyed the times that we were out together and that we always tell each other at such nights that those nights were too few. Thierry, it was nice sharing science or other things that busied our minds with you, either on the train or sporadically on the bike. Desalegn, Neli, Jimmy and Lemeng: I am very glad I got to know you guys a little bit better the last few months of our PhDs. It was always a pleasure, especially that time deep in the mud. Leo, Bas, Hanzi, Farzaneh, Wei, Rashid, Noorullah, Benyamin, Aaron, Anderson, Anna, Natalia, Elise, Alexandre, Esmer, Deborah, Yuanyuan, Jun, Tina, Ana-Carolina, Ya-Fen, Charles, Roxanne and many others from PPH and GEN: I am very happy to have met you all. Thanks also to the students that I supervised and that have made life easier at times: André, Peter, Matthijs and Ruth.

Then, there is a long list of people that collaborated with me on the thesis or on other publications. Willem, you have put an enormous effort in the GWAS, a big thank you for that. Henri, the LD tool that we developed together greatly helped all of us working in the GWA field. Ric, thank you for the discussions on metabolomics and your patience with me on many of my projects. I also thank Bert, Roland, Henriëtte and Harry. Frank, from Groningen, thank you for your input in the epigenetics stories. I would also like to thank the people in Germany where I really enjoyed the collaborations and discussions with Corina, Mark and Ronan. And also with Luis, Jia and Ales. I further really appreciate the discussions with Fred, and the collaborations with Paul, Sven and Elio.

In Australia, I had a great time and an awesome internship in Justin's lab. Justin, thanks for inviting me to your lab, introducing me to many people, and discuss science with me. Jared and Norman, it was great to hang with you guys all the time, drinking beers, eating pizza, listening to music, climbing mountains, canoeing and of course Jared, our trip to Melbourne. Further thanks go to Keng and Luciana, it was nice working with you girls. I would also like to thank Tim and Hank that I met again in Perth. It was very pleasant talking to both of you again, and Tim, thanks for giving me the opportunity to give a seminar.

I would also like to thank the people that I worked with during my MSc thesis in Utrecht. Rashmi, it was always very nice to meet you at seminars and conferences. Ronald, it was great hanging with you at the ASPB conference. I would also like to thank Divya, Kate, Wouter, Martijn, Diederik and Rens that I meet or talk to occasionally.

My friends back home: Thiemo, Nick, Luuk, Roel, Paul, Michel, Peter and many others from SDO Veldhoven. It is always great to see you guys, play some volleyball, football or basketball, and sadly with the years, my team loses more now than before, but at least we still drink way more beer than you guys. Our holidays together are always fun, with Budapest as the supreme highlight.

My roomies in Utrecht: Olaf, Lisa, Two Toms, Tatiana, Niki, Jurrian, Floor, Karijn, Lieveke, Anne and others. How great to sometimes lose myself with you guys after a week of hard study and delve myself into some industrial party, LGW, Dekmantel, Sziget, MELT or any other. The highlights were our two trips to Berlin, a memorable night in the Berghain, and the many after parties where we went home somewhere around midday. Besides that, I always enjoyed your company during my PhD and sharing thoughts on music, politics and philosophy.

My friends from the Swaffel United football team, and especially Maikel, Chiel and Vincent: You deserve some credits. The football was always great, and memorable was our first championship in the 1st class indoor competition. I greatly enjoyed the nights that we watched football, went for dinner or had a beer together.

Louis, we should really buy a new surfboard together. Man, we had a lot of fun trying our best to catch a wave. It was great seeing you again during my PhD, once back in Perth and once in Thailand with Nancy's family, where we rocked a mountain and got wasted on rice whisky because we were not allowed to say "no". And I will never forget that moment that you jumped out of the car late in the evening saying "Rik, take your camera" and you grabbed the tail of this huge python and took it back in front of my camera.

Federica, you are the first in a list of girls that needs some special attention. I had never thought that after you asked me to take a photograph of you and your friend, and the night of drinks that followed that we would see each other later in life. And then you showed up in Amsterdam. You have been my greatest model for the photo academy and we have had some weird nights at Ruigoord, dancing until early in the morning. And sipping wine from your balcony talking over politics, friends and conflicts after my classes

in Amsterdam wasn't so bad either. Jacquelyn, we have shared beautiful moments and went through some rough times, but I am grateful that you were there for some time during my PhD. Ciara, we have seen each other in so many different places: San Francisco, London, Dublin, Belfast and Amsterdam. From all these places we share great memories. Isabelle, we have smoked quite a bunch of greens together and we have laughed so much, rolling and rolling.

Natalia, my love is with you. You have been such a great support for me during the last months of my PhD. You have always put me first and really cared for me so that I could finish my PhD within time. Also before the end of my PhD, it always gave me a huge boost of energy to see your smiling face walking around the Radix. And, it was also you who helped me to realise that I should bring this PhD to a good end. I can really be myself with you, relax, let go and enjoy how time flies by. I hope we keep walking this same path together.

René, ever since you showed me your PSV season card at the introduction days for the BSc biology in Utrecht, we have been best friends and inseparable in many aspects of our lives. We shared the lab table, we were both in the Swaffel United football team, we both love whisky, Belgian beer, PSV and have the same under-the-rivers feeling. Our conversations will stay with me most of all and have kept me going strong during my PhD. The drives each morning from Utrecht to Wageningen are imprinted in my memory, where we talked about politics, philosophy, the life of a PhD, and most of all, about the women in our lives. It is amazing to have such a good friend like you on my side and I am not sure if this PhD would have come to a good end without you.

And my family: Luuk en Karlijn, you have always been there for me and have supported me always throughout my PhD. The moments that we got together in Kafé België in Utrecht, de Bierprofessor in Eindhoven or our tip to Copenhagen are great memories for me. It is nice that with family, things are as they are, and you can really be yourself. Mum and dad, thanks for always standing by my side and thinking with me about the paths that I should go. You came to visit me in Australia, and you have been collecting wild Arabidopsis for me all over the Netherlands. Mum, you have always been very sweet to me with small messages, e-mails and phone calls, wondering how I was doing. Dad, you have shared your enthusiasm for photography and images with me, and I really enjoyed the times that we were working together on photography.

Dear colleagues, friends and family: A massive thanks for your support during my PhD.

Curriculum Vitae

Rik Kooke was born on the 1st of August 1985 in Eindhoven, The Netherlands. After obtaining his highschool degree from "Sondervick College" te Veldhoven, the Netherlands, he moved to Utrecht in 2003 to study the Bachelor biology at Utrecht Universiy. After the completion of his BSc degree in 2007, he started a Master degree in Environmnetal Biology at Utrecht University. His first MSc thesis entitled "The role of microtubules during shade avoidance in *Arabidopsis thaliana*" was performed in the laboratory of Plant Ecophysiology under the supervision of Dr. R. Sasidharan and Dr. R. Pierik. The second MSc thesis entitled "On the ability of *Lotus tenuis* to regulate energy-dependent ion uptake in hypoxic, saline conditions" was performed in the school of Plant Biology at the University of Western Australia under the supervision of Dr. N. Ayers and Prof. T. Colmer. After his graduation in 2009, he shortly worked as a research assisant at Utrecht University.

In 2009, Rik started working as a PhD student in the Laboratories of Plant Physiology and Genetics under the supervision of Dr. D. Vreugdenhil and Prof. J.J.B. Keurentjes. The research focused on the genetic and epigenetic regulation of growth, morphology and metabolism in *Arabidopsis thaliana*. The results of the PhD project are described in this thesis.

Publications

Barboza L, Effgen S, Alonso-Blanco C, Kooke R, Keurentjes JJB, 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

Kooke R, Keurentjes JJB (2012) Multi-dimensional regulation of metabolic networks shaping plant development and performance. Journal of experimental botany 63: 3353-3365

Kooke R, Wijnker E, Keurentjes JB (2012) Backcross Populations and Near Isogenic Lines. Methods in Molecular Biology 871: 3-16

Sasidharan R, Keuskamp DH, Kooke R, Voesenek L, Pierik R (2014) Interactions between Auxin, Microtubules and XTHs Mediate Green Shade- Induced Petiole Elongation in Arabidopsis. Plos One 9: e90587

Education Statement of the Graduate School

Experimental Plant Sciences

Issued to: Rik Kooke
Date: 26 August 2014
Group: Plant Physiology, Wageningen University and Research Centre

1) Start-up phase	*date*
► **First presentation of your project**	
Genetical genomics to unravel regulation of plant development	Jun 28, 2010
► **Writing or rewriting a project proposal**	
► **Writing a review or book chapter**	
Multi-dimensional regulation of metabolic networks shaping plant development and performance, Journal of Exp. Botany (2011), pp 1-13	Jul 2011
Backcross Populations and Near Isogenic Lines, IN: Quantitatative Trait Loci: Mehods and Protocols (Scott A. Rifkin, ed.), Methods in Molecualr Biology 871, 2012, pp 3 - 16	Feb 2012
► **MSc courses**	
Genetic Analysis: Tools And Concepts (GEN-30306), Wageningen University	2009
Modern statistics for the Life Sciences (ABG-30806), Wageningen University	2010
Population and Quantitative Genetics (GEN-30806), Wageningen University	2010
► **Laboratory use of isotopes**	

Subtotal Start-up Phase *13.5 credits**

2) Scientific Exposure	*date*
► **EPS PhD student days**	
EPS PhD student day, Utrecht University	Jun 01, 2010
EPS PhD student day, Leiden University	Nov 29, 2013
► **EPS theme symposia**	
EPS theme 4 Day 'Genome Biology', Wageningen University	Dec 10, 2010
EPS theme 3 Day 'Metabolism and Adaptation', Utrecht University	Apr 27, 2012
EPS theme 3 Day Metabolism and Adaptation', University of Amsterdam	Mar 22, 2013
► **NWO Lunteren days and other National Platforms**	
ALW meeting 'Experimental Plant Sciences', Lunteren, NL	Apr 19-20, 2010
ALW meeting 'Experimental Plant Sciences', Lunteren, NL	Apr 04-05, 2011
ALW meeting 'Experimental Plant Sciences', Lunteren, NL	Apr 02-03, 2012
ALW meeting 'Experimental Plant Sciences', Lunteren, NL	Apr 22-23, 2013
► **Seminars (series), workshops and symposia**	
Invited seminar Nicolas Schauer, 'Metabolite QTL discovery and their inheritance in tomato breeding'	Dec 07, 2009
Invited seminar Justin Borevitz, 'Genetics of Adaptation: From model organisms to model ecosystems'	Jan 12, 2010
Plant Systems Biology Meeting (CBSG)	Feb 17, 2010
► **Seminar plus**	
Discussion with Justin Borevitz after seminar on genetics of adaptation	Jan 12, 2010

►	**International symposia and congresses**	*date*
	Regulation of plant growth - from cells to organs to organisms, Potsdam, Germany	Apr 12-14, 2010
	2nd Joint European retreat of PhD students in Plant Sciences, Cologne, Germany	Apr 15-17, 2010
	ASPB meeting, 2012, Austin, Texas	Jul 20-24, 2012
	ICAR 2013, Sydney, Australia	Jun 24-28, 2013
►	**Presentations**	
	The Next dimension in genetical genomics - Spatial and temporal regulation of Arabidopsis thaliana, EPS Lunteren (poster)	Apr 19-20, 2010
	The Next dimension in genetical genomics - Spatial and temporal regulation of Arabidopsis thaliana, CBSG conference in Arabidopsis cluster (oral)	Apr 2010
	Combining FSPM modelling and genetic mapping to unravel the regulation of growth and development in Arabidopsis thaliana, CBSG conference (poster)	Jan 2011
	GWAS in Arabidopsis thaliana, CBSG Summit 2012 (oral)	Mar 01, 2012
	QTNs - Come out, come out wherever you are - EPS Lunteren (oral)	Apr 03-04, 2012
	Genetic dissection of plant development using GWA and QTL analyses in Arabidopsis thaliana, ASPB meeting (poster)	Jul 20-24, 2012
	The epigenetic regulation of morphology and metabolism, EPS Theme 3 (oral)	Mar 22, 2013
	Genetic dissection of plant development using GWAS analyses in Arabidopsis thaliana, ICAR 2013 (invited speaker)	Jun 24-28, 2013
	Genetic dissection of plant development using GWAS analyses and transgenerational inheritance in Arabidopsis thaliana, Perth 2013 (invited speaker)	Jul 02, 2013
►	**IAB interview**	
	Meeting with a member of the International Advisory Board	Nov 15, 2012
►	**Excursions**	

Subtotal Scientific Exposure *19.0 credits**

3) In-Depth Studies		*date*
►	**EPS courses or other PhD courses**	
	Association mapping - Wageningen UR	Feb 23, 2012
	PhD Summerschool - Natural Variation of Plants, Wageningen UR	Aug 21-24, 2012
►	**Journal club**	
	member of the literature discussion group at Plant Physiology	2009-2013
►	**Individual research training**	
	Training at ANU, Canberra, Justin Borevitz lab	Jun 14-Jul 21, 2013

Subtotal In-Depth Studies *7.5 credits**

4) Personal development		*date*
►	**Skill training courses**	
	Scientific writing	Apr-Jun 2011
	Carreer Orientation Course	Feb-Mar 2014
►	**Organisation of PhD students day, course or conference**	
►	**Membership of Board, Committee or PhD council**	
►	**Skill training courses**	

Subtotal Personal Development *3.3 credits**

TOTAL NUMBER OF CREDIT POINTS*	**43.3**

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

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

The research described in this thesis was carried out at the Laboratories of Plant Physiology and Genetics at Wageningen University, Wageningen, The Netherlands and financially supported by the Centre of Biosystems Genomics (CBSG), Wageningen, The Netherlands.

Cover design by W.F.H. Kooke and R. Kooke
Thesis layout by R. Kooke

Printed by: Proefschriftmaken.nl || Uitgeverij BOXPress