Analysis of Plant Resistance against Herbivorous Insects with Molecular and Quantitative Genetics

Dissertation

zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)

vorgelegt dem Rat der Biologisch-Pharmazeutischen Fakultät
der Friedrich-Schiller-Universität Jena

von Diplom-Biologin Marina Pfalz

geboren am 3. Januar 1973 in Dresden

Referees:

1. Prof. Günther Theißen (Universität Jena)
2. Prof. Jonathan Gershenzon (MPICE, Jena)
3. Prof. Stefan Binder (Universität Ulm)

Public Defense: January 26th 2009

Table of Contents

1. Introduction

1.1 Plant defenses against insect herbivores

In their natural environment plants are exposed to a multitude of biotic challenges, inflicted by a diverse community of competitors, herbivores, pathogens, and other enemies. In response to herbivorous insects, plants have evolved a large variety of defense mechanisms (Stotz et al., 1999; Walling, 2000; Van Poecke, 2007; Howe & Jander, 2008). Plant defenses against insect herbivores may be constitutively present or are first induced upon attack. They can affect insect herbivory by influencing herbivore settling, feeding or oviposition. Furthermore, defenses often alter the physiology of insects and impede insect growth, development or reproduction.

Morphological structures such as trichomes and epicuticular waxes constitute an early barrier to ward off herbivores, contributing to plant fitness by preventing feeding damage (Eigenbrode & Espelie, 1995; Mauricio, 1998; Handley et al., 2005). Plants also utilize biochemical defenses. Major known chemical defenses comprise anti-digestive or -nutritional proteins. The latter constitute a quite diversified class of proteins, including proteases, lectins, amino acid deaminases, oxidases, and inhibitors of insect digestive enzymes (Chen, 2008). Furthermore, plants have evolved an immense variety of so-called secondary compounds. These plant secondary metabolites have diverse ecological functions. For example, they can serve as signal compounds to attract pollinating or seed-dispersing animals. More often, these compounds are important for plant defense. In many cases plant secondary metabolites are part of activated defense systems. These systems consist of metabolic precursor compounds and activators, usually enzymes. Both components, secondary metabolite and activator, can be stored in the plant tissue without a detrimental effect on the plant itself. Defense activation is initiated upon damage of plant tissue, bringing both components into contact with one another. The precursor is then metabolized to give rise to deterrent or toxic effector molecules. Activation in intact tissue is typically avoided by storing precursors and activators in different cell compartments or in different cell types. Morphological barriers and activated defenses are typical examples of defenses that are constitutively present, irrespective of herbivore presence.

Other defenses, however, are inducible upon herbivory. Plant damage inflicted by insects may involve transcriptional changes of a large and diverse set of genes, ultimately leading to the expression of plant defense compounds. Typically, there is a time delay of

several hours or even days between herbivory and deployment of defenses. Induction may also affect the expression of genes that are involved in the formation of constitutive traits. For example, trichome density can increase upon attack (Traw & Dawson, 2002). Likewise, nicotine levels in tobacco may increase as a result of insect herbivory (Baldwin, 1999). Inducible defenses are, in general, more pest-specific than constitutive defenses, and involve three conceptual stages: initial pest recognition, signal transduction and production of defensive compounds (Van Poecke & Dicke, 2004).

While some induced plant responses have a direct impact on insects, other defenses affect herbivores indirectly. Indirect defenses involve the attraction of the attacker's enemies, *e.g.* predators or parasitoids, which utilize and attack the herbivores for their own purposes, thereby reducing feeding damage on the host plant. Plant volatile emission and reward in form of nutrition play a major role in indirect plant defenses (Dicke, 1999; Van Poecke & Dicke, 2004).

These different defenses may contribute to plant resistance against a broad spectrum of herbivores and other enemies. However, the extent to which insects respond to specific defenses and how they respond may vary, dependent on the type of plant-insect interaction in question. Some insects are polyphagous, having a broad range of host plants. These 'generalists' have to cope with the disparate defensive mechanisms of many different plant species. Other insects are oligo- or even monophagus. These 'specialists' utilize only a small group of plant species as their hosts. Therefore, they have to overcome family or species-specific defenses and they have often evolved counteradaptations to cope with deterrent or toxic compounds produced by their host plants. Some specialist insects are even able to exploit host plant defenses for there own purposes, *e.g.* as oviposition cues or by sequestering plant defense compounds to ward off predators.

Hence, plant defenses against insect herbivores are diverse and complex. Model species, in particular species which are genetically tractable, are necessary for a thorough understanding of plant-insect interactions at the molecular level.

1.2 Arabidopsis as a model species for studying plant-insect interactions

Arabidopsis thaliana is an annual weed in the crucifer family (Brassicaceae). This species is native to Europe and central Asia and has become naturalized in many agricultural and disturbed habitats worldwide. *A. thaliana* is widely used as one of the model organisms in plant science. Its small size, usually 20-25 cm tall, and a relatively short life cycle of 2-3

months, combined with its simple requirements for growth conditions are some of the advantageous traits of this species. The flowers are naturally self-pollinating, but individuals can also be crossed for laboratory purposes. Therefore most plants collected in nature are practically homozygous (Koornneef et al., 2004). *A. thaliana* is diploid and has a small genome of about 150 million base pairs, containing approximately 26.000 genes. The entire genome sequence of the Arabidopsis accession[1] Columbia has been published in 2000 (Arabidopsis Genome Initiative, 2000). Several other accessions have been re-sequenced with novel sequencing technologies (Clark et al., 2007; Zeller et al., 2008) and current efforts are directed towards sequencing close and more distant Arabidopsis relatives, including *Arabidopsis lyrata* and *Capsella rubella* (www.jgi.doe.gov).

Molecular tools for Arabidopsis include whole-genome microarrays covering all genes known to be expressed in this species. Transcript profiling offers the potential to reveal information about the plant's transcriptome during development or in response to abiotic and biotic challenges, both in the laboratory and in nature. Arabidopsis can be transformed easily with *Agrobacterium* or by other means, and huge collections of mutants including loss-of-function and gain-of-function mutants are available from Arabidopsis stock centers[2]. Likewise, hundreds of Arabidopsis accessions have been sampled from the species' natural range and provide an additional resource for functional and evolutionary studies.

In their natural environment, Arabidopsis plants are attacked by various herbivores including flea beetles, aphids, leaf miners, weevils and lepidopteran larvae (Mauricio, 1998; Mitchell-Olds, 2001; Harvey et al., 2007; Arany et al., 2008). Many of these herbivores feed also on agriculturally important *Brassicas* such as mustard, radish, cauliflower, cabbage and rapeseed. Pest management and crop loss for *Brassica* species caused by insect herbivory are costly, and exceed $ 1 billion per year (Talekar & Shelton, 1993). Because *Brassica* and Arabidopsis are closely related, results from Arabidopsis research may be applicable in breeding and can help improve crop resistance against herbivorous insects.

[1] The term ‚accession' refers to Arabidopsis seeds that have been sampled at a particular location. For example, Je-0 is an accession that was collected in Jena and has been donated to the stock centers by Albert R. Kranz in 1993.

[2] Major Arabidopsis stock centers are located in Nottingham (United Kingdom), Columbus (Ohio), and Yokohama (Japan).

1.3 Natural genetic variation in Arabidopsis

Comparative studies among different Arabidopsis accessions have revealed natural genetic variation for a multitude of traits. Examples include developmental traits such as plant size and flowering time (Koornneef et al., 1998; Alonso-Blanco et al., 1999), primary and secondary metabolism (Kliebenstein et al., 2001a; Meyer et al., 2007; Lisec et al., 2008), biotic and abiotic stress resistance (Wilson et al., 2001; Kobayashi & Koyama, 2002; Kroymann et al., 2003) and molecular genetic mechanisms such as cytosine methylation (Riddle & Richards, 2002). This natural genetic variation has become an important resource for functional and evolutionary analyses (Koornneef et al., 2004), and complements conventional forward or reverse genetic approaches to identify genotype – phenotype relationships.

Complex traits such as plant resistance against herbivorous insects are controlled by the contribution of several or many genes in a quantitative manner (Kroymann & Mitchell-Olds, 2005). The genetic loci that contribute to these quantitative traits are referred to as quantitative trait loci (QTL). Recombinant inbred lines (RILs) have become an important tool in analyzing the genetic basis of quantitative traits. The generation of Arabidopsis RILs is initiated by a cross between two genetically distinct accessions, yielding a heterozygous F_1. Subsequent generations are obtained by repeated selfing. In these generations, parental genomes are mixed randomly by recombination. Furthermore, the level of heterozygosity declines on average by 50% in every subsequent generation, and drops to less than 1% in F_8 progeny. Individual lines from the F_8 generation are then genotyped with a set of genetic markers which exploit genetic polymorphisms between parental lines. RILs in the F_8 can be considered immortal because their genome composition remains (nearly) unchanged in subsequent generations when these are propagated by selfing. Therefore, genotyping is required only once while multiple traits can be analyzed in independent experiments. Typical RIL populations consist of more than 100 individual lines genotyped with 50 – 100 or more genetic markers. Meanwhile, more than 20 Arabidopsis RIL populations are available (Lister & Dean, 1993; Alonso-Blanco et al., 1998; Loudet et al., 2002; O'Neill et al., 2008), facilitating the mapping of QTL underlying many different complex traits (reviewed in Koornneef et al., 2004).

1.4 The glucosinolate-myrosinase system

The glucosinolate-myrosinase system is an activated defense system in the Brassicales (Rodman, 1991a; Rodman, 1991b). In this system, glucosinolates (mustard oil glycosides) serve as precursors, and myrosinases, enzymes with β-thioglucohydrolase activity, act as activators. In intact plant tissue, glucosinolates and myrosinases are stored in different cell types (Koroleva et al., 2000; Andreasson et al., 2001, Thangstad et al., 2004; Shroff et al., 2008). But upon tissue disruption, glucosinolates and myrosinase mix and myrosinase-catalyzed glucosinolate hydrolysis results in the formation of biologically active effector molecules, including isothiocyanates, nitriles, epithionitriles, thiocyanates, and others.

Glucosinolates are derived from a variety of amino acids but possess a common core structure, consisting of a sulfonated oxime and a β-thioglucose moiety. Depending on the precursor amino acid they can be classified as aliphatic, indole or aromatic glucosinolates. Aliphatic glucosinolates are derived from methionine, alanine, valine or leucine, indole glucosinolates have tryptophan as a precursor, and aromatic glucosinolates are synthesized from phenylalanine or tyrosine. Glucosinolate biosynthesis comprises two or three stages, elongation of the amino acid carbon chain, formation of glucosinolate core structure, and modification of the amino acid side chains (Figure 1). Glucosinolate carbon chain extension has been demonstrated for methionine (Chisholm & Wetter, 1964; Matsuo & Yamazaki, 1964; Graser et al., 2000) but occurs presumably also with phenylalanine (Underhill, 1968).

More than 120 different glucosinolates have been identified in the Brassicales (Daxenbichler et al., 1991; Fahey et al., 2001). *A. thaliana* synthesizes nearly 40 different glucosinolates but accessions vary extensively in glucosinolate composition and quantity (Kliebenstein et al., 2001a; Reichelt et al., 2002).

1.4.1 Aliphatic glucosinolates in Arabidopsis

In *A. thaliana*, all aliphatic glucosinolates are derived from methionine, representing the most common and abundant class of glucosinolates in the vegetative tissue of this species (Kliebenstein et al., 2001a). Biosynthesis of methionine-derived glucosinolates involves carbon chain elongation, generation of the glucosinolate core structure, and modifications of the carbon side chain.

Methylthioalkylmalate synthases encoded at the *MAM* locus catalyze the committed step in methionine carbon chain elongation. These enzymes carry out a condensation reaction

between a 2-oxo acid and acetyl-CoA. This 2-oxo acid originates from the transamination of methionine, catalyzed by members of a family of branched-chain amino acid transferases, BCATs (Schuster et al., 2006; Binder et al., 2007; Knill et al., 2008). Two further reactions, an isomerization and an oxidative decarboxylation, lead to the generation of a novel 2-oxo acid with a carbon chain extended by one methylene group. This chain-elongated reaction product can either be transaminated by BCATs to yield a methionine homolog which is utilized for glucosinolate core structure generation, or can re-enter the chain elongation pathway to undergo additional cycles of carbon chain extension. QTL mapping studies have revealed that the *MAM* is the major determinant for the length of the side chain of methionine-derived glucosinolates (Kroymann et al., 2001, Kroymann et al., 2003). In Arabidopsis, *MAM* gene composition is highly variable among accessions (Kroymann et al., 2003), and different MAM synthases differ in their substrate specificities (Benderoth et al., 2006; Textor et al., 2007), thereby explaining differences in the side chain lengths of aliphatic glucosinolates in *A. thaliana* (Figure 1).

The biosynthesis steps of the core structure control the formation of intermediates common to all glucosinolates (for a recent overview, see: Halkier & Gershenzon, 2006). Most of the enzymes involved in the formation of the glucosinolate core structure have been identified. The first two reactions are catalyzed by amino acid specific cytochrome P450 monooxygenases (CYPs), belonging to the CYP79 and CYP83 families, respectively. CYP79F1 and CYP79F2 convert chain elongated methionine derivatives into aldoximes (Hansen et al., 2001; Chen et al., 2003), while CYP83A1 and, to a lesser extent, CYP83B1 are responsible for the formation of *S*-alkyl thiohydroximates (Bak & Feyereisen, 2001; Naur et al., 2003). Subsequently, a *C*-*S* lyase catalyzes the conversion of *S*-alkyl thiohydroximates to thiohydroximates (Mikkelsen et al., 2004) which are then transformed into desulfoglucosinolates by an *S*-glucosyltransferase (Grubb et al., 2004). The final reaction that completes the glucosinolate core structure is catalyzed by desulfoglucosinolate-specific sulfotransferases (Piotrowski et al., 2004) which accept a broad range of different desulfoglucosinolates as substrates (Figure 1).

Finally, several genetic loci are involved in the modification of aliphatic glucosinolate side chain. The *GSL-AOP* locus encodes 2-oxo-acid-dependent dioxygenases which control the accumulation of either alkenyl or hydroxyalkenyl glucosinolates (Kliebenstein et al., 2001b). *GSL-OX* is responsible for the oxygenation of methylthioalkyl glucosinolate, and a crucifer specific family of flavin-monooxgenases has been proposed to carry out these reactions (Hansen et al., 2007). *GSL-OH* is another polymorphic locus responsible for

oxidative reactions on specific elongated side chains (Magrath et al., 1994; Kliebenstein et al., 2001a). Together, *MAM*, *GSL-AOP*, *GSL-OH* and *GSL-OX*, explain most of the structural variability in methionine-derived glucosinolates (Kliebenstein et al., 2001a; Kliebenstein et al., 2001b; Kroymann et al., 2003; Kliebenstein et al., 2005).

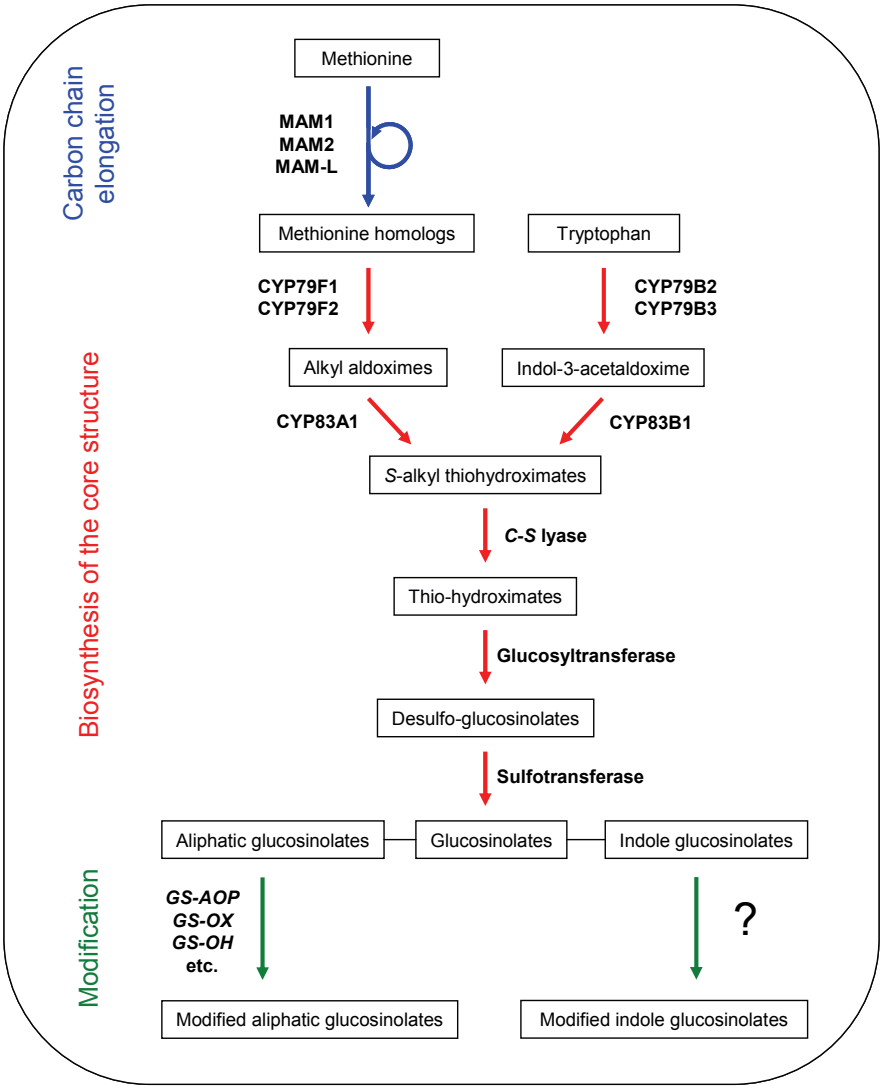

Figure 1. Biosynthesis of aliphatic and indole glucosinolates in *Arabidopsis thaliana*. Shown are amino acid precursors, reaction intermediates, and enzymes or genetic loci involved in glucosinolate biosynthesis.

1.4.2 Indole glucosinolates in Arabidopsis

Another class of glucosinolates is derived from tryptophan. These indole glucosinolates can accumulate to high concentrations in leaves and seeds (Kliebenstein et al., 2001a; Brown et al., 2003), and recent data suggest that specific indole glucosinolates may be important in the interaction between Arabidopsis and aphids (Kim & Jander, 2007).

Tryptophan is directly converted into the core glucosinolate structure without prior carbon chain elongation. As mentioned above, the intermediates of the core biosynthesis are common to all glucosinolate classes. However, in several cases indole glucosinolate-specific enzymes catalyze reactions equivalent to those in aliphatic glucosinolate biosynthesis. For example, the conversion of tryptophan to the aldoxime is catalyzed by CYP79B2 and CYP79B3 (Hull et al., 2000; Mikkelsen et al., 2000) and not by members of the CYP79F subfamily. The aldoxime-metabolizing enzymes are the same in aliphatic glucosinolate biosynthesis, CYP83A1 and CYP83B1, with CYP83B1 having a preference for indole substrates (Bak & Feyereisen, 2001; Naur et al., 2003). Likewise, *C*-*S* lyase, *S*-glucosyltransferase and desulfoglucosinolate-sulfotransferases function in both aliphatic and indole glucosinolate biosynthesis, although preliminary evidence suggests that further *S*-glucosyltransferases may be able to catalyze the conversion of thiohydroximates to desulfoglucosinolates (Grubb et al., 2004; Mikkelsen et al., 2004; Piotrowski et al., 2004).

The first intermediate in the core biosynthesis of indole glucosinolates, indole-3-acetaldoxime, has attracted particular attention because it serves as a precursor for other indolic metabolites, including the plant hormone indole-3-acetic acid (IAA) and indole alkaloids (Hull et al., 2000; Zhao et al., 2002; Glawischnig et al., 2004). Camalexin is the main indole alkaloid found in *A. thaliana* and has been shown to function as a phytoalexin in plant defense. Its accumulation is induced by a variety of microorganisms and contributes to resistance against fungal pathogens (Zhou et al., 1999; Glawischnig et al., 2004).

Similar to aliphatic glucosinolates, the initially formed glucosinolate can be subject to different types of modifications. These modifications occur at the indole ring by the insertion of hydroxy or methoxy groups (Figure 1). However, as yet neither enzymes nor genes responsible for these modification reactions have been identified.

1.5 The impact of the glucosinolate-myrosinase system on insect herbivores

The biological activities of glucosinolates as part of an activated defense system are, to a large extent, conferred by their hydrolysis products (Kliebenstein et al., 2005; Kliebenstein, 2008). The types of hydrolysis products which are formed during myrosinase-catalyzed glucosinolate hydrolysis depend on several factors. The precursor amino acid determines the general structure of the glucosinolate side chain. Carbon chain elongation in aliphatic glucosinolate biosynthesis, and side chain modification of both aliphatic and indole glucosinolates add to glucosinolate variability. Finally, formation of hydrolysis products

depends on the presence of various myrosinase-associated or -binding proteins. For example, plants with a functional epithiospecifier protein (ESP) produce primarily nitriles and epithionitriles, whereas plants impaired in the expression of this gene form primarily isothiocyanates (Lambrix et al., 2001). The epithiospecifier modifier protein (ESM1) favors the production of isothiocyanates (Zhang et al., 2006). Furthermore, a thiocyanate-forming protein was recently identified in *Lepidium sativum* (Burow et al., 2007).

Several field studies and laboratory tests have found the feeding behaviour of insect herbivores to be influenced by glucosinolate quantities and profiles. Plant damage and leaf glucosinolate concentration are typically negatively correlated for generalists such as *Spodoptera eridania*, *Spodoptera exigua* and *Trichoplusia ni* (Li et al., 2000; Kliebenstein et al., 2002; Kroymann et al., 2003; Kliebenstein et al., 2005; Arany et al., 2008). Many QTL for the glucosinolate-myrosinase system co-localize with QTL for resistance against generalist herbivores (Kliebenstein et al., 2002; Kroymann et al., 2003). Furthermore, studies with mutants impaired in glucosinolate biosynthesis or myrosinase activity revealed that these mutants sustained higher damage by generalist herbivores than mutants. For example, Beekwilder et al. (2008) generated a double mutant defective in the transcription factors MYB28 and MYB29 which are involved in the regulation of glucosinolate biosynthesis genes. This double mutant was devoid of aliphatic glucosinolates and suffered significantly more damage by *Mamestra brassicae* in comparison to wildtype plants. Likewise, a double mutant with lesions in the major Arabidopsis myrosinase genes, *TGG1* and *TGG2*, was particularly susceptible to *T. ni* and *Manduca sexta* (Barth & Jander, 2006). Several other studies have focused on variation in glucosinolate hydrolysis products and their effects on herbivores species. For example, isothiocyanates provided better resistance to *T. ni* and *Spodoptera littoralis* than nitriles (Lambrix et al., 2001; Burow et al., 2006; Zhang et al., 2006).

In contrast to generalist insect herbivores, glucosinolate concentration has been found to be positively correlated with herbivory of the specialist insect *Plutella xylostella* (Giamoustaris & Mithen, 1995; Kliebenstein et al., 2005). Likewise, high glucosinolate concentrations did not deter the specialist insect *Pieris brassicae* (Smallegange et al., 2007). In addition, volatile glucosinolate hydrolysis products mediate host recognition by specialist insect herbivores and have been found to stimulate oviposition by *P. xylostella* (Pivnick et al., 1994; Renwick et al., 2006). Nonetheless, laboratory experiments have shown that high concentrations of isothiocyanates are toxic to generalists as well as specialists (Li et al., 2000; Agrawal & Kurashige, 2003). However, crucifer specialist herbivores possess

counteradaptions that render the glucosinolate-myrosinase system ineffective. *P. xylostella* larvae express a glucosinolate sulfatase in their gut that removes the sulphate moiety from glucosinolates, thereby generating desulfoglucosinolates which cannot be hydrolyzed by myrosinase (Ratzka et al., 2002). A different counteradaptation was identified in *P. rapae* (Wittstock et al., 2004). Larvae express a nitrile-specifier protein (NSP) that promotes the production of nitriles instead of highly toxic isothiocyanates upon herbivory on crucifer plants. Therefore, interactions between Arabidopsis and crucifer specialists are exceptionally complex, and the genetic basis for defense against these specialists is largely unknown.

1.6 Study objectives

The objective of this thesis is to study plant defenses against different insect herbivores in the well established model species *Arabidopsis thaliana*.

1. The glucosinolate-myrosinase system protects cruciferous plants effectively against most herbivorous insects. Crucifer specialists, however, possess counteradaptations that render this activated defense system ineffective. Do cruciferous plants defend themselves against crucifer specialists? What is the genetic basis for resistance against specialist herbivores?

2. Herbivory by crucifer specialist insects leads to massive changes in the plant transcriptome. Does differential gene regulation contribute to plant fitness against these herbivores?

3. The glucosinolate-myrosinase system displays an enormous amount of qualitative and quantitative variation within and between species. Many genetic loci have been identified that determine variation in aliphatic glucosinolates. In contrast, little is known about the genes and enzymes which are responsible for variation in indole glucosinolates. Which genes control modifications of tryptophan-derived glucosinolates? What is their function? Which role do they play in plant-insect interactions?

2. List of Manuscripts: Contents and Author's Contributions

Manuscript I

Pfalz M, Vogel H, Mitchell-Olds T and Kroymann J (2007). Mapping of QTL for Resistance against the Crucifer Specialist Herbivore *Pieris brassicae* in a New Arabidopsis Inbred Line Population, Da(1)×Ei-2. *PLoS One* **2**, e578.

Several studies have investigated the genetic basis for Arabidopsis resistance against generalist insect herbivores but little is known about plant defense against specialists. This study detects natural genetic variation for resistance against the crucifer specialist insect herbivores *Pieris rapae* and *Plutella xylostella* among Arabidopsis accessions and in a new Arabidopsis inbred line population derived from a cross between the accessions Da(1)-12 and Ei-2. It also identifies QTL for glucosinolates, myrosinase activity and trichome density, traits known to contribute to resistance against generalist insect herbivores. Although some trichome, glucosinolate or myrosinase QTL co-localize with *Pieris* QTL, none of these traits explained the resistance QTL convincingly, indicating that resistance against specialist insect herbivores is influenced by other traits than resistance against generalists.

I conceived and designed the experiments in close collaboration with Heiko Vogel and Juergen Kroymann. I conducted all experiments including the genotyping of the Da(1)-12 × Ei-2 RIL population. Together with Juergen Kroymann, I analyzed the data and wrote the paper. Thomas Mitchell-Olds provided biological material.

Manuscript II

Böttcher C, Centeno D, Freitag J, Höfgen R, Köhl K, Kopka J, Kroymann J, Matros A, Mock HP, Neumann S, **Pfalz M**, von Roepenack-Lahaye E, Schauer N, Trenkamp S, Zubriggen M and Fernie AR (2008). Teaching (and learning from) metabolomics: The 2006 PlantMetaNet ETNA Metabolomics Research School. *Physiol. Plant.* **132**, 136-149.

This manuscript results from a Summer School in metabolomics hosted by the Max Planck Institute for Molecular Plant Physiology in Golm in 2006. It summarizes the content of this school and the results of a metabolomics experiment in *Arabidopsis thaliana*. This example experiment investigated the metabolic response of *A. thaliana* to the herbicide glyphosate with different analytical platforms.

The herbicide glyphosate is known to interfere with amino acid biosynthesis. Since glucosinolates are synthesized from several precursor amino acids, the glucosinolate composition was analyzed in response to herbicide treatment. While rapid and substantial changes in amino acid pools were observed, no substantial changes in glucosinolate profiles were observed over the time course of the experiment (8 hours).

Juergen Kroymann and I were responsible for the targeted analysis of glucosinolates with HPLC. Other participating groups focused on other aspects of Arabidopsis metabolism. Alisdair R. Fernie drafted the manuscript and wrote the paper with input from all other authors.

Manuscript III

Benderoth M, **Pfalz M**[*] and Kroymann J (2008). Methylthioalkylmalate synthases: genetics, ecology and evolution. *Phytochem. Rev.*, DOI 10.1007/s11101-008-9097-1.

This review focuses on a small gene family encoding methylthioalkylmalate synthases (MAM) which is central to the diversification of aliphatic glucosinolate structures in *Arabidopsis thaliana* and relatives. The manuscript elucidates the evolutionary forces that shape polymorphism at the *MAM* locus and investigates ecological consequences of this variation.

Markus Benderoth conducted phylogenetic analyses. I carried out ecological experiments. Juergen Kroymann, Markus Benderoth and I wrote the paper.

Manuscript IV

Pfalz M, Vogel H and Kroymann J. Analysis of Fitness Consequences of *Plutella*-Regulated Genes in Arabidopsis. *PLoS ONE*, *in review* (submitted October 23, 2008).

This manuscript investigates fitness consequences of Arabidopsis genes that respond with an alteration in transcript levels to herbivory by *Plutella xylostella* larvae. It compares the performance of Arabidopsis lines with T-DNA insertions to lines with wildtype alleles obtained from the same seed material. Very few lines display statistically significant differences in resistance to *P. xylostella*, and in all these cases, effects of gene disruption are in an unexpected direction. These findings indicate that the majority of *Plutella*-regulated genes do not contribute to defense against this insect and suggest that *Plutella* larvae may be able to manipulate host plant gene expression to their own benefit.

I conceived and designed the experiments in close collaboration with Heiko Vogel and Juergen Kroymann. I conducted all experiments and analyzed the data together with Juergen Kroymann. Juergen Kroymann and I wrote the manuscript.

Manuscript V

Pfalz M, Vogel H and Kroymann J. The Gene Controlling the *Indole Glucosinolate Modifier 1* QTL Alters Indole Glucosinolate Structures and Aphid Resistance in Arabidopsis. *Plant Cell*, *in review* (submitted September 08, 2008).

Previous work in Arabidopsis has mainly focused on genetic variation for aliphatic glucosinolates but neglected variation in indole glucosinolate structures. This work utilizes a combination of high-resolution QTL mapping and transcript profiling with whole-genome Arabidopsis microarrays to identify a cytochrome P450 gene that underlies *Indole Glucosinolate Modifier 1* (*IGM1*), a QTL that contributes to variation in indole glucosinolates. The analysis of Arabidopsis T-DNA insertion lines reveals that this gene influences proliferation of the aphid *Myzus persicae* but not herbivory by lepidopteran larvae, thereby providing evidence for different components of the glucosinolate-myrosinase complex acting against different enemies.

Juergen Kroymann, Heiko Vogel and I conceived and designed the experiments. I carried out all practical work, and analyzed data together with Juergen Kroymann. Juergen Kroymann and I wrote the manuscript, and Heiko Vogel commented on the manuscript.

[*] Markus Benderoth and Marina Pfalz contributed equally to this manuscript.

OPEN ACCESS Freely available online

PLoS one

Mapping of QTL for Resistance against the Crucifer Specialist Herbivore *Pieris brassicae* in a New Arabidopsis Inbred Line Population, Da(1)-12×Ei-2

Marina Pfalz[1], Heiko Vogel[2], Thomas Mitchell-Olds[1,3], Juergen Kroymann[1]*

1 Department of Genetics & Evolution, Max Planck Institute for Chemical Ecology, Jena, Germany, **2** Department of Entomology, Max Planck Institute for Chemical Ecology, Jena, Germany, **3** Department of Biology, Duke University, Durham, North Carolina, United States of America

Background. **In *Arabidopsis thaliana* and other crucifers, the glucosinolate-myrosinase system contributes to resistance against herbivory by generalist insects. As yet, it is unclear how crucifers defend themselves against crucifer-specialist insect herbivores.** ***Methodology/Principal Findings.*** **We analyzed natural variation for resistance against two crucifer specialist lepidopteran herbivores, *Pieris brassicae* and *Plutella xylostella*, among *Arabidopsis thaliana* accessions and in a new Arabidopsis recombinant inbred line (RIL) population generated from the parental accessions Da(1)-12 and Ei-2. This RIL population consists of 201 individual F_8 lines genotyped with 84 PCR-based markers. We identified six QTL for resistance against *Pieris* herbivory, but found only one weak QTL for *Plutella* resistance. To elucidate potential factors causing these resistance QTL, we investigated leaf hair (trichome) density, glucosinolates and myrosinase activity, traits known to influence herbivory by generalist insects. We identified several previously unknown QTL for these traits, some of which display a complex pattern of epistatic interactions.** ***Conclusions/Significance.*** **Although some trichome, glucosinolate or myrosinase QTL co-localize with *Pieris* QTL, none of these traits explained the resistance QTL convincingly, indicating that resistance against specialist insect herbivores is influenced by other traits than resistance against generalists.**

Citation: Pfalz M, Vogel H, Mitchell-Olds T, Kroymann J (2007) Mapping of QTL for Resistance against the Crucifer Specialist Herbivore *Pieris brassicae* in a New Arabidopsis Inbred Line Population, Da(1)-12×Ei-2. PLoS ONE 2(6): e578. doi:10.1371/journal.pone.0000578

INTRODUCTION

Arabidopsis thaliana recombinant inbred lines (RILs) have been widely used for mapping of quantitative trait loci (QTL) (reviewed in [1]). Taking advantage of RILs derived from crosses between the accessions Columbia (Col) and Landsberg *erecta* (L*er*) [2], and between L*er* and Cape Verdi Islands (Cvi) [3], several insect resistance QTL have been mapped and, subsequently, several were cloned and characterized [4–8]. In most cases, these studies involved lepidopteran species with a broad host range (generalists) such as *Spodoptera exigua* or *Trichoplusia ni*, and found that generalist insects were sensitive towards glucosinolate-based defenses. Glucosinolates (β-thioglucoside–N-hydroxysulfates) are amino acid-derived secondary plant metabolites that can be hydrolyzed by myrosinases, enzymes with β-thioglucoside glucohydrolase activity [9–11]. In *Arabidopsis thaliana*, three major classes of glucosinolates are known: aliphatic glucosinolates derived from chain-extended methionine homologues, indole glucosinolates derived from tryptophan, and benzyl glucosinolates originating from a phenylalanine precursor [12]. Two major loci, *AOP* [13,14] and *MAM* [7,15–17], and several minor loci [18] control composition and quantity of aliphatic glucosinolates. Methylthioalkylmalate synthases encoded at the *MAM* locus determine the side chain length of the methionine-derived precursors, while 2-oxoglutarate-dependent dioxygenases encoded at *AOP* modify the side chain structure. In intact plant tissue, glucosinolates and myrosinases are localized in separate cell types [19–22]. Upon tissue disruption, myrosinase-catalyzed glucosinolate hydrolysis results in the formation of bioactive products, including isothiocyanates, nitriles, thiocyanates and others [23]. The types of breakdown products formed depend on the glucosinolate structure, as well as on myrosinase-associated or –binding proteins that can direct the formation of breakdown products towards nitriles or isothiocyanates [5,8]. Typically, plant damage caused by generalist insect herbivores is negatively correlated with increasing glucosinolate concentration or myrosinase activity, and resistance QTL co-localize with glucosinolate biosynthesis or hydrolysis QTL, providing evidence for a major role of the glucosinolate-myrosinase system in the defense of cruciferous plants against generalist insect herbivores [5–9,24].

Entirely unclear, however, is how cruciferous plants defend themselves against specialist insect herbivores. Several counter-adaptations have been identified in crucifer specialist lepidopterans that render the glucosinolate-myrosinase system ineffective. *Plutella xylostella* (diamondback moth) larvae express a glucosinolate sulfatase in their gut that removes the sulfate moiety from glucosinolates, thereby preventing myrosinase-catalyzed hydrolysis and formation of toxic breakdown products [25]. *Pieris rapae* (cabbage white butterfly) possesses a nitrile-specifier protein (NSP) that redirects glucosinolate hydrolysis towards the formation of nitriles instead of highly toxic isothiocyanates when plant tissue is ingested by *Pieris* larvae [26]. Nonetheless, Arabidopsis accessions vary for resistance against specialist insect herbivores. In this paper, we analyze quantitative genetic variation for resistance against two crucifer specialist lepidopteran herbivores, *Pieris brassicae* and *Plutella xylostella*,

Academic Editor: Hany El-Shemy, Cairo University, Egypt

Received April 25, 2007; **Accepted** June 1, 2007; **Published** June 27, 2007

Funding: This work was supported by the Max Planck Society, the Bundesministerium für Bildung und Forschung (PTJ-BIO/0313063), the Deutsche Forschungsgemeinschaft (KR 2237/2-1 and KR 2237/2-2), and the European Commission (Contract QLRT-2000-01097).

Competing Interests: The authors have declared that no competing interests exist.

* **To whom correspondence should be addressed.** E-mail: kroymann@ice.mpg.de

among Arabidopsis accessions and in a new Arabidopsis recombinant inbred line, Da(1)-12×Ei-2, and we investigate whether variation in glucosinolate profiles, myrosinase activity or trichomes contributes to variation in resistance.

RESULTS

Natural Variation for Resistance against Crucifer Specialist Insect Herbivores among Arabidopsis Accessions

We analyzed 16 Arabidopsis accessions for natural variation in resistance against two crucifer specialist insects, *Pieris brassicae* and *Plutella xylostella*. We found substantial variation for resistance against *P. brassicae* ($F = 20.31$, $df = 15$, $N = 973$, $P < 0.00001$), with Col-0 being the most resistant and Tsu-0 the most susceptible accession (**Figure 1**). Variation for resistance against *P. xylostella* was less pronounced, but nonetheless statistically significant ($F = 2.38$, $df = 15$, $N = 912$, $P < 0.005$). Resistance to *P. brassicae* and *P. xylostella* was positively correlated ($r = 0.55$, $df = 14$, $P < 0.05$), suggesting that some determinants of plant resistance affect both specialists similarly.

We also analyzed natural genetic variation for trichome density, glucosinolate content and myrosinase activity, traits known to influence resistance against generalist insect herbivores. As expected, these traits varied among Arabidopsis accessions (**Figure 1**). However, no single trait alone could explain the observed variation in resistance to *P. xylostella* or *P. brassicae* among Arabidopsis accessions. We therefore chose to analyze quantitative variation for resistance and defense-related traits in a new Arabidopsis recombinant inbred line population, derived from a cross between the parental accessions Da(1)-12 and Ei-2 [5]. Although these lines did not represent the extreme phenotypes in the distribution of resistance against *P. brassicae*, they provided a variety of advantages regarding the composition of alleles at glucosinolate biosynthesis and hydrolysis loci compared to 'standard' mapping populations such as Col×L*er* (2) or L*er*×Cvi [3]. Leaves of Da(1)-12 and Ei-2 synthesize aliphatic and indole glucosinolates. Both lines accumulate similar quantities of glucosinolates in their leaf tissue (**Figure 1**). Furthermore, in both lines the predominant aliphatic glucosinolates are derived from a homo-methionine precursor, indicating the presence of an intact *MAM2* gene and the absence of a functional *MAM1* gene in the *MAM* gene cluster [7,27]. However, Da(1)-12 and Ei-2 differ in their alleles at the *AOP* locus [14]. Da(1)-12 possesses an *OHP* allele at *AOP*, while Ei-2 carries an *ALK* allele. Therefore, Da(1)-12 produces mainly 3-hydroxypropyl and 3-methylsulfinyl glucosinolates, and Ei-2 accumulates allyl glucosinolate. Finally, during glucosinolate hydrolysis, Da(1)-12 produces isothiocyanates whereas Ei-2 generates predominantly nitriles [5]. This combination of alleles at glucosinolate biosynthesis and hydrolysis loci helps reduce complexity in the investigation of potential impact of the glucosinolate-myrosinase system [9] on herbivory. It avoids epistatic interactions between known major biosynthesis loci, *AOP* and *MAM*, [12] while allowing analysis of potential effects of interactions between glucosinolate biosynthesis and hydrolysis loci on crucifer specialists. In addition, growth rates of both Da(1)-12 and Ei-2 were nearly identical (**Figure 1**), improving the accuracy of estimating the quantity of tissue removal during herbivory.

RIL Genotyping

F_9 progeny from 201 Da(1)-12×Ei-2 RILs was genotyped with 84 markers. Out of a total of 16,884 PCRs, only 126 failed or yielded ambiguous results. Residual heterozygosity was low, although the observed value of ca. 1.2% was larger than 0.4% which is expected for F_9 progeny obtained by repeated selfing, possibly indicating some heterozygote advantage. Also, a significant deviation from expected 1:1 genotype frequencies was observed for a large segment of chromosome 1, comprising markers *C1P12* to *B12* (**Figure 2**), with the most significant excess of the Da(1)-12 genotype at marker *F1K23_2* ($\chi 2 = 37.05$, $df = 1$, $P < 0.001$). Such distortion has also been observed for other RIL populations [2,3], and may have been caused by unintentional selection during RIL generation. Nonetheless, the order of all genetic markers in the Da(1)-12×Ei-2 RILs was compatible with their physical position in the Col-0 sequence.

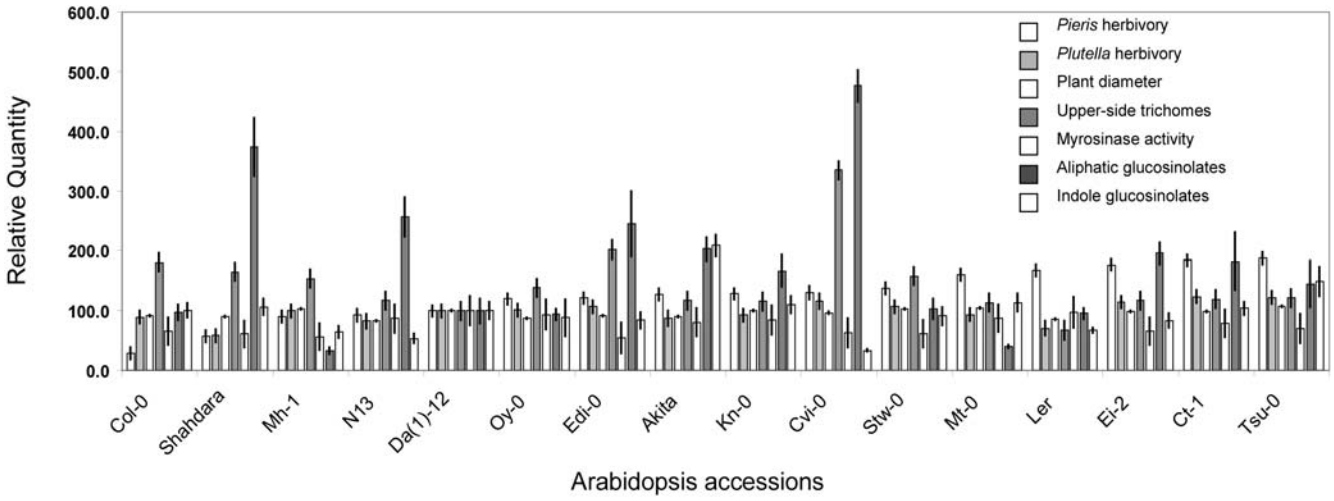

Figure 1. Natural variation among 16 Arabidopsis accessions for insect resistance, growth and defense-related traits. Accessions are ranked according to increasing susceptibility to herbivory by *Pieris brassicae* larvae. Shown are least squares means and standard errors (vertical bars) for *Pieris brassicae* herbivory, *Plutella xylostella* herbivory, plant diameter, leaf upper-side trichomes, myrosinase activity, aliphatic and indole glucosinolates. Values for Da(1)-12 are set as 100.
doi:10.1371/journal.pone.0000578.g001

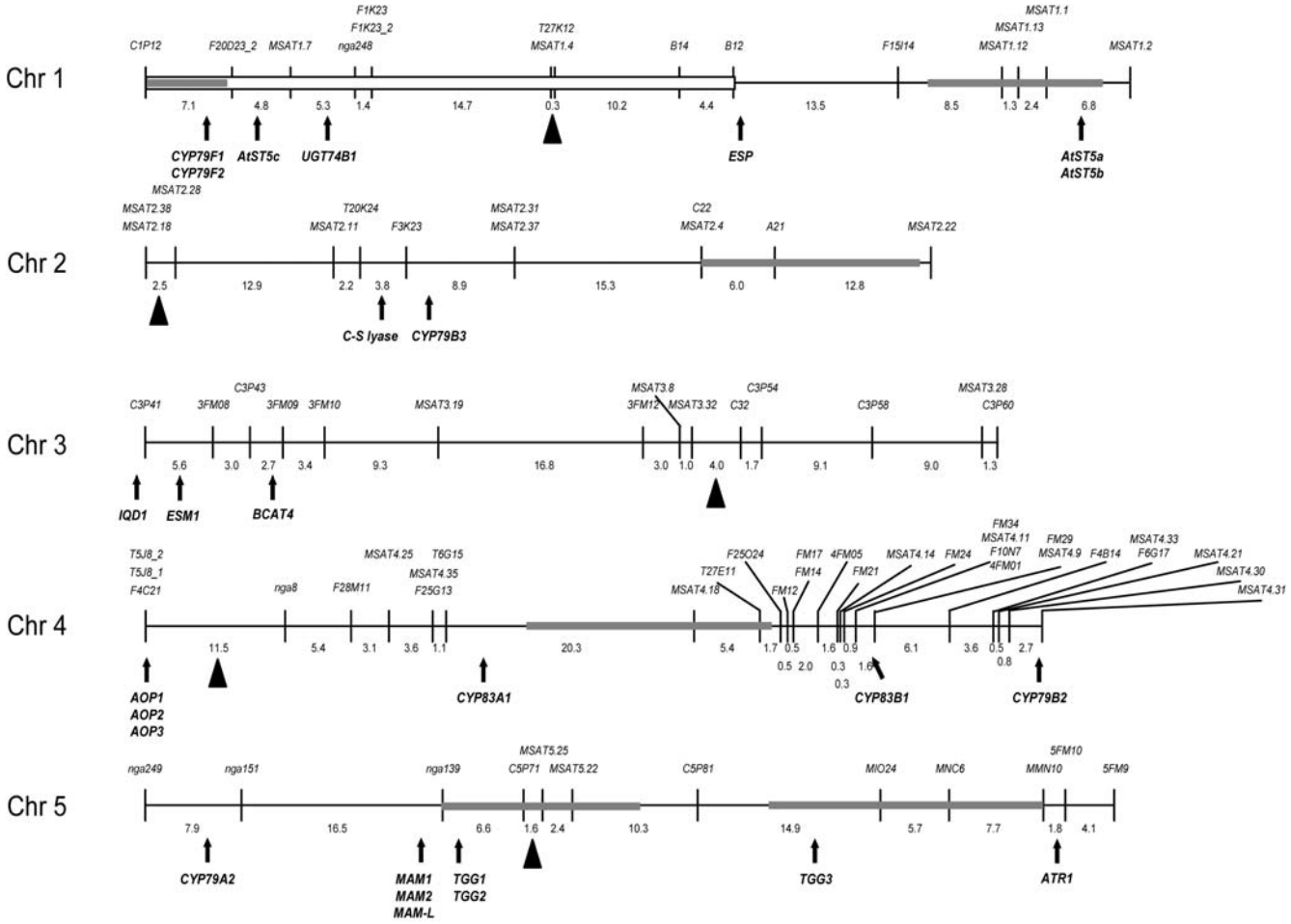

Figure 2. Genetic map of Da(1)-12×Ei-2 recombinant inbred lines. Shown are markers and genetic distances between adjacent markers along the chromosomes. Triangles indicate the approximate location of centromeres. The white bar for chromosome 1 indicates a region with marker distortion. Grey bars indicate 2-LOD support intervals for *Pieris* resistance QTL. Known genes involved in glucosinolate biosynthesis, hydrolysis, and gene regulation are shown below chromosomes; explanations, AGI numbers and references are given in **Table S2**.
doi:10.1371/journal.pone.0000578.g002

QTL for Resistance against Crucifer Specialist Insect Herbivores

We identified six QTL for resistance against *P. brassicae* herbivory, each two on chromosomes 1 and 5, and each one on chromosomes 2 and 4 (**Figure 3**). At all QTL, the Da(1)-12 genotype confers higher resistance to *P. brassicae*, and each Da(1)-12 allele increases resistance by 10–20%, consistent with higher resistance in the Da(1)-12 parental line (**Figure 1**). Together, these QTL explain nearly half of the phenotypic variance in our experiments ($R^2 = 48.4\%$). In contrast, we did not find any QTL for resistance against *P. xylostella* herbivory with composite interval mapping (CIM), while Bayesian interval mapping (BIM) indicated the presence of one weak QTL, located at the same position as the *Pieris* herbivory QTL on chromosome 2 (**Figure 3**).

We tested for epistasis between major herbivory QTL but detected a significant interaction only between markers *C1P12* and *MSAT1.1* ($F = 6.01$, $df = 1$, $N = 180$, $P < 0.05$) which correspond to the two QTL on chromosome 1. Here, a Da(1)-12 allele at *MSAT1.1* reduces plant damage by ca. 18% when plants carry the Da(1)-12 allele at *C1P12*, but only by ca. 4% when the allele at *C1P12* is Ei-2. *Vice versa*, a Da(1)-12 allele at *C1P12* reduces damage by ca. 20% in plants with a Da(1)-12 allele at *MSAT1.1* but only by ca. 5% in plants with an Ei-2 allele at *MSAT1.1*.

Confirmation of a *Pieris* Resistance QTL with a Heterogeneous Inbred Family Strategy

We used a heterogeneous inbred family (HIF) strategy [28] to confirm the *Pieris* resistance QTL near marker *C1P12*. This strategy utilizes residual heterozygosity in a RIL population, and compares the phenotypes of genotyped progeny from a line heterozygous at a QTL candidate marker. HIF allows the rapid generation of a plant family homozygous for the majority of the genome but segregating at the candidate QTL.

We chose RIL DE196 which was heterozygous at both *Pieris* QTL on chromosome 1 (**Table S1**). From progeny of this line, we selected plants that carried either only Da(1)-12 or only Ei-2 alleles at marker *MSAT1.1*, but segregated at *C1P12*. Our statistical model accounted for plant size, flat and position effects. As expected, plants with a Da(1)-12 allele at *C1P12* experienced significantly less damage in *P. brassicae* herbivory screens than plants with an Ei-2 allele at this locus when *MSAT1.1* was homozygous Da(1)-12 ($F = 18.70$; $df = 1$; $N = 130$, $P < 0.05$) but not

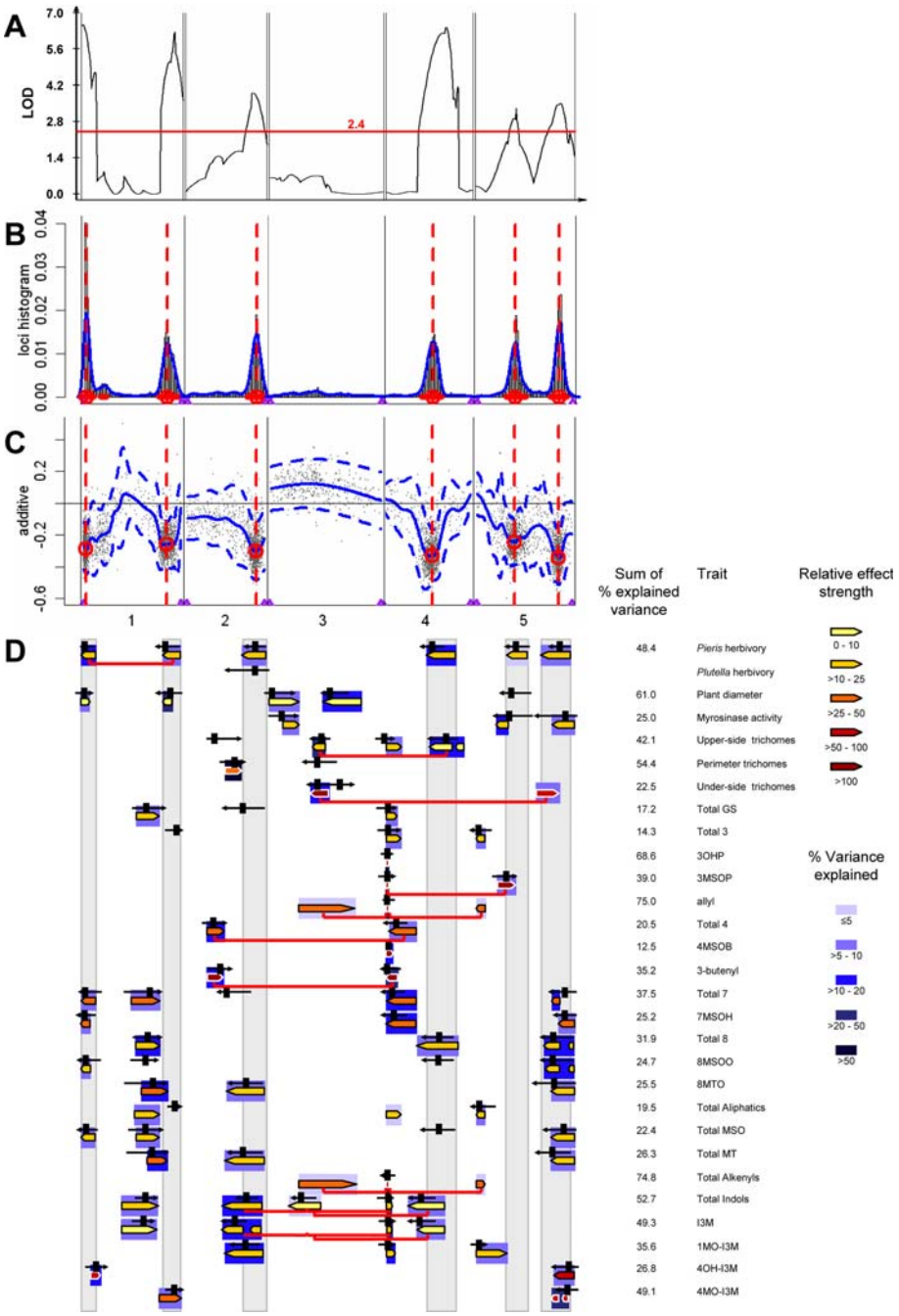

Figure 3. QTL for insect resistance and defense-related traits in Da(1)-12×Ei-2 recombinant inbred lines. A. QTL for resistance against *Pieris brassicae*, obtained with composite interval mapping (CIM). The horizontal red line indicates the significance threshold for this trait. **B.** Bayesian interval mapping (BIM) detects the same QTL as CIM. Horizontal red lines indicate high density probability regions in BIM, vertical red lines high density probability peaks. **C.** Additive effects in BIM, shown as a scatter plot with a smoothing spline fit (solid blue line) plus or minus two standard errors (dashed blue lines). For all six QTL, *Pieris* larvae cause greater damage when plants carry the Ei-2 alleles at the QTL. Hence, the Da(1)-12 alleles confer higher resistance. **D.** QTL for resistance against specialist lepidopterans, plant diameter, myrosinase activity, trichome density on the leaf upper and under-sides and perimeter, and for glucosinolates. For glucosinolates, QTL for individual compounds and for sum variables are given. Abbreviations are as follows: 3OHP = 3-hydroxypropyl; 3MSOP = 3-methylsulfinylpropyl; 4MSOB = 4-methylsulfinylbutyl; 7MSOH = 7-methylsulfinyl-heptyl; 8MSOO = 8-methylsulfinyloctyl; 8MTO = 8-methylthiooctyl; I3M = indol-3-yl-methyl; 1MO-I3M = 1-methoxy-indol-3-yl-methyl; 4OH-I3M = 4-hydroxy-indol-3-yl-methyl; 4MO-I3M = 4-methoxy-indol-3-yl-methyl. Total 3, 4, 7 and 8 are the sums of homomethionine-, di-homomethionine-, penta-homomethionine-, and hexa-homomethionine-derived glucosinolates, respectively. Total MSO, MT, and alkenyls are the sums of aliphatic glucosinolates with methylsulfinyl-, methylthio-, and alkenyl-groups, respectively. Total aliphatics and indoles are the sums of methionine- and tryptophan-derived glucosinolates, respectively, and total GS is the sum of all glucosinolates. Colored arrows correspond to 2-LOD support intervals for QTL identified with CIM, black arrows for high density probability regions in BIM, with vertical black bars showing the position of the high density probability peaks. Arrow directions correspond to effect directions; arrows pointing left indicate that the Ei-2 allele has a stronger effect on a particular trait. Arrow fill colors code for the relative effect strength of a QTL, arrow background colors for R^2, the percentage of variance explained by a QTL. Horizontal red lines connecting colored arrows indicate epistatic interactions between QTL.
doi:10.1371/journal.pone.0000578.g003

when *MSAT1.1* was homozygous Ei-2 ($F = 0.37$; $df = 2$, $N = 176$, n.s.).

QTL for Trichome Density

Leaf hairs, or trichomes, can contribute to plant defense against herbivorous insects in Arabidopsis [29,30] and related plant species [31]. Therefore, we mapped QTL for trichome numbers on the leaf upper and under-side surfaces and the leaf perimeter (**Figure 3**). We identified three QTL controlling trichome density on the leaf upper sides, one on chromosome 3 and two with opposing effect on chromosome 4. These QTL may correspond to trichome QTL that have been mapped in several other Arabidopsis recombinant inbred lines [32]. For trichomes on the leaf perimeter, we identified only one QTL on chromosome 2. This QTL maps to approximately the same position as a major trichome QTL previously identified in the Col×L*er* RIL population [33]. Two QTL with opposing effects control trichome density on the leaf under-side, located on chromosomes 3 and 5, with the one in the center of chromosome 3 sharing its position with a QTL for trichome density on the leaf upper-side. Finally, we found epistatic interactions for both upper- and under-side trichomes. For trichome numbers on the leaf upper-side, we found a significant interaction ($F = 4.13$, $df = 1$, $N = 94$, $P < 0.05$) between markers *3FM12*, close to the QTL LOD peak on chromosome 3, and *MSAT4.18*, which corresponds to the QTL near the center of chromosome 4. Trichome numbers were highest when both markers carried Ei-2 alleles and lowest when both markers had the Da(1)-12 genotype. A Da(1)-12 allele at *MSAT4.18* reduced trichome numbers on the leaf upper-side by ca. 17% when the allele at *3FM12* was Da(1)-12 but only by ca. 3.5% when *3FM12* had the Ei-2 allele. Likewise, the two QTL for leaf under-side trichomes interacted epistatically ($F = 5.49$, $df = 1$, $N = 92$, $P < 0.05$). Here, trichome numbers were highest when *3FM12* carried the Ei-2 allele and *MIO24*, on chromosome 5, had the Da(1)-12 allele. Substitution of the Da(1)-12 allele at *MIO24* with an Ei-2 allele resulted in a reduction of trichome numbers by ca. 64%, substitution of the Ei-2 allele at *3FM12* in a reduction by ca. 87%. Finally, substitution of both alleles with the reciprocal genotypes led to a reduction by ca. 96% such that trichomes were rarely detected on the leaf under-side of RILs with a Da(1)-12 allele at *3FM12* and an Ei-2 allele at *MIO24*.

QTL for Myrosinase Activity

We have identified three QTL for myrosinase activity that exceeded the significance threshold, one on chromosome 3, and two on chromosome 5 (**Figure 3**). At the QTL near the top of chromosome 3, the Ei-2 genotype confers higher myrosinase activity, while at the other QTL the Da(1)-12 alleles are more active. In a previous study, Mitchell-Olds and Pedersen [34] had identified two different myrosinase QTL in the Col×L*er* RIL population, one on chromosome 1 and the other near the center of chromosome 3. These QTL map to different locations than the ones identified in the present work. However, two of the known myrosinase genes in *Arabidopsis thaliana, TGG1* and *TGG2* [35,36], map close to the LOD peak of the first myrosinase QTL on chromosome 5 (**Figures 2, 3**), and likely cause this QTL. The third known myrosinase gene, *TGG3*, is a pseudogene in all accessions investigated so far [37], and does not map within the 2-LOD support interval of the second QTL on chromosome 5, although it is located in its vicinity. Hence, two of the QTL identified in Da(1)-12×Ei-2 represent novel myrosinase QTL.

QTL for Aliphatic Glucosinolates

Ei-2 leaves produce ca. 50 – 60% more total glucosinolates than Da(1)-12 leaves, but this difference is small compared to the variation present among *A. thaliana* accessions (**Figure 1**; [18]). Only few QTL control total glucosinolate quantity, one on chromosome 1 and one near the top of chromosome 4 (**Figure 3**), which very likely corresponds to *AOP* (**Figure 1**; [14]). Total aliphatic glucosinolate accumulation is influenced by three QTL; *AOP*, a QTL near the bottom of chromosome 1, and a QTL near the top of chromosome 5. While a QTL for aliphatic glucosinolates near the top of chromosome 5 has also been identified in the L*er*×Cvi [12] and Col×L*er* RILs [6], the QTL near the bottom of chromosome 1 was previously unknown. The *AOP* locus (or a closely linked gene) also constitutes a QTL for nearly all individual aliphatic glucosinolates except for the hexa-homomethionine-derived glucosinolates, 8-methylsulfinyloctyl and 8-methylthiooctyl glucosinolate.

Because both parental lines lack a functional *MAM1* gene, most aliphatic glucosinolates are homomethionine-derived, with only small quantities of di-homomethionine-derived glucosinolates being detectable. However, the genetic architecture underlying the biosynthesis of short chain aliphatic glucosinolates (*i.e.*, homo- and di-homomethionine derivatives) is nonetheless complex. The side chain structure of homomethionine-derived glucosinolates is modified by alleles at the *AOP* locus. RILs with the Da(1)-12 *OHP* allele accumulate 3-hydroxypropyl and 3-methylsulfinylpropyl glucosinolates, while lines with the Ei-2 *ALK* allele accumulate allyl glucosinolates. Therefore, QTL effects for 3-hydroxypropyl and 3-methylsulfinyl glucosinolates have the same direction, but are opposite to the QTL effect for allyl glucosinolate (**Figure 3**). The quantity of 3-methylsulfinylpropyl glucosinolate is also influenced by a QTL on chromosome 5, and by an epistatic interaction between this QTL and *AOP*. The QTL on chromosome 5 is located near marker *nga139*, in the vicinity of the *MAM* genes (**Figure 2**). Among the RILs that harbor the Da(1)-12 allele at *AOP* and are, thus, capable of producing 3-methylsulfinylpropyl glucosinolate, those lines with a Da(1)-12 allele at *nga139* accumulate two- to threefold more 3-methylsulfinylpropyl glucosinolate than lines with an Ei-2 allele at this marker (**Figure 4**).

Likewise, allyl glucosinolate accumulation is not only determined by *AOP* but also by two further QTL, on chromosome 3 (marker *3FM12*) and near the top of chromosome 5 (marker *nga249*), and by pairwise epistatic interactions between *AOP* and these QTL (**Figure 4**). Allyl glucosinolates are only synthesized when RILs harbor the Ei-2 allele at *AOP*, with a Da(1)-12 allele at marker *3FM12* increasing and a Da(1)-12 allele at marker *nga249* decreasing allyl glucosinolate accumulation. Since allyl glucosinolates account for the majority of alkenyl glucosinolates in Da(1)-12×Ei-2 leaves, total alkenyl glucosinolates follow the same pattern of QTL and epistatic interactions.

Within the *MAM* gene cluster, *MAM1* and/or *MAM2* control variation in short-chain aliphatic glucosinolate accumulation [7,27]. The third gene in this cluster, *MAM-L*, is essential for the production of long chain glucosinolates, and a *MAM-L* knock-out abolished the formation of long chain aliphatic glucosinolates [38]. With the possible exception of 3-methylsulfinylpropyl glucosinolate, the *MAM* locus has no detectable influence on glucosinolate profile variation in Da(1)-12×Ei-2 (**Figure 3**). Nonetheless, in other regions of the genome we found several QTL for long chain aliphatic glucosinolates, *i.e.* for penta- and hexa-homomethionine-derived glucosinolates. Two QTL are located on chromosome 1, two on chromosome 4, and one is positioned near the bottom of chromosome 5. Hence, these additional QTL control variation in the accumulation of long-chain aliphatic glucosinolates in Da(1)-12×Ei-2, even though the *MAM* locus does not contribute to this variability.

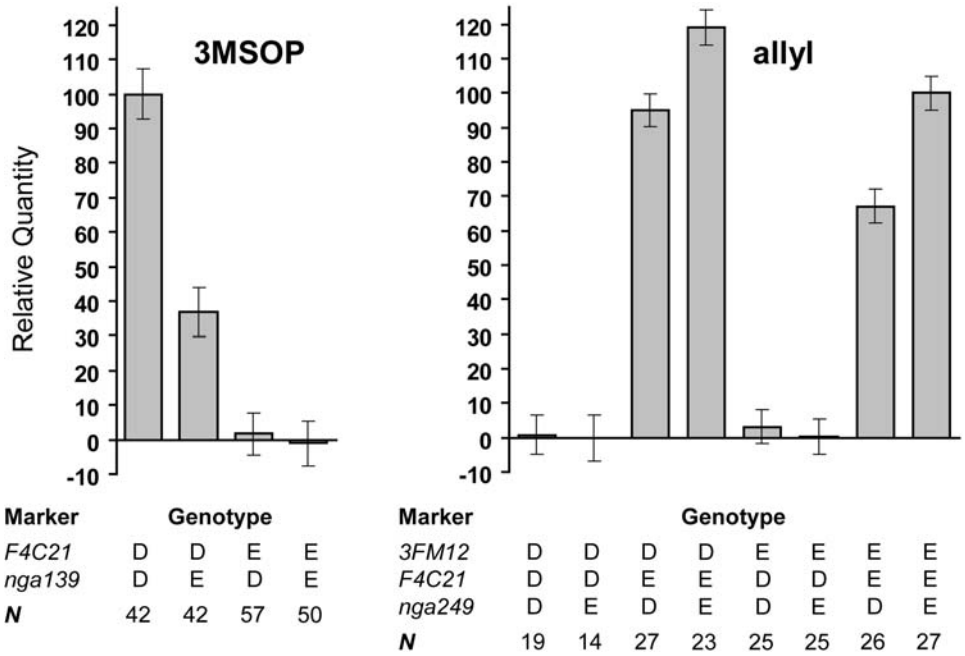

Figure 4. Epistatic interactions in the biosynthesis of homomethionine-derived glucosinolates. Left: 3-methylsulfinylpropyl glucosinolate (3MSOP) is produced when the genotype at the *AOP* locus (marker *F4C21*) is Da(1)-12. An Ei-2 allele at marker *nga139* reduces the accumulation of this glucosinolate by 60%. Right: allyl glucosinolate is produced when the genotype at *AOP* is Ei-2. A Da(1)-12 allele at *3FM12* increases and a Da(1)-12 allele at *nga249* decreases leaf allyl glucosinolate accumulation. *N* indicates the number of RILs with a particular combination of genotypes at the different markers.
doi:10.1371/journal.pone.0000578.g004

QTL for Indole Glucosinolates

We identified five QTL for total indole glucosinolate accumulation, each one on chromosomes 1, 2 and 3, and two on chromosome 4. The QTL near the top of chromosome 4 had the strongest effect on indole glucosinolate accumulation. This QTL co-localizes with the *AOP* locus, suggesting that *AOP* influences not only aliphatic glucosinolates but also indole glucosinolate accumulation, either directly by catalyzing a biosynthetic reaction step or indirectly via utilization of a pool of metabolites that is shared in aliphatic and indole glucosinolate biosynthesis. Alternately, a gene tightly linked to *AOP* could also explain this indole glucosinolate QTL.

We also detected complex epistatic interactions between the QTL for total indole glucosinolate accumulation. We found pairwise epistatic interactions between the QTL on chromosome 3 (marker *MSAT3.19*) and the QTL near the bottom of chromosome 2 (marker *C22*), between *MSAT3.19* and the QTL near the top on chromosome 4 (marker *F4C21*), and between *MSAT3.19* and the QTL at the center of chromosome 4 (marker *T6G15*), as well as a triple interaction between *C22*, *MSAT3.19* and *F4C21* (**Figure 3**, **Table 1**).

The QTL pattern for indol-3-yl-methyl glucosinolate, the most abundant indole glucosinolate in *A. thaliana* leaves [18], reflects largely the QTL pattern for total indole glucosinolate accumulation, except that the QTL for total indole glucosinolates on chromosome 3 did not exceed the significance threshold for indol-3-yl-methyl glucosinolate. Nonetheless, marker *MSAT3.19* showed the same pattern of epistatic interactions for indol-3-yl-methyl glucosinolates as for total indole glucosinolates, when we included it in our statistics models.

Da(1)-12×Ei-2 leaves synthesize three further indole glucosinolates, 1-methoxy-indol-3-yl-methyl, 4-hydroxy-indol-3-yl-methyl and 4-methoxy-indol-3-yl-methyl glucosinolate. 1-methoxy-indol-3-methyl glucosinolate shares two of its three QTL with indol-3-yl-methyl glucosinolate but has an additional QTL near the top of chromosome 5. This additional QTL might correspond to the position of a gene responsible for the generation of the methoxy-group at position 1 of the tryptophan moiety.

4-hydroxy-indol-3-yl-methyl glucosinolate accumulation is controlled by two QTL, one near the top of chromosome 1 and the second near the bottom of chromosome 5 (**Figure 3**). The position of this QTL on chromosome 5 and its effect direction are shared by one of the QTL controlling 4-methoxy-indol-3-yl-methyl glucosinolate accumulation, suggesting a biosynthetic connection between 4-hydroxy-indol-3-yl-methyl and 4-methoxy-indol-3-yl-methyl glucosinolate. However, the second QTL for 4-methoxy-indol-3-yl-methyl glucosinolate is located at a different position, near the bottom of chromosome 1, while the second 4-hydroxy-indol-3-yl-methyl glucosinolate QTL maps near the top of chromosome 1.

DISCUSSION

Is Herbivory by *Pieris brassicae* Influenced by Variation in Glucosinolates, Myrosinase Activity or Trichomes?

QTL for different traits may be considered to co-localize when their 2-LOD support intervals overlap. Co-localization of QTL does, of course, not prove that these QTL are caused by the same gene. Likewise, linked QTL for different traits may have the same cause, even when their 2-LOD support intervals do not overlap, due to the complexity of the statistics that guide QTL mapping. Nonetheless, co-localization of QTL for different traits may provide an approximation for comparing the genetic architecture underlying different but potentially related traits.

We have identified six QTL for resistance against *P. brassicae*, each two on chromosomes 1 and 5, and each one on chromosomes

Table 1. Epistatic interactions in glucosinolate biosynthesis

Trait	Source	*d.f.*	*F*-ratio	*P*
3-methylsulfinylpropyl	*F4C21*	1	102.54	0.00000
	nga139	1	23.81	0.00000
	F4C21×nga139	1	20.17	0.00001
	Error	187		
allyl	*3FM12*	1	8.59	0.00382
	F4C21	1	632.11	0.00000
	nga249	1	12.54	0.00051
	3FM12×F4C21	1	10.99	0.00111
	3FM12×nga249	1	0.20	0.65704
	F4C21×nga249	1	16.54	0.00007
	3FM12×F4C21×nga249	1	0.51	0.47812
	Error	178		
Total 4	*F3K23*	1	23.02	0.00000
	nga8	1	20.58	0.00001
	F3K23×nga8	1	9.88	0.00195
	Error	182		
3-butenyl	*F3K23*	1	20.77	0.00001
	F4C21	1	56.73	0.00000
	F3K23×F4C21	1	15.06	0.00015
	Error	181		
Total Indoles	*B12*	1	11.49	0.00090
	C22	1	17.21	0.00006
	MSAT3.19	1	14.35	0.00022
	F4C21	1	31.28	0.00000
	T6G15	1	6.02	0.01534
	C22×B12	1	1.68	0.19748
	MSAT3.19×B12	1	0.30	0.58454
	F4C21×B12	1	0.40	0.52740
	T6G15×B12	1	0.00	0.96366
	MSAT3.19×C22	1	8.33	0.00450
	F4C21×C22	1	0.04	0.84950
	T6G15×C22	1	0.01	0.94323
	F4C21×MSAT3.19	1	6.36	0.01279
	T6G15×MSAT3.19	1	7.27	0.00783
	T6G15×F4C21	1	0.55	0.46085
	MSAT3.19×C22×B12	1	0.00	0.96727
	F4C21×C22×B12	1	1.01	0.31703
	T6G15×C22×B12	1	0.00	0.96243
	F4C21×MSAT3.19×B12	1	0.17	0.68519
	T6G15×MSAT3.19×B12	1	0.73	0.39366
	T6G15×F4C21×B12	1	0.41	0.52310
	F4C21×MSAT3.19×C22	1	13.69	0.00031
	T6G15×MSAT3.19×C22	1	2.05	0.15448
	T6G15×F4C21×C22	1	0.23	0.63376
	T6G15×F4C21×MSAT3.19	1	0.69	0.40783
	Error	144		

doi:10.1371/journal.pone.0000578.t001

2 and 4. For many of the traits that we investigated because they are known to influence resistance against generalist insect herbivores, we found one or more QTL that co-localize with *Pieris* resistance QTL. However, one of the three myrosinase activity QTL does not co-localize with a *Pieris* resistance QTL. Only one of the three QTL for leaf upper-side trichomes, on chromosome 4, co-localizes with a resistance QTL but this QTL shows the wrong sign for its effect. One would expect that trichomes provide physical resistance against insect herbivory and, thereby, reduce plant damage. However, plant damage increases with increasing numbers of leaf trichomes controlled by this locus. None of the leaf perimeter trichome QTL co-localizes with a resistance QTL, and for QTL controlling leaf under-side trichomes, only one of two QTL co-localizes.

For total glucosinolate accumulation, we identified two QTL, but neither co-localizes with a *Pieris* resistance QTL. Likewise, none of the QTL for total aliphatic glucosinolate accumulation co-localizes with a resistance QTL, and only two of the five QTL for total indole glucosinolate accumulation co-localize with herbivory QTL. A similar picture emerges for QTL controlling sums of homomethionine, di-homomethionine, penta-homomethionine and hexa-homomethionine-derived glucosinolates or for QTL controlling total methylthio, methylsulfinyl or alkenyl glucosinolates (**Figure 3**). Hence, sum variables for glucosinolate classes do not account for resistance against *P. brassicae*. But also QTL for individual glucosinolates do not explain the QTL for *Pieris* resistance. The QTL with the largest impact on glucosinolate profiles, *AOP*, does not co-localize with any of the resistance QTL. And for all individual glucosinolates, either one or more QTL do not co-localize with herbivory QTL, or at least one QTL has a different sign for its effect than the others, while all *Pieris* QTL have the same sign for their effect, with the Da(1)-12 allele improving resistance against herbivory by *Pieris* larvae. Finally, none of the *Pieris* QTL maps near the two major loci that specify glucosinolate hydrolysis product identity, *ESP* [5] and *ESM1* [8] (**Figure 2**), although both parental lines, Da(1)-12 and Ei-2, display sequence polymorphisms in the *ESP* genomic region that correlate with *ESP* expression [5]. Hence, we conclude that none of the investigated traits, trichomes, myrosinase activity or glucosinolate accumulation, appears to cause the QTL for *Pieris* resistance in Da(1)-12×Ei-2 RILs. Further fine-mapping and, ultimately, cloning of the underlying genes causing the *Pieris* resistance QTL is necessary to help understand how Arabidopsis and other crucifers defend themselves against specialist insect herbivores such as *P. brassicae*.

Prospects of Improving Insect Resistance in Cruciferous Crops

Pieris brassicae and other Pieridae are some of the most serious pests on cruciferous crop plants such as rapeseed, cauliflower, or broccoli [39–41]. We have analyzed herbivory by *P. brassicae* larvae with a new RIL population, obtained from a cross between the parental lines Da(1)-12 and Ei-2. We found no detectable effect of glucosinolates or myrosinase activity on larval herbivory, indicating that the variation in the glucosinolate-myrosinase system that is present in our RIL population does not contribute to variation in plant damage caused by *Pieris* larvae. Nonetheless, the glucosinolate-myrosinase system does play a role in the interaction between *P. brassicae* and *A. thaliana* or other Brassicaceae: Adult *Pieris* females use glucosinolates and their hydrolysis products to locate host plants for oviposition, and hydrolysis products have a stimulating effect on oviposition for *P. brassicae* and other Pieridae [24,42–45]. Likewise, glucosinolate breakdown products serve as a stimulant for larval feeding initiation [46–48]. This may explain why herbivory by *Pieris rapae*, a close relative of *P. brassicae*, is significantly reduced in *tgg1 tgg2* double mutants

which have very low levels of Arabidopsis wild type myrosinase activity [22]. Hence, a reduction of glucosinolate levels or myrosinase activity in cruciferous crops could potentially reduce plant damage caused by *P. brassicae*. However, a decrease in the effectiveness of the glucosinolate-myrosinase system would very likely render crucifer crops more susceptible to generalist insect herbivores which are sensitive towards glucosinolate-based defenses [5–9,22]. Furthermore, such a manipulation of the glucosinolate-myrosinase system bears the risk that plants could become more attractive to herbivores which usually do not consume crucifers because these insect species have no effective means to withstand toxic products originating from glucosinolate hydrolysis. Thus, manipulating the glucosinolate-myrosinase system to increase resistance against insect herbivores may be problematic. The detection of QTL that appear to be independent of the glucosinolate-myrosinase system may provide a way to solve this dilemma. Manipulating the genes that underlie the detected resistance QTL could help increase crop protection against *P. brassicae*, without interfering with a complex defense system that protects crucifers effectively against most herbivorous insects.

MATERIALS AND METHODS

Plant and Insect Growth Conditions

Arabidopsis thaliana plants for RIL development were grown under continuous light, supplied by Osram Fluora L36/W77 neon bulbs with an intensity of 150 $\mu mol\ s^{-1}\ m^{-2}$ at 20°C and 60% relative humidity in an environment-controlled growth room. Plants for insect herbivory trials, leaf glucosinolate extraction, and myrosinase assays were grown in 11.5 h day/12.5 h night cycles at 22°C and 60% relative humidity (day conditions), and 16°C and 80% relative humidity (night conditions) in an environment-controlled growth room. Here, light was supplied by NH 360 FLX Sunlux ACE bulbs with an intensity of 200 $\mu mol\ s^{-1}\ m^{-2}$. Plants were grown in an autoclaved 1:3 vermiculate/potting soil mix with 20 ml time-release fertilizer (Osmocote) per flat. After sowing into damp potting medium, flats were covered with clear plastic grow domes, and seeds were stratified for 3–4 days at 6°C in the dark. In general, seeds germinated within 2–3 days, and grow domes were removed 5 days after transfer to the light. Then seedlings were transferred to fresh soil in 96-celled 32.5×51 cm2 flats at a density of 1 seedling per cell. All assays were carried out with 3-week old plants.

Pieris brassicae eggs were obtained from Seritech (Warwick, UK). After hatching, insects were reared on *Brassica napus var. oleifera* for 2–3 days before the experiments. *Plutella xylostella* eggs were obtained from New York State Agricultural Experiment Station Geneva, NY, USA, and a colony was maintained at the Max Planck Institute for Chemical Ecology, Jena, Germany. Insects were raised on artificial diet according to published procedures [49].

Generation of Da(1)-12×Ei-2 Recombinant Inbred Lines

Da(1)-12 (accession no. N917) and Ei-2 (N1124) accessions were obtained from the Arabidopsis stock center (Nottingham, U.K.). Except for the initial cross between both accessions, all following generations were propagated by selfing. 215 F_2 plants were randomly selected from the progeny of a single heterozygous F_1 plant. For every advanced generation, 4–8 seeds per line were planted, and a single plant was randomly chosen from each line for seed production. Finally, seeds from a single F_8 plant per line were bulk-collected resulting in a final set of 201 Da(1)-12×Ei-2 RILs. This new RIL population will be made available through the Arabidopsis stock centers.

DNA Extraction, Genotyping, Genetic Map

DNA was isolated as described in [17]. Molecular markers were obtained from publicly available sources (http://www.arabidopsis.org; http://www.inra.fr/qtlat, [49]), or were generated from microsatellite loci identified in the Col-0 genome sequence [51]. More than 150 potential markers were tested for polymorphism between parental lines, using DNA from Da(1)-12, Ei-2, and a 1:1 mixture of DNAs from both lines. 84 PCR products were found suitable for genotyping (**Table 2**), and allowed to distinguish between Da(1)-12 and Ei-2 genotypes, and between homozygous and heterozygous loci on 4% Metaphor agarose (Biozym diagnostics, Germany). PCR reactions contained, in general, ca. 30 ng DNA, 2.3 µl 10x PCR buffer (Qiagen, Germany), 4 nmol of each dNTP, 1.25 pmol of each of both primers, 70 nmol $MgCl_2$, and 0.15 U *Taq* DNA polymerase (Qiagen, Germany) in a 23 µl volume. Cycling conditions were 94°C for 2 min, followed by 38 cycles of 94°C for 15 s, 50°C or 55°C for 15 s, and 72°C for 30 s, with a final extension of 72°C for 2 min on an Applied Biosystems 9700 Thermocycler. Genotyping was carried out with DNA extracted from individual F_9 progeny. A genetic map was constructed with MAPMAKER/EXP Version 3.0 [52]. Genotype data for the Da(1)-12×Ei-2 RILs are available in **Table S1**.

Insect Herbivory Screens

P. brassicae herbivory screens were performed at 7 different times, and each experiment was carried out with at least 3 replicates per RIL. RIL replicates were completely randomized over 96-celled flats. At the beginning of each experiment, plant diameter was measured. One larva was placed on each plant rosette without prior starvation, and larvae were allowed to move freely. After 24 hours of herbivory, the leaf area removed by the insects was assessed visually, and an artificial scale was established to determine the percentage of removed rosette tissue. In total, more than 9000 data points were collected for *P. brassicae* herbivory. *P. xylostella* herbivory screens were performed similarly, except that larvae were starved for 6 hours prior transfer to plants, and larvae were allowed to feed for 2–3 days. Total sample size for *P. xylostella* herbivory exceeded 2400 plants.

Trichomes

Trichome analysis was carried out with 96 RILs chosen to include lines with a maximum number of recombination breakpoints. Per RIL, 4 replicates were analyzed, and trichomes from the 3rd to 6th true leaves of 3-week old plants were counted with a Stemi SV6 binocular (Carl Zeiss, Jena, Germany). Every leaf was placed underneath the binocular such that the leaf tip touched the border of the visual field. This way, only the upper half of the leaf was visible and used for trichome analysis. Trichome numbers were counted for the leaf upper and under-sides and for the leaf perimeter.

Glucosinolate Extraction and HPLC Analysis

Glucosinolates were extracted in a 96-well format as described in [18]. HPLC separation and identification of extracted desulfo-glucosinolates was carried out as described in [17]. In brief, HPLCs were run on a Hewlett Packard HP 1100 system (Agilent), equipped with a HP Lichrocart 250-4 RP18e 5 µm column. The elution was accomplished with a water (solvent A) – acetonitrile (solvent B) gradient using the following program: 1.5 – 5% (v/v) B (6 min), 5 – 7% B (2 min), 7 – 21% B (10 min), 21 – 29% B (5 min), 29 – 43% B (7 min), 43 – 92% B (0.5 min), 92% B (2.5 min), 92 – 1.5% B (0.5 min), 1.5% B (4.5 min). Desulfo-glucosinolates were identified according to retention time and UV spectra, and quantified from

Table 2. Markers used for genotyping of Da(1)-12×Ei-2 RILs

Marker	Chr	BAC/P1	Primer 1 (5′→3′)	Primer 2 (5′→3′)	Gel pattern
C1P12	1	F20B24	CTGGAAGTCCATACCATGAG	GTTCGTCGTTCGTGGTATTG	D>E
F20D23_2	1	F20D23	CCGTCACACCATTCACAATC	CCAACCCCTTATATATCGTTC	D>E
MSAT1.7[a]	1	F12K8	GCTTTTATCAGCTCAAACAT	ACTCTTACGTTTGGAGTTCA	D>E
NGA248[a]	1	F3H9	TCTGTATCTCGGTGAATTCTCC	TACCGAACCAAAACACAAAGG	D<E
F1K23	1	F1K23	GAACCAATAAGGAGGCTCAAC	CCATACGGAGAAACCTTCTTC	D>E
F1K23_2	1	F1K23	CAATTTCGAGTTTCGGATTTTC	CTTCACATCAATGCTTGTAATAG	D<E
MSAT1.4[a]	1	F28L22	CTAAACTAGAACCAGGGGTAA	ACAAAAATCGTGGTGATAATA	D<E
T27K12[a]	1	F7F22	GGAGGCTATACGAATCTTGACA	GGACAACGTCTCAAACGGTT	D>E
B14	1	F11F12	CCATTCTCGTCGTGTTATAAG	GAAATGTTAAGGCCAAAATACAG	D<E
B12	1	T18A20	CAACTCGTTATAACAGGTTTTAC	CCAAATACTAAAGAGGGAATTG	D<E
F5I14[a]	1	F5I14	CTGCCTGAAATTGTCGAAAC	GGCATCACAGTTCTGATTCC	D>E
MSAT1.12[a]	1	T26J14	TTAGAGATTCGCCAACCTC	CGTGTGCCCAACCA	D>E
MSAT1.13[a]	1	F24J5	GTCAAACCAGTTCAATCA	CAACCACCAGGCTC	D<E
MSAT1.1[a]	1	F20P5	ATACGATAAGATTTATTAGCA	CCCATGCTCTTTTTGTGAAA	D>E
MSAT1.2[a]	1	F22K20	TTGAGTGGTGCCGCTTG	ATATCTCCATCGCTGCAACC	D>E
MSAT2.38[a]	2	F18P14	TGTAACGCTAATTTAATTGG	CGCTCTTTCGCTCTG	D>E
MSAT2.18[a]	2	T4E14	TAGTCTCTTTTGGTGCGCATA	AGCCTCTCCAAGCTTAGGTCT	D>E
MSAT2.28[a]	2	T26I20	AATAGAAATGGAGTTCGACG	TGAACTTGTTGTGAGCTTTG	D<E
MSAT2.11[a]	2	F19F24	GATTTAAAAGTCCGACCTA	CCAAAGAGTTGTGCAA	D>E
T20K24	2	T20K24	CAATATTCGTGGGAGTTAGTC	GCTGTCGAATTACATTTCTTTAC	D<E
F3K23	2	F3K23	CTCGCAGCGTCTGCAAATTC	GAAGCGGAAGATGGAGAGAC	D>E
MSAT2.31[a]	2	T22F11	GCTCCTCTTTGCCGCTAG	GCGATTTCATCTTGTGCATC	D>E
MSAT2.37[a]	2	T19L18	GGTTGTTTCATCGAAAGCA	CATGGTCTCGCTGGTGTAT	D<E
C22	2	T26B15	CTTGGCAACTTCATTCAATTTC	GAAAGTAGAGAAGCATTTAGAC	D<E
MSAT2.4[a]	2	T26B15	TGGGTTTTTGTGGGTC	GTATTATTGTGCTGCCTTTT	D<E
A21	2	T1J8	CCATCTAAACTGCTTACGATG	GTGACCCATTCTTCTCTTTTC	D<E
MSAT2.22[a]	2	F17A22	CGATCCAATCGGTCTCTCT	TGGTAACATCCCGAACTTC	D<E
C3P41	3	F9F8	GGTCGTATCCTCTTATCGAAC	CTTGTGAGTGGTCTTATGAAAG	D<E
3FM08	3	K20I9	GGTTCGTATCCAAAAACCAAG	CCATCATTGGAGCAAGAGAC	D>E
C3P43	3	MRC8	CAATGTTGGCTTGGAAATAATG	CATTGCCGGTAAAAATGTTTTTC	D>E
3FM09	3	MAL21	CTAATTACTATGGCGGAGAATTC	CTAAAGAAATCTGCGGTCTTC	D>E
3FM10	3	MSD21	CATTACTTCACTGTTGCTTTAC	GACAGGTTATGGCTTGTTAATC	D>E
MSAT3.19[a]	3	K7M2	TAATTCGATCCAATTGACAT	TGGCTTGGCACAAAC	D<E
3FM12	3	K24A2	TAGGGAAGCATTTGTCTTGAG	TGCTTAAAGTGACGGTAAAATG	D: 1 band E: 2 bands
MSAT3.8[a]	3	K5K13	ATGTTAAAAACCCGTGTTGG	TTTAACCTTATCCGGGAAAG	D>E
MSAT3.32[a]	3	MXO21	GCACTTGCAGCTTAACTT	CGTGACTGTCAAACCG	D>E
C32	3	T15B3	GAAGAGGATGAACAAAGATAAG	CAAATCTGCCTCCTCCATAAG	D<E
MSAT3.28[a]	3	T26I12	TACAAGTCATAATAGAGGC	GGGTTTAGCATTTAGC	D<E
T5J8_2[b]	4	T5J8	CGATCATCGGTGTTCACCTT	GAAAATAAATCGTCATATGGTGTACTG	D>E
T5J8_1[b]	4	T5J8	GCCAAGACGCAGAAGAAGAG	TCTCATTATTCCCCACAATGC	D<E
F4C21	4	F4C21	GCGCTTCATCTAGTTACGCTTT	CCCGGACTGAACCAAACTAA	D>E
NGA8[a]	4	T32A17	GAGGGCAAATCTTTATTTCGG	TGGCTTTCGTTTATAAACATCC	D<E
F28M11	4	F28M11	CACCATATTGGCCTCAAATTG	CAAAAACCCGTTCCACCAAAC	D<E
MSAT4.25[a]	4	F25E4	GAATGGTTGTTGATAGTTGA	AAATTTCAGGAGGTGATAGA	D<E
F25G13	4	F25G13	GCCAGGTTCTTTTCATTTCTC	GGGCGTTTAATTTGCATTCTTC	D<E
MSAT4.35[a]	4	F25G13	CCCATGTCTCCGATGA	GGCGTTTAATTTGCATTCT	D<E
T6G15	4	T6G15	GTAGCCAGAGATGGAAGTTAC	GGGTCCTTACTGAGGCTTTG	D>E
MSAT4.18[a]	4	T12H17	TGTAAATATCGGCTTCTAAG	CTGAAACAAATCGCATTA	D<E
T27E11	4	T27E11	GTGATTCCCGTCTGCTAAAC	CCTCCTTCAGCATCATAGTG	D>E
F25O24	4	F25O24	CAATGTATTTGGATGTGTTTGTTC	GGATGGTAACACGGCTAAAC	D<E

Table 2. cont.

Marker	Chr	BAC/P1	Primer 1 (5′→3′)	Primer 2 (5′→3′)	Gel pattern
FM12	4	T16L4	GAAGCCCCTATGAGATGGTC	GTGAGGGAGTTAGGTAGCAAC	D<E
FM14	4	F9N11	TCAAGGGAGACTTGGAGAAC	GGGATCGTTATGCACTTGTTTG	D<E
FM17	4	F9N11	CTCCCTCTTCGGAGAAATTC	CATCTCTTATAGGCCTCTCTC	D>E
4FM05	4	F17I23	CTAACAGATTTGGTGAATAACAAG	TCATTTGATGTGCCAGTAAATC	D>E
FM21	4	T10C21	CTCAAGCGGTGGAAATTGGAG	GTAAAGAATGTCCAGGGCAG	D<E
MSAT4.14[a]	4	F8F16	GACCGTTTCTAGTGCTCACA	ACGGAATAAGCGGAGGA	D<E
FM24	4	F3L17	GAGCATCCGCGTAGGTTAAG	CACAGAGAGACTCAAAAATACTG	D<E
FM34	4	F11C18	TGGTCTCTCAACTCCAACAC	CATTGAGATTTGAGCCAAACAG	D<E
MSAT4.11[a]	4	F10N7	AAAAATCCGGTAGAGCATCC	CCAATTCCGAGCCAGTAA	D<E
F10N7	4	F10N7	GTTGCTCGAAACCTCTCAATC	GCCTCACCGATACGTTTCTG	D<E
4FM01	4	F8B4	AGTAGATACAATGCGTTGACC	GGAGCGTTAATAGTGTGTATG	D<E
FM29	4	L23H3	GTCCAGGTTGCTGAAGAGAAG	GTATTGTTTGGTTGGTATGAGC	D<E
MSAT4.9[a]	4	F4D11	GAAATCAACGGCTGAG	AAGTAATTAAGACGCTGAGA	D<E
F4B14	4	F4B14	CGTCGTTTATTTCACCACCAC	GGTACAAAGATGGGTTAAACTG	D<E
F6G17	4	F6G17	GACACGCAAACAAAGTAAAAGTC	GATGGTGACATAGACCCAATG	D>E
MSAT4.33[a]	4	F6G17	TTCTTTGACACGCAAACA	TGGTGACATAGACCCAATG	D>E
MSAT4.21[a]	4	F19F18	TTATGCTATGGCTGTTTGGT	CGAAATCTGTTCTTGCATTC	D>E
MSAT4.30[a]	4	F20D10	AGAGCACTCACCGTTCAT	TGTGTTCGTGGATTTACC	D<E
MSAT4.31[a]	4	T5J17	AGGGATATGGATTGAGA	GCCGTATAACTATTGGTT	D<E
NGA249[a]	5	MAH20	TACCGTCAATTTCATCGCC	GGATCCCTAACTGTAAAATCCC	D>E
NGA151[a]	5	F18022	GTTTTGGGAAGTTTTGCTGG	CAGTCTAAAAGCGAGAGTATGATG	D<E
NGA139[a]	5	K18P6	GGTTTCGTTTCACTATCCAGG	AGAGCTACCAGATCCGATGG	D>E
C5P71	5	T26D3	GACGATGGTGGAGTGATAAG	CTTTGACCTCAAACTTAAGTAG	D<E
MSAT5.25[a]	5	MOK9	GCTTAATTTGGGTTAAAT	GCACGCAAGTGACT	D<E
MSAT5.22[a]	5	MWP19	AGAACAAGTTAGGTGGCT	GGGACAAGAATGGAGT	D<E
C5P81	5	MFO20	GTCAAAGAGTTACTCCGTTAC	CGAGACAAGAGCATGTTATATG	D<E
MIO24	5	MIO24	GTACAATAATTTAGAGAGTATTTTG	CTAGCTCAACTTACTGCTTAATG	D<E
MNC6	5	MNC6	GTTTGGGTCCAATGATAAAATC	GCCTATTGGGCTGAGTTTTC	D>E
MMN10	5	MMN10	CAGTGTCGGCTAATTTCGAC	CAGTCGACATTTCAAAGGTTC	D<E
5FM10	5	MFB13	GATTTGACGACTGATTACATAAC	GCTTGAAATTTGTGTGTATTGTC	D>E
5FM09	5	MPA24	CAATTTCTTGTTATCTGCTTATG	CCATTGCCATATGTTTCCCTC	D<E

[a]: Markers are from http://www.inra.fr/qtlat/msat/index.php
[b]: Markers are from [14]
Bold-typed markers were used for QTL mapping.
doi:10.1371/journal.pone.0000578.t002

HPLC peak areas at A229 nm, using published response factors [53,54] to correct for different UV absorption capacities of individual glucosinolates.

Myrosinase Extraction and Analysis

Myrosinase extraction from 100 mg leaf tissue and UV-spectrophotometric activity assays were carried out as described in [22]. Two independent experiments were conducted, once with the complete RIL population, once with the 96 most informative RILs. Relative myrosinase activity was measured as a spectrophotometrical change at 227 nm through breakdown of sinigrin (allyl glucosinolate) within 15 min using a Multiskan Spectrum (Thermo Fisher Scientific, Germany) spectrophotometric plate reader.

Quantitative Analyses

Systat Version 10 (SPSS Inc.) was used to analyze natural variation for insect resistance and related traits in the following 16 Arabidopsis accessions: Akita (Akita, Japan, Versailles identification no. 252 AV), Col-0, Ct-1 (Catania, Italy, N1094), Cvi-0 (Cape Verdi Islands, N902), Da(1)-12 (Czech Republic or Slowakia, N917), Edi-0 (Edinburgh, UK, N1122), Ei-2 (Eifel, Germany, N1124), Kn-0 (Kaunas, Lithuania, N1286), L*er*, Shahdara (Shahdara river (Pamir), Tadjikistan, N929), Mh-1 (Muehlen, Poland, N1368), Mt-0 (Martuba/Cyrenaika, Lybia, N1380), N13 (Konchezero, Russia, CS22491), Oy-0 (Oystese, Norway, N1436), Stw-0 (Stobowa/Orel, Russia, N1538), and Tsu-0 (Tsu, Japan, N1564). ANOVA was used to obtain least squares means for each accession for *P. brassicae* and *P. xylostella* herbivory, for plant diameter in *P. brassicae* herbivory experiments, for myrosinase activity, and for trichomes on the leaf upper and under-side surfaces and the leaf perimeter. For *P. brassicae* (N = 973) and *P. xylostella* herbivory ($N = 912$), the ANOVA model was TRAIT = -CONSTANT+ACCESSION+EXPERIMENT+FLAT(EXPERIMENT)+COLUMN+ROW+PLANT DIAMETER. COLUMN and ROW are variables to control for position effects. These

variables are particularly important in *P. brassicae* herbivory screens to compensate for larval movement during experiments ($F = 19.18$, $df = 11$, $N = 973$, $P < 0.000001$ for COLUMN, and $F = 41.67$, $df = 7$, $P < 0.000001$ for ROW). EXPERIMENT accounts for variation between experiment replicates, FLAT(EXPERIMENT) for variation between flats within an experiment. Similarly, least squares means were obtained for plant diameter in *P. brassicae* ($N = 973$) and *P. xylostella* ($N = 912$) herbivory screens with the model TRAIT = CONSTANT+ACCESSION+EXPERIMENT+FLAT(EXPERIMENT)+COLUMN+ROW. Myrosinase activity ($N = 68$), glucosinolates ($N = 83$) and leaf trichome density ($N = 153$) were evaluated in only one experiment each and all accessions were grown completely randomized within one flat. Therefore, the model was TRAIT = CONSTANT+ACCESSION+COLUMN+ROW to obtain least squares means for myrosinase activity. In this model, COLUMN and ROW are variables used to control for the time delay that occurs during the processing of individual microtiter plate positions in our plate reader. The model for the analysis of all individual glucosinolates was TRAIT = CONSTANT+ACCESSION. Glucosinolate sum variables were generated by summing up least squares means for individual glucosinolates after correcting for different UV absorption capacities with published response factors [52,53]. Finally, least squares means for trichome density on the leaf upper and under-sides and the leaf perimeter were obtained with the model TRAIT = CONSTANT+ACCESSION+LEAF, with LEAF being a variable to control for a potential effect of leaf developmental stage on trichome density. For the Da(1)-12×Ei-2 RIL population, similar ANOVA models were used as described above, with a few exceptions: A variable FLAT was included in the glucosinolate and trichome models to account for variation between flats, and the variables EXPERIMENT and FLAT(EXPERIMENT) were included to control for variation between experimental replicates and between flats within an experiment replicate in the myrosinase assays. Sample sizes were $N = 9132$ for *P. brassicae* and $N = 2441$ for *P. xylostella* herbivory screens, $N = 930$ for myrosinase activity assays, $N = 1484$ for trichome density, and $N = 972$ for glucosinolates. Again, data from *P. brassicae* herbivory screens were also used to analyze variation in plant diameter.

QTL Mapping and Analysis

Windows QTL cartographer V2.5 [55] was used for composite interval mapping (CIM) of QTL. The standard model (Model 6) was used with forward regression, a window size of 10 cM, and 5 background control markers. QTL were scanned at a walk speed of 0.5 cM. Statistical significance of QTL for each trait was assessed by permuting each data set 1000 times, with a significance level of 0.05. Furthermore, 2-LOD support intervals [56] were obtained from the QTL cartographer output. For each QTL, the effect strength was estimated as the proportional difference, $(LSM_{Da(1)-12} - LSM_{Ei-2})/((LSM_{Da(1)-12} + LSM_{Ei-2})/2)$, where LSM is the ANOVA least squares mean at the marker closest to the QTL peak. Positive values indicate that the Da(1)-12 genotype has a stronger effect, negative values that Ei-2 has a stronger effect. Finally, the proportion of explained variance, R^2, was obtained from the QTL cartographer output.

In addition, Bayesian interval mapping (BIM) [57] as implemented in R/bim (http://www.stat.wisc.edu/~yandell/qtl/software/bmqtl) was used. For each trait, 400,000 Markov-Chain-Monte-Carlo steps were simulated and iterations were recorded at every 400^{th} step, with 1000 pre-burn-in and 20000 burn-in steps. Prior for the number of QTL was Poisson, with zero initial QTL.

CIM and BIM yielded very similar results, and the high density probability peak from BIM for a given QTL was usually found within the 2-LOD support interval for a QTL in CIM (**Figure 3**). The most notable difference between both methods was a weak QTL for *Plutella* herbivory that was identified with BIM but not with CIM. Since this QTL was located at the same position as one of the QTL for *Pieris* herbivory, we consider this QTL real (**Figure 3**).

Because epistatic interactions appear to be an important factor in the genetic architecture of complex traits [58] and have been documented for glucosinolate biosynthesis [12], the markers most closely linked to QTL peaks were tested for potential epistatic interactions. Based on this *a priori* expectation that main QTL might interact with one another, a significance threshold of 0.05 was chosen. First, for every trait all single markers and all pairwise interactions between these markers were included in the ANOVA models. If more than one significant interaction with a particular marker was detected, also higher-order interaction terms between markers were included.

SUPPORTING INFORMATION

Table S1 Genotype data for Da(1)-12×Ei-2 RILs
Found at: doi:10.1371/journal.pone.0000578.s001 (0.26 MB XLS)

Table S2 Genes of the glucosinolate-myrosinase system, AGI numbers, and references
Found at: doi:10.1371/journal.pone.0000578.s002 (0.06 MB DOC)

ACKNOWLEDGMENTS

The cross between Da(1)-12 and Ei-2 parental lines was carried out by Daniel J. Kliebenstein. We thank Andreas Weber and colleagues for help with RIL development, and Petra Eißmann, Andrea Hirsch, Tabea Birk, and Kathrin Eberhardt for technical assistance.

Author Contributions

Conceived and designed the experiments: JK HV MP. Performed the experiments: MP. Analyzed the data: JK MP. Contributed reagents/materials/analysis tools: TM. Wrote the paper: JK MP.

REFERENCES

1. Koornneef M, Alonso-Blanco C, Vreugdenhill D (2004) Naturally occurring genetic variation in *Arabidopsis thaliana*. Annu Rev Plant Biol 55: 141–172.
2. Lister C, Dean C (1993) Recombinant inbred lines for mapping RFLP and phenotypic markers. Plant J 4: 745–750.
3. Alonso-Blanco C, Peeters AJM, Koornneef M, Lister C, Dean C, et al. (1998) Development of an AFLP based linkage map of L*er*, Col and Cvi *Arabidopsis thaliana* ecotypes and construction of a L*er*/Cvi recombinant inbred line population. Plant J 14: 259–271.
4. Jander G, Cui J, Nhan B, Pierce NE, Ausubel FM (2001) The TASTY locus on chromosome 1 of Arabidopsis affects feeding of the insect herbivore *Trichoplusia ni*. Plant Physiol 126: 890–898.
5. Lambrix V, Reichelt M, Mitchell-Olds T, Kliebenstein DJ, Gershenzon J (2001) The Arabidopsis epithiospecifier protein promotes the hydrolysis of glucosinolates to nitriles and influences *Trichoplusia ni* herbivory. Plant Cell 13: 2793–2807.
6. Kliebenstein DJ, Pedersen D, Barker B, Mitchell-Olds T (2002) Comparative analysis of quantitative trait loci controlling glucosinolates, myrosinase and insect resistance in *Arabidopsis thaliana*. Genetics 161: 325–332.
7. Kroymann J, Donnerhacke S, Schnabelrauch D, Mitchell-Olds T (2003) Evolutionary dynamics of an Arabidopsis insect resistance quantitative trait locus. Proc Natl Acad Sci USA 100: 14587–14592.

8. Zhang Z, Ober JA, Kliebenstein DJ (2006) The gene controlling the quantitative trait locus *EPITHIOSPECIFIER MODIFIER1* alters glucosinolate hydrolysis and insect resistance in *Arabidopsis*. Plant Cell 18: 1524–1536.
9. Kliebenstein DJ, Kroymann J, Mitchell-Olds T (2005) The glucosinolate-myrosinase system in an ecological and evolutionary context. Curr Opin Plant Biol 8: 264–271.
10. Grubb CD, Abel S (2006) Glucosinolate metabolism and its control. Trends Plant Sci 11: 89–100.
11. Halkier BA, Gershenzon J (2006) Biology and biochemistry of glucosinolates. Annu Rev Plant Biol 57: 303–333.
12. 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.
13. Mithen R, Clarke J, Lister C, Dean C (1995) Genetics of aliphatic glucosinolates: III. Side chain structure of aliphatic glucosinolates in *Arabidopsis thaliana*. Heredity 74: 210–215.
14. Kliebenstein DJ, Lambrix VM, Reichelt M, Gershenzon J, Mitchell-Olds T (2001) Gene duplication and the diversification of secondary metabolism: side chain modification of glucosinolates in *Arabidopsis thaliana*. Plant Cell 13: 681–693.
15. Magrath R, Bano F, Morgner M, Parkin I, Sharpe A, et al. (1994) Genetics of aliphatic glucosinolates: I. Side chain elongation in *Brassica napus* and *Arabidopsis thaliana*. Heredity 72: 290–299.
16. Campos de Quiros H, Magrath R, McCallum D, Kroymann J, Schnabelrauch D, et al. (2000) α-Keto acid elongation and glucosinolate biosynthesis in *Arabidopsis thaliana*. Theor Appl Genet 101: 429–437.
17. Kroymann J, Textor S, Tokuhisa JG, Falk KL, Bartram S, et al. (2001) A gene controlling variation in *Arabidopsis thaliana* glucosinolate composition is part of the methionine chain elongation pathway. Plant Physiol 127: 1077–1088.
18. Kliebenstein DJ, Kroymann J, Brown P, Figuth A, Pedersen D, et al. (2001) Genetic control of natural variation in Arabidopsis glucosinolate accumulation. Plant Physiol 126: 811–825.
19. Koroleva OA, Davies A, Deeken R, Thorpe MR, Tomos AD, et al. (2000) Identification of a new glucosinolate-rich cell type in Arabidopsis flower stalk. Plant Physiol 124: 599–608.
20. Husebye H, Chadchawan S, Winge P, Thangstad OP, Bones AM (2002) Guard cell- and phloem idioblast-specific expression of thioglucoside glucohydrolase 1 (myrosinase) in Arabidopsis. Plant Physiol 128: 1180–1188.
21. Thangstad OP, Gilde B, Chadchawan S, Seem M, Husebye H, et al. (2004) Cell specific cross-species expression of myrosinases in *Brassica napus*, *Arabidopsis thaliana* and *Nicotiana tabacum*. Plant Mol Biol 54: 597–611.
22. Barth C, Jander G (2006) Arabidopsis myrosinases TGG1 and TGG2 have redundant function in glucosinolate breakdown and insect defense. Plant J 46: 549–562.
23. Wittstock U, Halkier BA (2002) Glucosinolate research in the *Arabidopsis* era. Trends Plant Sci 7: 263–270.
24. Raybold AF, Moyes CL (2001) The ecological genetics of aliphatic glucosinolates. Heredity 87: 383–391.
25. Ratzka A, Vogel H, Kliebenstein DJ, Mitchell-Olds T, Kroymann J (2002) Disarming the mustard oil bomb. Proc Natl Acad Sci USA 99: 11223–11228.
26. Wittstock U, Agerbirk N, Stauber EJ, Olsen CE, Hippler M, et al. (2004) Successful hervivore attack due to metabolic diversion of a plant chemical defense. Proc Natl Acad Sci USA 101: 4859–4864.
27. Benderoth M, Textor S, Windsor AJ, Mitchell-Olds T, Gershenzon J, et al. (2006) Positive selection driving diversification in plant secondary metabolism. Proc Natl Acad Sci USA 103: 9118–9123.
28. Tuinstra MR, Ejeta G, Goldsbrough PB (1997) Heterogeneous inbred family (HIF) analysis: a method for developing near-isogenic lines that differ at quantitative trait loci. Theor Appl Genet 95: 1005–1011.
29. Mauricio R, Rausher MD (1997) Experimental manipulation of putative selective agents provides evidence for the role of natural enemies in the evolution of plant defense. Evolution 51: 1435–1444.
30. Mauricio R (1998) Costs of resistance to natural enemies in field populations of the annual plant, *Arabidopsis thaliana*. Am Nat 151: 20–28.
31. Handley R, Ekbom B, Ågren J (2005) Variation in trichome density and resistance against a specialist insect herbivore in natural populations of *Arabidopsis thaliana*. Ecol Entomol 30: 284–292.
32. Symonds VV, Godoy AV, Alconada T, Botto JF, Juenger TE, et al. (2005) Mapping of quantitative trait loci in multiple populations of *Arabidopsis thaliana* identifies natural allelic variation for trichome density. Genetics 169: 1649–1658.
33. Mauricio R (2005) Ontogenetics of QTL: the genetic architecture of trichome density over time in *Arabidopsis thaliana*. Genetica 123: 75–85.
34. Mitchell-Olds T, Pedersen D (1998) The molecular basis of quantitative genetic variation in central and secondary metabolism in Arabidopsis. Genetics 149: 739–747.
35. Chadchawan S, Bishop J, Thangstad OP, Bones AM, Mitchell-Olds T, et al. (1993) Arabidopsis cDNA sequence encoding myrosinase. Plant Physiol 103: 671–672.
36. Xue J, Jorgensen M, Pihlgren U, Rask L (1995) The myrosinase gene family in *Arabidopsis thaliana*: gene organization, expression and evolution. Plant Mol Biol 27: 911–922.
37. Zhang J, Pontoppidan B, Xue J, Rask L, Meijer J (2002) The third myrosinase gene *TGG3* in *Arabidopsis thaliana* is a pseudogene specifically expressed in stamen and petal. Physiol Plant 115: 25–34.
38. Field BG, Cardon G, Traka M, Botterman J, Vancanneyt G, et al. (2004) Glucosinolate and amino acid biosynthesis in Arabidopsis. Plant Physiol 235: 828–839.
39. Bonnemail L (1965) Insect pests of crucifers and their control. Annu Rev Entomol 10: 233–258.
40. Samson PR, Geier PW (1983) Induction of crop damage by the cabbage white butterfly *Pieris rapae* (Lepidoptera, Pieridae) on cabbage. Protect Ecol 5: 199–233.
41. Hill DS (1987) Agricultural insect pests of temperate regions and their control Cambridge University Press, Cambridge. pp 411–413.
42. van Loon JJA, Blaakmeer A, Griepink FC, van Beek TA, Schoonhoven LM, et al. (1992) Leaf surface compound from *Brassica oleracea* (Cruciferae) induces oviposition by *Pieris brassicae* (Lepidoptera: Pieridae). Chemoecology 3: 39–44.
43. Huang XP, Renwick JAA (1994) Relative activities of glucosinolates as oviposition stimulants for *Pieris rapae* and *Pieris napi oleracea*. J Chem Ecol 20: 1025–1037.
44. Städler E, Renwick JAA, Radke CD, Sachdevgupta K (1995) Tarsal contact chemoreceptor response to glucosinolates and cardenolides mediating oviposition in *Pieris rapae*. Physiol Entomol 20: 175–187.
45. Miles CI, del Campo ML, Renwick JAA (2005) Behavioral and chemosensory responses to a host recognition cue by larvae of *Pieris rapae*. J Comp Physiol A – Neuroethol Sens Neur Behav Physiol 191: 147–155.
46. Schoonhoven LM (1969) Gustation and foodplant selection in some Lepidopterous larvae. Entomol Exp Appl 12: 555–561.
47. Renwick JAA, Lopez K (1999) Experience-based food consumption by larvae of *Pieris rapae*: addiction to glucosinolates? Entomol Exp Appl 91: 51–58.
48. Renwick JAA (2002) The chemical world of crucivores: lures, treats and traps. Entomol Exp Appl 104: 35–42.
49. Shelton AM, Cooley RJ, Kroening MK, Wilsey WT, Eigenbrode SD (1991) Comparative analysis of two rearing procedures for diamond-back moth *Lepidoptera plutellidae*. J Entomol Sci 26: 17–26.
50. Loudet OS, Chaillou S, Camilleri C, Bouchez D, Daniel-Vedele F (2002) Bay-0×Shahdara recombinant inbred line population: a powerful tool for the genetic dissection of complex traits in Arabidopsis. Theor Appl Genet 104: 1173–1184.
51. Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408: 796–815.
52. Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, et al. (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1: 174–181.
53. Buchner R (1987) Approach to determination of HPLC response factors for glucosinolates. In JP Wathelet, ed. Glucosinolates in Rapeseeds: Analytical Aspects. Martinus Nijhoff Publishers, Boston. pp 50–58.
54. 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.
55. Wang S, Basten CJ, Zeng ZB (2001) Windows QTL cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC, (http://statgen.ncsu.edu.qtlcart/WQTLCart.htm).
56. van Ooijen JW (1999) LOD significance thresholds for QTL analysis in experimental populations of diploid species. Heredity 83: 613–624.
57. Satagopan JM, Yandell BS, Newton MA, Osborn TC (1996) Markov chain Monte Carlo approach to detect polygene loci for complex traits. Genetics 144: 805–816.
58. Kroymann J, Mitchell-Olds T (2005) Epistasis and balanced polymorphism influencing complex trait variation. Nature 435: 95–98.

Physiologia Plantarum 132: 136–149. 2008

Teaching (and learning from) metabolomics: The 2006 PlantMetaNet ETNA Metabolomics Research School

Christoph Böttcher[a], Danilo Centeno[b], Jens Freitag[b,c], Rainer Höfgen[b], Karin Köhl[b], Joachim Kopka[b], Juergen Kroymann[d], Andrea Matros[c], Hans-Peter Mock[c], Stefan Neumann[a], Marina Pfalz[d], Edda von Roepenack-Lahaye[a], Nicolas Schauer[b], Sandra Trenkamp[b], Matias Zurbriggen[c,1] and Alisdair R. Fernie[b,*]

[a]Leibniz Institute of Plant Biochemistry, Weinberg 3, 06120 Halle/Saale, Germany
[b]Max-Planck Institute for Molecular Plant Physiology, Am Mühlenberg, 14476 Potsdam-Golm, Germany
[c]Leibniz Institut fur Pflanzengenetik und Kulturpflanzenforschung, Correnstrasse 3, 06466 Gatersleben, Germany
[d]Max-Planck Institute for Chemical Ecology, Carl-Zeiss-Promenade 10, 07745 Jena, Germany

Correspondence
*Corresponding author,
e-mail: fernie@mpimp-golm.mpg.de

Received 6 July 2007; revised 25 September 2007

doi: 10.1111/j.1399-3054.2007.00990.x

Under the auspices of the European Training and Networking Activity programme of the European Union, a 'Metabolic Profiling and Data Analysis' Plant Genomics and Bioinformatics Summer School was hosted in Potsdam, Germany between 20 and 29 September 2006. Sixteen early career researchers were invited from the European Union partner nations and the so-called developing nations (Appendix). Lectures from invited leading European researchers provided an overview of the state of the art of these fields and seeded discussion regarding major challenges for their future advancement. Hands-on experience was provided by an example experiment – that of defining the metabolic response of Arabidopsis to treatment of a commercial herbicide of defined mode of action. This experiment was performed throughout the duration of the course in order to teach the concepts underlying extraction and machine handling as well as to provide a rich data set with which the required computation and statistical skills could be illustrated. Here we review the state of the field by describing both key lectures given at and practical aspects taught at the summer school. In addition, we disclose results that were obtained using the four distinct technical platforms at the different participating institutes. While the effects of the chosen herbicide are well documented, this study looks at a broader number of metabolites than in previous investigations. This allowed, on the one hand, not only to characterise further effects of the herbicide than previously observed but also to detect molecules other than the herbicide that were obviously present in the commercial formulation. These data and the workshop in general are all discussed in the context of the teaching of metabolomics.

Abbreviations – ESI-MS, electron spray ionisation mass spectrometry; ETNA, European Training and Networking Activity; GABI, German Plant Genomics Research Program; GC-TOF, gas chromatography–time of flight; IPB, the Leibniz Institute for Plant Biochemistry; IPK, Leibniz Institute of Plant Genetics and Crop Plant Research; LC, liquid chromatograph; MPICE, Max-Planck Institute for Chemical Ecology; MPIMP, Max-Planck Institute for Molecular Plant Physiology; PlantMetaNet, Plant Metabolism Network; QTL, quantitative trait loci; UP, University of Potsdam.

[1]Present address: Instituto de Biología Molecular y Celular de Rosario, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Suipacha 531, S2002-LRK Rosario, Argentina

Introduction

Metabolite profiling or metabolomics has come a long way since its inception in the 1980s (Fernie et al. 2004, Kopka et al. 2004, Oksman-Caldentry and Saito 2005). In its earliest applications, metabolite profiling used metabolite composition as a diagnostic tool to ascertain, e.g. the metabolic response to herbicide (Ott et al. 2003, Sauter et al. 1988), the equivalence of genetically modified and conventional crops (Catchpole et al. 2005, Shepherd et al. 2006) and the classification of plant genotypes (e.g. Fiehn et al. 2000, Roessner et al. 2001, Schauer and Fernie 2006). Early descriptions of plant metabolomics tended to describe primary metabolites (Fiehn et al. 2000, Katona et al. 1999, Roessner et al. 2000, Sauter et al. 1988); however, broad methods for plant volatiles using headspace GC-MS (Song et al. 1997, Tikunov et al. 2005) and secondary metabolites using diverse variations of liquid chromatography (LC)-MS have recently been developed (Huhman and Sumner 2002, Keurentjes et al. 2006, von Roepenack-Lahaye et al. 2004, Tolstikov and Fiehn 2002).

In parallel to the much needed method development, the metabolomics approach has been much used in descriptions of the response of plants to a wide range of biotic or abiotic stresses (see, e.g. Baxter et al. 2007, Hirai et al. 2004, Kaplan et al. 2004, Nikiforova et al. 2005, Suzuki et al. 2005). In addition, metabolite profiling is also being used increasingly to decipher gene function (Fridman et al. (2005), Goossens et al. 2003, Hirai et al. 2005, Tieman et al. 2006, Tohge et al. 2005), to investigate metabolic regulation (Matsuda et al. 2005, Roessner et al. 2001) and (in combination with analysis of other molecular entities of the cell) as part of integrative analyses of the systemic response to environmental or genetic perturbations (see Sweetlove and Fernie 2005).

In addition to fundamental developments in the use of metabolite profiling to address biological questions, the cost of the analytical machinery required to practise this discipline has fallen dramatically in the last decade, rendering it a far more accessible technology. With this in mind, the Plant Metabolism Network (PlantMetaNet) consortium of plant research institutes in the eastern part of Germany applied to host a European Training and Networking Activity (ETNA) summer school in metabolic profiling in 2006. In this mini-review, we will give an overview of highlights of the presentations given by individuals of the host institutes, invited speakers and attendees of the conference (many of the talks can be downloaded at http://www.eu-summer-school.org/01_seiten/b_00.php). In addition, we will review the central experiment that was performed as an introduction to the methods of metabolite profiling, following the metabolic response of Arabidopsis to glyphosate treatment.

Workshop structure

As indicated above, the summer school took the form of a 10-day course held mainly at the Max-Planck Institute for Molecular Plant Physiology (MPIMP) in Potsdam-Golm; however, it was also jointly hosted by the University of Potsdam (UP), the Leibniz Institute for Plant Biochemistry (IPB) in Halle, the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) in Gatersleben and the Max-Planck Institute for Chemical Ecology (MPICE) in Jena. It comprised a mixture of lectures, hands-on wet laboratory work and computational analysis. The wet laboratory work consisted of a kinetic analysis of the metabolic response of *Arabidopsis thaliana* to treatment with the herbicide glyphosate (using the commercial formulation Roundup Ultra®, Monsanto Co., St Louis, MO). For this purpose, a large number of Arabidopsis plants from the Col-0 accession were grown for 33 days after sowing under highly controlled conditions (20/18°C, 16:8 h light–dark regime and 80/50% RH) prior to treatment with the herbicide at both low (2 ml l^{-1} Roundup Ultra) and high (20 ml l^{-1} Roundup Ultra) concentration. Samples were harvested at five different times post-treatment. The material was then snap-frozen in liquid nitrogen and subsequently distributed to the appropriate laboratories on dry ice for the various metabolite-profiling studies, as illustrated in Fig. 1. This experiment was performed twice – once 6 weeks before the summer school to ensure that the metabolic profiles were ready for analysis by the participants of the course and the second time during the course itself. Analyses and experimental work were, however, very much incorporated, with lectures to provide a comprehensive overview of this rapidly changing research field. Because of space limitations, we will only describe a few of the talks in detail – essentially focusing on those that highlighted recent advances and current limitations of metabolomics and bioinformatic analysis intimately associated with it.

The talks

The conference was opened with two lectures from Lothar Willmitzer (MPIMP, Potsdam-Golm). The first concerned the history, structure and remit of PlantMetaNet, a conglomerate of research institutes in the eastern part of Germany focused on plant metabolism that was set up in 2001 between IPK-Gatersleben, IPB-Halle and the MPIMP-Potsdam-Golm and the MPICE-Jena. The second one gave an overview of the concept of metabolite profiling (Fernie et al. 2004, Fiehn 2002), the gas chromatography–time of flight (GC-TOF) protocol being used in several of the laboratories in Golm (Lisec et al.

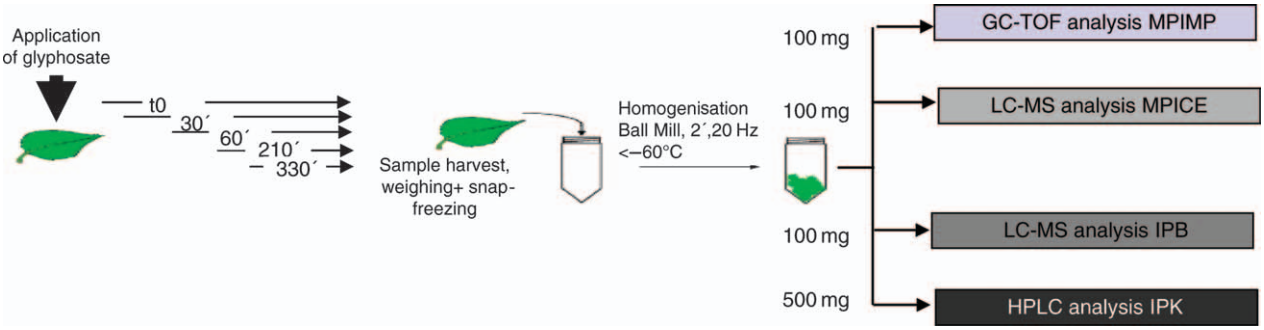

Fig. 1. The set up and experimental design of harvest, analytics and evaluation. Detailed bioinformatic evaluations will largely be covered in the accompanying paper (Steinfath et al. 2008).

2006) and several recent applications utilising tomato introgression lines and recombinant inbred lines of Arabidopsis (Meyer et al. 2007, Schauer et al. 2006). The programme continued with a detailed presentation of basic aspects of the application of MS in biology by Wolfram Weckwerth (MPIMP, Potsdam-Golm). In his lecture, Wolfram gave a broad overview of chromatography, ionisation and mass analysis as well as detailed descriptions of the various tandem MS methods in current use (Glinski and Weckwerth 2005). The next metabolomics-focused lecture was given by Maciej Stobiecki (Institute of Bioorganic Chemistry of Polish Academy of Science, Poznan, Poland), which began with an overview of the potential methods that can be used for profiling secondary metabolites (Sumner et al. 2003), before focusing on the profiling of phenylpropanoids and flavonoids. In particular, genistein conjugates and lupin isoflavones were used as case studies for discussion of chromatographic and MS behaviour, and a detailed literature survey was provided (key references included Bloor and Abrahams 2002, Hemm et al. 2004, Jones et al. 2003, Mehrtens et al. 2005, Stobiecki et al. 2006). This first day's lecture was followed by an entire day of practical demonstration (detailed in Hands-on sessions).

The third day started with database talks, talks on metabolic networks and on the combination of metabolomics with classical genetics. The lecture of Uwe Scholz (IPK, Gatersleben) gave a thought-provoking introduction to the power of correctly constructed databases as well as highlighting the immense diversity of databases currently available (Galperin 2006, Philippi and Kohler 2006). Scholz then went on to outline the added utility (and complexity) in integrating databases into data warehouses using work of his own group to illustrate this (Balko et al. 2004, Kunne et al. 2005), before finishing his lecture with the sobering statement that automated knowledge retrieval systems for biology remain elusive as yet. In his discussion of network modelling, Ralf Steuer (UP) gave a history of metabolic modelling involving kinetic, topological and stoichiometrical approaches. He illustrated this talk with work of his own on the TCA cycle (Steuer et al. 2007) and emphasised the importance of the analysis of stability, dynamics and robustness of biological systems (Steuer 2007). In the final talk of the session, Nigel Hardy (University of Wales, Aberystwyth) presented the current state of the art of standards for metabolomics and an insightful personal viewpoint on how these could (and should) be improved. Essentially, Hardy reported the now much held viewpoint that reporting standards for metabolomics need to be improved and also outlined steps that had been taken in this direction. His discussion included the proposals of MiaMet (Bino et al. 2004) and ArMet (Jenkins et al. 2005) for work in plants and specifically in Arabidopsis, respectively, as well as the systems such as standard metabolic reporting structure (SMRS) set up for clinical medicine (Lindon 2005, Lindon et al. 2005). The talk also introduced the then newly formed metabolomics standards initiative (Fiehn et al. 2006) and discussed two new tools, HiMet (Jenkins et al. 2005) and SetUpX (Fiehn et al. 2005), to facilitate database storage of metabolomics-derived information. An important conclusion of his presentation was the likely need to incorporate laboratory management system approaches into metabolomics. The theoretical talks of the morning were followed by detailed presentations of the PhD work of Jan Lisec and Nicolas Schauer (both MPIMP, Potsdam-Golm). The talk of Jan focused on the use of GC-TOF and novel annotation algorithms (Lisec et al. 2006) for the characterisation of recombinant inbred lines and introgression lines of Arabidopsis (J. Lisec, R. C. Meyer, M. Steinfath, M. Becher, H. Witucka-Wall, O. Fiehn, O. Törjék, J. Selbig, T. Altmann, L. Willmitzer [MPIMP and Universität Potsdam, Potsdam-Golm], submitted, Meyer et al. 2007), revealing a large number of QTL and several metabolic hotspots in the genome.

Nicolas's talk by contrast focused on the application of the same method to a tomato introgression population and discussed the prospects of using knowledge of the chemical composition of fruits of wide natural diversity (Schauer et al. 2005a) in the design of rational breeding strategies. The presentation also focussed on the linking of chemical composition data with morphological phenotypes, which corroborated with earlier findings (Fridman et al. 2004, 2005) in suggesting that the efficiency of assimilate partitioning to the fruit plays a major role in determining its final chemical composition (Schauer et al. 2006).

The fourth day began with a talk on carbon balances during diurnal cycles (Bläsing et al. 2005, Gibon et al. 2004, Osuna et al. 2007) and in response to nutrient supply (Morcuende et al. 2007) and global climate change (Ainsworth et al. 2007, Rogers et al. 2006). This talk was followed by an introduction to information theory given by Ivo Grosse (IPK, Gatersleben). This lecture gave a trajectory of the key developments in information theory and clear definitions and examples illustrating the importance of entropy, conditional entropy and mutual information. Heiko Schoof (Max-Planck-Institut für Zuchtungforschung [MPIZ], Köln) then presented tools for database interaction. His talk centred around the BioMoby initiative (Schoof 2003, Wilkinson et al. 2005) that aims to provide a simple, extensible platform through which the myriad of online biological databases and analytical tools can offer their information and analytical services in a fully automated and interoperable way. Schoof presented examples from his own work that illustrated the utility of this tool by giving examples of the aggregation of data on, e.g. a given gene (on the basis of its assigned AtG number) or cross species comparisons across genomes and illustrating the advantage of being able to utilise databases without having to store them locally on the researchers computer. The day concluded with an overview of tools for higher level statistical analysis given by Joachim Selbig (UP and MPIMP, Potsdam-Golm), the content of which is reviewed in another article in this issue (Steinfath et al. 2008).

Another perspective of network analysis was presented by Dirk Walter (MPIMP) on the fifth day as was an overview of the forces that drive RNA folding by Sebastian Doniach (Stanford University, USA; Das et al. 2003). In his lecture, Walter gave a complementary talk to that of Ralf Steuer, beginning by defining graph theory, connectivities and centralities and random vs scale-free networks (Barabasi and Oltvai 2004, Junker et al. 2006), before detailing their application to the analysis of biological systems. Walter ended his talk by presenting a list of hot topics and future challenges that included obtaining a greater understanding of the evolution of genetic networks (Walther et al. 2007) and their response to stress (Luscombe et al. 2004). The rest of this day was spent with the teaching of computational aspects of metabolite profiling (detailed in Hands-on sessions).

The next day consisted of four lectures and further computer-based tutorials. It was kicked off with an overview of structural elucidation of secondary metabolites, which was presented by Ales Svatos (MPICE, Jena). In his talk, Svatos gave several nice examples of compound identification including indolics and phenylpropanoids of Arabidopsis and camalexin as well as identification of hydrocarbons in the sheep tick (Saman et al. 2006, Schuhegger et al. 2006, Tan et al. 2004). The analytical approaches he discussed included difference fragmentation by the application of a range of ionisation sources and single-ion monitoring and tandem MS approaches. Complementary approaches such as gas-phase Fourier transform infrared and NMR were also discussed with respect to lipid analysis. The second lecture was more applications based, being concerned with the use of metabolomics in food quality and safety assessments. It was given by Paul Fraser (Royal Holloway, UK) and ranged from definitions of the compounds associated with quality traits in foods and the technique of choice by which to evaluate them, through detailed descriptions of carotenoids and flavonoids, including description of a novel matrix-assisted laser desorption ionisation–time of flight mass spectrometry method for rapid quantification of the former (Fraser et al. 2007), to the concept and practicalities of substantial equivalence testing. In his discussion of substantial equivalence, Fraser used transgenic tomato modified in a number of enzymes of carotenoid metabolism (Romer and Fraser 2005), to illustrate the role that metabolomics could play in evaluation of food safety. Juergen Kroymann (MPICE, Jena) presented an overview of how natural variation can be exploited to integrate metabolomics with genomics. He described how they were able, in their own work, to map genes responsible for qualitative and quantitative differences in glucosinolate biosynthesis among Arabidopsis accessions. Subsequent comparative genomic approaches allowed them to conclude that a locus central to diversification in glucosinolate profiles, *MAM*, evolved dynamically within and between species (Heidel et al. 2006, Kliebenstein et al. 2005, Kroymann et al. 2003, Windsor et al. 2005). The final talk in this session was presented by Edda von Roepenack-Lahaye (IPB, Halle). In her talk, von Roepenack-Lahaye focused on Metabolomic Profiling of secondary metabolites. At the IPB, a LC–electron spray ionisation mass spectrometry (ESI-MS) platform has been developed based on capillary LC and

quadrupole time of fight technology, allowing the combination of accurate mass and high sensitivity. This system enables to cover five of the six secondary metabolite classes in Arabidopsis (glucosinolates, indolics, phenylpropanoids, flavonoids and anthocyanins). von Roepenack-Lahaye highlighted the importance of system and data validation, stressing the point of matrix effects in ESI-MS. A 'proof of concept experiment' was presented using the transparent testa mutants in Arabidopsis with defects in the already well-elucidated flavonoid pathway. The talk concluded with examples of metabolomic analyses in Arabidopsis treated with biotic and abiotic stress.

The sixth day of the workshop started with a talk from Winnie Weschke (IPK, Gatersleben) describing a genomics-based analysis of seed development in barley (Borisjuk et al. 2004, Sreenivasulu et al. 2006, Zhang et al. 2004). It continued with a talk outlining the emerging field of evolutionary genomics presented by Karl Schmidt (MPICE, Jena). In his lecture, Schmidt attempted to address two major questions – 'how does metabolism evolve?' and 'why?'. He began by showing phylogenetic diversity of alkaloid production (see Facchini et al. 2004) and then detailed which genetic mutations produce metabolic diversity (Ober 2005). Particular focus was on the evolution of functionalities following gene duplication, returning to the subject of glucosinolate biosynthesis (Benderoth et al. 2006, Kliebenstein et al. 2001). However, detailed explanations of the analysis of natural vs adaptive selection were also presented using the floral scent gene isoeugenol-*O*-methyltransferase as an example (Barkman 2003). This talk was followed by a talk entitled 'Beyond Metabolomics', given by Steffen Neumann, which first covered the processing steps for LC-MS-based metabolomics and gave an overview of the vendor-independent XML exchange standard mzData for MS peak data. He emphasised the need for open-source MS and metabolomics frameworks. Finally, he suggested that future developments should allow the connection of distributed databases and web services to facilitate integration of different -omics data sources and pathway databases. The session was concluded with an outlook into molecular ecology by Ian Baldwin (MPICE, Jena) in which the central role of understanding performance of organisms in their natural environments to the process of understanding gene function was highlighted (Kessler et al. 2004, Pandey and Baldwin 2007). Following this session, the workshop travelled a short distance to the functional genomics company Metanomics GmbH. The session held there provided the students with a deep insight into the applications of metabolomics at an industrial scale. High-throughput metabolite profiling for genome-wide gene function analysis was introduced and exemplified as a key approach for lead gene discovery, both for compositional and agronomic traits. Future trends were highlighted, indicating the increasing value of metabolomics as a key component of systems biology, in models and crops, for lead gene selection and commercial crop development.

The final day of the workshop contained only a single lecture, with the rest of the time reserved for hands-on experimentation and computation. Fittingly, this lecture was given by Sean May (Nottingham Arabidopsis Stock Centre [NASC], Nottingham, UK) and covered standards and ontologies for Arabidopsis research. May presented a highly detailed survey, starting with the problem of hypervariability in biological replicates as well as covering good laboratory practice, the need for controlled vocabularies and how to search the NASC arrays. He also underlined the take-home message from Nigel Hardy's talk that standards are crucial in all postgenomics research.

Hands-on sessions

As briefly defined above, the experimental set up was to spray 33-day-old Arabidopsis plants with Roundup Ultra (360 g glyphosate l^{-1}) and determine its impact on metabolism. The treatment was started 4 h into the light period, and plants were harvested, washed and snap-frozen in liquid nitrogen, at time points 1, 2, 4, 6 and 8 h post-treatment. Two herbicide concentrations (2 and 20 ml l^{-1}) and untreated control samples were harvested for each time point. This experiment was performed two times, once during the preparation of the course by the technical staff of the MPIMP and once by the participants during the course, allowing verification of the results. The frozen material was homogenised using a ball mill, and appropriate aliquots were distributed to the contributing institutes for measurement on the various analytical platforms used in the course.

Experimental design, sample harvest, extraction and, where necessary, derivatization were practised, and additional injection of the samples onto the GC-MS was demonstrated.

Following the wet laboratory work, a detailed teaching of GC-MS chromatogram evaluation (and the pitfalls inherent in it) was provided as were overviews of LC-MS and traditional HPLC approaches, not only with particular emphasis on peak alignment, peak quantification and biological interpretation, but also including fundamental lesions on computer scripts and an introduction into the R software suite. Finally, an introduction and practical demonstration of the LCQ-Orbitrap was given by Patrick Giavalisco (MPIMP, Potsdam-Golm), which

gave a glimpse of what can be expected to emerge soon as a highly powerful tool in plant metabolomics.

Experimental results

Soil grown Arabidopsis plants were treated with the herbicide formulation, Roundup Ultra, 3 weeks after germination. A scheme of the experimental design and the set up of harvest, analyses and evaluation is presented in Fig. 1. The results obtained from the experiment are presented in Figs 2–4. In targeted analysis by HPLC-photodiode array detector (PDA), the absolute values of 18 amino acids were determined. During the time course of the experiment, significant changes were observed in the levels of almost every metabolite, with the exception of aspartate and serine (Fig. 2). The content of individual amino acids was mostly increased, with the exception of reduced values for phenylalanine during herbicide treatment. In general, the changes in amino acid concentration were gradual and became more pronounced during the time following treatment and with higher amount of herbicide treatment; however, phenylalanine, methionine, arginine and the branched-chain amino acids, leucine, isoleucine and valine, all displayed differences within the first hour of treatment. The decrease in phenylalanine and increase in amino acids of competing pathways fitted the known mode of action of the herbicide (Amrhein et al. 1980, DeMaria et al. 2006) as well as previous descriptions of its consequences on metabolism (Aliferis and Chrysayi-Tokousbalides 2006, Oikawa et al. 2006, Ott et al. 2003). However, in contrast to the large impact on amino acid levels, glyphosate had little influence on glucosinolates. Targeted HPLC profiling revealed that while there were minor changes in the quantities of individual glucosinolates, there were very little differences in the total levels of indole and aliphatic glucosinolates, as well as in the level of total glucosinolates during the experiment (Fig. 2). The application of non-biased approaches of GC-MS and LC-MS profiling and subsequent analysis by independent component analysis (Scholz et al. 2004) revealed that using GC-MS profiling, it was relatively difficult to discriminate (Fig. 3A) treated from untreated samples in contrast to the targeted approach (Fig. 3B). Results of the independent component analysis indicate that the use of LC-MS (Fig. 3C) on first glance appears to give far better discriminatory power relative to the GC-MS approach. Metabolite phenotypes indicate a non-linear dose- and time-dependent response to the herbicide treatment. However, a close analysis of the raw data (see *Supplementary material* Tables S1–S4) reveals that profiling is dominated by herbicide formulation additives, which superimpose classification according to metabolites. Returning to the clustering of the GC-MS data, an interesting point to note is that among the most highly defining components are changes in aromatic amino acids and shikimate (Fig. 3A). This is in accordance with the known mode of action of Roundup, as well as being in accordance with the results of the targeted amino acid analysis. A more detailed comparison of the targeted exact quantifications of individual amino acids using HPLC-PDA and metabolite fingerprinting using GC-MS is shown in Fig. 4. In general, similar results were obtained when comparing average values normalised to control values 1 h after incubation. However, differences in amino acid kinetic profiles were clearly less pronounced when using GC-MS profiling (Fig. 4B) compared with HPLC-based targeted analysis (Fig. 4A). In addition, RSD levels were higher for results coming from GC-MS analysis relative to relative standard deviation (RSD) levels coming from HPLC data, leading to reduced significance values for kinetic profiles from GC-MS analysis. Our findings within the experiments discussed above are in accordance to the results described by Noctor et al. (2007). In this study, a comparison of amino acid quantification based on GC-TOF-MS and HPLC-fluorescence detection was performed for Arabidopsis leaf extracts. As a result, similar levels of biological variation have been observed for both techniques, with significantly smaller RSD values for results obtained by HPLC analysis. In addition, more variable recovery rates for standards were observed when using GC-MS analysis relative to HPLC analysis, and accurate relative quantification gave consistent values for a number of amino acids but differs for some others. Besides this drawback, a clear advantage of the non-targeted approach is the detection of less commonly occurring amino acid residues, e.g. homoserine and pyroglutamate as well as low-abundant amino acids such as methionine.

When taken together, the results of this study also suggest that glucosinolate levels are not overly influenced by the supply of amino acid precursors. This finding is highly interesting, however, not without precedent because there are several other examples in which levels of secondary metabolites do not appear to be regulated by substrate supply (see, e.g. Tieman et al. 2006). The observation that the LC-MS results were drastically effected by using the commercial formulation Roundup (Monsanto) was on one hand unfortunate; on the other hand, it clearly demonstrates the power of metabolomics as a tool in experimental design. Essentially, without the LC-MS analyses, we would not have picked up this basic point. Nevertheless, in practice, glyphosate is applied in formulation, and thus, the changes in 'metabolites' other than those applied are of value in understanding detailed short-term metabolic consequences of the application of this herbicide. As the experiment was designed as

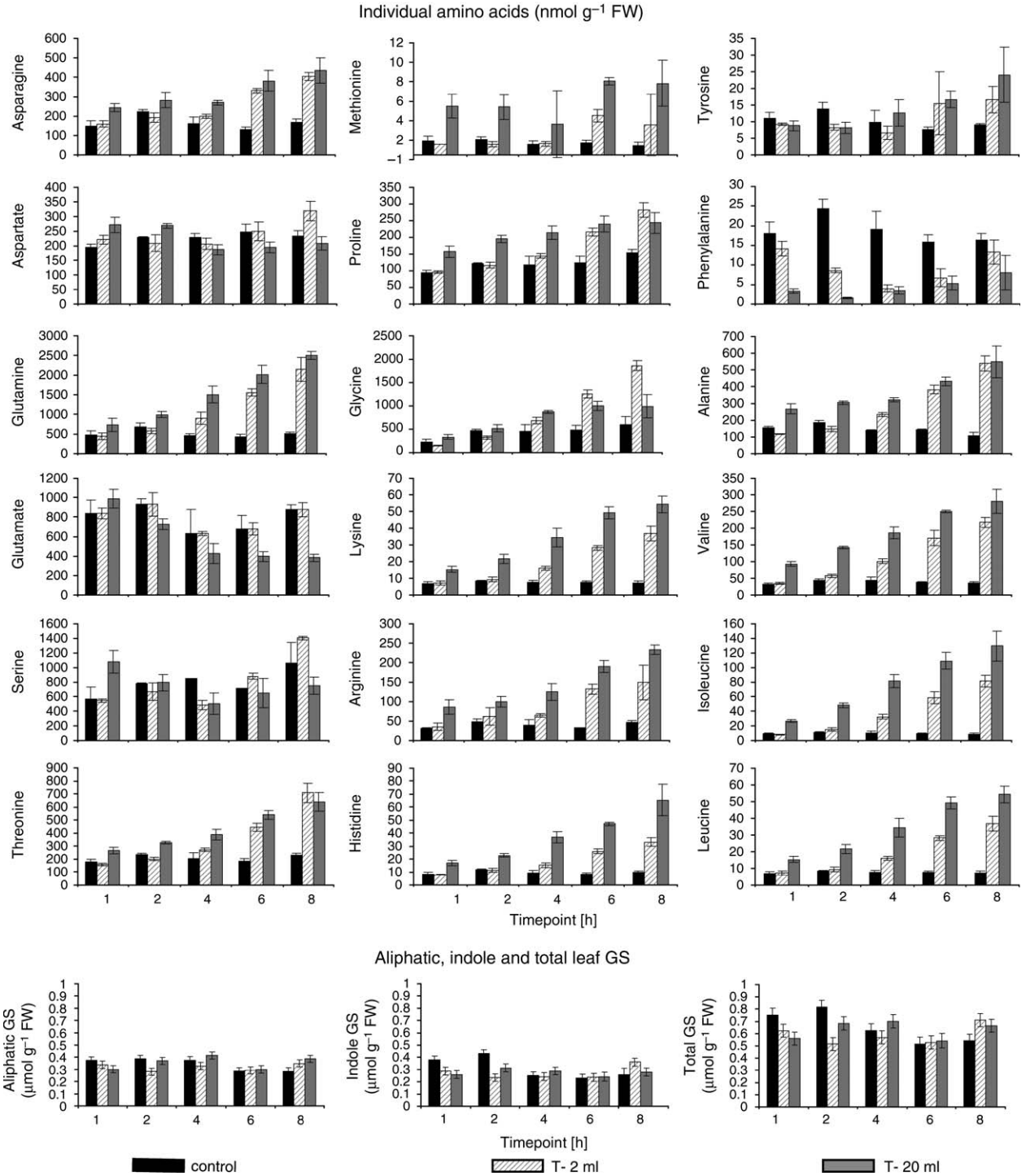

Fig. 2. Kinetic changes of amino acid and selected glucosinolate profiles following spraying with different concentrations of Roundup. Amino acids were determined from ethanol/water extracts by HPLC-PDA, as described in Geigenberger et al. (1996). Values represent means ± RSD of three independent experiments and are given in nmol g^{-1} fresh weight. GS were determined in six biological replicates following HPLC separation, as described in Kroymann et al. (2001). Values represent least squares means ± SE and are given in µmol g^{-1}. GS, glucosinolates; T, treatment.

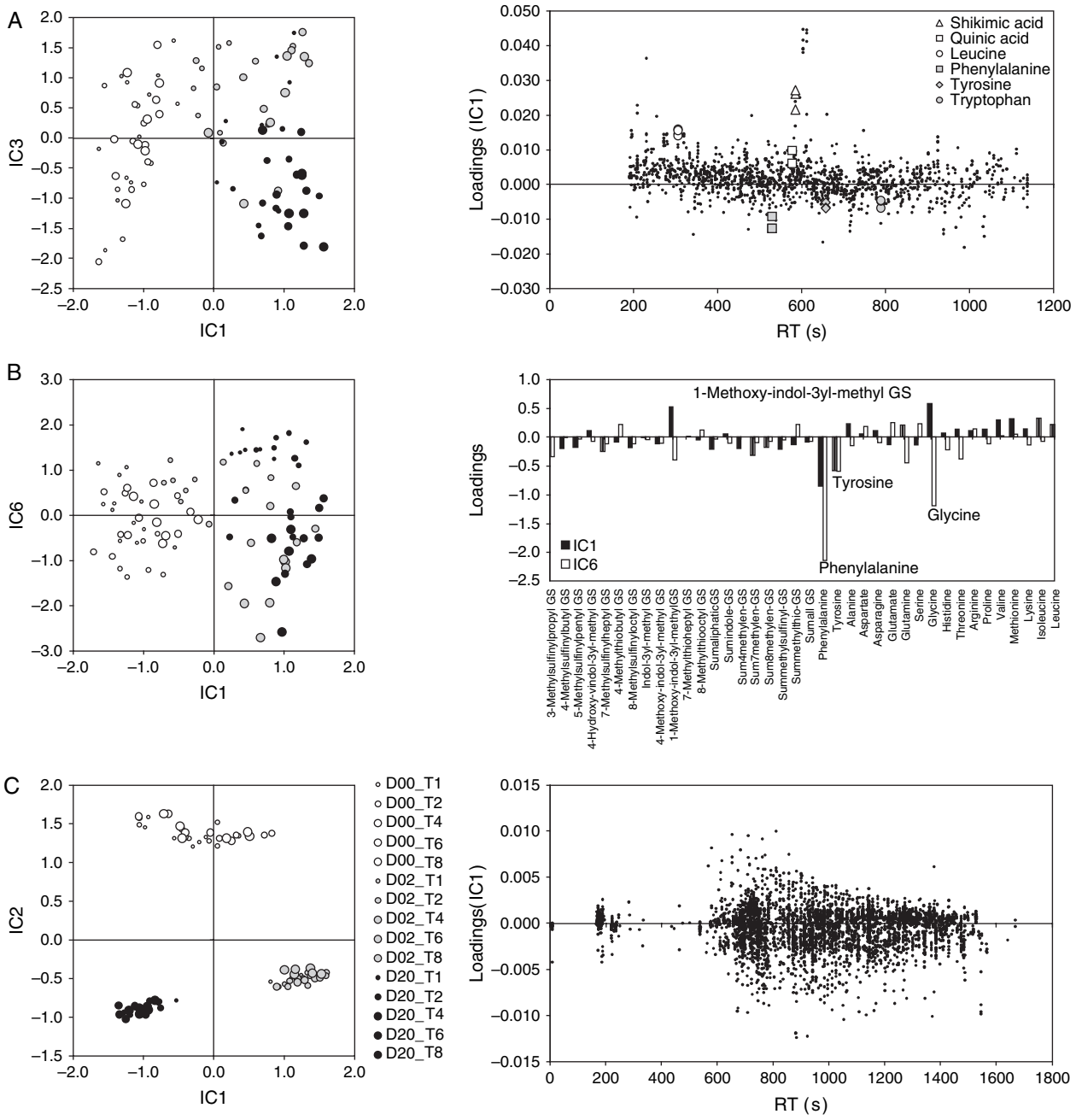

Fig. 3. Independent component analysis of a GC-TOF-MS (A) and a LC-ESI-quadrupole time of fight (QTOF)-MS/MS (C) metabolite profile compared with the set of metabolite-targeted exact quantifications (B). Soil grown Arabidopsis plants were treated with the herbicide formulation, Roundup Ultra, 3 weeks after germination. Sample harvest and targeted metabolite analysis was performed as described in Figs 1 and 2. Metabolite profiling by GC-TOF-MS and LC-ESI-QTOF-MS/MS was performed according to Lisec et al. (2006) and von Roepenack-Lahaye et al. (2004). Independent component analysis (ICA; Scholz et al. 2004) was applied to the metabolite fingerprints, i.e. all mass spectral information available prior to metabolite identification from both GC-MS and LC-MS profiling experiments. Missing value substitution was as described by Scholz et al. (2005). ICA is publicly available through the MetaGenAlyse Web service (Daub et al. 2003; http://metagenealyse.mpimp-golm.mpg.de).

a classroom exercise, as opposed to a scientific study, we will not dwell on the novel results here but rather invite the interested reader to download the raw data from *Supplementary material* (Tables S1–S4).

Concluding remarks

We feel that we can safely state that the 2006 PlantMetaNet ETNA Metabolomics Research School

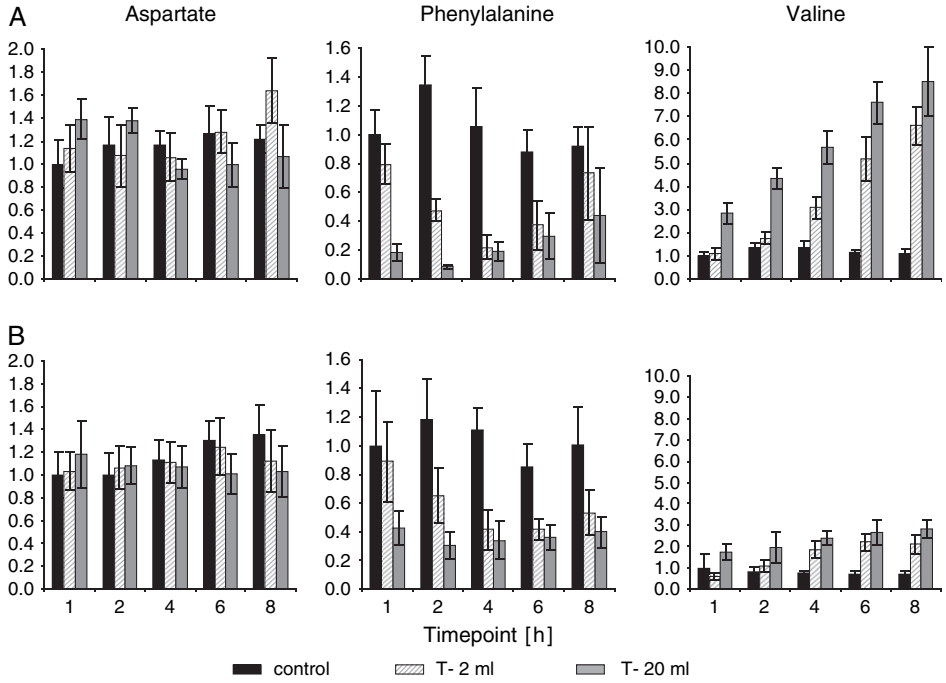

Fig. 4. Comparison of targeted exact quantifications of individual amino acids using HPLC-PDA (A) and metabolite fingerprinting using GC-TOF-MS (B). Kinetic profiles of selected amino acids (aspartate, phenylalanine and valine) following spraying with different concentrations of the herbicide formulation, Roundup are shown. Bars represent average values normalised to control values 1 h after incubation RSD.

was a highly successful gathering of expert researchers and interested and highly motivated early career scientists. Despite current expectations for a metabolite profiling experiment to be performed over a far greater time period, it was possible using a simple experimental design to conduct the experiment in its entirety during the 10 day framework of the course. Indeed, feedbacks from the participants were highly positive, indicating that they felt the course aided them both in understanding aspects of metabolite profiling from plant to data interrogation as well as to make important contacts, both between themselves and within the metabolomics community in general. The forum set up by the participants, http://www.biosolutions.nl/forum/forum.cgi, reflects this, with over 1600 accesses as of 1 October 2007. Moreover, the first joint experiment encompassing all the institutes of the PlantMetaNet was achieved. It is furthermore likely, given the range of instrumentation used in acquiring it, that the hierarchical data set obtained from these studies could be a rich data set for future data-mining studies. Not uncommonly in biological science, the course also reminded us that however simple we believe an experiment to be, there is often still a lot to be learnt.

Acknowledgements – Sponsors, partners, lecturers, participants and supporters became the heart of the summer school ETNA. Many people became involved in the logistics and the organisation of the course, providing material, resources, equipment, know-how, etc. We are not able to thank all of them individually, but we are aware that without them the training and network activity would not be a success. In research as well as in training, the driving force is enthusiasm and the fuel is the financial resource provided. Therefore, we would like to thank the European Commission Research Directorate General (Directorate E – Life Sciences: Biotechnology, Agriculture and Food Research) for the financial support of the summer course on metabolite profiling and bioinformatics. ETNA (http://www.eu-summer-school.org) is financially supported under the European Framework Program (FP6-2004-Food-3-C Proposal No. 31740). Furthermore, we would like to thank the Max-Planck Institute of Molecular Plant Physiology and UP for hosting the summer course and supporting the idea from the beginning. The Plant Meta Net, a network of research institutions dedicated on plant research in the area, became the active partner during the course in 2006 (http://www.plantmetanet.de). The German Plant Genomics Research Program (GABI) was involved in the organisation of ETNA from the beginning. Therefore, we would like to thank all partners of the national research programme as well as the funding agency Bundesministerium für Bildung und Forschung and P. T. Jülich and the partners from industry supporting the GABI programme as the root stock for our European networking. Last, but not least, we would like to thank the 16 participants from Europe and beyond for their enthusiasm and active participation. We also

learnt a lot from our next generation of plant researchers and firmly believe that the knowledge and experiences transferred during the summer school is in the right hands to gain new insights and to answer relevant biological questions.

References

Ainsworth EA, Rogers A, Leakey ADB, Heady LE, Gibon Y, Stitt M, Schurr U (2007) Does elevated atmospheric [CO_2] alter diurnal C uptake and the balance of C and N metabolites in growing and fully expanded soybean leaves? J Exp Bot 58: 579–591

Aliferis KA, Chrysayi-Tokousbalides M (2006) Metabonomic strategy for the investigation of the mode of action of the phytotoxin (5S,8R,13S, 16R)-(-)-pyrenophorol using 1H nuclear magnetic resonance fingerprinting. J Agric Food Chem 54: 1687–1692

Amrhein N, Schab J, Steinrucken HC (1980) The mode of action of glyphosate. Naturwissenschaft 67: 356–357

Balko S, Lange M, Schnee R, Scholz U (2004) BioDataServer: an applied molecular biological data integration service. In: Leser U, Naumann F, Eckman B (eds) Data Integration in the Life Sciences. Proceedings Lecture Notes in Bioinformatics. Springer, Berlin, pp140–155

Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nat Rev Genet 5: 105–115

Barkman TJ (2003) Evidence for positive selection on the floral scent gene isoeugenol-O-methyltransferase. Mol Biol Evol 20: 168–172

Baxter CJ, Redestig H, Schauer N, Repsilber D, Patil KR, Nielsen J, Selbig J, Liu JL, Fernie AR, Sweetlove LJ (2007) The metabolic response of heterotrophic Arabidopsis cells to oxidative stress. Plant Physiol 143: 312–325

Benderoth M, Textor S, Windsor AJ, Mitchell-Olds T, Gershenzon J, Kroymann J (2006) Positive selection driving diversification in plant secondary metabolism. Proc Natl Acad Sci USA 103: 9118–9123

Bino RJ, Hall RD, Fiehn O, Kopka J, Saito K, Draper J, Nikolau BJ, Mendes P, Roessner-Tunali U, Beale MH, Trethewey RN, Lange BM, Wurtele ES, Sumner LW (2004) Potential of metabolomics as a functional genomics tool. Trends Plant Sci 9: 418–425

Bläsing OE, Gibon Y, Gunther M, Hohne M, Morcuende R, Osuna D, Thimm O, Usadel B, Scheible WR, Stitt M (2005) Sugars and circadian regulation make major contributions to the global regulation of diurnal gene expression in Arabidopsis. Plant Cell 17: 3257–3281

Bloor SJ, Abrahams S (2002) The structure of the major anthocyanin in Arabidopsis thaliana. Phytochemistry 59: 343–346

Borisjuk L, Rolletschek H, Radchuk R, Weschke W, Wobus U, Weber H (2004) Seed development and differentiation: a role for metabolic regulation. Plant Biol 6: 375–386

Catchpole GS, Beckmann M, Enot DP, Mondhe M, Zywicki B, Taylor J, Hardy N, Smith A, King RD, Kell DB, Fiehn O, Draper J (2005) Hierarchical metabolomics demonstrates substantial compositional similarity between genetically modified and conventional potato crops. Proc Natl Acad Sci USA 102: 14458–14462

Das R, Kwok LW, Millett IS, Bai Y, Mills TT, Jacob J, Maskel GS, Seifert S, Mochrie SGJ, Thiyagarajan P, Doniach S, Pollack L, Herschlag D (2003) The fastest global events in RNA folding: electrostatic relaxation and tertiary collapse of the tetrahymena ribozyme. J Mol Biol 332: 311–319

Daub CO, Kloska S, Selbig J (2003) MetaGeneAlyse: analysis of integrated transcriptional and metabolite data. Bioinformatics 19: 2332–2333

DeMaria N, Becerril JM, Garcia-Plazaola JI, Hernandez A, De Felipe MR, Fernandez-Pascual M (2006) New insights on glyphosate mode of action in nodular metabolism: role of shikimate accumulation. J Agric Food Chem 54: 2621–2628

Facchini PJ, Bird DA, St-Pierre B (2004) Can Arabidopsis make complex alkaloids? Trends Plant Sci 9: 116–122

Fernie AR, Trethewey RN, Krotzky A, Willmitzer L (2004) Metabolite profiling: from diagnostics to systems biology. Nat Rev Mol Cell Biol 5: 763–769

Fiehn O (2002) Metabolomics – the link between genotypes and phenotypes. Plant Mol Biol 48: 155–171

Fiehn O, Kopka J, Dormann P, Altmann T, Trethewey RN, Willmitzer L (2000) Metabolite profiling for plant functional genomics. Nat Biotechnol 18: 1142–1161

Fiehn O, Wohlgemuth G, Scholz M (2005) Setup and annotation of metabolomic experiments by integrating biological and mass spectrometric metadata. In: Leser U, Naumann F, Eckman B (eds) Data Integration in the Life Sciences. Springer, Berlin, pp 224–239

Fiehn O, Kristal B, Van Ommen BV, Sumner LW, Sansone S-A, Taylor C, Hardy N, Kaddurah-Daouk R (2006) Establishing reporting standards for metabolomic studies: a call for participation. OMICS 10: 158–163

Fraser PD, Enfissi EMA, Goodfellow M, Eguchi T, Bramley PM (2007) Metabolite profiling of plant carotenoids using the matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Plant J 49: 552–564

Fridman E, Carrari F, Liu YS, Fernie AR, Zamir D (2004) Zooming in on a quantitative trout for tomato yield using interspecific introgressions. Science 305: 1786–1789

Fridman E, Wang JH, Iijima Y, Froehlich JE, Gang DR, Ohlrogge J, Pichersky E (2005) Metabolic, genomic, and biochemical analyses of glandular trichomes from the wild tomato species Lycopersicon hirsutum identify a key enzyme in the biosynthesis of methylketones. Plant Cell 17: 1252–1267

Galperin MY (2006) The molecular biology database collection: 2006 update. Nucleic Acids Res 34: D3–D5

Geigenberger P, Lerchl J, Stitt M, Sonnewald U (1996) Phloem-specific expression of pyrophosphatase inhibits long-distance transport of carbohydrates and amino acids in tobacco plants. Plant Cell Environ 19: 43–55

Gibon Y, Bläsing OE, Hannemann J, Carillo P, Höhne M, Hendriks JHM, Palacios N, Cross J, Selbig J, Stitt M (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

Glinski M, Weckwerth W (2005) The role of mass spectrometry in plant systems biology. Mass Spectrom Rev 25: 173–214

Goossens A, Hakkinen ST, Laakso I, Seppanen-Laakso T, Biondi S, De Sutter V, Lammertyn F, Nuutila AM, Soderlund H, Zabeau M, Inze D, Oksman-Caldentey KM (2003) A functional genomics approach toward the understanding of secondary metabolism in plant cells. Proc Natl Acad Sci USA 100: 8595–8600

Heidel AJ, Clauss MJ, Kroymann J, Savolainen O, Mitchell-Olds T (2006) Natural variation in MAM within and between populations of Arabidopsis lyrata determines glucosinolate phenotype. Genetics 173: 1629–1636

Hemm MR, Rider SD, Ogas J, Murry DJ, Chapple C (2004) Light induces phenylpropanoid metabolism in Arabidopsis roots. Plant J 38: 765–778

Hirai MY, Yano M, Goodenowe DB, Kanaya S, Kimura T, Awazuhara M, Arita M, Fujiwara T, Saito K (2004) Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in Arabidopsis thaliana. Proc Natl Acad Sci USA 101: 10205–10210

Hirai MY, Klein M, Fujikawa Y, Yano M, Goodenowe DB, Yamazaki Y, Kanaya S, Nakamura Y, Kitayama M, Suzuki H, Sakurai N, Shibata D, Tokuhisa J, Reichelt M, Gershenzon J, Papenbrock J, Saito K (2005) Elucidation of gene-to-gene and metabolite-to-gene networks in Arabidopsis by integration of metabolomics and transcriptomics. J Biol Chem 280: 25590–25595

Huhman DV, Sumner LW (2002) Metabolic profiling of saponins in Medicago sativa and Medicago truncatula using HPLC coupled to an electrospray ion-trap mass spectrometer. Phytochemistry 59: 347–360

Jenkins H, Johnson H, Kular B, Wang T, Hardy N (2005) Toward supportive data collection tools for plant metabolomics. Plant Physiol 138: 67–77

Jones P, Messner B, Nakajima JI, Schaffner AR, Saito K (2003) UGT73C6 and UGT78D1, glycosyltransferases involved in flavonol glycoside biosynthesis in Arabidopsis thaliana. J Biol Chem 278: 43910–43918

Junker BH, Koschutzki D, Schreiber F (2006) Exploration of biological network centralities with CentiBiN. BMC Bioinformatics 7: 219

Kaplan F, Kopka J, Haskell DW, Zhao W, Schiller KC, Gatzke N, Sung DY, Guy CL (2004) Exploring the temperature-stress metabolome of Arabidopsis. Plant Physiol 136: 4159–4168

Katona ZF, Sass P, Molnar-Perl I (1999) Simultaneous determination of sugars, sugar alcohols, acids and amino acids in apricots by gas chromatography-mass spectrometry. J Chromatogr A 847: 91–102

Kessler A, Halitschke R, Baldwin IT (2004) Silencing the jasmonate cascade: induced plant defences and insect populations. Science 305: 665–668

Keurentjes JJB, Fu JY, de Vos CHR, Lommen A, Hall RD, Bino RJ, van der Plas LHW, Jansen RC, Vreugdenhil D, Koornneef M (2006) The genetics of plant metabolism. Nat Genet 38: 842–849

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

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

Kopka J, Fernie AR, Weckwerth W, Gibon Y, Stitt M (2004) Metabolite profiling in plant biology: platforms and destinations. Genome Biol 5: 109

Kroymann J, Textor S, Tokuhisa JG, Falk KL, Bartram S, Gershenzon J, Mitchell-Olds T (2001) A gene controlling variation in Arabidopsis thaliana glucosinolate composition is part of the methionine chain elongation pathway. Plant Physiol 127: 1077–1088

Kroymann J, Donnerhacke S, Schnabelrauch D, Mitchell-Olds T (2003) Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait locus. Proc Natl Acad Sci USA 100: 14587–14592

Kunne C, Lange M, Funke T, Miehe H, Thiel T, Grosse I, Scholz U (2005) CR-EST: a resource for crop ESTs. Nucleic Acids Res 33: D619–D621

Lindon JC (2005) Standardisation of reporting methods for metabolic analyses: a draft policy document from the standard metabolic reporting structures (SMRS) group. Available at http://www.smrsgroup.org/documents/SMRS_policy_draft_v2.3.pdf

Lindon JC, Nicholson JK, Holmes E, Keun HC, Craig A, Pearce JTM, Bruce SJ, Hardy N, Sansone SA, Antti H, Jonsson P, Daykin C, Navarange M, Beger RD, Verheij ER, Amberg A, Baunsgaard D, Cantor GH, Lehman-McKeeman L, Earll M, Wold S, Johansson E, Haselden JN, Kramer K, Thomas C, Lindberg J, Schuppe-Koistinen I, Wilson ID, Reily MD, Robertson DG, Senn H, Krotzky A, Kochhar S, Powell J, van der Ouderaa F, Plumb R, Schaefer H, Spraul M (2005) Summary recommendations for standardization and reporting of metabolic analyses. Nat Biotechnol 23: 833–838

Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR (2006) Gas chromatography mass spectrometry for metabolite profiling in plants. Nat Protoc 1: 387–396

Luscombe NM, Babu MM, Yu HY, Snyder M, Teichmann SA, Gerstein M (2004) Genomic analysis of regulatory network dynamics reveals large topological changes. Nature 431: 308–312

Matsuda F, Yamada T, Miyazawa H, Miyagawa H, Wakasa K (2005) Characterization of tryptophan-overproducing potato transgenic for a mutant rice anthranilate synthase alpha-subunit gene (OASA1D). Planta 222: 535–545

Mehrtens F, Kranz H, Bednarek P, Weisshaar B (2005) The Arabidopsis transcription factor MYB12 is a flavonol-specific regulator of phenylpropanoid biosynthesis. Plant Physiol 138: 1083–1096

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

Morcuende R, Bari R, Gibon Y, Zheng WM, Pant BD, Blasing O, Usadel B, Czechowski T, Udvardi MK, Stitt M, Scheible WR (2007) Genome-wide reprogramming of metabolism and regulatory networks of Arabidopsis in response to phosphorus. Plant Cell Environ 30: 85–112

Nikiforova VJ, Kopka J, Tolstikov V, Fiehn O, Hopkins L, Hawkesford MJ, Hesse H, Hoefgen R (2005) Systems rebalancing of metabolism in response to sulfur deprivation, as revealed by metabolome analysis of Arabidopsis plants. Plant Physiol 138: 304–318

Noctor G, Bergot G, Mauve C, Thominet D, Lelarge-Trouverie C, Prioul JL (2007) A comparative study of amino acid measurement in leaf extracts by gas chromatography-time of flight-mass spectrometry and high performance liquid chromatography with fluorescence detection. Metabolomics 3: 161–174

Ober D (2005) Seeing double: gene duplication and diversification in plant secondary metabolism. Trends Plant Sci 10: 444–449

Oikawa A, Nakamura Y, Ogura T, Kimura A, Suzuki H, Sakurai N, Shinbo Y, Shibata D, Kanaya S, Ohta D (2006) Clarification of pathway-specific inhibition by Fourier transform ion cyclotron resonance/mass spectrometry-based metabolic phenotyping studies. Plant Physiol 142: 398–413

Oksman-Caldentry KM, Saito K (2005) Integrating genomics and metabolomics for engineering plant metabolic pathways. Curr Opin Biotechnol 16: 174–179

Osuna D, Usadel B, Morcuende R, Gibon Y, Blasing OE, Hohne M, Gunter M, Kamlage B, Trethewey R, Scheible WR, Stitt M (2007) Temporal responses of transcripts, enzyme activities and metabolites after adding sucrose to carbon-deprived Arabidopsis seedlings. Plant J 49: 463–491

Ott KH, Aranibar N, Singh BJ, Stockton GW (2003) Metabonomics classifies pathways affected by bioactive compounds. Artificial neural network classification of NMR spectra of plant extracts. Phytochemistry 62: 971–985

Pandey SP, Baldwin IT (2007) RNA-directed RNA polymerase 1 (RdR1) mediates the resistance of Nicotiana attenuata to herbivore attack in nature. Plant J 50: 40–53

Philippi S, Kohler J (2006) Addressing the problems with life-science databases for traditional uses and systems biology. Nat Rev Genet 7: 482–488

Roessner U, Wagner C, Kopka J, Trethewey RN, Willmitzer L (2000) Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. Plant J 23: 131–142

Roessner U, Luedemann A, Brust D, Fiehn O, Linke T, Willmitzer L, Fernie AR (2001) Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. Plant Cell 13: 11–29

Rogers A, Gibon Y, Stitt M, Morgan PB, Bernacchi CJ, Ort DR, Long SP (2006) Increased C availability at elevated carbon dioxide concentration improves N assimilation in a legume. Plant Cell Environ 29: 1651–1658

Romer S, Fraser PD (2005) Recent advances in carotenoid biosynthesis, regulation and manipulation. Planta 221: 305–308

Saman D, Cvacka J, Svatos A, Bouman EAP, Kalinova B (2006) Structural identification of an anthrasteroid hydrocarbon from the sheep tick Ixodes ricinus. J Nat Prod 69: 1203–1205

Sauter H, Lauer M, Fritsch H (1988) Metabolic profiling of plants – a new diagnostic-technique. Abstr Pap Am Chem Soc 195: 129. AGRO Part 1

Schauer N, Fernie AR (2006) Plant metabolomics: toward biological function and mechanism. Trends Plant Sci 11: 508–516

Schauer N, Zamir D, Fernie AR (2005a) Metabolic profiling of leaves and fruit of wild species tomato: a survey of the Solanum lycopersicum complex. J Exp Bot 56: 297–307

Schauer N, Semel Y, Roessner U, Gur A, Balbo I, Carrari F, Pleban T, Perez-Melis A, Bruedigam C, Kopka J, Willmitzer L, Zamir D, Fernie AR (2006) Genetics of metabolite content in fruits of interspecific introgressions of tomato. Nat Biotechnol 24: 447–454

Scholz M, Gatzek S, Sterling A, Fiehn O, Selbig J (2004) Metabolite fingerprinting: detecting biological features by independent component analysis. Bioinformatics 20: 2447–2454

Scholz M, Kaplan F, Guy CL, Kopka J, Selbig J (2005) Non-linear PCA: a missing data approach. Bioinformatics 21: 3887–3895

Schoof H (2003) Towards interoperability in genome databases: the MAtDB (MIPS Arabidopsis thaliana database) experience. Comp Funct Genomics 4: 255–258

Schuhegger R, Nafisi M, Mansourova M, Petersen BL, Olsen CE, Svatos A, Halkier BA, Glawischnig E (2006) CYP71B15 (PAD3) catalyzes the final step in camalexin biosynthesis. Plant Physiol 141: 1248–1254

Shepherd LVT, McNicol JW, Razzo R, Taylor MA, Davies HV (2006) Assessing the potential for unintended effects in genetically modified potatoes perturbed in metabolic and developmental processes. Targeted analysis of key nutrients and anti-nutrients. Transgenic Res 15: 409–425

Song J, Gardner BD, Holland JF, Beaudry RM (1997) Rapid analysis of volatile flavour compounds in apple fruit using SPME and GC/time-of-flight mass spectrometry. J Agric Food Chem 45: 1801–1807

Sreenivasulu N, Radchuk V, Strickert M, Miersch O, Weschke W, Wobus U (2006) Gene expression patterns reveal tissue-specific signaling networks controlling programmed cell death and ABA-regulated maturation in developing barley seeds. Plant J 47: 310–327

Steuer R (2007) Computational approaches to the topology, stability and dynamics of metabolic networks. Phytochemistry 68: 2139–2151

Steuer R, Nunes-Nesi A, Fernie AR, Gross T, Blasius B, Slbig J (2007). From structure to dynamics of metabolic pathways: application to the plant mitochondrial TCA cycle. Bioinformatics 23: 1378–1385

Stobiecki M, Skirycz A, Kerhoas L, Kachlicki P, Muth D, Einhorn J, Mueller-Roeber B (2006) Profiling of phenolic glycosidic conjugates in leaves of Arabidopsis thaliana using LC/MS. Metabolomics 2: 197–219

Sumner LW, Mendes P, Dixon RA (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics era. Phytochemistry 62: 817–836

Suzuki H, Reddy MSS, Naoumkina M, Aziz N, May GD, Huhman DV, Sumner LW, Blount JW, Mendes P, Dixon RA (2005) Methyl jasmonate and yeast elicitor induce differential transcriptional and metabolic re-programming in cell suspension cultures of the model legume Medicago truncatula. Planta 220: 696–707

Sweetlove LJ, Fernie AR (2005) Regulation of metabolic networks: understanding metabolic complexity in the systems biology era. New Phytol 168: 9–23

Tan JW, Bednarek P, Liu HK, Schneider B, Svatos A, Hahlbrock K (2004) Universally occurring phenylpropanoid and species-specific indolic metabolites in infected and uninfected Arabidopsis thaliana roots and leaves. Phytochemistry 65: 691–699

Tieman D, Taylor M, Schauer N, Fernie AR, Hanson AD, Klee HJ (2006) Tomato aromatic amino acid decarboxylases participate in synthesis of the flavor volatiles 2-phenylethanol and 2-phenylacetaldehyde. Proc Natl Acad Sci USA 103: 8287–8292

Tikunov Y, Lommen A, de Vos CHR, Verhoeven HA, Bino RJ, Hall RD, Bovy AG (2005) A novel approach for nontargeted data analysis for metabolomics. Large-scale profiling of tomato fruit volatiles. Plant Physiol 139: 1125–1137

Tohge T, Nishiyama Y, Hirai MY, Yano M, Nakajima J, Awazuhara M, Inoue E, Takahashi H, Goodenowe DB, Kitayama M, Noji M, Yamazaki M, Saito K (2005) Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. Plant J 42: 218–235

Tolstikov VV, Fiehn O (2002) Analysis of highly polar compounds of plant origin: combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry. Anal Biochem 301: 298–307

von Roepenack-Lahaye E, Degenkolb T, Zerjeski M, Franz M, Roth U, Wessjohann L, Schmidt J, Scheel D, Clemens S (2004) Profiling of Arabidopsis secondary metabolites by capillary liquid chromatography coupled to electrospray ionization quadrupole time-of-flight mass spectrometry. Plant Physiol 134: 548–559

Walther D, Brunnemann R, Selbig J (2007) The regulatory code for transcriptional response diversity and its relation to genome structural properties in A. thaliana. PLoS Genet 3: 216–229

Wilkinson M, Schoof H, Ernst R, Haase D (2005) BioMOBY successfully integrates distributed heterogeneous bioinformatics web services. The PlaNet exemplar case. Plant Physiol 138: 4–16

Windsor AJ, Reichelt M, Figuth A, Svatoš A, Kroymann J, Kliebenstein DJ, Gershenzon J, Mitchell-Olds T (2005) Geographic and evolutionary diversification of glucosinolates among near relatives of Arabidopsis thaliana (Brassicaceae). Phytochemistry 66: 1321–1333

Zhang HN, Sreenivasulu N, Weschke W, Stein N, Rudd S, Radchuk V, Potokina E, Scholz U, Schweizer P, Zierold U, Langridge P, Varshney RK, Wobus U, Graner A (2004) Large scale analysis of the barley transcriptome based on expressed sequence tags. Plant J 40: 276–290

Supplementary material

Four supplementary files are provided for this paper, which represent the underlying numerical data used for the independent component analyses of Fig. 3 and the comparison shown in Fig. 4.

Tables S1–S3 These files represent non-paired analyses of samples taken from the same time course and dose-dependency experiments. Each file contains underlying numeric experimental data (profile sheet) containing $\log_{10}$-transformed response ratios based on row medians. Sample labels of all participating laboratories and respective metabolite or mass fragment identifications are included.

Table S4 This file comprises underlying numerical data used for the comparison of targeted and non-targeted amino acid analysis represented in Fig. 4. Data sheets

contain average values normalised to control values 1 h after incubation and RSD.

This material is available as part of the online article from: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1399-3054.2007.00990.x

(This link will take you to the article abstract).

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary material supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Appendix. List of participants in the school

Spela Baebler	spela.baebler@nib.si	Slovenia	National Institute of Biology, Slovenia
Katarina Cankar	katja.cankar@nib.si	Slovenia	National Institute of Biology, Slovenia
Hua Wang	wanghua@ensat.fr	France/China	National Institute of Polytechnics, France
Magdalena Czemplik	czemplik@ibmb.uni.wroc.pl	Poland	University of Wroclaw, Poland
Liezel Gouws	plantb24@sun.ac.za	South Africa	University of Stellenbosch, South Africa
Julia Hofmann	julia.hofmann@boku.ac.at	Austria	
Julia Krasensky	julia.krasensky@gmi.oeaw.ac.at	Austria	Gregor Mendel Institute of Molecular Plant Biology, Austria
Moco Sofia Moco	sofia.moco@wur.nl	Portugal	Wageningen University, the Netherlands
Dorota Muth	dorotus@ibch.poznan.pl	Poland	Institute of Bioorganic Chemistry of Polish Academy of Science, Poland
Dunia Pino Del Carpio	dunia.pino@wur.nl	Holland, Peru	Wageningen University, the Netherlands
Gerard Bannenberg	gbannenberg@cnb.uam.es	Holland	Centro Nacional de Biotecnologia, Spain
Pawel Stachowiak	pws500@york.ac.uk	UK, Poland	University of York
Ruben tKindt	ruben.tkindt@ugent.be	Belgium	Ghent University, Belgium
Wessel van Leeuwen	wessel.vanleeuwen@wur.nl	Holland	Wageningen University
Marcel van Verk	m.v.verk@chem.leidenuniv.nl	Holland	Leiden University, the Netherlands
Alvaro Cuadros-Inostroza	inostroza@mpimp-golm.mpg.de	Germany, Chile	MPI-MP, Potsdam-Golm

Edited by C. Guy

Phytochem Rev
DOI 10.1007/s11101-008-9097-1

Methylthioalkylmalate synthases: genetics, ecology and evolution

Markus Benderoth · Marina Pfalz · Juergen Kroymann

Received: 30 January 2008 / Accepted: 10 April 2008

Abstract Glucosinolates display an enormous amount of structural variation, both within and between species. This diversity is thought to have evolved in response to challenges imposed on plants by their biotic environment. During the past decade, glucosinolates and myrosinase-catalyzed glucosinolate hydrolysis have become excellent examples for understanding functional diversification in plant secondary metabolism and plant defence. Methylthioalkylmalate (MAM) synthase genes and enzymes are central to the diversification of aliphatic glucosinolate structures in *Arabidopsis thaliana* and related plants. This review summarizes efforts to elucidate how *MAM*-mediated diversity in aliphatic glucosinolate structures is generated and maintained. It also attempts to put variability in methionine carbon chain elongation during glucosinolate biosynthesis into an ecological and evolutionary context.

Keywords Complex traits · Evolutionary dynamics · Glucosinolate metabolism · Natural variation · Plant–insect interactions

The authors M. Benderoth and M. Pfalz have contributed equally to the manuscript.

M. Benderoth · M. Pfalz · J. Kroymann (✉)
AG Genetics and Evolution, Max Planck Institute for Chemical Ecology, Hans-Knöll-Str. 8, 07745 Jena, Germany
e-mail: kroymann@ice.mpg.de

Abbreviations

IPMS	Isopropylmalate synthase
MAM	Methylthioalkylmalate synthase
NIL	Near-isogenic line
QTL	Quantitative trait locus/loci
RIL	Recombinant inbred line

Introduction

Glucosinolates and myrosinases are almost exclusively found in plants from the order Capparales. This order consists of more than a dozen families, including the Brassicaceae and the Capparaceae (Rodman 1991a, b; Rodman et al. 1996). More than 130 different glucosinolates have been characterized; however glucosinolate composition varies remarkably between and within species (Daxenbichler et al. 1991; Fahey et al. 2001; Kliebenstein et al. 2001a; Windsor et al. 2005; Heidel et al. 2006). In *Arabidopsis thaliana*, nearly 40 different glucosinolates have been identified (Kliebenstein et al. 2001a; Reichelt et al. 2002). These glucosinolates are generated from methionine, tryptophan or phenylalanine, with methionine-derived (aliphatic) glucosinolates being the predominant glucosinolate class represented in *A. thaliana*.

Comparative analyses of glucosinolate profiles among Arabidopsis accessions reveal extensive qualitative and quantitative variation (Kliebenstein et al. 2001a; Pfalz et al. 2007). Quantitative trait locus

(QTL) mapping studies in various Arabidopsis recombinant inbred line (RIL) populations (Magrath et al. 1994; Mithen et al. 1995; Campos de Quiros et al. 2000; Kliebenstein et al. 2001b, c; Keurentjes et al. 2006; Wentzell et al. 2007) show consistently that two major genetic loci, methylthioalkylmalate (*MAM*) and *AOP* (Fig. 1), account for most of the variability in aliphatic glucosinolate structures. Enzymes encoded at *AOP* are responsible for the modification of the side chain structure (Kliebenstein et al. 2001b), while the *MAM* locus (also referred to as *ELONG*, *GS-ELONG* or *GSL-ELONG*) controls variability in aliphatic glucosinolate carbon chain length.

MAM encodes a small family of MAM synthases (Kroymann et al. 2001, 2003; Benderoth et al. 2006) in *A. thaliana* and related Brassicaceae. MAM synthases catalyze the condensation of ω-methylthio-2-oxoalkanoic acids derived from methionine with acetyl-CoA to form methylthioalkylmalic acids. Subsequent isomerization and oxidative decarboxylation reactions lead to the generation of ω-methylthio-2-oxoalkanoic acids with extended carbon chains (Fig. 2). Together, condensation, isomerization and oxidative decarboxylation are referred to as the methionine carbon chain elongation cycle, with an extension of one methylene group per cycle. MAM substrate specificity (Table 1) determines whether the reaction products of a given cycle enter the biosynthetic pathway generating the glucosinolate core structure or whether they undergo additional cycles of carbon chain elongation. *A. thaliana* MAM1 and MAM2 are involved in the formation of aliphatic glucosinolates with short carbon chains (Kroymann et al. 2001; Textor et al. 2004; Benderoth et al. 2006), while MAM3 catalyzes condensation reactions in the biosynthesis of aliphatic glucosinolates with long carbon chains (Field et al. 2004; Textor et al. 2007; Knoke et al. 2008). MAM synthases determine variability during the earliest stage of aliphatic glucosinolate biosynthesis and play a central role in glucosinolate diversity. Hence, evolutionary forces acting on *MAM* have a fundamental impact on aliphatic glucosinolate composition.

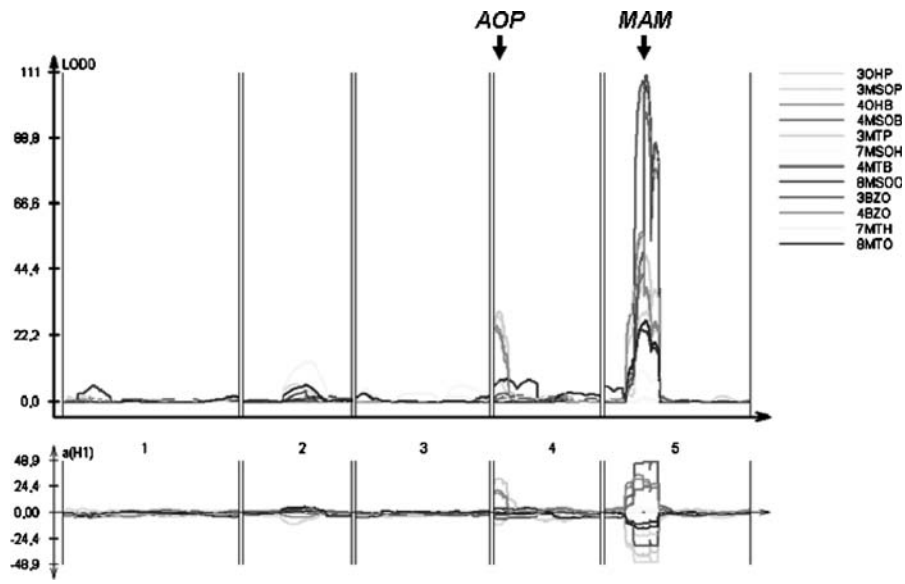

Fig. 1 QTL for aliphatic glucosinolates in seeds from Col-0 × L*er* RILs. Top: The *MAM* locus on chromosome five influences the quantity of all aliphatic glucosinolates. 3-hydroxypropyl (3OHP), 3-methylsulfinylpropyl (3MSOP), 3-methylthiopropyl (3MTP) and 3-benzoyloxypropyl (3BZO) originate from homomethionine (3C), 4-hydroxybutyl (4OHB), 4-methylsulfinylbutyl (4MSOB), 4-methylthiobutyl (4MTB) and 4-benzoyloxybutyl (4BZO) from dihomomethionine (4C), 7-methylsulfinylheptyl (7MSOH) and 7-methylthioheptyl (7MTH) from pentahomomethionine (7C) and 8-methylsulfinylocytyl (8MSOO) and 8-methylthiooctyl (8MTO) from hexahomomethionine (8C). Bottom: Additive genetic effects for homo- and pentahomomethionine-derived glucosinolates are opposite to those for dihomo- and hexahomomethionine-derived glucosinolates. QTL mapping was conducted with Windows QTL cartographer 2.5 (Wang et al. 2001–2004) and is based on log-transformed HPLC data from 297 Col-0 × L*er* RILs (John D'Auria and Juergen Kroymann, unpublished data). For all traits, the significance threshold was ~2.5 LOD units, based on 1,000 permutations of the data

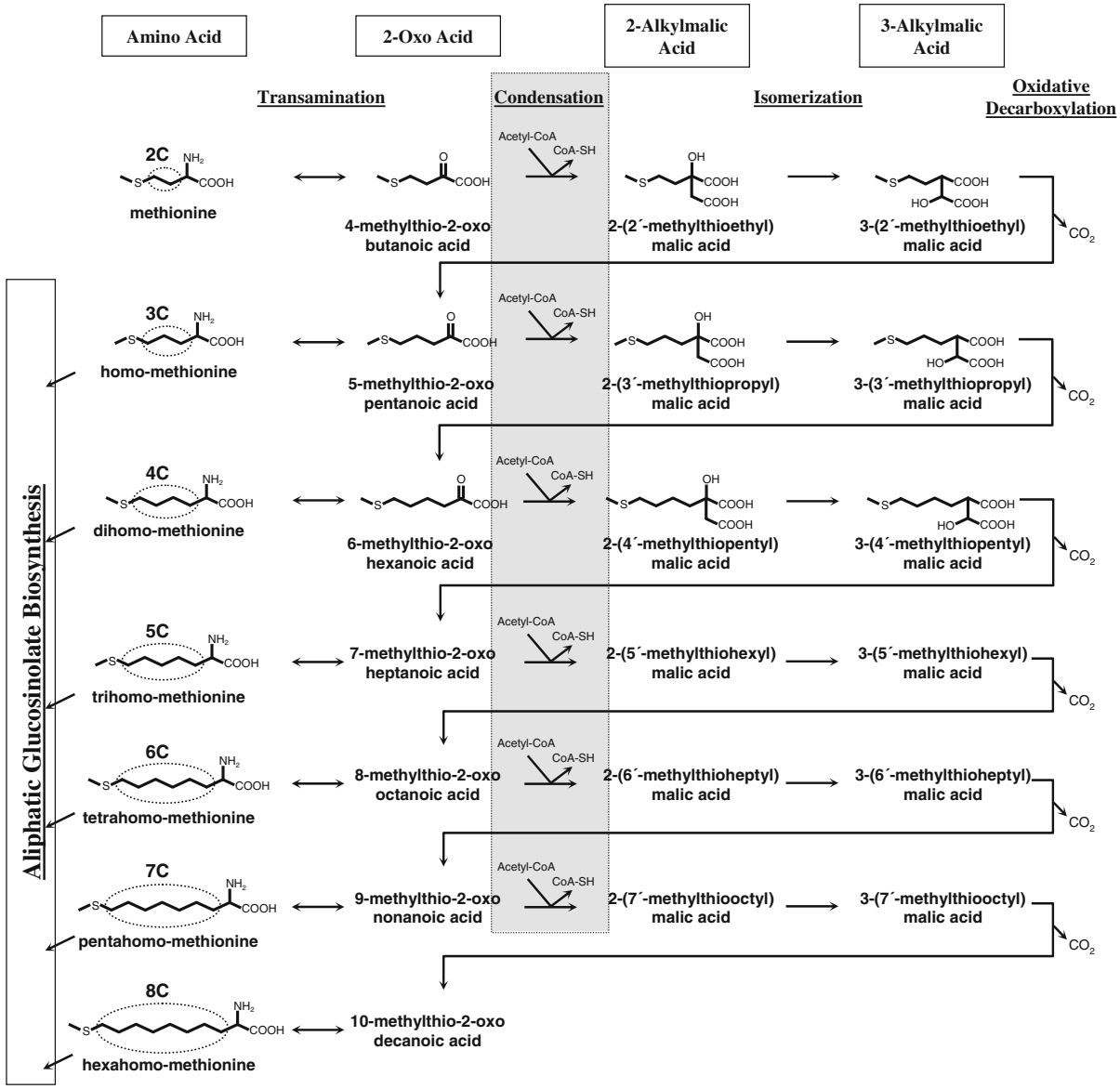

Fig. 2 The methionine carbon chain elongation cycle. MAM synthases encoded at the *MAM* locus catalyze the committed step in methionine carbon chain elongation. Two subsequent reactions, an isomerization and an oxidative decarboxylation complete a cycle. MAM substrate specificity (Table 1) determines whether ω-methylthio-2-oxoalkanoic acids remain in the carbon chain elongation cycle or enter the biosynthetic pathway that generates the glucosinolate core structure. Here, all reactions occurring in *A. thaliana* are depicted; in other Brassicaceae fewer or more cycles are possible

Cloning of the *MAM* QTL

The widely used *A. thaliana* accession Landsberg *erecta* (L*er*) accumulates homomethionine-derived (3C) glucosinolates. In contrast, aliphatic glucosinolates in Columbia (Col-0), recognized for having provided the blueprint of the Arabidopsis genome (Arabidopsis Genome Initiative 2000), originate mainly from dihomomethionine (4C). This biochemical difference between Col-0 and L*er* enabled initial mapping of the responsible genetic locus to a region of approximately 140 kb on chromosome 5 (Magrath et al. 1994; Campos de Quiros et al. 2000). Within this region, two tandemly arranged genes were identified as

Table 1 MAM substrate specificities in glucosinolate biosynthesis

Assays were conducted with heterologously expressed *MAM* genes and generic and/or artificial substrates

MAM synthase	Substrates	Elongation cycle(s)	References
A. thaliana MAM1	2C, 3C, (4C)	1, 2, (3)	Kroymann et al. (2001), Textor et al. (2004), Benderoth et al. (2006)
A. thaliana MAM2	2C	1	Benderoth et al. (2006)
A. thaliana MAM3	2C–7C	1–6	Textor et al. (2007), Knoke et al. (2008)
A. lyrata MAMa	2C	1	Benderoth et al. (2006)
B. stricta MAMa			

candidates, based on their sequence similarity with isopropylmalate synthase (IPMS) genes. IPMS catalyzes the condensation of 2-oxoisovalerate with acetyl-CoA to form isopropylmalate in leucine biosynthesis, a reaction similar to the MAM-catalyzed condensation of ω-methylthio-2-oxoalkanoic acids with acetyl-CoA in glucosinolate biosynthesis. The two candidate genes were termed *MAM1* and *MAM-L* (meanwhile often referred to as *MAM3*), respectively (Kroymann et al. 2001).

In the Col-0 accession, *MAM1* and *MAM3* are separated by ca. 11.5 kb of intervening sequence. High-resolution mapping was employed to separate functional effects of these candidate genes. Col-0 was crossed with CL5, an RIL from the Col-0 × L*er* population (Lister and Dean 1993). This particular line was chosen because it shared approximately 70% of its genome with the Col-0 accession but has the L*er* allele at the *MAM* locus, thus enabling fine-mapping with near-isogenic lines (NILs). In Col-0 × CL5 F_2 progeny one line was identified that had recombined between *MAM1* and *MAM3*. This recombinant line had the Col-0 *MAM1* genotype and was heterozygous at *MAM3*. Its glucosinolate phenotype resembled the parental Col-0 profile closely, with 4C glucosinolates predominating. Likewise, progeny from this line produced mainly 4C glucosinolates when they had the Col-0 *MAM1* and the L*er MAM3* genotype. Hence, the biochemical difference in short-chain aliphatic glucosinolate composition was attributable to *MAM1* and not *MAM3* (Kroymann et al. 2001). Further evidence for the role of MAM1 in methionine carbon chain elongation was obtained with *MAM1* mutants, deficient in dihomomethionine-derived glucosinolates (Haughn et al. 1991), and with biochemical assays of heterologously expressed *MAM1*, which showed that the encoded protein has the capacity of condensing ω-methylthio-2-oxoalkanoic acids with acetyl-CoA (Kroymann et al. 2001; Textor et al. 2004; Benderoth et al. 2006).

Origin of *MAM* genes

Several groups of enzymes catalyze condensation reactions between 2-oxo acids and acetyl-CoA. These enzymes belong to enzyme class EC 2.3.3.-. Examples are citrate synthase (EC 2.3.3.1), which condenses oxaloacetate with acetyl-CoA in the TCA cycle, malate synthase (EC 2.3.3.9), which is responsible for the condensation of glyoxylate with acetyl-CoA in the glyoxylate cycle, and 2-IPMS (EC 2.3.3.13) involved in leucine biosynthesis.

Among the genes that encode enzymes of class EC 2.3.3.-, *MAM* genes share the highest degree of sequence similarity with 2-*IPMS* genes. Phylogenetic analyses reveal a close relationship between *MAM* and *IPMS* genes, with *MAM*s having evolved from *IPMS* (Fig. 3). Similar gene structures of *MAM* and *IPMS* corroborate this close relationship; most intron positions are identical in *MAM* and *IPMS* (Kroymann et al. 2001). However, the first *IPMS* exon corresponds to the first two exons in *MAM* genes, and the last four *IPMS* exons are fused to a single exon in *MAM*s. Exon fusion was accompanied with a net loss of ca. 250 bp in the 3′ portion of *MAM* genes, and may have functional significance since this region encodes a conserved allosteric leucine binding site in IPMS (de Kraker et al. 2007). As a consequence, Arabidopsis *IPMS* genes consist of 12 exons and *MAM* genes of 10 exons, with *IPMS* exons 2–8 and *MAM* exons 3–9 corresponding to one another (Fig. 4). Furthermore, all *MAM* genes encode an N-terminal plastid targeting signal (Kroymann et al. 2001), suggesting that methionine carbon chain elongation occurs—like leucine biosynthesis (Hagelstein and Schultz 1993)—in the plastids, a

Fig. 3 Bayesian phylogenetic analysis of MAM and IPMS sequences. Amino acid sequences were aligned with ClustalW. Target peptides and alignment gaps were excluded. The tree was constructed with MrBayes v 3.1.2 (Huelsenbeck and Ronquist 2001) using a mixed amino acid substitution model approach. Data structure was best explained by the wag-model (Whelan and Goldman 2001) with a posterior probability of 1.000. The Markov-Chain-Monte-Carlo simulation was performed in three parallel runs with four chains each for 1,000,000 iterations with a sample frequency of 200 and a burnin fraction of 0.05. Convergence diagnostic over all three runs (PSRF) was 1.000. Shown are posterior probabilities for all internal nodes. Note that *Medicago truncatula IPMS* is incorrectly annotated in the databases; here, a corrected annotation was assembled from clone mth2-19b5 for alignment and tree construction

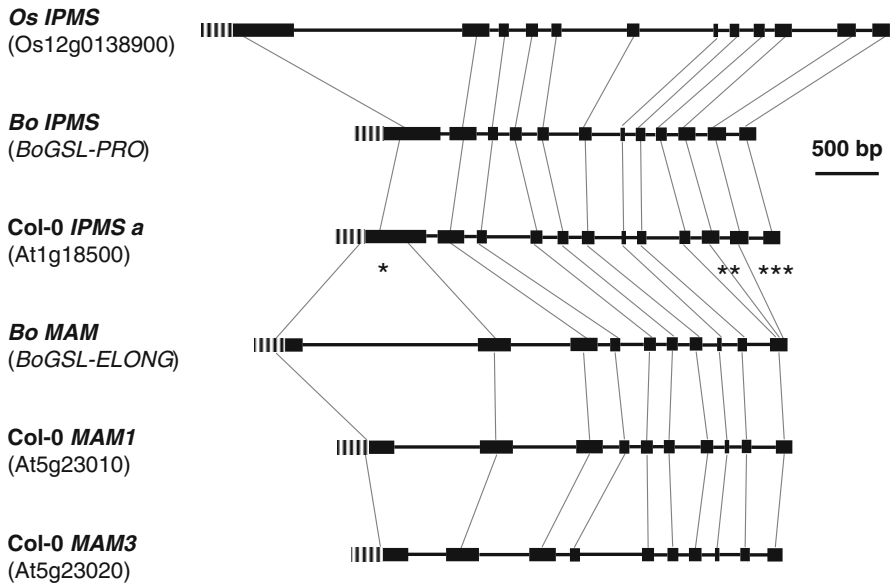

Fig. 4 Exon-intron structure of *IPMS* and *MAM* genes. *IPMS* genes from higher plants consist of 12, *MAM* genes of ten exons. Differences between Arabidopsis *IPMS* and *MAM* gene structures involve the split of *IPMS* exon I (*), the fusion and partial loss of *IPMS* exons IX, X and XI (**), and the loss of *IPMS* exon XII (***). Plastid targeting sequences are represented by dashed lines. Os, *Oryza sativa*; Bo, *B. oleracea*; Col, *A. thaliana* accession Col-0

prediction corroborated by the demonstration of MAM activity in chloroplast-enriched extracts of *Eruca sativa* leaves (Falk et al. 2004) and by recent immunolocalization experiments with MAM3 antibodies (Textor et al. 2007).

Database searches reveal that *IPMS* genes are present in eubacteria and archaebacteria, in green algae, higher plants, and fungi, but are absent in animals. Among the prokaryotes, *IPMS* genes from cyanobacteria appear to be most closely related to those from higher plants and green algae (Fig. 5), suggesting that plant *IPMS* genes originate from the cyanobacterial endosymbionts that gave rise to plastids, with a transfer of the cyanobacterial/plastid copy to the plant nucleus (Martin and Herrmann 1998).

Variability in the genetic composition of the *MAM* locus in *A. thaliana*

To elucidate the genomic basis of carbon chain length control, *MAM* genes and flanking DNA were sequenced from 25 Arabidopsis accessions randomly selected from the species' natural distribution (Kroymann et al. 2003). This comparative analysis revealed extensive variation in the molecular composition of the *MAM* locus among Arabidopsis accessions (Fig. 6). In addition to *MAM1* and *MAM3*, which are present in Col-0, other accessions harbored a third gene, referred to as *MAM2* (Kroymann et al. 2003). Sequence comparison enabled the reconstruction of the archetypical configuration of the *MAM* locus in *A. thaliana*, consisting of tandemly arranged *MAM2*, *MAM1* and *MAM3* genes, all transcribed in the same direction. However, secondary gene deletions have occurred frequently in the history of the *A. thaliana MAM* cluster. In some accessions (including Col-0) the *MAM2* gene was deleted. In other accessions, the *MAM1* gene was deleted, either partially, involving promoter and 5′-portion of the coding sequence (e.g. in L*er*), or completely. Finally, the Lm-2 accession contains a chimerical gene with 5′ *MAM2* fused to 3′ *MAM1* sequence, possibly caused by a deletion of the intervening region. Nonetheless, in contemporary accessions at least one of these genes has been retained. Likewise, all tested natural accessions contain a functional *MAM3* gene.

Sequence exchange between *MAM1* and *MAM2* genes is another factor contributing to variability in the gene composition of the *MAM* locus. In some accessions, sequence transfer has occurred from *MAM2* to *MAM1*, in others sequence portions have been shifted from *MAM1* to *MAM2*. In the most extreme cases, represented by two accessions from Tajikistan (Hodja and Condara), the gene at the position originally occupied by *MAM1* has been almost completely converted to a *MAM2*-like gene. However, another accession from Tajikistan, Sorbo, was not affected by gene deletion or conversion events. This accession has functional *MAM2* and

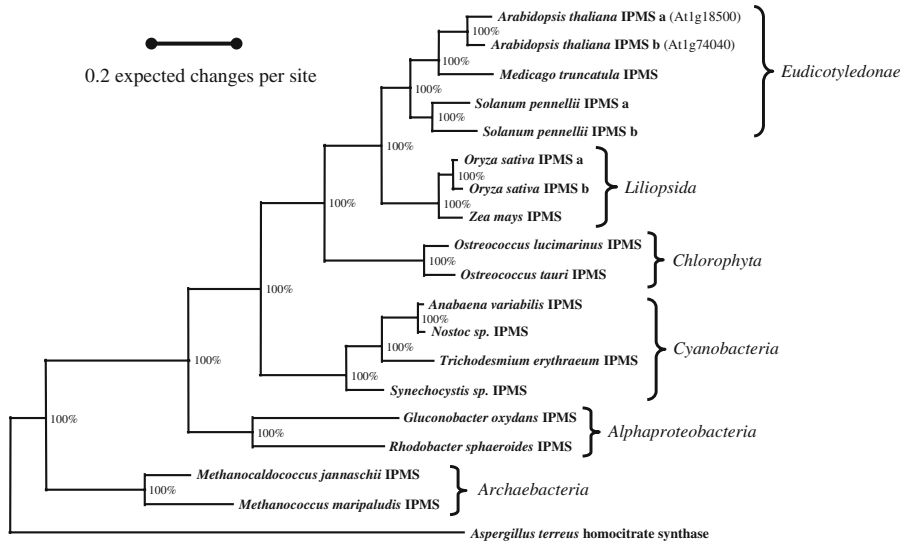

Fig. 5 Evolutionary origin of plant IPMS sequences inferred with Bayesian phylogenetic analysis. Calculations were based on ClustalW-aligned amino acid sequences. Target peptides and alignment gaps were excluded. The analysis was conducted as explained in the legend to Fig. 3. Again, data structure was best explained by the wag-model (Whelan and Goldman 2001) with a posterior probability of 0.999. Convergence diagnostic over all three runs (PSRF) was 1.000. Posterior probabilities for all internal nodes are shown

MAM1 genes and accumulates 4C glucosinolates, indicating that MAM1 overrides MAM2 function. Likewise, all accessions with at least one functional *MAM1*-like gene accumulate short chain aliphatic glucosinolates generated from dihomomethionine, while accessions without a functional *MAM1* produce short-chain aliphatic glucosinolates almost exclusively from homomethionine. Taken together, these data show that the *MAM* cluster in *A. thaliana* is subject to dynamic evolutionary change.

Quantitative effects of the *MAM1/MAM2* polymorphism on glucosinolate profiles

The consequences of the *MAM1/MAM2* polymorphism on Arabidopsis glucosinolate profiles were investigated in a set of NILs, which were developed from Col-0 × CL5 progeny with recombination breakpoints in the vicinity of the *MAM* QTL (Kroymann et al. 2001, 2003). All individual aliphatic glucosinolates were found to be influenced by the *MAM1/MAM2* polymorphism (Fig. 7). Col-0 *MAM1* genotypes produced more aliphatic glucosinolates derived from dihomo- (4C), trihomo- (5C), tetrahomo- (6C) and pentahomomethionine (7C), while L*er MAM2* genotypes accumulated larger quantities of homo- (3C) and hexahomomethionine (8C) derivatives. Moreover, the *MAM1/MAM2* polymorphism also caused differences in the concentration of total aliphatic glucosinolates, with L*er MAM2* genotypes accumulating 60% more aliphatic glucosinolates in leaves and 20% more in seeds than Col-0 *MAM1* genotypes (Kroymann et al. 2003; Fig. 8).

Even though MAM1 and MAM2 lack the capacity to catalyze the condensation reactions in advanced methionine carbon chain elongation cycles (Benderoth et al. 2006), the QTL peaked in all cases at an interval containing *MAM1* or *MAM2*, but not *MAM3*. Hence, the *MAM1/MAM2* polymorphism influenced not only aliphatic glucosinolates with short carbon chains but also affected glucosinolates with long carbon chains. The biochemical basis for this effect of *MAM1* versus *MAM2* on long-chain aliphatic glucosinolate accumulation is not yet understood.

A complex influence of the *MAM* locus on glucosinolate profiles was also seen in studies with another Arabidopsis RIL population, Bay-0 × Sha (Kliebenstein et al. 2006; Wentzell et al. 2007). eQTL mapping and network analysis (further explained in Kliebenstein 2008) suggest that this influence is, to a certain degree, caused by changes in transcript levels.

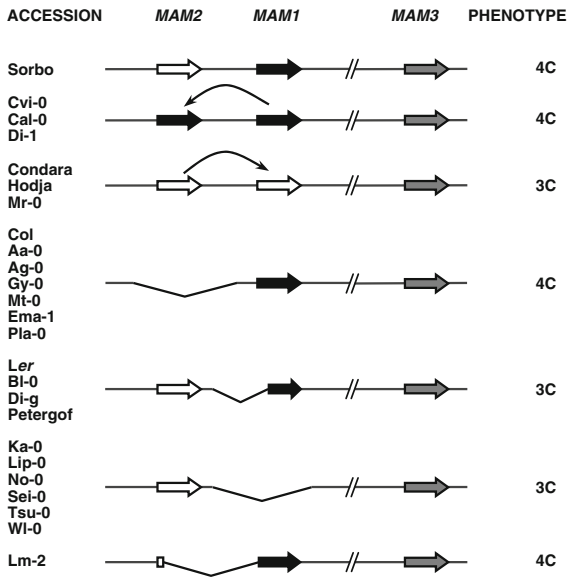

Fig. 6 Schematic representation of the *MAM* cluster in Arabidopsis accessions. The organization of the *MAM* cluster is highly variable. Partial or complete deletions of *MAM2* or *MAM1* have occurred frequently. In Lm-2, 5′-*MAM2* sequence is fused to 3′-*MAM1* sequence. Furthermore, sequence information has been transferred between loci. In Cvi-0, Cal-0 and Di-1, the gene at the *MAM2* position has been partially converted into a *MAM1*-like sequence; in Condara, Hodja and Mr-0, *MAM1* has been converted into a *MAM2*-like gene. All tested accessions have a *MAM3* gene. The predominant aliphatic glucosinolate class is indicated in the right column (3C: homomethionine-derived, 4C: dihomomethionine-derived)

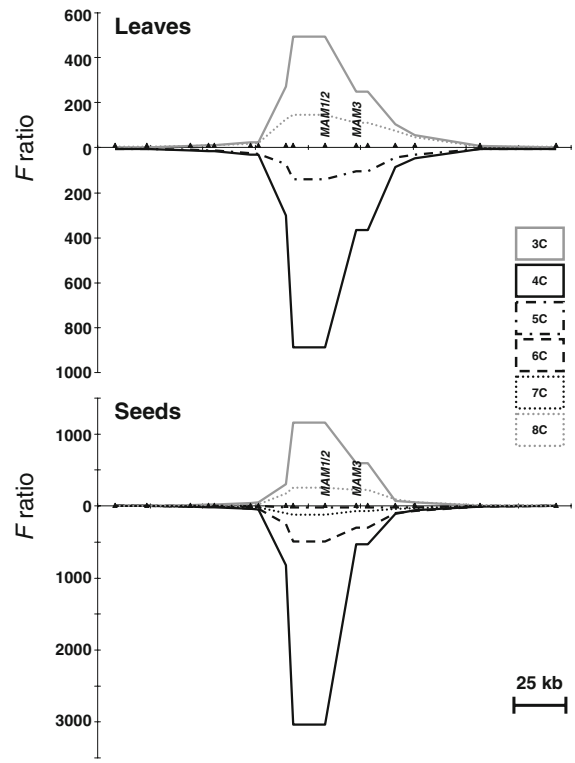

Fig. 7 Single-marker QTL analysis of aliphatic glucosinolates with different carbon chain lengths in a 205-kb region of the Arabidopsis genome. Fifty-eight NILs with recombination breakpoints in the *MAM* region were assayed for trait values in leaves (top) and seeds (bottom). QTL for homo- (3C), dihomo- (4C), trihomo- (5C), tetrahomo- (6C), pentahomo- (7C) and hexahomomethionine (8C)-derived glucosinolates are centred at a 15-kb non-recombinant region that contains Col-0 *MAM1* or L*er MAM2*, but not *MAM3*. Col-0 *MAM1* genotypes produce more 4C–7C derivatives than L*er MAM2* genotypes. By contrast, L*er MAM2* genotypes accumulate higher quantities of 3C and 8C derivatives. Small black triangles show marker positions based on the Col-0 sequence; the position of *MAM1*/*MAM2* and *MAM3* is indicated. *F* ratios quantify statistical significance of a QTL at a given marker position. Note that leaves contain lower concentrations of glucosinolates than seeds; therefore, low-abundant glucosinolates are difficult to detect by HPLC in leaves

Ecological consequences of the *MAM1*/*MAM2* polymorphism

Two lepidopteran insects, *Spodoptera exigua* and *Plutella xylostella*, were tested for differences in larval performance contingent on the genotype at the *MAM* locus in Col-0 × CL5 NILs. *S. exigua* is referred to as a generalist due to possession of a broad host range; it has the ability to feed on a variety of plants from different families. In contrast, *P. xylostella* has a narrow host range utilizing almost exclusively plants from the crucifer family. *S. exigua* was found to respond to the *MAM1*/*MAM2* polymorphism. Its performance was lower on the *MAM2* genotype, and larvae caused approximately 17% less damage than on the *MAM1* genotype (Kroymann et al. 2003). Again, single marker analysis showed that the resistance QTL peaked in the interval containing *MAM1*/*MAM2* but not *MAM3* (Fig. 8). For *P. xylostella*, no significant difference was found with ANOVA. Nonetheless, larvae performed on average better on *MAM2* genotypes (Kroymann et al. 2003). Furthermore, re-analysis of this data set with regression showed that *Plutella* larval herbivory was positively correlated with leaf aliphatic glucosinolate content (Kliebenstein et al. 2005). Hence, the *MAM1*/*MAM2* polymorphism had contrasting effects on *S. exigua* versus *P. xylostella*. *MAM2* genotypes were

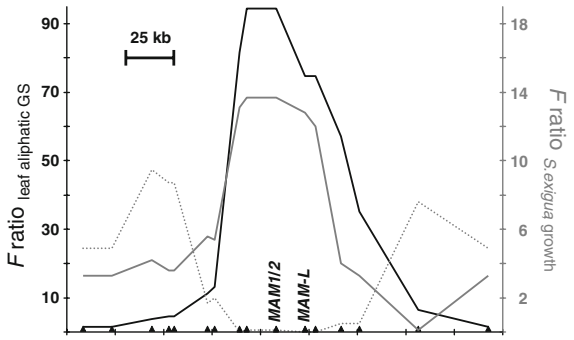

Fig. 8 Single-marker QTL analysis of leaf total aliphatic glucosinolates (black line), resistance against *S. exigua* (solid grey line) and growth rate (dotted grey line). QTL for leaf aliphatic glucosinolates and for resistance against *S. exigua* centre at a 15-kb non-recombinant interval containing L*er MAM2* or Col-0 *MAM1*. NILs with a L*er MAM2* accumulate more aliphatic glucosinolates and are more resistant against *Spodoptera* than the *MAM1* genotype. Two QTL for biomass accumulation were detected upstream and downstream of the *MAM* cluster, but there is no indication for a growth rate QTL at the *MAM* genes (Kroymann et al. 2003; Kroymann and Mitchell-Olds 2005)

better defended against the generalist insect herbivore, whereas *MAM1* genotypes appeared to suffer less damage from the specialist. However, although the *MAM1*/*MAM2* polymorphism controls these differences in insect performance, it is not yet clear which biochemical parameter ultimately accounts for the observed effects, due to the complex influence of the *MAM* locus on aliphatic glucosinolate profiles. Differences in resistance could be attributable to leaf total aliphatic glucosinolate concentration, levels of individual glucosinolates, differences in glucosinolate composition, or any combination of these factors.

Since *MAM2* genotypes produced approximately 60% more leaf and 20% more seed glucosinolates than *MAM1* genotypes, this raised the question whether increased glucosinolate production in the *MAM2* genotype was paid for with a reduction in plant growth rate. Such allocation costs can occur when defenses are energetically expensive, so that genotypes with strong defenses have fewer resources to invest in growth and reproduction (Purrington 2000; Tian et al. 2003). Therefore, dry weight of plant rosettes was measured at the pre-reproductive state. Quantitative analyses indicated the presence of growth rate QTL upstream and downstream of the *MAM* locus, but there was no trace of a significant genotype effect on biomass accumulation at the *MAM* locus itself (Fig. 8, Kroymann et al. 2003; Kroymann and Mitchell-Olds 2005). Thus, allocation costs appear to be of minor importance for the *MAM1*/*MAM2* polymorphism in *A. thaliana*.

Generation of glucosinolate diversity

Sequencing of the MAM locus from the close *A. thaliana* relatives *A. petraea* ssp. *lyrata* (=*A. lyrata*), *Arabidopsis petraea* ssp. *petraea* (= *A. petraea*) and *Arabidopsis cebennensis*, and the more distant relative *Boechera divaricarpa* revealed that these species all possessed three MAM genes in the MAM cluster. These genes were termed MAMa, MAMb and MAMc (Fig. 9, Benderoth et al. 2006). Phylogenetic analyses showed that *A. thaliana* MAM1 and MAM2 originated from the duplication of a MAMa gene after *A. thaliana* diverged from its congeners (Fig. 3). *MAMb* genes were found to be orthologous to *A. thaliana* MAM3. A MAMc does not exist in the Arabidopsis genome, but is present in other Arabidopsis species and in *Boechera*. Hence, this gene was lost in *A. thaliana* after divergence from its congeners.

Analyses of nucleotide substitution patterns revealed that *MAM1* had accumulated an excess of non-synonymous nucleotide substitutions (i.e., substitutions that alter codon meaning) after the *MAMa* duplication, indicating positive (Darwinian) selection. The biochemical characteristics of MAM1, MAM2 and MAMa were investigated to infer which properties of *MAM1* were targeted by positive selection. *MAMa* from *A. lyrata* and from *Boechera stricta* (a close relative of *B. divaricarpa*), *MAM2* from the Arabidopsis accession L*er*, and *MAM1* from the accession Sorbo were heterologously expressed in *Escherichia coli*. Enzyme assays were carried out to investigate substrate specificity, ATP-, metal ion- and pH dependence (Benderoth et al. 2006). Major differences were found only for the enzymes' substrate specificities (Table 1). *A. petraea* and *B. stricta MAMa*, and L*er MAM2* all utilized 4-methylthio-2-oxobutanoic acid (2C) for condensation with acetyl-CoA, but none accepted ω-methylthio-2-oxoalkanoic acids with more than two methylene groups as a substrate. By contrast, Sorbo MAM1 accepted 4-methylthio-2-oxobutanoic acid (2C), 5-methylthio-2-oxopentanoic acid (3C) and, with low activity, also 6-methylthio-2-oxohexanoic

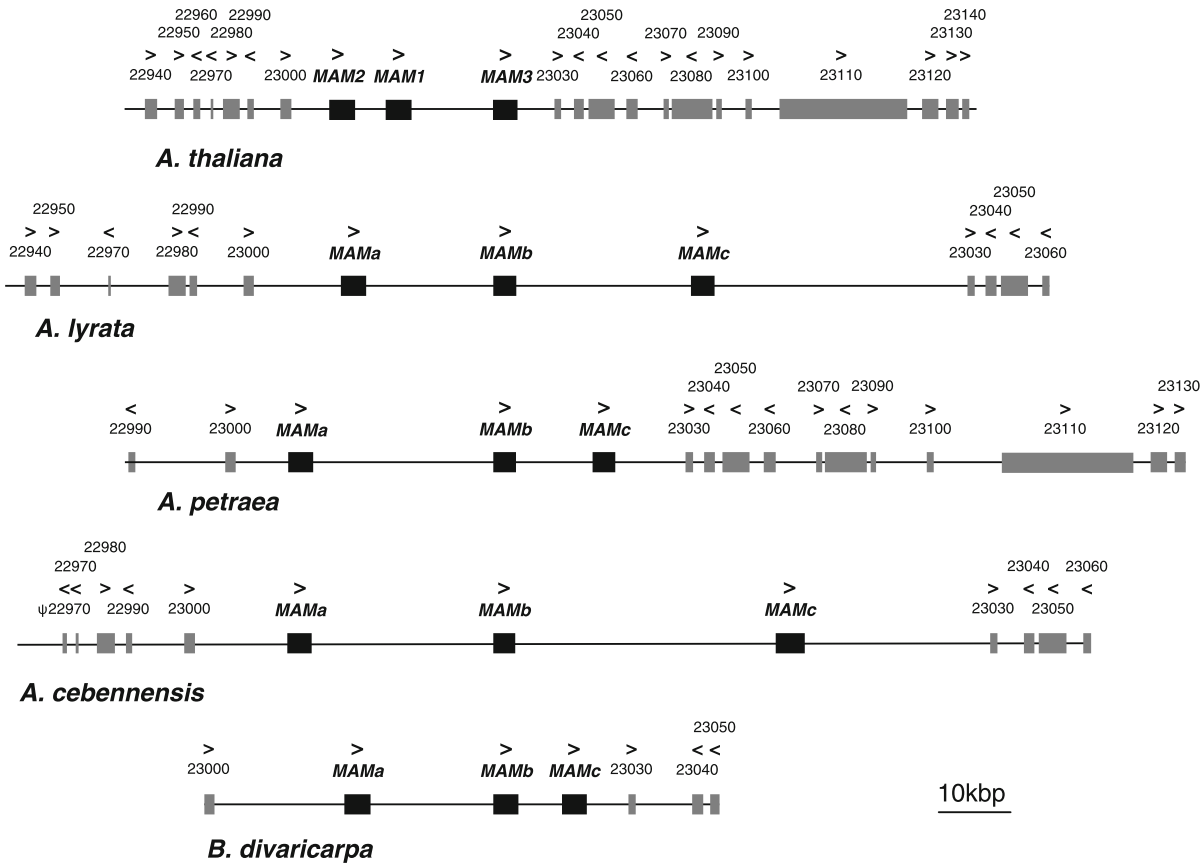

Fig. 9 Organization of the *MAM* region in *A. thaliana* relatives. The *MAM* cluster contains three genes (*MAMa*, *MAMb* and *MAMc*) in *A. lyrata*, *A. petraea*, *A. cebennensis* and *B. divaricarpa*. The regions flanking *MAM* are largely co-linear. For comparison, the archetypical configuration for *A. thaliana* is included. Numbers refer to AGI annotations (Arabidopsis Genome Initiative 2000), and the direction of transcription is indicated

acid (4C) as a substrate. Thus, MAMa and MAM2 function only in the first cycle of carbon chain extension, whereas MAM1 has acquired additional capacity to carry out condensation reactions in subsequent chain elongation cycles, equivalent to a biochemical neofunctionalization. MAM1 and MAM2 substrate specificities are matched by *A. thaliana* glucosinolate phenotypes. Accessions with a functional *MAM1* gene accumulate 4C glucosinolates, whereas accessions without a functional *MAM1* (but with a functional *MAM2*) generate mainly 3C aliphatic glucosinolates (Kroymann et al. 2003). Thus, gene duplication, biochemical neofunctionalization and positive selection account for the generation of diversity in the carbon chain lengths of aliphatic glucosinolates in *A. thaliana*.

The same processes are duplicated along the basal branches of the *MAM* gene tree (Fig. 3). *MAMa*, *MAMb* and *MAMc* have originated by gene duplication events, and are functionally diversified. *MAMa* controls short-chain aliphatic glucosinolates (Benderoth et al. 2006) and *MAMb* is, like its *A. thaliana* ortholog *MAM3* (Textor et al. 2007; Knoke et al. 2008), presumably involved in the biosynthesis of long-chain aliphatic glucosinolates. However, the function of *MAMc* is not yet known. Analyses of nucleotide substitution patterns indicate positive selection along the deep branches of the *MAM* tree, connecting *MAMa*, *MAMb* and *MAMc* (Benderoth et al. 2006). Hence, the same events that explain diversity in short-chain glucosinolates, i.e., gene duplication and biochemical neofunctionalization, driven by positive selection, appear to account for the diversification in aliphatic glucosinolate carbon chain lengths in general.

Dihomomethionine (4C) differs by only one methylene group from homomethionine (3C). At first

glance, this seems to be a minor difference. But the consequences of this small chemical difference can be profound when taken within the context of the glucosinolate-myrosinase system. After methionine carbon chain elongation and glucosinolate core structure generation, enzymes encoded at other genetic loci act to modify the carbon chain (Kliebenstein et al. 2001a, b). These loci are, like *MAM*, also polymorphic in *A. thaliana*, i.e., they harbour alleles whose gene products have diverse biochemical activities and cause different modifications of the aliphatic carbon chain. Likewise, modifying proteins like ESP (Lambrix et al. 2001) and ESM1 (Zhang et al. 2006) can alter glucosinolate breakdown identity during myrosinase-catalyzed glucosinolate hydrolysis. Taken together, different combinations of alleles at glucosinolate biosynthesis and hydrolysis loci generate quite different blends of glucosinolate hydrolysis products (Kliebenstein et al. 2005), and these may, in turn, cause different responses in attacking insect herbivores and other enemies.

Maintenance of glucosinolate diversity

While gene duplication, neofunctionalization and positive selection contribute to the generation of metabolic diversity, these processes do not satisfyingly explain how and why this diversity is maintained. Positive selection on *MAM1* indicates that this gene has provided a fitness advantage to its carriers after duplication of an ancestral *MAMa* gene. MAM1 overrides MAM2 function, and MAM1 has retained the capacity to function in the first methionine carbon chain elongation cycle (Kroymann et al. 2003; Benderoth et al. 2006). Thus, the MAM2 is not required to sustain MAM1 activity. Why then is *MAM2* still present in *A. thaliana*?

It could be possible that *MAM2* is in the process of becoming a pseudo-gene, but that the period after MAM1 neofunctionalization was too short for a complete degeneration of *MAM2*. For two reasons, this is not likely. First, estimates based on the analysis of nucleotide substitution rates in the Brassicaceae (Yang et al. 1999; Koch et al. 2001) suggest that the *MAMa* duplication occurred more than 10^5 generations ago. Second, a degenerating gene is expected to accumulate nucleotide substitutions since selection no longer acts to eliminate deleterious mutations. Mutations in open reading frames are called synonymous when the amino acid sequence remains unchanged and non-synonymous when the codon usage is altered. Because of the nature of the genetic code, a coding sequence has more non-synonymous than synonymous positions. Therefore, in a degenerating gene non-synonymous changes are more likely to occur than synonymous substitutions. But after correction for the number of non-synonymous and synonymous positions, the ratio between non-synonymous and synonymous changes in a degenerating gene is expected to be close to 1. However, when *MAM2* was compared to *A. thaliana MAM1* or *A. lyrata MAMa*, *MAM2* exhibited an excess of synonymous relative to non-synonymous changes, indicating purifying selection (Benderoth et al. 2006). In conclusion, MAM2 function appears to be preserved in *A. thaliana*.

What else could explain why *MAM2* was retained in *A. thaliana*? Insect herbivory assays had shown that NILs with the L*er MAM2* performed ca. 17% better against *S. exigua* larvae than lines with the Col-0 *MAM1* (Kroymann et al. 2003). Thus, *MAM2* can provide a selective advantage over *MAM1* under certain conditions. In addition, secondary gene deletion events and exchange of sequence information between paralogous *MAM1* and *MAM2* loci have occurred frequently in the history of *A. thaliana* (Fig. 6). Deletion of *MAM1* or conversion of *MAM1* into a *MAM2*-like sequence both result in a switch from a 4C glucosinolate profile to a profile dominated by 3C glucosinolates. If such a switch occurred in a local population of plants that were otherwise genetically uniform, this could confer a—temporal—selective advantage to the novel genotypes, provided that the local herbivore community was accustomed to a particular blend of glucosinolates. Since *A. thaliana* propagates mostly by selfing and local populations were founded frequently in the history of this species, such a scenario is not unlikely, and might account for the large proportion of derived genotypes at the *MAM* locus.

Furthermore, the composition of local herbivore communities varies temporally, with different classes of insects—specialists, generalists, and non-feeders (i.e., herbivorous insects not utilizing glucosinolate-containing plants as a host)—occurring in different frequencies over time. This may lead to fluctuating selection on glucosinolate profiles, with periods during which phenotypes with a particular glucosinolate

composition or with high glucosinolate levels increase in frequency, alternating with periods during which other types of glucosinolate profiles or low levels are advantageous. As a consequence, one would expect some form of equilibrium between different glucosinolate phenotypes. Indeed, among 51 Arabidopsis accessions whose glucosinolate profiles have been analyzed (Kliebenstein et al. 2001a; Kroymann et al. 2003; Pfalz et al. 2007), 29 produced aliphatic glucosinolates predominantly from homomethionine (3C) and 22 from dihomomethionine (4C). These data are not significantly different from the hypothesis that both phenotypes have equal frequencies in *A. thaliana* ($N = 51$; $df = 1$, $\chi^2 = 1.27$, n.s.). Also, two independent statistical tests of molecular population genetics found evidence for balancing selection acting on the *MAM2* gene (Kroymann et al. 2003). Balancing selection refers to evolutionary scenarios that maintain more genetic variation in a population than expected under neutrality (Nordborg and Innan 2002). First, a positive Tajima's *D* indicated significantly more intermediate frequency nucleotide polymorphisms segregating at *MAM2* than expected (Tajima 1989). Second, a McDonald and Kreitman test (McDonald and Kreitman 1991) showed that too many amino acids segregated in *A. thaliana* MAM2, when compared to *MAMa* from *A. lyrata*. However, this *MAM2* polymorphism has no impact on glucosinolate identity (Kroymann et al. 2003; Benderoth et al. 2006), suggesting that non-neutrality at *MAM2* is caused by selection on glucosinolate quantity and not quality, a hypothesis that remains to be tested.

While gene duplication, biochemical neofunctionalization and positive selection account for the generation of metabolic diversity at the *MAM* locus, secondary gene deletions, gene conversion and balancing selection appear to maintain biochemical diversity. Of course, deletion of *MAM1* and conversion of *MAM1* into a *MAM2*-like sequence are both one-way streets from an archetypical *MAM2–MAM1–MAM3* configuration. Likewise, deletion of *MAM2* or conversion of *MAM2* into a *MAM1*-like gene both prevents future switches of chain-length phenotypes. Therefore, the Arabidopsis *MAM* locus appears to be in a process during which paralogous genes are being sorted among lineages, ultimately leading to plants that harbor, in combination with *MAM3*, either a *MAM1* or a *MAM2* gene.

Similarities and differences in chain-length variation between *Arabidopsis* and *Brassica*

In *Brassica oleracea* and other members of the genus, aliphatic glucosinolates can be generated from homo-, dihomo- and trihomomethionine (Velasco and Becker 2000). As in *A. thaliana*, there is natural variation for homo- versus dihomomethionine-derived glucosinolates among different accessions. However, in contrast to *A. thaliana*, accumulation of homomethionine- and accumulation of dihomomethionine-derived glucosinolates do not mutually exclude each other. Hence, *B. oleracea* accessions can produce aliphatic glucosinolates from homomethionine (3C), dihomomethionine (4C) or from homomethionine *and* dihomomethionine (3C + 4C). This biochemical polymorphism is caused by variation at two different genetic loci, *BoGSL-ELONG* and *BoGSL-PRO* (Magrath et al. 1994; Li et al. 2001). *BoGSL-ELONG* harbors a typical *MAM* gene with ten exons (Fig. 4), and, in addition, a *MAM* pseudogene (Li et al. 2001; Gao et al. 2005). The *B. oleracea MAM* gene is closely related to *MAM* genes from other Brassicaceae (Fig. 3). Genetic data indicate that *BoGSL-ELONG* is responsible for the generation of dihomomethionine-derived glucosinolates (Li and Quiros 2002). However, *B. oleracea MAM* does not belong to any of the *MAMa*, *b* or *c* subclades found in *A. thaliana* and close relatives (Fig. 3). The second *Brassica* locus, *BoGSL-PRO*, contains a gene with a typical 12-exon *IPMS* structure (Fig. 4), and phylogenetic analyses support a close phylogenetic relationship with other *IPMS* genes (Fig. 3). This gene is supposedly involved in the generation of homomethionine-derived glucosinolates (Gao et al. 2006). This suggests that (i) the gene duplication events leading to *MAMa*, *b* and *c* occurred after *Arabidopsis*, *Boechera* and *Brassica* diverged from a common ancestor, (ii) the ability to utilize 3C precursors for carbon chain elongation evolved independently in *Brassica* and *Arabidopsis* and (iii) the *Brassica* MAM function responsible for generating 3C glucosinolates evolved de novo from a *MAM* progenitor gene. To confirm these hypotheses it will be necessary to sample further *MAM* (and also *IPMS*) genes from close and distant *Arabidopsis* and *Brassica* relatives and to analyse the biochemical properties of the encoded proteins.

Conclusions

The composition of genes at the *MAM* locus varies between and within cruciferous species, causing substantial diversity in glucosinolate profiles. Different types of selection act on *MAM* gene family members, and different factors account for generating glucosinolate variability and for maintaining this diversity. Comparative analyses suggest that particular MAM substrate specificities have evolved repeatedly in different genera of glucosinolate-producing plants. The model plant *A. thaliana* has been invaluable for making progress in dissecting the genetic, functional and ecological basis of glucosinolate diversity. These discoveries have already facilitated the identification and cloning of genes from the glucosinolate-myrosinase system in cruciferous species for which an ab initio approach proves much more difficult, such as crop plants or wild Arabidopsis relatives. Further functional and evolutionary studies, involving additional species from the Brassicaceae but also from the Capparaceae, the second large family in the Capparales order that is capable of using methionine homologs as glucosinolate precursors, will help to better understand the complexity and evolutionary dynamics of variation in plant secondary metabolism.

Acknowledgements We thank John D'Auria for critical reading of the manuscript and gratefully acknowledge financial support from the Max Planck Society and the DFG (grants KR2237/2-1, KR2237/2-2 and KR2237/2-3 to JK).

References

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

Benderoth M, Textor S, Windsor AJ, Mitchell-Olds T, Gershenzon J, Kroymann J (2006) Positive selection driving diversification in plant secondary metabolism. Proc Natl Acad Sci USA 103:9118–9123

Campos de Quiros H, Magrath HCR, McCallum D, Kroymann J, Schnabelrauch D, Mitchell-Olds T, Mithen R (2000) Quantitative trait loci and the biochemical genetics of glucosinolate biosynthesis in *Arabidopsis thaliana*. Theor Appl Genet 101:429–437

Daxenbichler ME, Spencer GF, Carlson DG, Rose GB, Brinker AM, Powell RG (1991) Glucosinolate composition of seeds from 297 species of wild plants. Phytochemistry 30:2623–2638

de Kraker JW, Luck K, Textor S, Tokuhisa JG, Gershenzon J (2007) Two *Arabidopsis thaliana* genes (*IPMS1* and *IPMS2*) encode isopropylmalate synthase, the branchpoint step in the biosynthesis of leucine. Plant Physiol 143: 970–986

Fahey JW, Zalcmann AT, Talalay P (2001) The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. Phytochemistry 56:5–51

Falk KL, Vogel C, Textor S, Bartram S, Hick A, Pickett JA, Gershenzon J (2004) Glucosinolate biosynthesis: demonstration and characterization of the condensing enzyme of the chain elongation cycle in *Eruca sativa*. Phytochemistry 65:1073–1084

Field B, Cardon G, Traka M, Botterman J, Vancanneyt G, Mithen R (2004) Glucosinolate and amino acid biosynthesis in Arabidopsis. Plant Physiol 135:828–839

Gao M, Li G, McCombie RW, Quiros CF (2005) Comparative analysis of a transposon-rich *Brassica oleracea* BAC clone with its corresponding sequence in *A. thaliana*. Theor Appl Genet 111:949–955

Gao M, Li G, Potter D, McCombie RW, Quiros CF (2006) Comparative analysis of methylthioalkylmalate synthase (MAM) gene family and flanking DNA sequences in *Brassica oleracea* and *Arabidopsis thaliana*. Plant Cell Rep 25:592–598

Hagelstein P, Schultz G (1993) Leucine synthesis in spinach chloroplasts: partial characterization of 2-isopropylmalate synthase. Biol Chem Hoppe-Seyler 374:1105–1108

Haughn GW, Davin L, Giblin M, Underhill EW (1991) Biochemical genetics of plant secondary metabolites in *Arabidopsis thaliana*: the glucosinolates. Plant Physiol 97:217–226

Heidel A, Clauss M, Kroymann J, Savolainen O, Mitchell-Olds T (2006) Natural variation in *MAM* within and between populations of *Arabidopsis lyrata* determines glucosinolate phenotype. Genetics 173:1629–1636

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17:754–755

Keurentjes JJB, Fu J, de Vos CHR, Lommen A, Hall RD, Bino RJ, van der Plas LHW, Jansen RC, Vreugdenhil D, Koornneef M (2006) The genetics of plant metabolism. Nat Genet 38:842–849

Kliebenstein DJ (2008) A quantitative genetics and ecological model system: understanding the aliphatic glucosinolate biosynthesis network via QTLs. Phytochem Rev. doi: 10.1007/s11101-008-9102-8

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

Kliebenstein DJ, Lambrix VM, Reichelt M, Gershenzon J, Mitchell-Olds T (2001b) Gene duplication in the diversification of secondary metabolism: tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in Arabidopsis. Plant Cell 13:681–693

Kliebenstein DJ, Gershenzon J, Mitchell-Olds T (2001c) Comparative quantitative trait loci mapping of aliphatic,

indolic and benzylic glucosinolate production in *Arabidopsis thaliana* leaves and seeds. Genetics 159: 359–370

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

Kliebenstein DJ, West MAL, van Leeuwen H, Loudet O, Doerge RW, St. Clair DA (2006) Identification of QTLs controlling gene expression networks defined a priori. BMC Bioinformatics 7:308

Knoke B, Textor S, Gershenzon J, Schuster S (2008) Modelling aliphatic glucosinolate diversity in *Arabidopsis thaliana*. Phytochem Rev (this issue)

Koch MA, Haubold B, Mitchell-Olds T (2001) Molecular systematics of the Brassicaceae: evidence from coding plastidic *matK* and nuclear *Chs* sequences. Am J Bot 88:534–544

Kroymann J, Donnerhacke S, Schnabelrauch D, Mitchell-Olds T (2003) Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait locus. Proc Natl Acad Sci USA 100:14587–14592

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

Kroymann J, Textor S, Tokuhisa JG, Falk KL, Bartram S, Gershenzon J, Mitchell-Olds T (2001) A gene controlling variation in *Arabidopsis thaliana* glucosinolate composition is part of the methionine chain elongation pathway. Plant Physiol 127:1077–1088

Lambrix V, Reichelt M, Mitchell-Olds T, Kliebenstein DJ, Gershenzon J (2001) The *Arabidopsis* epithiospecifier protein promotes the hydrolysis of glucosinolates to nitriles and influences *Trichoplusia ni* herbivory. Plant Cell 13:2793–2807

Li G, Quiros CF (2002) Genetic analysis, expression and molecular characterization of *BoGSL-ELONG*, a major gene involved in the aliphatic glucosinolate pathway of Brassica species. Genetics 162:1937–1943

Li G, Riaz S, Goyal S, Abel S, Quiros CF (2001) Inheritance of three major genes involved in the synthesis of aliphatic glucosinolates in *Brassica oleracea*. J Amer Soc Hort Sci 126:427–431

Lister C, Dean C (1993) Recombinant inbred lines for mapping RFLP and phenotypic markers. Plant J 4:745–750

Magrath R, Bano F, Morgner M, Parkin I, Sharpe A, Lister C, Dean C, Lydiate D, Mithen RF (1994) Genetics of aliphatic glucosinolates. I. Side chain elongation in *Brassica napus* and *Arabidopsis thaliana*. Heredity 72:290–299

Martin W, Herrmann RG (1998) Gene transfer from organelles to the nucleus: how much, what happens, and why? Plant Physiol 118:9–17

McDonald JH, Kreitman M (1991) Adaptive evolution at the *Adh* locus in *Drosophila*. Nature 251:652–654

Mithen R, Clarke J, Lister C, Dean C (1995) Genetics of aliphatic glucosinolates. III. Side chain structure of aliphatic glucosinolates in *Arabidopsis thaliana*. Heredity 74:210–215

Nordborg M, Innan H (2002) Molecular population genetics. Curr Opin Plant Biol 5:69–73

Pfalz M, Vogel H, Mitchell-Olds T, Kroymann J (2007) Mapping of QTL for resistance against the crucifer specialist insect herbivore *Pieris brassicae* in a new Arabidopsis inbred line population, Da(1)-12 × Ei-2. PLoS ONE 2:e578

Purrington CB (2000) Costs of resistance. Curr Opin Plant Biol 3:305–308

Reichelt M, Brown PD, Schneider B, Oldham NJ, Stauber E, Tokuhisa J, Kliebenstein DJ, Mitchell-Olds T, Gershenzon J (2002) Benzoic acid glucosinolate esters and other glucosinolates from *Arabidopsis thaliana*. Phytochemistry 59:663–671

Rodman JE (1991a) A taxonomic analysis of glucosinolate-producing plants, part 1: phenetics. Syst Bot 16:598–618

Rodman JE (1991b) A taxonomic analysis of glucosinolate-producing plants, part 2: cladistics. Syst Bot 16:619–629

Rodman JE, Karol KG, Price RA, Sytsma KJ (1996) Molecules, morphology, and Dahlgren's expanded order Capparales. Syst Bot 21:289–307

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

Textor S, Bartram S, Kroymann J, Falk KL, Hick A, Pickett JA, Gershenzon J (2004) Biosynthesis of methionine-derived glucosinolates in *Arabidopsis thaliana*: recombinant expression and characterization of methylthioalkylmalate synthase, the condensing enzyme of the chain elongation cycle. Planta 218:1026–1035

Textor S, de Kraker JW, Hause B, Gershenzon J, Tokuhisa G (2007) MAM3 catalyzes the formation of all aliphatic glucosinolate chain lengths in Arabidopsis. Plant Physiol 144:60–71

Tian D, Traw MB, Chen J, Kreitman M, Bergelson (2003) Pleiotropic fitness cost of *Rpm1* resistance in *Arabidopsis thaliana*. Nature 423:74–77

Velasco L, Becker HC (2000) Variability for seed glucosinolates in a germplasm collection of the genus *Brassica*. Genet Res Crop Evol 47:231–238

Wang S, Basten CJ, Zeng ZB (2001–2004) Windows QTL cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC. www.statgen.ncsu.edu.qtlcart/WQTLCart.htm

Wentzell AM, Rowe HC, Hansen BG, Ticconi C, Halkier BA, Kliebenstein DJ (2007) Linking metabolic QTLs with network and *cis*-eQTLs controlling biosynthetic pathways. PLoS Genet 3:e162

Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 18:691–699

Windsor AJ, Reichelt M, Figuth A, Svatoš A, Kroymann J, Kliebenstein DJ, Gershenzon J, Mitchell-Olds T (2005) Geographic and evolutionary diversification of glucosinolates among near relatives of *Arabidopsis thaliana* (Brassicaceae). Phytochemistry 66:1321–1333

Yang YW, Lai KN, Tai PY, Li WH (1999) Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between *Brassica* and other angiosperm lineages. J Mol Evol 48:597–604

Zhang Z, Ober JA, Kliebenstein DJ (2006) The gene controlling the quantitative trait locus *epithiospecifier modifier1* alters glucosinolate hydrolysis and insect resistance in *Arabidopsis*. Plant Cell 18:1524–1536

Analysis of Fitness Consequences of *Plutella*-Regulated Genes in Arabidopsis

Marina Pfalz, Heiko Vogel, Juergen Kroymann*

Max Planck Institute for Chemical Ecology, Jena, Germany

***To whom correspondence should be addressed:** Juergen Kroymann, Genetics & Evolution, Max Planck Institute for Chemical Ecology, Hans-Knoell-Str. 8, 07745 Jena, Germany. Fax: +49 3641 571498. E-mail: kroymann@ice.mpg.de

Keywords: Transcript Profiling, Arabidopsis, Plant-Insect Interactions, *Plutella xylostella*, T-DNA Insertion Lines

ABSTRACT

Transcript profiling experiments with Arabidopsis whole-genome microarrays have shown that hundreds or even thousands of plant genes change their expression in response to damage inflicted by herbivorous insects. The biological significance of these transcriptome changes is largely unknown. In particular, fitness consequences of differential gene regulation have not been investigated systematically. Here, we use Arabidopsis T-DNA and transposon insertion lines to test whether genes that are differentially regulated upon herbivory by larvae from the crucifer specialist insect herbivore *Plutella xylostella* contribute to defense against this insect. Our results indicate that the majority of genes have no impact on plant resistance against *Plutella*. Moreover, fitness consequences of T-DNA insertions in several genes are contrary to our expectations based on gene expression patterns, suggesting that *P. xylostella* larvae may be able to manipulate Arabidopsis plants to their own benefit.

INTRODUCTION

Plants have evolved a broad arsenal of defenses to ward off herbivorous insects [1-3]. These defenses include preformed morphological or biochemical barriers which are expressed regardless of whether a plant is being attacked or not. Other defenses are inducible, and are switched on only when a plant is challenged by herbivorous insects. Several methods have been employed to identify genes that contribute to Arabidopsis resistance against herbivorous insects, including quantitative trait locus (QTL) mapping and transcript profiling.

QTL mapping with Arabidopsis recombinant inbred lines has been successful in identifying genetic loci and, ultimately, genes that contribute to resistance against generalist insect herbivores such as *Trichoplusia ni* or *Spodoptera spec.* [5-9]. These herbivores feed not only on Arabidopsis and relatives but also on members from distantly related plant families. Most QTL for generalists were found to co-localize with QTL that control variation in the glucosinolate-myrosinase system, an activated defense system in cruciferous plants [9]. In brief, this system consists of precursor molecules, glucosinolates, and corresponding breakdown enzymes, myrosinases. In intact tissue, glucosinolates and myrosinases are stored separate from each other but upon tissue disruption, they mix and myrosinase-catalyzed glucosinolate hydrolysis leads to the formation of glucosinolate breakdown products which are toxic or deterrent to herbivorous insects [10]. Insects specialized on plants from the crucifer family, including *Plutella xylostella* and *Pieris spec.*, have evolved counteradaptations that render the glucosinolate-myrosinase system ineffective. These counteradaptations prevent the formation of glucosinolate-hydrolysis products [11] or divert hydrolysis towards (almost) non-toxic products [12]. Consequently, QTL for the glucosinolate-myrosinase system do not constitute QTL for resistance against specialist insects. Nonetheless, there is considerable natural genetic variation for resistance against specialist insects among Arabidopsis accessions, but the genetic basis of defense against these specialists is not yet known [13].

Transcript profiling with Arabidopsis whole-genome microarrays provides a direct means to identify genes that are differentially regulated upon herbivory. Thus, this method appears to be particularly suited to investigate inducible defenses. Indeed, transcript profiling experiments have revealed that hundreds or even thousands of genes respond with altered transcript levels to various insects, including *Plutella xylostella* [14-17]. However, it is unknown which role these

genes serve in the plant response to herbivory. In particular, it is not clear whether differential gene expression contributes to plant fitness during the encounter with herbivorous insects.

Here, we investigate fitness effects of a subset of genes that are differentially regulated upon herbivory by *Plutella xylostella* larvae in herbivory screens with the same insect. In these screens we compare the damage caused by *P. xylostella* larvae on available Arabidopsis T-DNA insertion or transposon-tagged lines [18, 19] with wildtype lines obtained from the same seed material. We mostly focus on genes that are induced upon *Plutella* herbivory but also include several lines with T-DNA insertions in downregulated genes. In general, we expect an insertion of foreign DNA into genes that are upregulated upon *Plutella* herbivory to cause a decrease in plant performance while insertions into downregulated genes should lead to an increase in plant fitness, if up- or downregulation of the respective genes contribute to plant defense against *P. xylostella*.

Our results are surprising because most insertions do not alter plant resistance against *P. xylostella* larvae. In the few cases where we observe an impact on plant fitness, effects are in an unexpected direction, suggesting that *P. xylostella* larvae may be manipulating plant responses to their own benefit.

MATERIALS AND METHODS

Plant and Insect Culture Conditions

Arabidopsis SALK T-DNA insertion lines with a Columbia (Col) background [18] were obtained from the Nottingham Arabidopsis Stock Centre (NASC), Nottingham (UK), and the Arabidopsis Biological Resource Centre (ABRC), Columbus (Ohio). A transposon-tagged line with a Nossen (No-0) background was obtained from RIKEN, Genomic Science Center (GSC), Japan [19]. Selected genes are described in Table 1, T-DNA insertion and transposon-tagged lines are listed in Table 2, together with primers used for tests of the presence/absence of insertions and for semiquantitative RT-PCR (below).

Plants were grown in temperature and humidity-controlled growth chambers as described in [13]. From each SALK and RIKEN line, approximately 30 seeds were planted and screened for the presence/absence of the T-DNA or transposon insertion in the gene of interest. In general, we aimed at obtaining 3 different homozygous mutant lines and 2 different homozygous wildtype lines per SALK or RIKEN line.

Plutella xylostella (diamond backmoth) eggs were originally provided from the New York State Agricultural Experimental Station (Geneva, NY), and a colony was maintained at the Max Planck Institute for Chemical Ecology in Jena. All larvae were reared for 6 days on artificial diet according to published procedures [20] and were starved for 6 hours before the onset of herbivory experiments.

Herbivory Screens

Herbivory assays were carried out in 96-celled flats with 3-week old plants as described in [13]. Mutant and wildtype lines were completely randomized, with 100 to 300 plants per experiment and gene. Prior to herbivory experiments, the diameter of each plant was measured. Subsequently, one larva was placed into the center of each plant rosette. During experiments, larvae were allowed to move freely. After 48 hours the leaf area that was consumed by larvae was assessed visually and evaluated on an artificial scale.

Genotyping of SALK T-DNA Insertion and RIKEN Transposon-Tagged Lines

DNA was extracted from leaf material of 3 week old plants as described in [21]. Plant genotypes were determined by PCR. PCR reactions were performed in general with ca. 30 ng DNA, 1 x PCR buffer (Qiagen, Germany), 4 nmol of of each dNTP, 1.25 pmol of each of both primers and 0.15 U *Taq* DNA polymerase (Qiagen). The following PCR program was applied: Initial denaturation for 2 min at 94°C, followed by 35 cycles of 20 s at 94 °C, 20 s at 55 °C, 30 s at 72 °C, and a final extension at 72 °C for 2 min. Each SALK T-DNA insertion line was tested with two primer combinations. The first combination consisted of forward and reverse gene-specific primers flanking the location of the T-DNA insertion. This primer combination amplified PCR products from wildtype alleles. The second combination consisted of a T-DNA specific primer, LBb1 (5'-TCAAACAGGATTTTCGCCTGCT-3'), and either the forward or the reverse gene-specific primer, depending on the orientation of the insertion [18]. This second primer combination yielded PCR products for mutant alleles. Gene-specific primer sequences are listed in Table 2. For line PST18451, a transposon-tagged line from the RIKEN Arabidopsis transposon mutant collection [19], a similar approach was used with a transposon-specific primer, Ds3-2a (5'-CCGGATCGTATCGGTTTTCG-3'), instead of primer LBb1.

RT-PCR

To confirm transcription patterns of *Plutella*-regulated genes, five 3rd instar larvae were placed on 3 week old Arabidopsis plants and allowed to feed for 8 hours. Control plants were grown under the same environmental conditions but did not receive insects. Rosette leaves were harvested from insect-treated and control plants and immediately frozen in liquid nitrogen. Total RNA was extracted with Trizol (Invitrogen) according to the manufacturer's instructions. After DNAse (Ambion) digestion RNA was purified with the RNeasy MinElute Cleanup Kit (Qiagen). First strand cDNA synthesis was carried out with Superscript III First-Strand Synthesis Super mix (Invitrogen) from 1 µg total RNA in a volume of 20 µl. For each sample three parallel reactions were carried out, mixed and diluted to a total volume of 600 µl with distilled water. This dilution served as a template for all subsequent RT-PCRs. Reaction mixes contained 1 µl cDNA, 5 µl 10 x PCR buffer (Qiagen), 1 µl dNTP mix (10 mM), 7 pmol of each of both primers, and 2 U Taq DNA polymerase (Qiagen) in a volume of 50 µl. All PCR primers were chosen such that reaction products had a size of approximately 300 bp. Cycling conditions were 94 °C for 3 min,

followed by 37 cycles of 94 °C for 15 s, 55 °C for 15 s and 72 °C for 20 s. After 22, 27 and 32 cycles, respectively, the reaction was interrupted and 10 µl from the reaction were stored on ice. Subsequently PCR products obtained after 22, 27, 32, and 37 PCR cycles were analyzed on 1.5 % agarose gels and photographed. A similar protocol was used to investigate the effect of T-DNA (or transposon) insertions on the expression of target genes.

Southern Hybridizations

The number of T-DNA insertions was analyzed with Southern blot hybridization. Total genomic DNA was extracted from rosette leaves of 4 week old plants using the Genomic DNA buffer set (Qiagen) according to the manufacturer's instructions. 2 µg *Bcl*I-digested DNA were separated by gel electrophoresis and blotted onto positively charged Biodyne Plus nylon membranes (Biodyne). A ca. 300 bp T-DNA specific PCR product was amplified with primers T-DNA_L1F (5'-TCCAGTTTGGAACAAGAGTCCA-3') and T-DNA_L1R (5'-CCTGGCGTTACCCAACTTAATC-3') to serve as a probe. Probe labelling and hybridization procedure with the ECL Direct Nucleic Acid Labeling and Detection Systems (Amersham Biosciences) followed the manufacturer's instructions.

DNA Walking

A DNA walking method (DNA Walking SpeedUp Premix Kit, Seegene) was applied according to the manufacturer's instructions to determine the genomic location of previously unreported T-DNA insertions in SALK_082089. This method involves three PCR steps using a combination of unique DNA walking primers (Seegene) and different T-DNA or Ti-Plasmid specific primers (Table 3) for each round of amplification. Purified product from the first PCR step serves as a template in the second step, and purified products from the second step as a template for the third PCR step. PCR products from the third round were run on agarose gels. All visible bands were extracted, cloned into a TOPO TA cloning vector (pCR 2.1 Cloning Kit, Invitrogen) and sequenced.

Statistical Analyses

Systat version 10 (SPSS Inc.) was used for statistical analyses of herbivory screens with mixed model ANOVA. Genotype was treated as a fixed effect, and flat and flat×genotype interaction as random effects to obtain mean squares (MS) for genotype and flat×genotype. In general, the following model was used: HERBIVORY = CONSTANT + COLUMN + ROW + PLANT DIAMETER + FLAT + GENOTYPE + FLAT×GENOTYPE. COLUMN and ROW are variables to control for position effects. The variable GENOTYPE compares mutant *versus* wildtype genotypes, FLAT accounts for variation between flats within an experiment, and PLANT DIAMETER controls for variation in plant size. *F* ratios were calculated as $MS_{GENOTYPE}/MS_{FLAT\times GENOTYPE}$. In several cases, herbivory screens with the same T-DNA insertion line were repeated at a later date. In these cases, an additional variable EXPERIMENT was included in the ANOVA model to account for variation between experiments, and used a nested variable FLAT(EXPERIMENT) instead of FLAT. Finally, analogous models were applied to investigate the effect of the T-DNA insertions on plant biomass accumulation.

RESULTS

T-DNA Insertion Lines

Initially, we screened progeny from more than 80 T-DNA insertion lines for the presence/absence of T-DNA insertions in candidate genes that were differentially regulated upon *Plutella* herbivory. Our goal was to obtain from each line progeny carrying T-DNA insertions in both alleles and progeny with homozygous wildtype alleles to serve as adequate controls in our herbivory experiments. For more than half of the lines we were not able to identify T-DNA insertions or to obtain homozygous mutants, or we experienced other problems. We excluded these lines from further analyses. Thus, we finally worked with 35 T-DNA insertion lines representing 33 genes (Table 1). These genes encode a diverse set of proteins, including transcription factors, pathogenesis-related proteins, proteins involved in general defense responses and stress, in the metabolism of amino acids and aromatic compounds, and several proteins with unknown function. The majority of insertions was located in the coding sequence

(CDS) (14) or in the 3'UTR (13), 2 insertions were located in the promoter or in the 5'UTR, 3 in internal introns, and 1 T-DNA was inserted 2 bases 3' of the 3'UTR (Figure 1).

Next, we tested expression patterns for a subset of these genes with RT-PCR (Figure 2). We compared plants infested with 5 *Plutella* larvae for 9 hours *versus* controls, each with two biological replicates. In all cases, we were able to confirm results from previous microarray experiments.

Finally, we investigated the effects of T-DNA or transposon insertions on the expression of target genes. We analyzed *Plutella*-treated and unmanipulated control plants in comparison to wildtype plants grown under the same experimental conditions. 16 out of 25 tested lines showed a molecular phenotype (Figure 3), 9 lines yielded ambiguous results and 10 lines remained untested. Most lines with insertions of foreign DNA into the coding sequence and many lines with insertions into the 3'UTR had less transcript than wildtypes. In one case (SALK_104816) the PCR product was larger in insertion lines than in wildtypes, possibly caused by a splice defect that resulted in the incorporation of an intron. In the single case where we tested a line with an insertion within an intron (SALK_006603), no PCR product was visible in the mutant. Lastly, in SALK_030723, with a T-DNA insertion in the promoter, mutant lines deviated from the wildtype transcript pattern, with a strong reduction of the transcript level in response to *Plutella*-treatment compared to wildtype.

Herbivory Screens

For 33 genes which displayed differential regulation in response to *Plutella* herbivory, we analyzed the effects of T-DNA (or transposon) insertions compared to plants with wildtype alleles. We tested for differences in plant performance against *P. xylostella* larvae but investigated also whether insertions influenced plant biomass accumulation. We found a significant effect on *P. xylostella* larval herbivory only for three genes, At2g23350 ($N = 247$; $df = 1,2$; $F = 28.52$; $p < 0.05$), At2g39050 ($N = 966$; $df = 1,10$; $F = 12.25$; $p < 0.01$) and At3g04720 ($N = 1094$; $df = 1,12$; $F = 6.57$; $p < 0.025$), coding for a putative polyadenylate-binding protein, a protein with an unknown function, and a hevein-like protein precursor, respectively (Table 4). For At2g23350, plants with insertions in both alleles experienced approximately 10 % less damage than plants with wildtype alleles. For the other two genes, plants with insertions were ca.

10 % and 15 % more resistant than plants without T-DNA insertions. We tried to confirm these results with other lines available from the stock centers. For At2g23350, we found no suitable T-DNA insertion or transposon-tagged line. Likewise, for At2g39050 we obtained a SAIL line with an appropriate insertion but in a sample of ca. 30 seeds we did not identify any homozygous mutants. Finally, for At3g04720 we got another SALK line (SALK_130680) with an insertion in the 3'UTR and a third mutant line (PST18451) with a transposon inserted into the open reading frame. Both lines had no genotype effect in herbivory screens. Hence, we could not confirm the genotype effects found in our initial screens with independent experiments.

We further analyzed SALK_082089. With Southern blot hybridizations we identified a second insertion in this line and used DNA walking to identify the location of this second insertion. After several rounds of walking we were still obtaining DNA completely identical to agrobacterial Ti-Plasmid sequences. Therefore, we changed our strategy and utilized primers designed to bind at various locations of the Ti-Plasmid. All amplifications were successful but yielded only Ti-Plasmid sequences, indicating that the second insertion in SALK_082089 consisted not only of T-DNA but of a large proportion of the Ti-Plasmid, and perhaps even of concatenated Ti-Plasmid sequences.

Furthermore, we routinely noted the diameter of each plant before our herbivory assays to account for differences in plant size, enabling us to also investigate the impact of T-DNA or transposon insertions on plant biomass accumulation. The majority of insertions had no effect on plant biomass, indicating that T-DNA insertions alone are not costly. However, for 8 lines we found significant genotype effects on plant biomass accumulation (Table 4). In most cases effect strengths were well below 10 %, except for SALK_036887 where the difference between wildtypes and mutants was nearly 13 %. In some cases, wildtype lines performed better than mutants (At2g14750, At2g20340, At2g38870, and At3g12500), in other cases mutant lines accumulated more biomass than wildtypes (At2g32150, and At2g38470). In one case (At4g04720) genotype effects on plant biomass were reversed with the genetic background. In the Col-0 background (SALK_082089) mutants performed better than wildtypes, whereas in the No-0 background (PST18451) mutant plants grew on average to a smaller size than wildtype plants. Both SALK_082089 and PST18451 had foreign DNA inserted into an exon. In a third line (SALK_130680) with a T-DNA insertion in the 3'UTR, we found no genotype effect on plant biomass accumulation.

DISCUSSION

In this study we investigated fitness effects of genes that are differentially regulated upon herbivory by the crucifer specialist *P. xylostella*. We compared the damage caused by larvae on Arabidopsis lines with T-DNA (or transposon) insertions to lines with wildtype alleles from the same seed material. The genes represented a broad spectrum of functional categories (Table 1) and also a broad range of transcript levels, both base line and upon *P. xylostella* herbivory (Figure 2). Nearly half of the insertions were located in the CDS, but also lines with insertions in introns or in the 3'UTR had significant effects on plant biomass accumulation, indicating that gene function was effectively disrupted in the majority of T-DNA insertion lines. Altogether, we included more than 11,000 plants and *Plutella* larvae in our herbivory assays.

Only three out of 33 genes displayed a fitness effect in *Plutella* herbivory experiments. By contrast, 8 lines (representing 7 genes) showed significant differences in plant biomass accumulation when T-DNA insertion lines were compared to wildtypes. However, for these three genes we were not able to replicate fitness differences upon *Plutella* herbivory with independent insertion lines, in part because other suitable lines were not available. In the case of At3g04720, coding for a hevein-like gene, we are not entirely certain that the effect that we observed with SALK_082089 was caused by the disruption of the gene At3g04720 itself. This line contained a second insertion, consisting of a very large proportion of the Ti-Plasmid and perhaps even of concatenated Ti-Plasmid sequences, that prevented us from identifying the location of this insertion. Nonetheless, the number of lines which showed a difference in resistance to *Plutella* was altogether remarkably low. We, therefore, conclude that the majority of genes that respond to *P. xylostella* herbivory with an alteration in transcript levels do not contribute to plant defense against *Plutella*.

Which other purpose could these genes serve? Several of the investigated genes respond also to other triggers, such as pathogen infection (*e.g.*, At3g04720, At3g12500, At3g57260 – these genes are known to encode pathogenesis related proteins PR4, PR3 and PR2, respectively) – or drought (*e.g.*, At1g54100, At2g47180). Hence, increased expression of these genes may be important to minimize secondary harmful effects caused by a potential pathogen infection or by water loss through the wounded tissue. In other cases differential gene regulation could be a chance effect, due to the presence of regulatory elements that respond to the complex network of signals that is triggered upon herbivory, but without serving any purpose in the context of insect

herbivory. However, in all three lines which showed a difference in resistance to *Plutella* larvae, effects were in an unexpected direction. At2g39050 and At3g04720 were induced upon *Plutella* herbivory, but the corresponding lines with T-DNA insertions, SALK_037116 and SALK_082089, were more resistant than wildtypes, contrary to expectations. Likewise, expression of At2g23350 was reduced after *Plutella* herbivory, but in SALK_113383 progeny, plants with wildtype alleles were more resistant than plants with T-DNA insertions in the open reading frame. Hence, plant performance did apparently not benefit from a downregulation of At2g23350 in response to *Plutella* herbivory. On the contrary, differential regulation of these genes upon herbivory by *Plutella* rendered plants more susceptible to the larvae, suggesting that *Plutella* larvae may be able to manipulate the Arabidopsis transcriptome in a way that serves their own benefit. Indeed, several other studies provide evidence for herbivore manipulation of induced plant defenses. In *Helicoverpa zea* larvae a glucose oxidase was identified as a principle salivary enzyme which suppressed induced defenses in *N. attenuata* and prevented the induction of nicotine [22]. The silverleaf whitefly was found to be able to repress jasmonic acid-regulated defenses that deter nymph development [23]. Finally, jasmonic acid-mediated upregulation of defense genes was suppressed in response to *P. rapae* [24]. These findings indicate that the adaptation of specialist herbivores to their host plants is more substantial than previously thought. Apparently, specialist insect herbivores have not only evolved counteradaptations against preformed activated defenses such as the glucosinolate-myrosinase system but also appear to possess counteradaptations against induced defenses, thus adding a further level of complexity to interactions between specialist insect herbivores and their host plants.

REFERENCES

1. Stotz HU, Pittendrigh BR, Kroymann J, Weniger K, Fritsche J, Bauke A, Mitchell-Olds T (2000) Induced plant defense responses against chewing insects. Ethylene signaling reduces resi stance of Arabidopsis against Egyptian worm but not diamondback moth. Plant Physiol 124: 1007-1017.

2. Walling LL (2000) The myriad plant responses to herbivores. J Plant Growth Reg 19: 195-216.

3. Van Poecke RMP (2007). Arabidopsis-Insect Interactions. In CR Somerville, EM Meyerowitz, eds, The Arabidopsis Book. American Society of Plant Biologists, Rockville, MD, doi: 10.1199/tab.0107, http://www.aspb.org/publications/arabidopsis/

4. Jander G, Cui J, Nhan B, Pierce NE, Ausubel FM (2001) The *TASTY* locus on chromosome 1 of Arabidopsis affects feeding of the insect herbivore *Trichoplusia ni*. Plant Physiol 126: 890-898.

5. Lambrix V, Reichelt M, Mitchell-Olds T, Kliebenstein DJ (2001) The Arabidopsis epithiospecifier protein promotes the hydrolysis of glucosinolates to nitriles and influences *Trichoplusia ni* herbivory. Plant Cell 13: 2793-2807.

6. Kliebenstein DJ, Pedersen D, Barker B, Mitchell-Olds T (2002) Comparative analysis of quantitative trait loci controlling glucosinolates, myrosinase and insect resistance in *Arabidopsis thaliana*. Genetics 161: 325-332.

7. Kroymann J, Donnerhacke S, Schnabelrauch D, Mitchell-Olds T (2003) Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait locus. Proc Natl Acad Sci USA 100: 14587-14592.

8. Zhang Z, Ober JA, Kliebenstein DJ (2006) The gene controlling the quantitative trait locus *EPITHIOSPECIFIER MODIFIER1* alters glucosinolate hydrolysis and insect resistance in *Arabidopsis*. Plant Cell 18: 1524-1536.

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

10. Halkier BA, Gershenzon J (2006) Biology and biochemistry of glucosinolates. Annu Rev Plant Biol 57: 303-333.

11. Ratzka A, Vogel H, Kliebenstein DJ, Mitchell-Olds T, Kroymann J (2002) Disarming the mustard oil bomb. Proc Natl Acad Sci USA 99: 11223-11228.

12. Wittstock U, Agerbirk N, Stauber EJ, Olsen CE, Hippler M, Mitchell-Olds T, Gershenzon J, Vogel H (2004) Successful herbivore attack due to metabolic diversion of a plant chemical defense. Proc Natl Acad Sci USA 101: 4859-4864.

13. Pfalz M, Vogel H, Mitchell-Olds T, Kroymann J (2007) Mapping of QTL for resistance against the crucifer specialist herbivore *Pieris brassicae* in a new Arabidopsis inbred line population, Da(1)-12×Ei-2. PLoS ONE 2: e578.

14. Reymond P, Bodenhausen N, Van Poecke RMP, Krishnamurthy V, Dicke M, Farmer EE (2004) A conserved transcript pattern in response to a specialist and a generalist herbivore. Plant Cell 16: 3132-3147.

15. De Vos M, Van Oosten V, Van Poecke R, Van Pelt J, Pozo MJ, Mueller MJ, Buchala AJ, Métraux JP, Van Loon LC, Dicke M, Pieterse CMJ (2005) Signal signature and transcriptome changes of Arabidopsis during pathogen and insect attack. Mol Plant Microbe Interact 18: 923-937.

16. Vogel H, Kroymann J, Mitchell-Olds T (2007) Different transcript patterns in response to specialist and generalist herbivores in the wild Arabidopsis relative *Boechera divaricarpa*. PLoS ONE 2: e1081.

17. Ehlting J, Chowrira SG, Mattheurs N, Aeschliman DS, Arimura GI, Bohlmann J (2008) Comparative transcriptome analysis of *Arabidopsis thaliana* infested by diamond back moth (*Plutella xylostella*) larvae reveals signatures of stress response, secondary metabolism and signaling. BMC Genomics 9: 154.

18. Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, Gadrinab C, Heller C, Jeske A, Koesema E, Meyers CC, Parker H, Prednis L, Ansari Y, Choy Nathan, Deen H, Geralt M, Hazari N, Hom E, Karnes M, Mulholland C, Ndubaku R, Schmidt I, Guzman P, Aguilar-Henonin L, Schmid M, Weigel D, Carter DE, Marchand T, Risseeuw E, Brogden D, Zeko A, Crosby WL, Berry CC, Ecker JR (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science 301: 653–657.

19. Kuromori T, Hirayama T, Kiyosue Y, Takabe H, Mizukado S, Sakurai T, Akiyama K, Kamiya A, Ito T, Shinozaki K (2004) A collection of 11,800 single-copy *Ds* transposon insertion lines in Arabidopsis. Plant J 37: 897-905.

20. Shelton AM, Cooley RJ, Kroening MK, Wilsey WT, Eigenbrode SD (1991) Comparative analysis of two rearing procedures for diamondback moth, *Plutella xylostella* (Lepidoptera: Plutellidae). J Entomol Science 26: 17-26.

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

22. Musser RO, Hum-Musser SM, Eichenseer H, Peiffer M, Ervin G, Murphy JB, Felton GW (2002) Herbivory: caterpillar saliva beats plant defenses. Nature 416: 599-600.

23. Zarate SI, Kempema LA, Walling LL (2007) Silverleaf whitefly induces salicylic acid defenses and suppresses effectual jasmonic acid defenses. Plant Physiol: 143: 866-875.

24. Bodenhausen N, Reymond P (2007) Signaling pathways controlling induced resistance to insect herbivores in Arabidopsis. Mol Plant Microbe Interact: 20: 1406-1420.

Figure Legends

Figure 1. Gene Structures and T-DNA (or Transposon) Insertions. Gene annotations were taken from the Arabidopsis Information Resource (http://www.arabidopsis.org). Thick horizontal lines correspond to exons (including 5' and 3'UTR), thin lines to introns or noncoding sequences. Thin vertical lines show positions of start and stop codons. Note that for At5g44430 no information on UTRs is available. Green vertical lines indicate the position of T-DNA or transposon insertions. Small red arrows correspond to the position of primer LBb1 in T-DNA or its equivalent in transposons (Table 2), and show the orientation of the insertion.

Figure 2. Analysis of Transcript Levels in Control and Herbivory-Treated Plants with RT-PCR. Three sets of genes were investigated, genes upregulated upon *Plutella* herbivory, downregulated genes, and control genes. Two biological replicates were used in herbivory experiments (H) and as control plants (C). The same samples were used for all RT-PCRs. Numbers in brackets indicate the number of PCR cycles (22, 27, 32 or 37) which gave satisfying results for the various genes. Control genes At1g34030, At4g05320 and At5g09810 code for ribosomal protein 18B (RPS18B), ubiquitin 10 (UBQ10) and actin 7 (ACT7), respectively.

Figure 3. Analysis of T-DNA Insertion Lines with RT-PCR. Transcript levels of target genes were analyzed in T-DNA insertion (or transposon-tagged) lines and wildtype plants after *Plutella* herbivory (H) and in unmanipulated controls (C) to investigate the effect of insertions on gene expression. UBQ10 and ACT7 served as loading controls. Numbers in brackets indicate the number of PCR cycles (27, 32 or 37). Because cDNA quantity was limited, loading controls for wildtype were applied at informative positions of the agarose gels and are indicated by lower-case letters (a – h). Note that transposon-tagged and control plants for PST18451 have a No-0 background, whereas all other plants have a Col-0 background. Note also that the At1g54020 PCR product is larger in insertion than in wildtype lines (indicated by arrows).

Figure 1

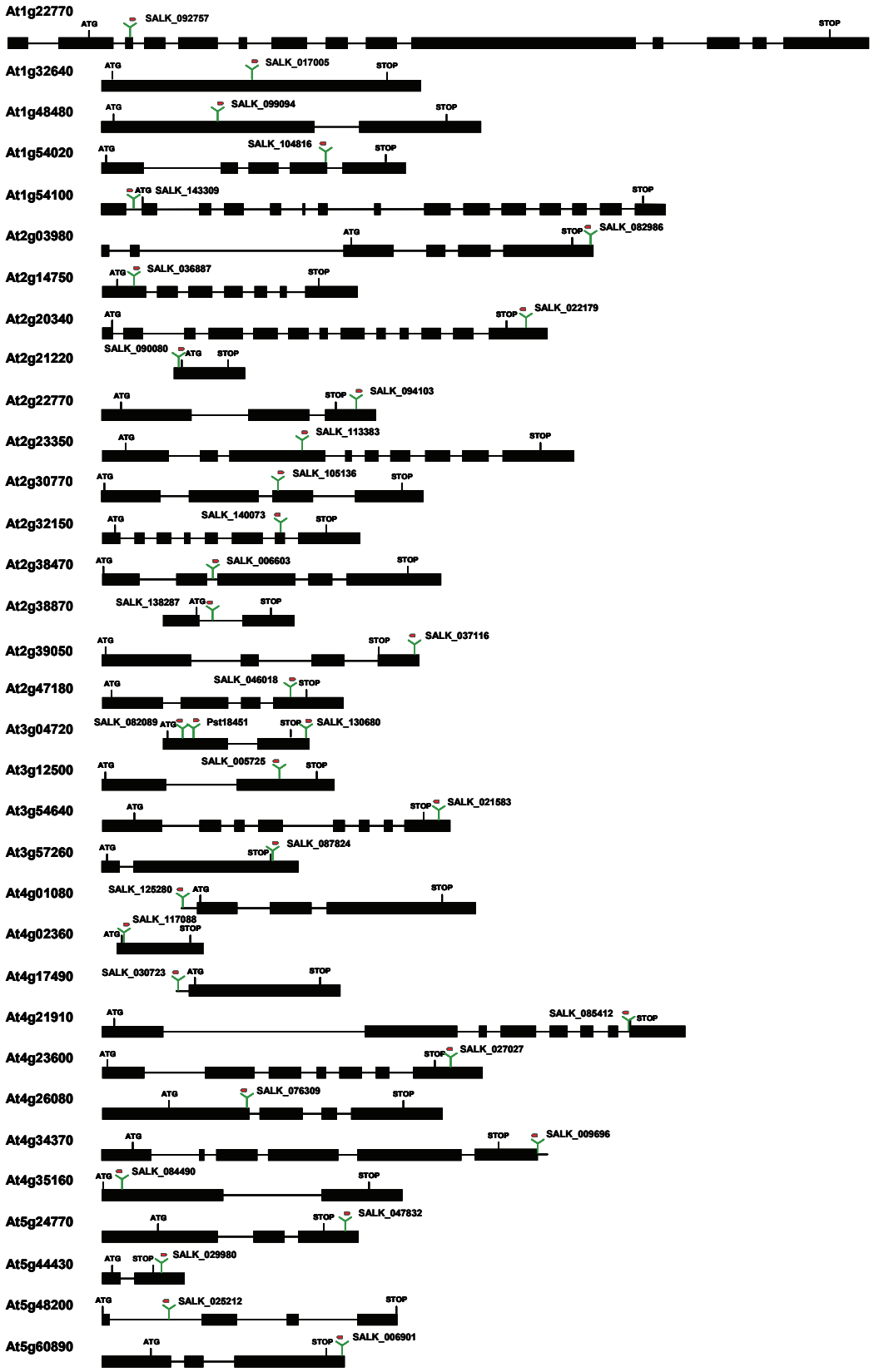
At1g22770
ATG
SALK_092757
STOP
At1g32640
ATG
SALK_017005
STOP
At1g48480
ATG
SALK_099094
STOP
At1g54020
ATG
SALK_104816
STOP
At1g54100
ATG
SALK_143309
STOP
At2g03980
ATG
STOP
SALK_082986
At2g14750
ATG
SALK_036887
STOP
At2g20340
ATG
STOP
SALK_022179
At2g21220
SALK_090080
ATG
STOP
At2g22770
ATG
STOP
SALK_094103
At2g23350
ATG
SALK_113383
STOP
At2g30770
ATG
SALK_105136
STOP
At2g32150
ATG
SALK_140073
STOP
At2g38470
ATG
SALK_006603
STOP
At2g38870
SALK_138287
ATG
STOP
At2g39050
ATG
STOP
SALK_037116
At2g47180
ATG
SALK_046018
STOP
At3g04720
SALK_082089
ATG
Pst18451
STOP
SALK_130680
At3g12500
ATG
SALK_005725
STOP
At3g54640
ATG
STOP
SALK_021583
At3g57260
ATG
STOP
SALK_087824
At4g01080
SALK_125280
ATG
STOP
At4g02360
SALK_117088
ATG
STOP
At4g17490
SALK_030723
ATG
STOP
At4g21910
ATG
SALK_085412
STOP
At4g23600
ATG
STOP
SALK_027027
At4g26080
ATG
SALK_076309
STOP
At4g34370
ATG
STOP
SALK_009696
At4g35160
ATG
SALK_084490
STOP
At5g24770
ATG
STOP
SALK_047832
At5g44430
ATG
STOP
SALK_029980
At5g48200
ATG
SALK_025212
STOP
At5g60890
ATG
STOP
SALK_006901

Figure 2

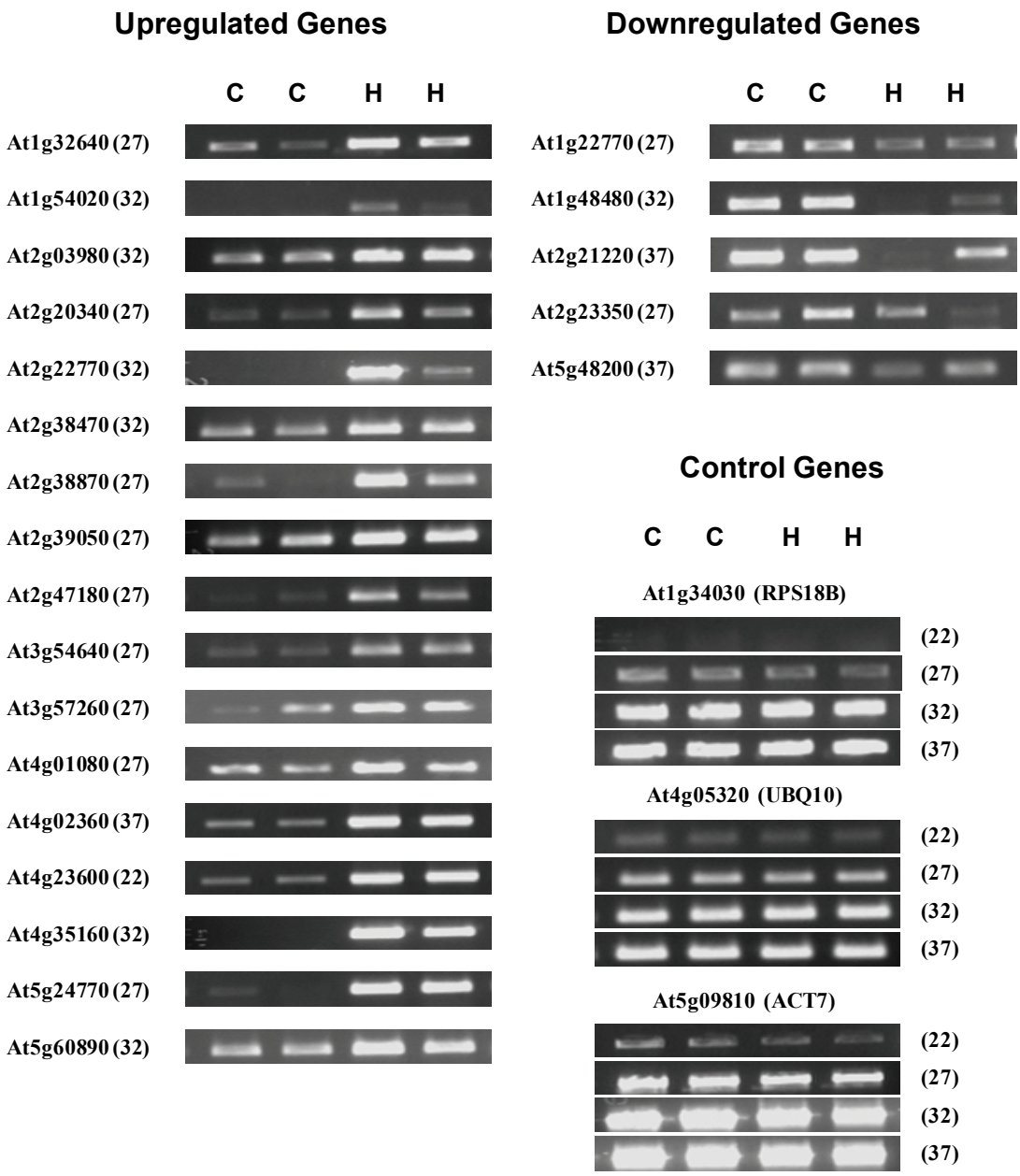
Upregulated Genes
C C H H
At1g32640 (27)
At1g54020 (32)
At2g03980 (32)
At2g20340 (27)
At2g22770 (32)
At2g38470 (32)
At2g38870 (27)
At2g39050 (27)
At2g47180 (27)
At3g54640 (27)
At3g57260 (27)
At4g01080 (27)
At4g02360 (37)
At4g23600 (22)
At4g35160 (32)
At5g24770 (27)
At5g60890 (32)
Downregulated Genes
C C H H
At1g22770 (27)
At1g48480 (32)
At2g21220 (37)
At2g23350 (27)
At5g48200 (37)
Control Genes
C C H H
At1g34030 (RPS18B)
(22)
(27)
(32)
(37)
At4g05320 (UBQ10)
(22)
(27)
(32)
(37)
At5g09810 (ACT7)
(22)
(27)
(32)
(37)

Figure 3

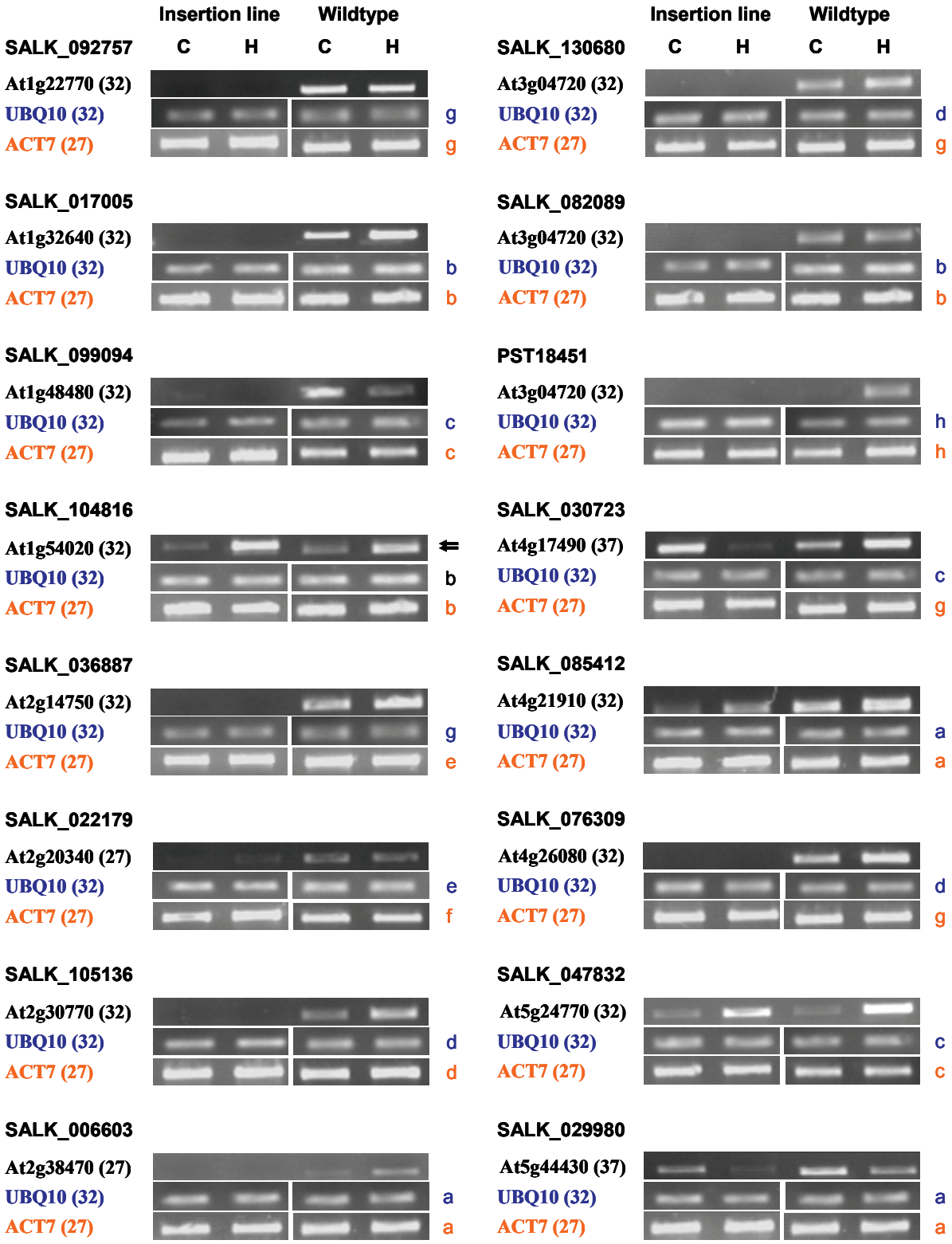
Insertion line
Wildtype
C
H
C
H
SALK_092757
At1g22770 (32)
UBQ10 (32)
ACT7 (27)
g
g
SALK_130680
At3g04720 (32)
UBQ10 (32)
ACT7 (27)
d
g
SALK_017005
At1g32640 (32)
UBQ10 (32)
ACT7 (27)
b
b
SALK_082089
At3g04720 (32)
UBQ10 (32)
ACT7 (27)
b
b
SALK_099094
At1g48480 (32)
UBQ10 (32)
ACT7 (27)
c
c
PST18451
At3g04720 (32)
UBQ10 (32)
ACT7 (27)
h
h
SALK_104816
At1g54020 (32)
UBQ10 (32)
ACT7 (27)
b
b
SALK_030723
At4g17490 (37)
UBQ10 (32)
ACT7 (27)
c
g
SALK_036887
At2g14750 (32)
UBQ10 (32)
ACT7 (27)
g
e
SALK_085412
At4g21910 (32)
UBQ10 (32)
ACT7 (27)
a
a
SALK_022179
At2g20340 (27)
UBQ10 (32)
ACT7 (27)
e
f
SALK_076309
At4g26080 (32)
UBQ10 (32)
ACT7 (27)
d
g
SALK_105136
At2g30770 (32)
UBQ10 (32)
ACT7 (27)
d
d
SALK_047832
At5g24770 (32)
UBQ10 (32)
ACT7 (27)
c
c
SALK_006603
At2g38470 (27)
UBQ10 (32)
ACT7 (27)
a
a
SALK_029980
At5g44430 (37)
UBQ10 (32)
ACT7 (27)
a
a

Table 1. Description of *Plutella*-Related Genes with TAIR Identifiers and Response to Herbivory (↑: upregulation; ↓: downregulation).

Gene	Response to *Plutella* herbivory	Other Name(s)	Description/Function
At1g22770	↓	GIGANTEA	regulation of transcription
At1g32640	↑	MYC2, JAI1, JIN1	bHLH transcription factor
At1g48480	↓	RKL1	receptor-like protein kinase
At1g54020	↑		putative myrosinase-associated protein
At1g54100	↑	ALDH7B4	aldehyde dehydrogenase
At2g03980	↑		GDSL-motif lipase/hydrolase family protein
At2g14750	↑	APK	APS kinase
At2g20340	↑		putative tyrosine decarboxylase
At2g21220	↓		auxin-responsive protein
At2g22770	↑	NAI1	bHLH transcription factor
At2g23350	↓	PAB4	putative polyadenylate-binding protein
At2g30770	↑	CYP71A13	putative cytochrome P450
At2g32150	↑		haloacid dehalogenase-like hydrolase family protein
At2g38470	↑	WRKY33	WRKY transcription factor
At2g38870	↑		putative serine protease inhibitor
At2g39050	↑		hydroxyproline-rich glycoprotein family protein
At2g47180	↑	ATGOLS1	galactinol synthase 1
At3g04720	↑	PR4	hevein-like protein
At3g12500	↑	PR3	basic chitinase
At3g54640	↑	TRP3, TSA1	tryptophan synthase alpha chain
At3g57260	↑	PR2	beta 1,3-glucanase
At4g01080	↑		protein of unknown function
At4g02360	↑		protein of unknown function
At4g17490	↑	ATERF6	ERF/AP2 transcription factor family member
At4g21910	↑		MATE efflux family protein
At4g23600	↑	CORI3, JR2	cystine lyase
At4g26080	↑	ABI1	abscisic acid signal transduction
At4g34370	↑		IBR domain-containing protein
At4g35160	↑		O-methyltransferase family 2 protein
At5g24770	↑	VSP2	acid phosphatase activity; vegetative storage protein 2
At5g44430	↑	PDF1.2c	plant defensin
At5g48200	↓		protein of unknown function
At5g60890	↑	MYB34, ATR1	myb-like transcription factor

Table 2. T-DNA Insertion and Transposon-Tagged Lines. Primers used for testing of the presence/absence of insertions and for RT-PCR are listed.

TAIR identifier	T-DNA Insertion line	Test for T-DNA Insertion		RT-PCR	
		Forward Primer (5'→3')	Reverse Primer (5'→3')	Forward Primer (5'→3')	Reverse Primer (5'→ 3')
At1g22770	SALK_092757	TTTCCCGAATCATTTGATGTCTTT	GCCCATTGCTCCGAATAGTCA	GCCCAAGTAGTGAGAATGAC	CCACTACACCATCGGAAATAG
At1g32640	SALK_017005	CGGAAGCTTCTAGATCGGAGT	AGATAAAGCAAACCCGAGACG	CTTCACTTCAATCTCCATCC	GTAGAGAAGAGCCACTAAACC
At1g48480	SALK_099094	TCTCTGTTTTCCTCTCCCTCC	GTCAAGCACTGCCTTATACGC	CTCTTACATGGAAACAGAGG	CGTTTCGGATCGGTTACTTC
At1g54020	SALK_104816	AATATGAGCACAAAGCAGGTCA	TCAACGTAAGTATTTTCATCCAA	GTCACGGAAAACCGGTAGTTC/ CAAGTGTACGACGAATTTAC	AGAGTGATATCAAACTTGTTG/ CTTTGGCTCTGAATCAACAAG
At1g54100	SALK_143309	TTCAATCCACAAACAAACCAACC	CAACACAATCAGATCATGCGACA		
At2g03980	SALK_082986	ACGCAAAAATTCATCAAAACG	TCATAAACGAATAACACCTCTTT	GACAGGAAGGCAAATAGGAAC	CCTAACTTGTGCAATCTTTC
At2g14750	SALK_036887	TGGCGTGATTAGTTCCACAAC	AGACCGGTGACCCAAATCACA	CTTAGTGGTTCAGGGAAGAG	GTACATCCATGAACACCTCAAC
At2g20340	SALK_022179	TTGTTCTTGGCTTCTCTGCGG	GGTATTGGTCTTCTTGAGCAAGTC/	CCAAGTTCTTGATGATGTTC	CTCCACCTCCATTTCCTTTG
At2g21220	SALK_090080	GCAATTTTTAGGTAAAGTACACGCA	GGGTACGATGTACCGGCTCCT	CGGTTAAGAGATCTTCAAAG	GGATGTTAAAGAGCGGAAAAC
At2g22770	SALK_094103	AAGAAGCTGGAGGAGGAGCGA	GAGCCCATAGAGGTCAACTGAA	CAAAAAGACGGACAAGGCAAC	GTCCATCTTAGTGAGAATGG
At2g23350	SALK_113383	TCTGCCCGTAGAAGTGGGGTT	GGCAGCAGAGAAGGCAACAAA	CACCTCTTGCAAGGTTATGC	CCAAATCCAGGCTGGTGAGG
At2g30770	SALK_105136	TTGCAAATGATAATTAACCGTGA	ATCTGCTCACCAACAAAATGG	GTTGGATCGCCCAAGCATTG	GAAGTCCAAAGAGTATGAAG
At2g32150	SALK_140073	TTCGGACAAGAATCATGCGGT	CCGCAGAAAACTGCGTTTGAC	GCTGTCAAGAAAGAATATTGAC	CCAAGTTTCTTGAGGACCTTC
At2g38470	SALK_006603	CCTTCTCTTGTCTCTCCTTCCA	AGGTTTAGGATGGTTGTGGCT	CTCCTCCTCTGCTAACGTTC	CTTCTTCTGTTGTTCCTGAG
At2g38870	SALK_138287	GCCACACGTGTTTTTCTCTT	TAATGTCAACGAAGGGGCTA	GCCGGATTTGGGGGTTTTG	CGACCGAATGTCCTAGGAAG
At2g39050	SALK_037116	CGTCGTGGTTCGATTCGATTT	CCGCAGCCACAGAAAAACAAG	CTGATGAAGCACAGCATTGG	GATTTGCTGTCACCATGATATG
At2g47180	SALK_046018	CTGATCAAATGTGTGAAGAAGACC	TCAGATATTGCCAACAATGCC	CATGTTCAAAAAATCCTGTTC	CCGTATCTGGAAGTTTGTGG
At3g04720	SALK_082089	ACACGTTGACGTTGGTAAATG	TCTCGTAAAACGAGCATATTACATTG	GTGTTCTTCACCCTTAAACAC	GAACTTGTCCCGGTAACATC
At3g04720	SALK_130680	AGAGCAATTGTTGGGGAAGTG	AGCCTAACAAAACGCTTGGAG	GTGTTCTTCACCCTTAAACAC	GAACTTGTCCCGGTAACATC
At3g04720	PST18451(15-2245-1)	TAAACCGTCCGATCAGATGAC	GAATCCATTTATACCCACCGG	GTGTTCTTCACCCTTAAACAC	GAACTTGTCCCGGTAACATC
At3g12500	SALK_005725	ACGAAACGTGTCGGCAGAGTC	CGCAAATAACCAATCAAGGAA		
At3g54640	SALK_021583	CGCGGATAGAGATCGGGAAGT	CAAAGCCGGAGCATGTGAAAC	CTCCAATTGAGCTCACAAGG	GATGTTGGATAAGGTTGTTCC
At3g57260	SALK_087824	GGTGAGAAGAAGTTCGCCAAGA	CGTTGGAAATGAGGTGAAACCC	GTCTTTGCGTTTTCCACACTC	CTCCCTTGCTCGTGAATCTC
At4g01080	SALK_125280	TCCAATCTCCGGTGAAAAGATCA	TTCTTCTTTCTTGTCGTTGTTTGTG/	CTTCCACCTGAGAGAGAATG	CTCCCGGAAACGTTTCATCAG
At4g02360	SALK_117088	CAAGAGGGTCGAGGCAAAACC	GCCGTCAAGAGACACTTCAGCA	CTTGTTCTATCGTCTTCTTTC	CCTTAACACTAACTCCTTTC
At4g17490	SALK_030723	ATCCACTTGGTTCGGTTCGG	TCCGACGTACGTGAAGTAAAACCA	GTCAAATCCCAGTCATCTATC	CGTCGGGACGTTTGAGACG
At4g21910	SALK_085412	TCATTGGTTAGATTACTGACCCG	AGGGAATATGGACCGGGATGA	CGCTTTTTCCACCTCTTTATC	ACCTGTCTTGTCCGGAGTGG
At4g23600	SALK_027027	GCCCTCTCTACAGTTTCAACACA	GTCTTACTATTTATTGGCACAAA	CCAAAAGTGCCGTAGCAGAG	CCAGTTCAGCCAGCTGTTTG
At4g26080	SALK_076309	ATGGCGAGAACACCGAAAACA	GCAGCTGCTGATATAGTCGTCG	CTATCCGGTTTATGGTCAACG	TTCTCAGGTAGCGAACTATTG
At4g34370	SALK_009696	CAAGGAGCTCGGAAGACACCA	TTTGTTTTTATTTAGTGGGCTCTTG/	GCTTCTAGCAGCACAGAGAG	CTCCGTCCAACAATTATTGC
At4g35160	SALK_084490	ACTCCTTCACAAGCATCCCCA	GTGGTCTGAACTCTATCTTATG	CGTGTAACACCCACTTGATG	GAGGCCATGGCTTGTGATGC
At5g24770	SALK_047832	CGTGACTCACAAAATGGTGACA	GGTTGTGGTTAGGGACCGGAG	GAACCATTAGGCTTCAATATG	GGATACGGAACAGAGAAGAC
At5g44430	SALK_029980	TAAAGCTTGTCCCAATGTCCA	GGAGAAACCAGCCATCAAGTA	ACATGGGACGTAACAGATAC	GCTAAGTCTGCTACCATCATC
At5g48200	SALK_025212	GCATCAAACTCTCAAATCTTTAGGC	CGCAACAGTCATTTGCGTTAAA	CATGAATGAAGCATTTGGTG	CAAGCAACGAGCCACAATTC
At5g60890	SALK_006901	CGACTGTCGATAATTTTGGGT	TTGCTTTTTCTCTTTCCAATG	GAAAAAGCTGGATTGAAGAG	GTTGAATTGATCGGTTTGTG
LBb1		GCGTGGACCGCTTGCTGCAACT	"Left Border Primer"		
Ds5-2a		TCCGTTCCGTTTTCGTTTTTAC	"Transposon-specific Primer"		
Ds3-2a		CCGGATCGTATCGGTTTTCG	"Transposon-specific Primer"		
At1g34030 (RPS18b)				ATTGTTGCAAACCCAAGACAG	CAGACCACGATGGTTTCTGAT
At4g05320 (UBQ10)				AAAGAGATAACAGGAACGGAAACATAGT	GGCCTTGTATAATCCCTGATGAATAAG
At5g09810 (ACT7)				GACATGGAAAAGATATGGCATCACAC	AGATCCTTCCTGATATCGACATCA

Table 3. Primers for Identification of the Second Insertion in SALK_082089.

Primer 1	Sequence (5'→3')	Primer 2	Sequence (5'→3')	Primer 3	Sequence (5'→3')
T_DNA_walk1R	CAACCCTATCTCGGGCTATT	T_DNA_walk2R	TCGGAACCACCATCAAACAG	T_DNA_walk3R	ACCGCTTGCTGCAACTCTCT
neuT_walk1R	GACAGATGCGCCTTGTGAG	neuT_walk2R	CCAATGATCTCGCCGTCGATG	neuT_walk3R	CACGGCATCTCGCAACCGTTC
2neuT_walk1R	CCTGCTGCCACACCAGTTC	2neuT_walk2R	CTTCACGTCCTTGTTGACGTG	2neuT_walk3R	CTCGCGCGGGATTTTCTTGTTG
3neuT_walk1R	GCATCCTCGGCGGAAAAC	3neuT_walk2R	CAGTTCTTGCCTGTATGCCTTC	3neuT_walk3R	CGTCCGATTCATTCACCCTCCTTG
4newT_walk1R	CTTTTCAGGGCTTTGTTCATC	4newT_walk2R	CTTCATACTCTTCCGAGCAAAG	4newT_walk3R	CACTCATGAGCAGATTGCTCCAG
6neuT_walk1R	CAGGTTAAAAGACAGGTTAGC	6neuT_walk2R	GGAAACCCTTGCAAATGCTGG	6neuT_walk3R	CTGTGGACAGCCCCTCAAATGTC
7tDNA_1	GTATGAATCCCGCCTGAAGG	7tDNA_2	CGTAATCAGCAAGCGATATACG	7tDNA_3	GAATTGAGCGGCATAACCTGAATC
8tDNA_1	CTATGCCAAGAATGCCCAAG	8tDNA_2	GAAGTCCGTGAATGCCCCGAC	8tDNA_3	CTGGTCGCTGAATGTCGATG
9tDNA_1	CGATATACAGGATTTTGCCAAAGG	9tDNA_2	CTTTCCTTGGTGTATCCAACGG	9tDNA_3	GATAGGTGAAGTAGGCCCAC
10tDNA_1	GCAGAGCCATGTAGACAACATCC	10tDNA_2	CTTTTTCATGCCCTGCCCTAG	10tDNA_3	CCTAGCGTCCAAGCCTCAC
11tDNA_3	CAGCGAGACGAGCAAGATTG	11tDNA_2	CCCGAAACGATCCGACAGC	11tDNA_1	CACCAACGCATACAGCGCCAG
12tDNA_1	GGATTGAGAGTGAATATGAGAC	12tDNA_2	CGAAACGATCCAGATCCGGTG	12tDNA_3	GTGCAATCCATCTTGTTCAATC
13tDNA_1	CGACGAGATCATCGCCGTC	13tDNA_2	GATATTCGGCAAGCAGGCATCG	13tDNA_3	CAGTCGATGAATCCAGAAAAG
14tDNA_1	GCGATATCTTGCTGCGTTC	14tDNA_2	CGGATATTTTCGTGGAGTTCC	14tDNA_3	CCGATCGTTCAAACATTTGG
15tDNA_1	GATTCTGTCGCTACTGATTAC	15tDNA_2	GCTATCGATGGTTTCATTGGTG	15tDNA_3	GGTAATGGTGCTACTGGTG
16tDNA_1	GAGCAATAATCTCCAGGAAATC	16tDNA_2	GTCAAAAGATTCAGGACTAACTG	16tDNA_3	CAGAAGTACTATTCCAGTATGG
17tDNA_1	GATGTGATATCTCCACTGACG	17tDNA_2	CGCACAATCCCACTATCCTTC	17tDNA_3	GAAGTTCATTTCATTTGGAGAG
18tDNA_1	CTTTAGGGTTCCGATTTAGTG	18tDNA_2	CTTGATTTGGGTGATGGTTCAC	18tDNA_3	CCTTTGACGTTGGAGTCCAC
19tDNA_1	GGAGGGTAGCATGTTGATTG	19tDNA_2	GTAACGATGACAGAGCGTTGCTG	19tDNA_3	GCAGAGATCCGAATTATCAGC
20tDNA_1	CCTGTAGGCATCGGGATTG	20tDNA_2	CTGGAACGAGCAGAAGTCCTC	20tDNA_3	GTAAAGGTGAGCAGAGGCAC
tDNA-24.1	CGTATAAGGTTATTGTCCTGG	tDNA-24.2	GTTTCAAGCATTAGTCCATGCAAG	tDNA-24.3	CTGAAATCCTTACATACGGC
tDNA-25.1	CAGCCCTGGTTAAAAACAAG	tDNA-25.2	GTTAAAAGACAGGTTAGCGGTG	tDNA-25.3	CTTGCAAATGCTGGATTTTCTG
tDNA-34.1	GTAATCCAATTCGGCTAAGC	tDNA-34.2	CAGTTTTCGCAATCCACATCGG	tDNA-34.3	CTTTAAATGGAGTGTCTTCTTC
tDNA-35.1	GATTCATTCACCCTCCTTGC	tDNA-35.2	GTTCTTGCCTGTATGCCTTCC	tDNA-35.3	CGATGGTTTCGGCATCCTC
RBwalk1R	CTCATGATCAGATTGTCGTTTC	RBwalk2R	GCGGGTAAACCTAAGAGAAAAGA	RBwalk3R	GGGCGTGAAAAGGTTTATCCGTTC

Table 4. Results of the *Plutella* Herbivory Screens. Location of insertion, number of experimental replicates, number of 96-well flats, sample sizes (*N*), *F*-ratios and *P*-values for damage and plant size are indicated. Relative differences (in per cent) in herbivory and in plant size were calculated as the differences between mutant and wildtype lines, divided by the damage on wildtype lines. *P*-values are given for two-tailed tests.

Gene	T-DNA Insertion Line	Insertion	Experiments	# flats	*N*	*df*	F_{Damage}	P_{Damage}	Δ_{Damage}	F_{Size}	P_{Size}	Δ_{size}
At1g22770	SALK_092757	Exon	1	2	187	1, 1	0.1	n.s.	0.8	2.5	n.s.	-6.0
At1g32640	SALK_017005	Exon	1	2	174	1, 1	0.1	n.s.	-2.2	0.0	n.s.	0.4
At1g48480	SALK_099094	Exon	1	2	180	1, 1	0.5	n.s.	3.0	0.0	n.s.	0.9
At1g54020	SALK_104816	Exon	1	2	175	1, 1	0.5	n.s.	-8.4	2.1	n.s.	0.8
At1g54100	SALK_143309	Intron in 5'UTR	1	3	160	1, 2	1.4	n.s.	-7.7	3.2	n.s.	-7.2
At2g03980	SALK_082986	3'UTR	1	2	150	1, 1	21.6	n.s.	16.2	61.2	n.s.	-6.8
At2g14750	SALK_036887	Exon	1	2	148	1, 1	0.0	n.s.	-0.7	236.4	**< 0.05**	**-12.2**
At2g20340	SALK_022179	3'UTR	1	2	183	1, 1	2.3	n.s.	-3.6	415.6	**< 0.05**	**-6.7**
At2g21220	SALK_090080	5'UTR	1	2	179	1, 1	1.4	n.s.	-6.8	0.1	n.s.	0.2
At2g22770	SALK_094103	3'UTR	1	2	187	1, 1	0.0	n.s.	-2.0	110.5	n.s.	7.9
At2g23350	SALK_113383	Exon	1	3	247	1, 2	28.5	**< 0.05**	**8.9**	2.8	n.s.	1.5
At2g30770	SALK_105136	Exon	1	2	176	1, 1	0.0	n.s.	-1.3	0.2	n.s.	1.2
At2g32150	SALK_140073	Exon	3	9	790	1, 8	2.2	n.s.	10.0	22.3	**< 0.005**	**4.1**
At2g38470	SALK_006603	Intron	1	2	170	1, 1	0.7	n.s.	-12.9	33383.4	**< 0.01**	**5.1**
At2g38870	SALK_138287	Intron	2	6	555	1, 5	0.7	n.s.	4.4	25.0	**< 0.005**	**-4.6**
At2g39050	SALK_037116	3'UTR	3	11	966	1, 10	12.3	**< 0.01**	**-8.9**	0.1	n.s.	0.3
At2g47180	SALK_046018	3'UTR	1	3	230	1, 2	0.2	n.s.	5.4	4.0	n.s.	-1.9
At3g04720	SALK_082089	Exon	3	13	1094	1, 12	6.6	**< 0.025**	**-15.3**	12.1	**< 0.005**	**4.8**
	SALK_130680	3'UTR	3	9	700	1, 8	0.6	n.s.	7.0	0.1	n.s.	-0.4
	PST18451	Exon	1	4	321	1, 3	0.0	n.s.	-1.2	35.3	**< 0.01**	**-3.4**
At3g12500	SALK_005725	3'UTR	2	6	496	1, 5	0.3	n.s.	4.4	11.7	**< 0.025**	**-3.8**
At3g54640	SALK_021583	3'UTR	1	3	264	1, 2	0.3	n.s.	-4.8	10.2	n.s.	-2.2
At3g57260	SALK_087824	3'UTR	1	3	263	1, 2	0.2	n.s.	4.0	0.1	n.s.	1.2
At4g01080	SALK_125280	Promoter	1	3	223	1, 2	0.1	n.s.	-2.8	2.7	n.s.	2.5
At4g02360	SALK_117088	Exon	1	3	222	1, 2	0.6	n.s.	7.7	2.9	n.s.	2.6
At4g17490	SALK_030723	Promoter	1	2	180	1, 1	0.7	n.s.	9.3	1.7	n.s.	3.5
At4g21910	SALK_085412	Exon	1	3	245	1, 2	4.0	n.s.	-10.7	9.4	n.s.	9.7
At4g23600	SALK_027027	3'UTR	1	2	115	1, 1	1.0	n.s.	20.2	0.0	n.s.	1.2
At4g26080	SALK_076309	Exon	2	7	588	1, 6	1.0	n.s.	5.7	3.6	n.s.	-7.0
At4g34370	SALK_009696	3' of 3'UTR	1	3	242	1, 2	0.1	n.s.	-4.9	2.8	n.s.	0.8
At4g35160	SALK_084490	Exon	1	2	190	1, 1	4.3	n.s.	-7.7	5.0	n.s.	4.2
At5g24770	SALK_047832	3'UTR	1	3	242	1, 2	0.5	n.s.	7.0	6.3	n.s.	5.6
At5g44430	SALK_029980	3'UTR	1	2	177	1, 1	0.3	n.s.	6.8	14.5	n.s.	5.7
At5g48200	SALK_025212	Intron	3	10	797	1, 9	1.0	n.s.	5.6	0.0	n.s.	0.1
At5g60890	SALK_006901	3'UTR	1	2	178	1, 1	0.0	n.s.	0.5	0.5	n.s.	-1.5

The Gene Controlling the *Indole Glucosinolate Modifier 1* QTL Alters Indole Glucosinolate Structures and Aphid Resistance in Arabidopsis

Marina Pfalz, Heiko Vogel, Juergen Kroymann*

Max Planck Institute for Chemical Ecology; Hans-Knoell-Str. 8; D-07745 Jena

To whom correspondence should be addressed: Juergen Kroymann;
kroymann@ice.mpg.de

Running title: Arabidopsis CYP81F2

Estimated length of the article: 14 printed pages

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Juergen Kroymann (kroymann@ice.mpg.de)

Glucosinolates are defensive secondary compounds which display large structural diversity in Arabidopsis and related plants. Much attention has been paid to variation in the biosynthesis of methionine-derived aliphatic glucosinolates and its ecological consequences but little is known about the genes that cause qualitative and quantitative differences in tryptophan-derived indole glucosinolates. We use a combination of QTL fine mapping and microarray-based transcript profiling to identify *CYP81F2* (At5g57220), encoding a cytochrome P450 monooxygenase, as the gene underlying *Indole Glucosinolate Modifier 1* (*IGM1*), a metabolic QTL for the accumulation of two modified indole glucosinolates, 4-hydroxy-indole-3-yl-methyl and 4-methoxy-indole-3-yl-methyl glucosinolate. We verify *CYP81F2* function with two SALK T-DNA insertion lines, and show that CYP81F2 catalyzes the conversion of indole-3-yl-methyl to 4-hydroxy-indole-3-yl-methyl glucosinolate. We further show that the *IGM1* QTL is largely caused by differences in *CYP81F2* expression, which results from a combination of *cis* and *trans* acting expression QTL (eQTL) different from known regulators of indole glucosinolate biosynthesis. Finally, we elucidate a potential ecological function of *CYP81F2* and find that CYP81F2 contributes to defence against the green peach aphid (*Myzus persicae*), but not to resistance against herbivory by larvae from four lepidopteran species.

INTRODUCTION

Activated defenses are among the most fascinating adaptations that plants have evolved to ward off natural enemies. These defense systems consist of precursor molecules and activators, usually enzymes. The precursor molecules can be stored in the plant tissue without harmful effects to the plant, but have the potential to give rise to highly toxic or deterrent effectors upon activation. The glucosinolate-myrosinase system is an activated defense system which protects plants from the Brassicaceae and related families effectively against most herbivorous insects. This system is often referred to as 'the mustard oil bomb' to characterize its basic functional principle (Matile, 1980; Lüthy and Matile, 1984). The mustard oil bomb consists of glucosinolates, amino acid-derived *β*-thioglucoside-N-hydroxysulfates, and matching breakdown enzymes, myrosinases, which have *β*-thioglucoside glucohydrolase activity. In intact plant tissue, glucosinolates and myrosinases are stored in separate cell types (Koroleva et al., 2000; Husebye et al., 2002; Thangstad et al., 2004). However, upon tissue disruption myrosinases gain access to glucosinolates and hydrolyze their *β*-thioglucoside ester bond. The resulting aglycone is instable and rearranges to form a variety of breakdown products such as isothiocyanates, thiocyanates, nitriles, epithionitriles and others, depending on reaction conditions and the presence (or absence) of modifying proteins (Lambrix et al. 2001; Bones and Rossiter, 2006; Zhang et al., 2006; Burow et al., 2007). These hydrolysis products have diverse ecological functions. They may serve as oviposition and feeding stimulants for insects specialized on glucosinolate-containing plants, but usually act as toxins or deterrents towards other insect herbivores (Raybould and Moyes, 2001; Kliebenstein et al., 2005).

More than 120 different glucosinolate structures have been identified in the Capparales (Daxenbichler et al., 1991; Fahey et al., 2001). They share a chemical core, consisting of a sulfonated oxime and a *β*-thioglucose moiety, but differ in their side chain structures. Depending on the precursor amino acid, glucosinolates are grouped into different classes. Aliphatic glucosinolates originate from alanine, methionine, leucine, isoleucine, or valine, aromatic glucosinolates from phenylalanine or tyrosine, and indole glucosinolates are derived from tryptophan. In Arabidopsis, ca. 40 different glucosinolates have been identified, derived from methionine,

phenylalanine or tryptophan (Kliebenstein et al., 2001a; Reichelt et al., 2002). This enormous structural variety is thought to have evolved in response to challenges imposed by a large and diverse community of herbivores and other enemies (Kliebenstein et al., 2001a; Kroymann et al., 2003; Benderoth et al., 2006; Benderoth et al., 2008).

Natural genetic variation among Arabidopsis accessions has been exploited to map glucosinolate QTL in various recombinant inbred line (RIL) populations (Campos de Quiros et al., 2000; Kliebenstein et al., 2001b; Kliebenstein et al., 2002; Keurentjes et al., 2006; Pfalz et al., 2007). Efforts to clone metabolic QTL (mQTL) have mainly focused on aliphatic glucosinolates, the largest and most diverse glucosinolate class in Arabidopsis. As a result, several genes underlying major aliphatic glucosinolate biosynthesis QTL have been identified and ecological consequences of natural variation in these genes have been investigated (Kliebenstein et al., 2001c; Kroymann et al., 2001; Kroymann et al., 2003; Kroymann and Mitchell-Olds, 2005).

Less attention has been paid to variation in indole glucosinolate structures. Here, research has mainly focused on the biosynthesis of the glucosinolate core from tryptophan because an early intermediate, indole-3-acetaldoxime, also serves as a precursor for auxin (Bak and Feyereisen, 2001; Bak et al., 2001; Zhao et al., 2002; Ljung et al., 2005) and camalexin biosynthesis (Glawischnig *et al.*, 2004), such that alterations in the indole glucosinolate core pathway often feedback on plant growth and development (Boerjan et al., 1995; Celenza et al., 1995; King et al., 1995; Lehman et al., 1996; Delarue et al., 1998; Barlier et al., 2000; Bak et al., 2001; Smolen and Bender, 2002). Nonetheless, recent studies demonstrate that indole glucosinolates and their breakdown products influence ecological interactions, contributing to defense against aphids and deterring oviposition by the European cabbage butterfly (*Pieris rapae*) (Kim and Jander, 2007; De Vos et al., 2008; Kim et al., 2008). Arabidopsis accessions contain typically four different indole glucosinolates (Kliebenstein et al., 2001a; Brown et al., 2003), indole-3-yl-methyl (I3M), 1-methoxy-indole-3-yl-methyl (1MO-I3M), 4-hydroxy-indole-3-yl-methyl (4OH-I3M) and 4-methoxy-indole-3-yl-methyl (4MO-I3M) glucosinolate. I3M is the most abundant indole glucosinolate in Arabidopsis. 1MO-I3M, 4OH-I3M and 4MO-I3M bear various modifications of the indole ring and are present in lower quantities than I3M.

However, the genes that are responsible for indole glucosinolates structures are not yet known.

In a recent study, we have mapped QTL for indole glucosinolates in a RIL population derived from a cross between the Arabidopsis accessions Da(1)-12 and Ei-2, and found a complex genetic architecture underlying variation in indole glucosinolates (Pfalz et al., 2007). To start dissecting this architecture, we focused on the bottom of chromosome 5, where a 4OH-I3M QTL co-localized with a 4MO-I3M QTL, termed *Indole Glucosinolate Modifier 1* (*IGM1*) (Figure 1). With a combination of QTL fine-mapping in near isogenic lines (NILs) and transcript profiling with whole-genome Arabidopsis microarrays we identified candidate genes for *IGM1* on chromosome 5, and used T-DNA insertion lines to verify that a single gene, *CYP81F2* (At5g57220) underlies this QTL. The product of this gene was a cytochrome P450 monooxygenase and we showed that CYP81F2 catalyzed the conversion of I3M to 4OH-I3M. We further showed that metabolic variation in modified indole glucosinolates is largely attributable to variation in gene expression, with *cis*- and *trans*-acting factors controlling *CYP81F2* transcript abundance. Finally, we investigated ecological effects of *CYP81F2* on Arabidopsis-insect interactions, and found that the gene did not affect herbivory by several generalist and specialist lepidopterans but, instead, contributed to resistance against the green peach aphid (*Myzus persicae*), a generalist phloem-feeding herbivore.

RESULTS

Fine Mapping of QTL for Modified Indole Glucosinolates in DE089 × Ei-2 NILs

DE089 is a recombinant inbred line (RIL) from the Da(1)-12 × Ei-2 population (Pfalz et al., 2007). Most of its genome originates from the Ei-2 parental accession except for a region corresponding to ca. 7 Mbp in Col-0 at the bottom of chromosome 5, where DE089 has a Da(1)-12 genotype. Therefore, DE089 was ideally suited to generate near isogenic lines (NILs) for simultaneous fine-mapping of the 4OH-I3M and 4MO-I3M QTL near the bottom of chromosome 5.

We crossed DE089 with Ei-2, and analyzed 500 F_2 progeny for recombinants in the QTL candidate region. We genotyped with markers *MNC6* and *5FM9*, located near the borders of the chromosomal segment segregating in DE089 × Ei-2 progeny, and with further markers between *MNC6* and *5FM9*, spaced at an average distance of ca. 600 kb (Figure 2A). Analysis of glucosinolate profiles confirmed that the indole glucosinolate QTL were segregating in F_2 progeny.

From F_2 progeny we selected two series of recombinant plants for further fine-mapping. One series of F_2 progeny was homozygous Da(1)-12 at one side of the QTL candidate region, the other series was homozygous Ei-2. In both series, the remainder of the QTL candidate region was heterozygous, and segregated in the F_3 progeny (Figure 2A). We selected 25 families of recombinant F_2 plants such that their recombination breakpoints allowed us to break down the QTL candidate region into intervals of approximately equal size. From each family we grew 16 progeny randomized over five 96-celled flats, analyzed glucosinolate profiles at 3 weeks age, and genotyped plants with the same marker set as before. We also added two further markers, *5FM39* and *5FM42*, to improve resolution. We found high statistical support for a 4MO-I3M QTL between markers *5FM39* and *MMN10* (Figure 2B). This QTL peaked near marker *5FM42* (N = 251, df = 1, F = 140.1, $P < 10^{-7}$ for Da(1)-12 *versus* Ei-2 genotypes, and N = 351, df = 2, F = 72.8, $P < 10^{-7}$ when heterozygote genotypes were included). Likewise, we found evidence for a 4OH-I3M QTL. Here, the QTL peaked near marker *5FM39* (N = 157, df = 1, F = 9.5, $P < 0.003$, and N = 203; df = 2, F = 4.1, $P < 0.02$, respectively). Ei-2 genotypes accumulated more 4OH-I3M and 4MO-I3M than Da(1)-12 genotypes, and heterozygotes had an intermediate glucosinolate phenotype (Figure 3).

Next, we visually inspected the data from F_3 progeny that had recombined near or within the QTL candidate region between *5FM39* and *MMN10*. 4OH-I3M quantity was too low to reliably detect variation on a within-family basis. Therefore, we focused on 4MO-I3M. The 4MO-I3M QTL was segregating in families 6A2, 5G10, 8C2, and 5H5, but not in families 5C7, 7E10, 7A10, and 6G9. Visual inspection of variation in 4MO-I3M accumulation within families was consistent with our previous QTL mapping results, and placed the 4MO-I3M QTL between *5FM39* and *5FM42*, in an interval corresponding to ca. 1.1 Mbp of the Col-0 genome (Please refer to Supplemental Results for details).

At the final stage we fine-mapped recombination breakpoints in families 5G10, 6A2, and 8C2. The 4MO-I3M QTL segregated in these families with recombination breakpoints in close vicinity of the presumptive QTL. We placed three markers, *5FMJ01*, *5seq01*, and *5seq02*, in the interval between *5FM39* and *5FM15*, and two further markers, *5seq03* and *5seq05*, between *5FM15* and *5FM42* (Figure 2A). Family 5G10, with an Ei-2 genotype at *5FM39*, turned out to segregate at *5FMJ01* and thus provided no further information on the QTL position. In family 6A2 alleles segregated at markers *5seq01* and *5seq02* but not at *5FMJ01*, indicating that the QTL was located downstream of *5FMJ01*. Finally, in family 8C2, marker *5seq03* segregated but not marker *5seq05*; hence, *5seq05* marked the right-hand border of the QTL candidate region.

In summary we were able to place the 4MO-I3M QTL between *5FMJ01* and *5seq05*. This interval corresponds to 283 kb in Col-0 and contains nearly 70 genes. Low 4OH-I3M quantity precluded a comparably exact delimitation of the 4OH-I3M QTL. Nonetheless, results from single marker analyses suggested that the 4OH-I3M QTL was located near marker *5FM39*, close to the 4MO-I3M QTL.

Identification of Candidate Genes with Transcript Profiling

QTL can be caused by structural variation in the underlying genes, *i.e.* by differences in the amino acid sequence of the gene products, or by regulatory variation causing differences in gene expression. Transcript profiling with microarrays is a convenient means to monitor differences in transcript levels of a large number of genes simultaneously. We used Agilent Arabidopsis 3 arrays to compare DE089 and Ei-2 transcriptomes under two different experimental conditions, each with three biological replicates per plant line. In the first set of hybridizations, we measured transcripts of unmanipulated DE089 and Ei-2. In the second set of hybridizations, we compared transcript levels after 9-hour herbivory with one 3-day old larva of the Cabbage white butterfly (*Pieris brassicae*) per plant.

After RNA-isolation, hybridization and data processing, we focused on the genes in the QTL candidate interval between markers *5FMJ01* and *5seq05*. Most genes had similar transcript levels in Ei-2 and DE089, but there were two notable exceptions (Figure 2C). In both control and herbivory experiments, At5g57123 and At5g57220 had much higher transcript levels in Ei-2 than in DE089. At5g57123

encodes an unknown protein with a putative DNA-binding domain, and At5g57220 is annotated as a gene for a cytochrome P450 monooxygenase of the CYP81F subfamily, *CYP81F2*.

T-DNA Insertions in *CYP81F2* Cause Reduced Levels of Modified Indole Glucosinolates

Because cytochromes P450 are often involved in the biosynthesis of plant secondary metabolites and typically catalyze oxygenation reactions (Werck-Reichhart and Feyereisen, 2000; Guengerich, 2001; Werck-Reichhart et al., 2002), *CYP81F2* appeared to be a promising candidate gene for the QTL. We obtained two SALK lines with T-DNA insertions in *CYP81F2* (Alonso et al., 2003) from the Arabidopsis stock centers, SALK_123882 and SALK_005861. In SALK_123882, the T-DNA is located in the second exon of *CYP81F2*, and in SALK_005861, the T-DNA is inserted in the third exon very close to the stop codon (Figure 4). Because we could not determine the exact position of the insertion in SALK_005861 from the sequence deposited in the T-DNA insertion database (http://signal.salk.edu/cgi-bin/tdnaexpress), we re-sequenced the T-DNA insertion site and found the insertion located 10 nucleotides upstream of the *CYP81F2* stop codon. As a result, the gene product encoded by SALK_005861 had an altered composition of the 4 original C-terminal amino acids and an extension of 26 amino acids.

To identify plants with homozygous mutant *CYP81F2* alleles, we planted several seeds from each line and screened for the presence/absence of the T-DNA insertion. From SALK_123882, we obtained homozygous ($CYP81F2^{\Delta/\Delta}$) and heterozygous ($CYP81F2^{Col/\Delta}$) mutant progeny and also progeny without a T-DNA insertion in At5g57220 ($CYP81F2^{Col/Col}$). From SALK_005861, we obtained progeny with homozygous mutant and homozygous wildtype alleles but no heterozygotes. We also examined the number of T-DNA insertions with DNA gel blots. Hybridization experiments were compatible with the presence of a single T-DNA insertion in SALK_123882, but SALK_005861 had multiple insertions.

All SALK_123882 and SALK_005861 genotypes had a typical Col-0 glucosinolate profile (Kliebenstein et al., 2001a), with two notable exceptions. Compared to $CYP81F2^{Col/Col}$, $CYP81F2^{\Delta/\Delta}$ plants from SALK_123882 progeny accumulated significantly less 4MO-I3M (N = 68, df = 1, F = 763.0, P < 0.000001)

and 4OH-I3M (N = 55, df = 1, F = 61.4, P < 0.000001). Likewise, *CYP81F2*$^{\Delta/\Delta}$ plants from SALK_005861 progeny produced significantly less 4MO-I3M (N = 68, df = 1, F = 881.6, P < 0.000001) and 4OH-I3M (N = 57, df = 1, F = 20.9, P < 0.00003) than control plants (Figure 3). For all other glucosinolates we did not detect any significant difference between *CYP81F2*$^{Col/Col}$ and *CYP81F2*$^{\Delta/\Delta}$ genotypes. In particular, the other indole glucosinolates, I3M and 1MO-I3M, did not show significant differences between mutant and wildtype genotypes (all: N = 68, df = 1, F < 1.5, P > 0.2). Hence, in comparison to the *CYP81F2* wildtype, T-DNA insertion mutants showed a substantial reduction in both 4OH-I3M and 4MO-I3M, indicating that a single gene encoding CYP81F2 controlled both the 4OH-I3M and the 4MO-I3M QTL.

The T-DNA Insertion in SALK_123882 Mutants Causes Degradation of *CYP81F2* Transcripts

In a separate experiment, we included heterozygous SALK_123882 progeny (*CYP81F2*$^{Col/\Delta}$) and repeated our glucosinolate analyses. This experiment confirmed our previous results; *CYP81F2*$^{\Delta/\Delta}$ genotypes had significantly reduced 4OH-I3M and 4MO-I3M levels. However, *CYP81F2*$^{Col/\Delta}$ genotypes showed an unexpected phenotype, accumulating only slightly more of modified indole glucosinolates than homozygous mutants (*CYP81F2*$^{\Delta/\Delta}$). This was in strong contrast to heterozygous F_2 and F_3 progeny from DE089 × Ei-2 which had intermediate 4OH-I3M and 4MO-I3M phenotypes (Figure 3). Therefore, we investigated *CYP81F2* transcript levels in SALK_123882 and SALK_005861 progeny with RT-PCR. Two control transcripts, coding for RPS18B and actin7, respectively, amplified equally well in all progeny, independent of the *CYP81F2* genotype. Likewise, *CYP81F2* transcripts had approximately equal levels in *CYP81F2*$^{\Delta/\Delta}$ and *CYP81F2*$^{Col/Col}$ genotypes from SALK_005861 progeny. However, in SALK_123882 progeny we detected a clearly visible PCR product only when plants carried two wildtype *CYP81F2* alleles (Figure 5). We found no PCR product in homozygous mutants, indicating that the inserted T-DNA caused a degradation of the *CYP81F2* transcripts. In heterozygotes we sometimes found very weak signals, but the amount of PCR product was always much lower than half the product intensity of homozygous wildtypes. Thus, in heterozygotes, the T-DNA insertion affected not only transcripts of the mutant allele but also triggered degradation of transcripts from the functional allele, probably by

some form of nonsense-mediated RNA decay. Hence, in SALK_123882 progeny, RT-PCR patterns of the different *CYP81F2* genotypes matched glucosinolate phenotypes, with heterozygotes having minimal levels of *CYP81F2* transcript and only slightly higher quantities of 4OH-I3M and 4MO-I3M than homozygous mutants.

CYP81F2 Catalyzes the Conversion of I3M to 4OH-I3M

The current model of indole glucosinolate biosynthesis (http://www.arabidopsis.org) suggests that 4MO-I3M is generated from I3M *via* 4OH-I3M as an intermediate. Variation in a single gene, *CYP81F2*, causes quantitative difference in the accumulation of both 4OH-I3M and 4MO-I3M among Arabidopsis accessions and in T-DNA insertion lines. This observation led us to postulate that CYP81F2 catalyzes the conversion of I3M to 4OH-I3M. To test this hypothesis, we expressed *CYP81F2* in insect cells, and conducted enzyme assays with the intact I3M, isolated from seeds of Dyer's woad (*Isatis tinctoria*), as a substrate. Because of an N-terminal membrane anchor in the mature protein, we used the microsome fraction of the insect cells for our assays. We included two appropriate controls, microsomes isolated from *Sf*9 cells expressing a gene from the European cabbage butterfly (*Pieris rapae*), and a buffer control. In all samples, the peak corresponding to I3M was clearly visible with HPLC (Figure 6). However, we obtained a second major peak in the samples with heterologously expressed *CYP81F2* but not in the controls. We converted the respective compound to its desulfo form and subjected it to liquid chromatography-mass spectrometry (LC-MS). The mass spectrum of this reaction product showed a base peak of *m/z* 385 $[M+H]^+$ and a typical fragment of *m/z* 223 $[M\text{-Glucose}+H]^+$. Hence, the molecular weight of this reaction product was +16 compared to the substrate, I3M (desulfo form *m/z* 369 $[M+H]^+$), corresponding to the addition of a hydroxy-group. Furthermore, we compared the retention time of the reaction product with known glucosinolate profiles from different Arabidopsis accessions (Kliebenstein et al., 2001a; Pfalz et al., 2007) and found it to elute at exactly the same retention time as 4OH-I3M. We also tested desulfo-I3M but CYP81F2 had no activity with this substrate. Thus, CYP81F2 was indeed capable of catalyzing the suspected reaction, the conversion of I3M to 4OH-I3M.

CYP81F2 Sequence Variation between Col-0, Da(1)-12 and Ei-2

At5g57220 is annotated as a member of the *CYP81F* subfamily of cytochrome P450s (http://www.arabidopsis.org). The gene has three exons and two introns (Figure 4), and encodes a protein of 491 amino acids. TargetP predicts an N-terminal signal peptide of 27 amino acids in the translated sequence, serving as a membrane anchor (Emanuelsson et al., 2000; Nielsen et al., 1997). ScanProsite identified a cytochrome P450 cysteine heme-iron ligand signature, comprising amino acids 422 – 431 (www.expasy.ch/tools/scanprosite/). Furthermore, we detected a 'proline-rich membrane hinge' next to the signal peptide, an 'I-helix' involved in oxygen binding and activation (aa 295 – 299), and an 'E-R-R triad' (aa 352 – 355) with a 'PERF' consensus (aa 407 – 410) in the primary protein sequence. These features are typical and conserved structures in cytochrome P450s (Paquette et al., 2000; Werck-Reichhart et al., 2002).

We sequenced a ~ 3.9 kb fragment containing *CYP81F2* from Da(1)-12 and Ei-2. This fragment included approximately 1,200 base pairs of sequence upstream of the start and 900 base pairs downstream of the stop codon. The Da(1)-12 sequence was nearly identical to the Col-0 sequence, except for one single nucleotide polymorphism (SNP) in the intergenic region between At5g57220 and At5g57230 (Figure 4). By contrast, we identified numerous sequence differences between Ei-2 and Col-0 or Da(1)-12. However, only four nucleotide substitutions altered the meaning of codons in the open reading frame, and all caused conservative amino acid exchanges. For example, two non-synonymous substitutions were found in the N-terminal membrane anchor, and in both cases a hydrophobic amino acid was replaced by another hydrophobic amino acid (Figure 4). In the promoter region, we found more substantial differences. This region contained ca. 25 SNPs, and several insertion/deletion polymorphisms, the largest consisting of a stretch of 35 nucleotides absent from the Ei-2 sequence. These data suggest that the indole glucosinolate QTL near the bottom of chromosome 5 is controlled by variation in the *CYP81F2* promoter, although we cannot entirely exclude the possibility that structural variation in the amino acid sequence causes differences in the biochemical properties of CYP81F2 variants.

Two Expression QTL Control *CYP81F2* Transcript Levels

Microarray hybridization experiments enabled us to identify *CYP81F2* as a candidate gene underlying the 4OH-I3M and 4MO-I3M QTL, because DE089 and Ei-2 displayed strong differences in *CYP81F2* transcript accumulation under two different experimental conditions. In both experiments, Ei-2 accumulated substantially more transcript than DE089 (Figure 2C). This difference in expression correlated with glucosinolate phenotypes (Figure 3). Likewise, Da(1)-12 × Ei-2 RILs with an *CYP81F2* Ei-2 genotype produced on average more 4OH-I3M and 4MO-I3M than Da(1)-12 genotypes (Figure 1; Pfalz et al., 2007).

To confirm our microarray data, we analyzed *CYP81F2* transcript levels in Da(1)-12, Ei-2 and DE089 with quantitative RT-PCR. We included three biological replicates per line and compared Δ*C*t values between lines (Supplemental Table 1). As expected, we found low *CYP81F2* expression in DE089, corresponding to ca. 31 % of the Ei-2 transcript level. Surprisingly, the other parental line, Da(1)-12, had *CYP81F2* transcript levels comparable to Ei-2. This led us to suspect that additional factors, encoded outside of the chromosomal segment segregating in DE089 × Ei-2 progeny, contributed to *CYP81F2* expression control.

To test this hypothesis we analyzed *CYP81F2* transcript levels in Da(1)-12 × Ei-2 RILs with quantitative RT-PCR to map expression QTL (eQTL). We grew plants randomized in 96-celled flats, with 1 replicate per RIL, and carried out RT-PCR in 96-well PCR plates, with control (*RPS18B*) and test (*CYP81F2*) gene assays on the same plate. Altogether, we analyzed 181 RILs and, thus, each marker/genotype combination was replicated ~90 times. For each RIL, we calculated $\Delta C t_{(\text{Test Gene} - \text{Control Gene})}$. To control for plate-to-plate variation in RT-PCR efficiency, we normalized data such that all plates had identical sums of Δ*C*t values. These data were then used as the input for QTL cartographer (Wang et al., 2001-2004).

As expected, we found a *cis*-acting eQTL near *CYP81F2* but we also detected an eQTL in *trans* near the bottom of chromosome 1 (Figure 1). This eQTL co-localized with another metabolic QTL (mQTL) that affected 4MO-I3M but not 4OH-I3M accumulation (Pfalz et al., 2007). To test for a potential epistatic interaction between both eQTL, we used markers *MSAT1.1*, closely linked to the eQTL on chromosome 1, and *MMN10*, near *CYP81F2*, with the following statistical model:

$$\Delta Ct = \text{CONSTANT} + MSAT1.1 + MMN10 + MSAT1.1 * MMN10$$

Both markers, *MSAT1.1* (N = 174, df = 1, F = 22.80, P < 0.00001) and *MMN10* (N = 174, df = 1, F = 27.36, P < 0.000001), were significantly associated with *CYP81F2* expression. The interaction term *MSAT1.1 * MMN10* was not significant (N = 174, df = 1, F = 0.28, P = 0.60), indicating independence of the two eQTL. Furthermore, eQTL effects had opposite directions. For *MSAT1.1*, we obtained $\Delta\Delta Ct_{(Ei\text{-}2\ -\ Da(1)\text{-}12)}$ = 0.74, indicating higher expression in RILs with a Da(1)-12 genotype. For *MMN10*, a $\Delta\Delta Ct_{(Ei\text{-}2\ -\ Da(1)\text{-}12)}$ value of –0.80 indicated higher expression in the Ei-2 genotype. Hence, eQTLs were additive and eQTL effects compensated each other in Da(1)-12 and Ei-2, resulting in nearly equal *CYP81F2* expression in these accessions.

Finally, we tested for an association between *CYP81F2* transcript quantity and indole glucosinolate accumulation in Da(1)-12 × Ei-2 RILs. Indole glucosinolate data were taken from a previously published analysis (Pfalz et al., 2007). ΔCt values were negatively correlated with 4OH-I3M (r_G = –0.22, t = –2.93, P = 0.0038) and 4MO-I3M (r_G = –0.39, t = –5.62, P < 0.00001) but not with I3M and 1MO-I3M (both: P > 0.35; n.s.). Hence, because large ΔCt values indicate low and small values high transcript levels, the expression of *CYP81F2* is positively correlated with the production of both 4OH-I3M and 4MO-I3M.

CYP81F2 Contributes to Resistance against Aphids but not Lepidopteran Herbivores

To determine a potential ecological function of *CYP81F2*, we tested the performance of several insects on Col-0 wildtype and on SALK_005861 and SALK_123882 progeny with $CYP81F2^{Col/Col}$ and $CYP81F2^{\Delta/\Delta}$ genotypes. We used larvae from four lepidopteran species, the crucifer specialists Diamondback moth (*Plutella xylostella*) and *Pieris brassicae*, and the generalist herbivores Cabbage looper (*Trichoplusia ni*) and Beet armyworm (*Spodoptera exigua*). We conducted herbivory assays to examine whether larvae fed differentially on mutant and wildtype plants, but found that *CYP81F2* genotype did not significantly influence any of the lepidopterans (Table 1).

We also investigated the performance of the green peach aphid (*Myzus persicae*), a generalist phloem-feeding insect (Pollard, 1972), and measured its proliferation contingent on plant genotype. Here, we reared aphids for several

generations either on Col-0 or on *CYP81F2*$^{\Delta/\Delta}$ genotypes before we transferred the aphids to test plants and started our experiments. We included this factor 'origin' in our statistical models. Plant genotype had a significant effect on aphid proliferation (Figure 7). We found fewer aphids on *CYP81F2*$^{Col/Col}$ than on *CYP81F2*$^{\Delta/\Delta}$ genotypes (N = 82, df = 1, F = 9.19, P = 0.0033) irrespective of whether aphids had been previously reared on plants with wildtype or with mutant *CYP81F2* alleles. Furthermore, the pre-experimental rearing procedure strongly influenced aphid performance. Aphids that originated from wildtype plants proliferated less well than aphids that came from mutant plants (N = 82, df = 1, F = 13.04, P = 0.00056). However, we found no interaction between both variables (N = 82, df = 1, F = 0.89, P = 0.37, n.s.), indicating that plant genotype effects during the experiment were independent of genotype effects in the pre-experimental treatment. This showed that aphids did not adapt to the genotype on which they had been reared.

DISCUSSION

The Role of CYP81F2 in the Biosynthesis of Modified Indole Glucosinolates

Most reaction steps in the biosynthesis of I3M from tryptophan are known and the underlying genes have been cloned (Figure 8). Tryptophan is synthesized in the plastids (Li et al., 1995; Zhao and Last, 1995; Bak et al., 1998), and CYP79B2 and CYP79B3 which convert tryptophan to indole-3-acetaldoxime, contain plastid transit peptides (Hull et al., 2000; Mikkelsen et al., 2000). All other enzymes of the indole glucosinolate core pathway have no discernable plastid targeting sequences, indicating that subsequent reaction steps take place in other cell compartments. CYP83B1 *N*-hydroxylates indole-3-acetaldoxime to the corresponding *aci*-nitro compound, 1-*aci*-nitro-2-indolyl-ethane (Bak and Feyereisen, 2001; Bak et al., 2001; Hansen et al., 2001; Naur et al., 2003). This compound is conjugated with an *S*-donor, most likely cysteine (Wetter and Chisholm, 1968). The respective genes and enzymes are not yet known (Halkier and Gershenzon, 2006). Next, a *C*-*S*-lyase converts *S*-indolyl-thiohydroximate to the corresponding thiohydroximate (Mikkelsen et al., 2004), followed by a glycosylation which is catalyzed by an *S*-

glycosyltransferase to form desulfo-I3M (Grubb et al., 2004). The final reaction in the core structure pathway is carried out by a family of three desulfoglucosinolate sulfotransferases, with AtST5a preferably acting on desulfo-I3M (Piotrowski et al., 2004).

The current model of indole glucosinolate biosynthesis (http://www.arabidopsis.org) suggests that 4OH-I3M and 4MO-I3M originate from I3M by a common pathway. Detached leaves of Arabidopsis *cyp79B2 cyp79B3* double mutants (Zhao et al., 2002), blocked early in the indole glucosinolate pathway, can convert artificially supplied I3M to 4MO-I3M but only when they are infested with aphids (Kim and Jander, 2007). We show that CYP81F2 catalyzes the first reaction in this pathway, the hydroxylation at position 4 of the indole ring, which results in the formation of 4OH-I3M. Generation of 4MO-I3M, however, requires the subsequent methylation of the hydroxy group, which is probably catalyzed by an as yet unidentified methyltransferase. The respective gene appears to be not variable in Da(1)-12 × Ei-2, and therefore remained undetected during QTL mapping. Nonetheless, this reaction sequence explains why natural variation in a single gene, *CYP81F2*, causes both a 4OH-I3M and a 4MO-I3M QTL and why mutations in this gene affect both compounds simultaneously; reduction of the metabolic flux from I3M to the 4OH intermediate reduces the quantity of the substrate available for the methylation reaction and, hence, the accumulation of the end product, 4MO-I3M. Levels of 1MO-I3M, the third modified indole glucosinolate in Arabidopsis, remained unaffected by mutations in *CYP81F2*, corroborating our previous finding that different QTL control natural variation in 1MO-I3M *versus* 4OH-I3M and 4MO-I3M accumulation in Da(1)-12 × Ei-2 RILs (Pfalz et al., 2007). This is consistent with 1MO-I3M originating from I3M by another pathway than 4OH-I3M and 4MO-I3M. However, in contrast to the indole glucosinolate biosynthesis model (http://www.arabidopsis.org) we surmise that I3M is not directly converted to 1MO-I3M but rather *via* a 1OH intermediate, in analogy to the reaction scheme for the generation of 4MO-I3M.

Natural Variation in the Quantitative Genetic Architecture of Modified Indole Glucosinolate Biosynthesis

T-DNA insertions in *CYP81F2* strongly impair 4OH-I3M and 4MO-I3M production. RT-PCRs failed to amplify *CYP81F2* cDNA in SALK_123882 *CYP81F2*$^{\Delta/\Delta}$ genotypes, while control genes had wildtype levels of PCR product. Heterozygotes, with one intact copy of *CYP81F2*, accumulated only minimal amounts of *CYP81F2* cDNA (Figure 5), and the glucosinolate profile of *CYP81F2*$^{Col/\Delta}$ genotypes was almost indistinguishable from *CYP81F2*$^{\Delta/\Delta}$ plants. Hence, the T-DNA insertion in SALK_123882 causes a complete knockout of *CYP81F2* function. Nonetheless, even homozygous mutant plants produced basic levels of 4OH-I3M and 4MO-I3M. This suggests that other gene products are also capable of catalyzing the oxygenation of I3M to 4OH-I3M, a hypothesis that we are currently testing. Candidate genes are the three other members of the Arabidopsis *CYP81F* subfamily, At4g37400, At4g37410 and At4g37430 which form a small gene cluster on *A. thaliana* chromosome 4.

To identify the molecular basis for the *IGM1* QTL, we investigated differences between the Col-0, Da(1)-12 and Ei-2 *CYP81F2* sequences. While the sequenced region was almost identical in Col-0 and Da(1)-12, Ei-2 deviated at numerous positions, both in the promoter and in the coding region. However, non-synonymous substitutions were rare and all involved conservative amino acid replacements, suggesting that the 4OH-I3M and 4MO-I3M QTL are most likely caused by quantitative variation in *CYP81F2* transcript levels. This matches the observation of Wentzell et al. (2007) who found that many mQTL in glucosinolate biosynthesis are caused by differences in transcript abundance of the underlying genes and, hence, co-localize with corresponding eQTL. Indeed, *CYP81F2* transcript levels correlated well with 4OH-I3M and 4MO-I3M quantity in Da(1)-12 × Ei-2 RILs, and mapping revealed the presence of two eQTL, one in *cis* and another one in *trans* on chromosome 1. These QTL compensated each other in the parental lines, Da(1)-12 and Ei-2, which had nearly equal *CYP81F2* steady-state transcript levels. Transcript levels were highest in RILs with a combination of Ei-2 alleles at the *cis* and Da(1)-12 alleles at the *trans* eQTL, and lowest in the reciprocal combination, as exemplified by the DE089 RIL. This suggests that the Ei-2 *CYP81F2* promoter contains an activating element that is absent from Da(1)-12 (or, *vice versa*, lacks a repressing element present in Da(1)-12). The corresponding transcription factor remains to be identified. It is, however, not likely that this factor is encoded by the gene that underlies the

trans eQTL in Da(1)-12 × Ei-2, because *cis* and *trans* eQTL act additively and are, hence, independent of each other. Likewise, the gene causing the *trans* eQTL represents a novel regulator specifically acting on *CYP81F2* expression. Most other regulators known to be involved in the formation of indole glucosinolates, ATR1/MYB34 (Bender and Fink, 1998; Celenza et al., 2005), IQD1 (Levy et al., 2005), and HIG1/MYB51 (Gigolashvili et al., 2007), appear to affect all indole glucosinolates equally and are encoded by genes whose position does not match the QTL location. A notable exception is AtDof1.1, whose overexpression causes an increase in I3M but a decrease in 4MO-I3M (Skirycz et al., 2006), but again, the location of this gene does not correspond to the eQTL position.

Ecological Consequences of Variation in Indole Glucosinolates

Several studies have indicated that glucosinolates can play an important role in Arabidopsis defense against aphids. High *IQD1* expression which increases total glucosinolate levels made Arabidopsis less attractive to *Myzus persicae* in choice assays (Levy et al., 2005). Likewise, relative performance of two aphid species, *Myzus persicae* and the cabbage aphid (*Brevicoryne brassicae*), was negatively correlated with total glucosinolate content in Arabidopsis genotypes with mutations in different phytohormone signaling pathways (Mewis et al., 2005). Subsequently, this inhibitory effect on aphid proliferation was more specifically attributed to indole glucosinolates and their breakdown products (Kim and Jander, 2007; Kim et al., 2008), although a contribution of aliphatic glucosinolates to aphid resistance cannot yet be excluded. Aphid-induced changes in Arabidopsis leaf glucosinolate profiles are largely negligible, with some variation in short-chain (Mewis et al., 2005; Mewis et al., 2006) or long-chain (Kim and Jander, 2007) aliphatic glucosinolates, but with a high and statistically significant increase in 4MO-I3M levels (Kim and Jander, 2007). Our own analyses were, in general, consistent with these results, but in addition to elevated 4MO-I3M levels upon aphid infestation we also found a higher quantity of 4OH-I3M, which was not measured in the other studies. Changes in the *CYP81F2* mutant glucosinolate profiles in response to aphid feeding paralleled the changes in wildtype, except for 4OH-I3M and 4MO-I3M which remained, as expected, unchanged and at low levels in the mutants.

In assays with artificial diet supplemented with I3M, 4MO-I3M, 1MO-I3M, or the aliphatic glucosinolate sinigrin, 4MO-I3M turned out as the strongest inhibitor of aphid proliferation. Addition of myrosinase, the plant glucosinolate breakdown enzyme, increased the inhibitory effect of I3M and 1MO-I3M to a level equivalent to artificial diet supplemented with 4MO-I3M alone (Kim and Jander, 2007). Thus, 4MO-I3M has a particularly strong impact on *Myzus persicae* proliferation *in vitro*. Our assays, comparing aphid proliferation on *CYP81F2* wildtype *versus* mutant plants show that this also holds true *in planta*, and demonstrate that CYP81F2 is an important player in Arabidopsis defense against aphids.

Perspectives for Targeted Manipulation of Glucosinolate Profiles

Most transcription factor genes known to be involved in the regulation of the indole glucosinolate pathway affect all indole glucosinolates similarly. Likewise, mutations in functional genes typically disturb this entire class of glucosinolates. For example, *cyp79B2 cyp79B3* double knock-outs are almost completely devoid of indole glucosinolates (Zhao et al., 2002), *cyp83B1* was found to have reduced levels of all indole glucosinolates (Naur et al., 2003), and impaired *C-S* lyase function blocked glucosinolate synthesis completely (Mikkelsen et al., 2004). Furthermore, alterations in the indole glucosinolate core pathway have often side effects including aberrant plant growth because pathways for the synthesis of the plant hormone auxin and for indole glucosinolate biosynthesis partially overlap. Thus, genetic alteration of known transcription factors or of structural genes involved in the indole glucosinolate core pathway does not yet allow a targeted manipulation of specific glucosinolates and is rarely likely to yield the desired results, a healthy plant with increased herbivore resistance. The identification of *CYP81F2* controlling the *IGM1* QTL may offer a new perspective for metabolic engineering of glucosinolates against aphids, because mutations in *CYP81F2* specifically act on 4OH-I3M and 4MO-I3M while the levels of all other glucosinolates do not change detectably and plants appear to be perfectly normal. It remains to be seen, however, whether this is also the case when *CYP81F2* expression is artificially increased above wildtype levels.

METHODS

Plant Material and Growth Conditions

The Da(1)-12 × Ei-2 RIL population was developed at the Max Planck Institute for Chemical Ecology, Jena, and has been described previously (Pfalz et al., 2007). All other plant material was obtained from the Nottingham Arabidopsis Stock Center. Seeds were sowed into damp potting medium, covered with clear plastic grow domes, and seeds were stratified for 3-4 days at 6 °C in the dark. Grow domes were removed 5 days after transfer to the light. Seedlings were transferred to 1:3 vermiculate/potting soil mix with 20 ml time release fertilizer (Osmocote) per flat. Plants were grown in 11.5 h day/12.5 h night cycles at 22 °C and 60 % relative humidity (day), and 16 °C and 80 % relative humidity (night) in an environment-controlled growth room. Light was supplied by NH 360 FLX Sunlux ACE bulbs with an intensity of 200 µmol s^{-1} m^{-2}. Assays were, in general, carried out with 3-week old plants.

Glucosinolate Extraction and Analysis

For glucosinolate analyses from DE089 × Ei-2 progeny, 100 mg fresh leaf material was harvested and immediately frozen in liquid nitrogen, lyophilized to dryness and ground to a fine powder with seven 2.3 mm ball bearings in a paint shaker. Glucosinolates were extracted in a 96-well format as described in Kliebenstein et al. (2001a). HPLC separation and identification of extracted desulfo-glucosinolates were carried out according to Kroymann et al. (2001). For glucosinolate analyses of T-DNA insertion line progeny, 1 g fresh leaf material was used and glucosinolates were extracted according to a protocol modified from Brown et al. (2003), to accommodate for reduced 4OH-I3M and 4MO-I3M levels.

Genotyping and DNA Gel Blot Analyses

For mapping purposes, DNA was extracted from 1–2 freeze-dried leaves as described in Kroymann et al. (2001). Genotyping was performed with PCR-based markers (Supplemental Table 2). PCR reactions contained, in general, ca. 30 ng DNA, 2.3 µl 10x PCR buffer (Qiagen), 4 nmol of each dNTP, 1.25 pmol of each of

both primers, 70 nmol $MgCl_2$, and 0.15 U *Taq* DNA polymerase (Qiagen) in a 23 µl volume. Cycling conditions were 94 °C for 2 min, followed by 38 cycles of 94 °C for 15 s, 50 °C or 55 °C for 15 s, and 72 °C for 30 s, with a final extension of 72 °C for 2 min on an Applied Biosystems 9700 Thermocycler. PCR products obtained with primers 5seq01f/r, 5seq02f/r, 5seq03f/r, and 5seq05f/r harbor single nucleotide polymorphisms (Accession nos. AM412009 – AM412016). These products were gel purified with QiaQuick columns (Qiagen) and sequenced directly. The other PCR products were separated on 4 % MetaPhor (BMA) agarose gels.

Progeny from SALK T-DNA insertion lines SALK_123882 and SALK_005882 (Alonso et al., 2003) was tested for the presence of T-DNA insertions in *CYP81F2* (At5g57220). Primers 123882-f and 123882-r produced a PCR product when SALK_123882 progeny carried a *CYP81F2* wildtype allele, and primers LBb1 and SALK_123882-r when plants carried a mutant allele. Similarly, in SALK_005861 progeny primers 005861-f and 005861-r generated a PCR product from the wildtype and primers LBb1 and 005861-f from the mutant allele (Figure 4).

The number of T-DNA insertions in SALK_005861 and SALK_123882 was analyzed with DNA gel blots, using 2 µg X*ba*I-restricted DNA per plant and a fragment of ca. 300 bp amplified from the left border region of the T-DNA with primers T-DNA_L1F and T-DNA-L1R (Supplemental Table 2) as a hybridization probe.

RNA Isolation for Hybridization of Agilent Arrays and for Quantitative Real-Time PCR

Per plant, three leaves were used for RNA extraction. Leaf material was ground to a fine powder in liquid N_2, and total RNA was isolated using the TRIzol Reagent (Invitrogen) according to the manufacturers´ instructions. A DNAse (Turbo DNAse, Ambion) treatment followed to eliminate any contaminating DNA. A second purification step was performed with RNeasy MinElute columns (Qiagen) to remove DNAse and any contaminating polysaccharides and proteins. RNA integrity was verified on an Agilent 2100 Bioanalyzer using RNA Nano chips. RNA quantity was determined on a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies) or with a BioPhotometer 6131 (Eppendorf).

Amplification, Labeling and Hybridization

Total RNA was amplified using the Agilent low input linear amplification kit according to the process outlined by the manufacturer (Agilent). 1-5 µg of amplified target cRNA was labeled with either cy5 or cy3 using the Micromax kit (Applied Biosystems). The labeled material was passed through Zymo RNA Clean-up Kit-5 columns (Zymo Research Corporation) to remove any unincorporated label and eluted in 15-20 µl of RNAse-free water (Ambion). Concentration of labeled cRNA and label incorporation was determined by Nanodrop-1000 spectrophotometer analysis. Labeled material was hybridized onto Agilent Arabidopsis 3 Arrays o/n in a rotating oven at 60 °C and slides were processed as outlined in the Agilent processing manual. Arrays were scanned using the Agilent G2565BA fluorescent microarray scanner. Agilent's feature extraction software (Version 7.5) was used for extracting array data. Further analysis was done with Rosetta Luminator (at MOgene LC) and with GeneSifter (VizXlabs) software.

Quantitative Real-Time PCR

500 ng of DNA-free total RNA was converted into single-stranded cDNA using a mix of random and oligo-dT20 primers according to the ABgene protocol (ABgene). Real-time PCR oligonucleotide primers were designed using the online Primer3 internet based interface (http://frodo.wi.mit.edu). Gene-specific primers were designed on the basis of sequence obtained for *CYP81F2* and several additional genes (At1g54270 = EIF-4A2, At1g34030 = RPS18B, At2g18110 = EF1-beta, At2g29550 = Tubulin beta-7 chain, At5g09810 = Actin7) as potential house-keeping genes to serve as the endogenous control (normalizer). Primer design, qRT-PCR conditions, dissociation curve analysis, tests of dynamic range and quantification were carried out as described previously (Freitak et al., 2007).

The comparative quantitation method (ΔΔ*C*t) was used to contrast the different treatments and tissues (Livak and Schmittgen 2001). *C*t values quantify the number of PCR cycles necessary to amplify a template to a chosen threshold concentration, Δ*C*t values quantify the difference in *C*t values between a test and a control gene for a given sample, and ΔΔ*C*t values are used for the comparison between two samples. ΔΔ*C*t values were transformed to absolute values with $2^{-\Delta\Delta Ct}$ for obtaining relative fold changes. Except for the RIL population, all assays were run in triplicate (biological

replication) and duplicate (technical replication) to control for overall variability. Relative fold changes for each gene were set to 1 for the control.

Expression Constructs and Heterologous Expression in *Sf*9 cells

Total RNA was extracted from Arabidopsis Col-0 with TRIzol (Invitrogen) and reverse-transcribed with SuperScript III Reverse Transcriptase (Invitrogen) according to the manufacturer's instructions. *CYP81F2* cDNA was amplified with AccuPrime Taq polymerase (Invitrogen) with primers: IGM1f and IGM1r (Supplemental Table 2). Primers were chosen such that the PCR product contained a Kozak translation initiation sequence in addition to the start codon. The native stop codon was omitted to include a C-terminal peptide encoded by the vector from the pIB/V5-His TOPO TA Expression Kit (Invitrogen) which enables detection with a V5 antibody. PCR was as follows: 94 °C for 1 min, 30 cycles at 94 °C for 15 s, 55 °C for 30 s, and 72 °C for 1.45 min, followed by a final elongation at 72 °C for 5 min on a PE Applied Biosystems 9700 thermal cycler. Purified product was cloned into a pIB/V5-His TOPO vector and transformed into *E. coli* TOP-10 cells (Invitrogen). Plasmids were isolated with the HiPure Plasmid Filter Midiprep Kit (Invitrogen) and sequenced to verify correct cDNA sequence, reading frame and cloning direction.

*Sf*9 cells (Invitrogen) were grown at 27 °C in *Sf*-900 II SFM with 50 µg/ml Gentamycin (both: Gibco). Per culture dish, 660 µl SFM was mixed with 66 µl Insect GeneJuice Transfection Reagent (Novagen). In parallel, 12 µg of expression construct was added to 660 µl SFM. Plasmid containing solution and transfection reagent were combined, left for 15 min at ambient temperature, and added to *Sf*9 cells. After 4 h incubation at 27 °C SFM was replaced once.

After 48 h cells were suspended in their culture medium, centrifuged with 500 g at 4 °C for 10 min, and washed twice with ice-cold 1 x PBS. Next, cells were re-suspended in 2 ml hypotonic buffer (20 mM Tris-HCl, pH 7.5, 5 mM EDTA, 1 mM DTT and 1x protease inhibitor cocktail (Pierce)) and kept on ice for 20 min. Cells were homogenized and the lysate was mixed with an equal volume of sucrose buffer (20 mM Tris-HCl, pH 7.5, 5 mM EDTA, 1 mM DTT, 500 mM sucrose and protease inhibitors), and centrifuged with 1,200 g at 4 °C for 10 min. The supernatant was stored on ice and the homogenization step was repeated once with the pellet. Supernatants were pooled and centrifuged with 10,000 g at 4 °C for 15 min. The

supernatant from this step was centrifuged with 100,000 g at 4 °C for 1 h. The resulting pellet was re-suspended in 0.5 ml 100 mM phosphate buffer, pH 7.4, with 20 % (v/v) glycerol and protease inhibitors. Finally, heterologous expression was confirmed with SDS/PAGE and Western blot analysis.

Isolation of Native Indol-3-yl-methyl Glucosinolate from Seeds of Dyer's woad (*Isatis tinctoria*)

Isatis tinctoria seeds contain high amounts of I3M (Mohn et al., 2007), and were therefore used for the isolation of intact glucosinolates. 50 g seeds (Saatzucht Quedlinburg GmbH, Germany) were homogenized with 250 ml 80 % (v/v) aqueous methanol using a Polytron PT3100. After centrifugation with 7,000 g at ambient temperature for 10 min, the supernatant was collected. The extraction step was repeated with 200 ml 80 % (v/v) aqueous methanol, and the combined supernatant was loaded onto an anion-exchange column filled with 1 g DEAE-Sephadex A-25. The column was washed three times with 5 ml of a 3 : 2 : 5 (v/v/v) mixture of formic acid, isopropanol and water, and four times with 5 ml water. Glucosinolates were eluted with 25 ml 0.5 M K_2SO_4/3 % (v/v) aqueous isopropanol and collected in 25 ml ethanol (Thies, 1988). This first elution step preferably elutes aliphatic glucosinolates. For isolation of indole glucosinolates an additional 75 ml of eluent was added to the column and the eluate was collected in 100 ml ethanol. This eluate was centrifuged with 3,000 g at ambient temperature for 10 min, and the supernatant was lyophilized with a rotary evaporator. Indole glucosinolates were dissolved in 3 x 0.5 ml water and fractionated on an Agilent HP 1100 Series system, equipped with a SUPELCOSIL LC-18-DB SEMI-PREP 250 x 10 mm, 5 µm, column (Supelco) and a fraction collector. The program for the separation of intact I3M from other glucosinolates was as follows: start with 5 % solvent B (acetonitrile) and 95 % A (0.02 % trifluoroacetic acid), 8 min 13 % B, 8.1 min 95 % B, 10 min 95 % B, 10.1 min 5 % B and 15 min 5 % B. The fraction containing intact I3M was split into several aliquots, lyophilized with a Genevac HT-4X Series II vacuum evaporator (Ipswich), and stored at -20 °C. Prior to enzyme assays, I3M was dissolved in 100 mM KPi (pH 7.4).

Enzyme Assays

CYP81F2 activity was measured in microsomes isolated from *Sf*9 insect cells. Microsomes isolated from *Sf*9 cells expressing a major allergen gene from the European cabbage butterfly (*Pieris rapae*), amplified with primers MA-ExF and MA-ExR (Supplemental Table 2), served as a control (Fischer et al., 2008). A second control contained phosphate buffer instead of microsomes but was otherwise treated the same. Microsome protein concentration was determined with a Bio-Rad DC Protein Assay. Typical protein concentrations of microsomes transfected with *CYP81F2* were 2.5 µg/µl. Concentration of microsomes with control plasmids was adjusted accordingly.

50 µl of transfected microsomes were incubated o/n at ambient temperature with I3M in 100 mM KPi, pH 7.4, 2 mM NADPH, 2 mM glucose-6-phosphate, and 0.1 units of glucose-6-phosphate dehydrogenase in a total volume of 200 µl. The reaction mixture was loaded onto an anion-exchange column filled with DEAE-Sephadex A-25 and washed 3 times with 1 ml 67 % (v/v) aqueous methanol, 2 times with 1 ml water, and 3 times with 1 ml MES buffer, pH 5.2. 50 µl (30 u) sulfatase (Sigma) were added to the column. After 6 h incubation at ambient temperature, samples were eluted with 6 x 0.5 ml water and concentrated on a Genevac HT-4X Series II vacuum evaporator. Finally, the reaction products were dissolved in 220 µl water and 95 µl were injected in an Agilent HP 1100 Series system with the following program: start with 1.5 % solvent B (acetonitrile) and 98.5 % A (water), 1 min 1.5 % B, 6 min 5 % B, 8 min 7 % B, 18 min 21 % B and 23 min 29 % B, 30 min 43 % B, 30.5 min 100 % B, 33 min 100 % B, 33.1 min 1.5 % B, and 38 min 1.5 % B. Identity of the reaction product was confirmed by liquid chromatography-mass spectrometry (LC-MS).

Herbivory Screens with Lepidopterans

Herbivory assays were essentially performed as described in Pfalz et al. (2007). Diamondback moth (*Plutella xylostella*) was obtained from New York State Agricultural Experiment Station Geneva, NY, USA, and a colony was maintained at the Max Planck Institute for Chemical Ecology, Jena, Germany. Cabbage looper (*Trichoplusia ni*) eggs were obtained from Benzon Research (Carlisle, PA), and Beet armyworm (*Spodoptera exigua*) from Bayer Crop Sciences. All larvae were reared for

5–6 days on artificial diet (Shelton et al., 1991) prior to experiments. Cabbage white butterfly (*Pieris brassicae*) eggs were obtained from Seritech (Warwick, UK). After hatching, larvae were pre-experimentally reared on *Brassica napus var. oleifera* for 2–3 days.

Lepidopteran performance was analyzed separately for SALK_005861 and SALK_123882 progeny. Each experiment included several homozygous mutant and wildtype lines and Col-0 wildtype, randomized over 96-celled flats. Plant diameter was recorded prior to herbivory screens. Each plant received one larva and larvae were allowed to move freely. *Plutella xylostella*, *Trichoplusia ni* and *Spodoptera exigua* fed for 2 days, *Pieris brassicae* for 24 hours. Afterwards, the leaf area removed by the insects was assessed visually, and an artificial scale was established to quantify tissue damage, as described in Stotz et al. (2000). In general, the following statistical model was used:

PLANT DAMAGE = CONSTANT + EXPERIMENT + FLAT(EXPERIMENT) + COLUMN + ROW + PLANT DIAMETER + MUTWT

In this linear ANOVA model, COLUMN and ROW are variables to control for position effects which may result from larval movement during the experiments. EXPERIMENT accounts for variation between experimental replicates, FLAT(EXPERIMENT) for variation between flats within an experiment. PLANT DIAMETER controls for potential effects of plant size on insect herbivory. MUTWT estimates the genotype effect, mutant or wildtype, on herbivory.

Aphid Proliferation Screens

A colony of the green peach aphid (*Myzus persicae*) was maintained since 2005 on Col-0 wildtype at the Max Planck Institute for Chemical Ecology in Jena. Prior to experiments aphids were either raised for several generations on 4–5 week old Col-0 wildtype plants or on SALK_123882 *CYP81F2*$^{\Delta/\Delta}$ mutant progeny. Adult aphids were collected and transferred to fresh 5 week old plants of the same genotypes as before. Adults were allowed to proliferate for 2 days. Afterwards, adults were removed and only the aphid progeny remained on the plants. This procedure ensured that all aphids were approximately of the same age (+/- 1 day old) at the time of the experiment.

After 7 days aphids were transferred to 18-day old test plants, with 3 aphids of the same origin per test plant. After 5 further days, *i.e.* within the linear proliferation phase, the number of aphids per plant was counted with the help of a magnifying glass. Test plants were grown randomized in 96-celled flats, with 48 plants per flat and every other cell left empty. Each experiment included three lines derived from homozygous mutant progeny of SALK_005861, two lines from homozygous wildtype progeny of SALK_005861, three lines from homozygous mutant progeny of SALK_123882, two lines derived from homozygous wildtype progeny of SALK_123882, and Col-0 wildtype as test plants, with a sample size 6 - 9 plants per line. Here, the following ANOVA model was used:

APHID NUMBER = CONSTANT + ORIGIN + MUTWT + MUTWT * ORIGIN

Because position effects were absent, it was not necessary to include COLUMN or ROW variables. Also, plant size had no detectable effect on aphid number. ORIGIN accounts for pre-experimental and MUTWT for genotype effects during the experiment.

Sequencing of Genomic At5g57220 from Da(1)-12 and Ei-2

The At5g57220 region was obtained as a single fragment using primers 5-CYP and 3-CYP (Supplemental Table 2). PCR products were gel purified with QiaQuick columns (Qiagen) and cloned into TOPO XL vectors (Invitrogen). Four clones each were sequenced on an automated Applied Biosystems 3730xl DNA Analyzer using BigDye terminators version 3.1 with universal primers m13u and m13f, and with insert-specific primers 23204480F, 23205533R, 23206263R, 23206821R, 23207360R, Seq57720F, and Seq57720R (Supplemental Table 2). The DNAstar software package (DNASTAR Inc.) was used for sequence assembly and alignment. Sequence data are deposited at EMBL (accession nos. FM208178, FM208179).

Statistics Programs

Systat Version 10 (SPSS Inc.) was used for ANOVA. Expression QTL were mapped with Windows QTL Cartographer V2.5 (Wang et al., 2001-2004), using the standard model (Model 6) for composite interval mapping with forward regression, a window size of 10 cM, and 5 background control markers. QTL were scanned at a walk

speed of 0.5 cM. Statistical significance of QTL was assessed by permuting data 1,000 times, with a significance level of 0.05.

Accession Numbers

Sequence data from this article can be found in the EMBL/GenBank data libraries under accession numbers AM412009 – AM412016, FM208178, FM208179.

Supplemental Data

Supplemental Results. QTL Fine-mapping with DE089 × Ei-2 F_3 Progeny
Supplemental Table 1. Calculation of *CYP81F2* Transcript Abundance in Da(1)-12, Ei-2, and DE089.
Supplemental Table 2. Primer Sequences Used in this Work.

Acknowledgements

We thank Grit Kunert for sharing her *Myzus persicae* colony with us, John D'Auria, Michael Reichelt and Jonathan Gershenzon for inspiring discussions and advice on indole glucosinolate biochemistry, and for comments on the manuscript. We are also grateful to Choon Wei (Jeffrey) Wee for his help with heterologous expression of cytochrome P450s in insect cells and Yannick Pauchet for his protocol on microsome isolation. This work was supported by the Max Planck Society and by the Université Paris-Sud in Orsay, France.

FIGURE LEGENDS

Figure 1. Metabolic and Expression QTL for Modified Indole Glucosinolates.
Shown are chromosomes (1 and 5) with significant QTL for 4OH-I3M (dark grey), 4MO-I3M (light grey) and eQTL for *CYP81F2* expression (black). Markers from Pfalz et al. (2007) are indicated on the horizontal axis, and the vertical axis quantifies statistical significance as LOD scores. Horizontal lines represent significance thresholds (~ 2.5 LOD), based on 1,000 permutations of the data. Vertical arrows indicate QTL effect directions; upwards arrows correspond to higher values for the Da(1)-12 allele. Note that higher Δ*C*t values indicate lower expression.

Figure 2. Identification of Candidate Genes for the *IGM1* QTL.
(A) Near isogenic lines used for fine-mapping of *IGM1*. Markers are indicated on top, with distances in Mbp corresponding to the marker position in Col-0. NIL families have fixed genotypes (either D = Da(1)-12 or E = Ei-2) on one side but segregate (seg) at the other side. As an example, family 5C7 is Ei-2 at markers *5FM39* to *5FM09* and segregates at markers *MNC6* and *5FM38*. One NIL family turned out to have only Ei-2 alleles and is not included in the panel.
(B) Statistical support for the 4MO-I3M QTL, given as *F* ratios for the comparison between Da(1)-12 and Ei-2 genotypes ($F_{D, E}$) and when heterozygotes are included ($F_{D, H, E}$).
(C) Differences in transcript abundance between DE089 and Ei-2 in the QTL candidate region under two different experimental conditions. Values are given as average fold changes for Ei-2 *versus* DE089 from three biological replicates each for non-manipulated plants (light grey bars) and herbivory-treated plants (dark grey).

Figure 3. Modified Indole Glucosinolate Content in SALK T-DNA Insertion Lines and in DE089 × Ei-2 NILs Contingent on *CYP81F2* Genotype.
4MO-I3M (light grey) and 4OH-I3M (dark grey) quantity differ in SALK_123882 and SALK_005861 and in DE089 × Ei-2 NILs contingent on the genotype at *CYP81F2*. Vertical lines indicate standard errors. Note that 4MO-I3M and 4OH-I3M content are significantly reduced in plants with mutant alleles but are not completely absent. Note also that heterozygotes from SALK_123882 contain only a little more 4MO-I3M and

4OH-I3M than homozygous mutants, in contrast to heterozygous NILs which display a phenotype intermediate between parental lines.

Figure 4. *CYP81F2* in Different Arabidopsis Accessions.
Small black vertical lines show nucleotide and insertion/deletion polymorphisms and grey vertical lines show amino acid polymorphisms between Ei-2 and Col-0 (top) and between Da(1)-12 and Col-0 (bottom). Positive numbers correspond to an insertion and negative numbers to a deletion relative to the Col-0 reference. Positions of N-terminal membrane anchor and cysteine heme-iron ligand structure are indicated, as well as location and direction of T-DNA insertions in SALK lines. Positions of selected primers for RT-PCR are also shown.

Figure 5. RT-PCR of SALK_lines.
SALK_123882 lines with a T-DNA insertion in At5g57220 lack *CYP81F2* transcript. Shown are results for primers 123882-f and Cyt2R, in comparison to control genes. Each genotype is represented by two biological replicates. We obtained the same result with primer combinations 123882-f/123882-r and Cyt2F/123882-r (Figure 4). By comparison, we found similar *CYP81F2* cDNA quantities for different *CYP81F2* genotypes from SALK_008561.

Figure 6. CYP81F2 Catalyzes the Conversion of I3M to 4OH-I3M.
Compared to buffer controls **(A)** or heterologously expressed major allergen **(B)**, HPLC runs of I3M incubated with heterologously expressed *CYP81F2* **(C)** have an additional peak, corresponding to hydroxylated I3M and having the same retention time as 4OH-I3M in an extract of leaf glucosinolates from Col-0 **(D)**, as shown with a mix of I3M incubated with CYP81F2 and Col-0 leaf glucosinolates **(E)**. All HPLC runs were done with desulfo-glucosinolates. The Figure shows representative examples from several independent experimental replicates. Note that Col-0 leaves contain much more 4MO-I3M than 4OH-I3M. mAU = milli Absorption Units at 226 nm wavelength.

Figure 7. Aphid Proliferation Contingent on *CYP81F2* Genotype.
Shown are aphid numbers (± standard error) with different combinations of pre-experimental rearing (‘Origin’, either on Col-0 or on homozygous SALK_123882

mutants) and experimental test plants (‘Test Plants’, either with *CYP81F2* wildtype alleles (WT) or with mutant alleles (M)). We conducted these assays three times with the same results.

Figure 8. Biosynthesis of Indole Glucosinolates from Tryptophan.
Shown are intermediates and known biosynthesis enzymes. GT: *S*-glucosyltransferase, ST: sulfotransferase. Glucosinolates are boxed. I3M = indol-3-yl-methyl, 4OH-I3M = 4-hydroxy-indol-3-yl-methyl, 4MO-I3M = 4-methoxy-indol-3-yl-methyl, 1OH-I3M = 1-hydroxy-indol-3-yl-methyl, and 1MO-I3M = 1-methoxy-indol-3-yl-methyl glucosinolate. Note that we postulate a two-step synthesis of 1MO-I3M from I3M with 1OH-I3M (grey) as an intermediate.

Insect Species	SALK_005861			SALK_123882		
	N	$F_{df=1}$	*P*	*N*	$F_{df=1}$	*P*
Plutella xylostella	685	0.880	0.349	479	0.300	0.584
Pieris brassicae	88	1.129	0.292	392	0.121	0.728
Trichoplusia ni	516	3.502	0.062	207	0.481	0.489
Spodoptera exigua	246	0.911	0.341	413	0.220	0.639

Table 1. Performance of Lepidopteran Larvae Contingent on *CYP81F2* Genotype.

REFERENCES

Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., Jeske, A., Koesema, E., Meyers, C.C., Parker, H., Prednis, L., Ansari, Y., Choy, N., Deen, H., Geralt, M., Hazari, N., Hom, E., Karnes, M., Mulholland, C., Ndubaku, R., Schmidt, I., Guzman, P., Aguilar-Henonin, L., Schmid, M., Weigel, D., Carter, D.E., Marchand, T., Risseeuw, E., Brogden, D., Zeko, A., Crosby, W.L., Berry, C.C., and Ecker, J.R. (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science **301:** 653-657.

Bak, S., Nielsen, H.L., and Halkier, B.A. (1998). The presence of CYP79 homologues in glucosinolate-producing plants shows evolutionary conservation of the enzymes in the conversion of amino acid to aldoxime in the biosynthesis of cyanogenic glucosides and glucosinolates. Plant Mol. Biol. **38**: 725-734.

Bak, S., and Feyereisen, R. (2001). The involvement of two P450 enzymes, CYP83B1 and CYP83A1, in auxin homeostasis and glucosinolate biosynthesis. Plant Physiol. **127:** 108-118.

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

Barlier, I., Kowalczyk, M., Marchant, A., Ljung, K., Bhalerao, R., Bennett, M., Sandberg, G., and Bellini, C. (2000). The SUR2 gene of *Arabidopsis thaliana* encodes the cytochrome P450 *CYP83B1*, a modulator of auxin homeostasis. Proc. Natl. Acad. Sci. USA **97:** 14819–24.

Bender, J., and Fink, G.R. (1998). A Myb homologue, *ATR1*, activates tryptophan gene expression in Arabidopsis. Proc. Natl. Acad. Sci. USA **95**: 5655–60.

Benderoth, M., Textor, S., Windsor, A.J., Mitchell-Olds, T., Gershenzon, J., and Kroymann, J. (2006). Positive selection driving diversification in plant secondary metabolism. Proc. Natl. Acad. Sci. USA **103:** 9118-9123.

Benderoth, M., Pfalz, M., and Kroymann, J. (2008). Methylthioalkylmalate synthases: genetics, ecology and evolution. Phytochem. Rev. DOI 10.1007/s11101-008-9097-1.

Boerjan, W., Cervera, M.T., Delarue, M., Beeckman, T., Dewitte, W., Bellini, C., Caboche, M., Oncklen, H.V., Montagu, M.V., and Inzé, P. (1995). *Superroot*, a recessive mutation in Arabidopsis, confers auxin overproduction. Plant Cell **7:** 1405–19.

Bones, A.M., and Rossiter, J.T. (2006). The enzymatic and chemically induced decomposition of glucosinolates. Phytochemistry **67:** 1053-1067.

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

Burow, M., Bergner, A., Gershenzon, J., and Wittstock, U. (2007). Glucosinolate hydrolysis in *Lepidium sativum* – identification of the thiocyanate-forming protein. Plant Mol. Biol. **63:** 49-61.

Campos de Quiros, H., Magrath, R., McCallum, D., Kroymann, J., Schnabelrauch, D., Mitchell-Olds, T., and Mithen, R. (2000). α-Keto acid elongation and glucosinolate biosynthesis in *Arabidopsis thaliana*. Theor. Appl. Genet. **101:** 429-437.

Celenza, J.L., Grisafi, P.L., and Fink, G.R. (1995). A pathway for lateral root-formation in *Arabidopsis thaliana*. Genes Dev. **9:** 2131–42.

Celenza, J.L., Quiel, J.A., Smolen, G.A., Merrikh, H., Sivestro, A.R., Normanly, J., and Bender, J. (2005). The Arabidopsis ATR1 Myb transcription factor controls indolic glucosinolate homeostasis. Plant Physiol. **137:** 253-262.

Daxenbichler, M.E., Spencer, G.F., Carlson, D.G., Rose, G.B., Brinker, A.M., and Powell, R.G. (1991). Glucosinolate composition of seeds from 297 species of wild plants. Phytochemistry **30:** 2623-2638.

Delarue, M., Prinsen, E., van Onckelen, H., Caboche, M., and Bellini, C. (1998). *Sur2* mutations of *Arabidopsis thaliana* define a new locus involved in the control of auxin homeostasis. Plant J. **14:** 603–11.

De Vos, M., Kriksunov, K.L., and Jander, G. (2008). Indole-3-acetonitrile production from indole glucosinolates deters oviposition by *Pieris rapae*. Plant Physiol. **146:** 916-926.

Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol. **300:** 1005-1016.

Fahey, J.W., Zalcmann, A.T., and Talalay, P. (2001). The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. Phytochemistry **56:** 5-51.

Freitak, D., Wheat, C.W., Heckel, D.G., and Vogel, H. (2007). Immune system responses and fitness costs associated with consumption of bacteria in larvae of *Trichoplusia ni*. BMC Biology **5:** 56.

Fischer, H.M., Wheat, C.W., Heckel, D.G., and Vogel, H. (2008). Evolutionary origins of a novel host plant detoxification gene in butterflies. Mol. Biol. Evol. **25:** 809-820.

Gigolashvili, T., Berger, B., Mock, H.P., Müller, C., Weisshaar, B., and Flügge, U.I. (2007). The transcription factor HIG1/MYB51 regulates indolic glucosinolate biosynthesis in *Arabidopsis thaliana*. Plant J. **50:** 886-901.

Glawischnig, E., Hansen, B.G., Olsen, C.E., and Halkier, B.A. (2004). Camalexin is synthesized from indole-3-acetaldoxime, a key branching point between primary and secondary metabolism in *Arabidopsis*. Proc. Natl. Acad. Sci. USA **101:** 8245-8250.

Grubb, C.D., Zipp, B.J., Ludwig-Müller, J., Masuno, M.N., Molinski, T.F., and Abel, S. (2004). Arabidopsis glucosyltransferase UGT74B1 functions in glucosinolate biosynthesis and auxin homeostasis. Plant J. **40:** 893–908.

Guengerich, F.P. (2001). Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. Chem. Res. Toxicol. **14:** 611-650.

Halkier, B.A., and Gershenzon, J. (2006). Biology and biochemistry of glucosinolates. Annu. Rev. Plant Biol. **57:** 303-333.

Hansen, C.H., Du, L.C., Naur, P., Olsen, C.E., Axelsen, K.B., Hick, A.J., Pickett, J.A., and Halkier, B.A. (2001). CYP83B1 is the oxime-metabolizing enzyme in the glucosinolate pathway in *Arabidopsis*. J. Biol. Chem. **276:** 24790–24796.

Hull, A.K., Vij, R., and Celenza, J.L. (2000). *Arabidopsis* cytochrome P450s that catalyze the first step of tryptophan-dependent indole-3-acetic acid biosynthesis. Proc. Natl. Acad. Sci. USA **97:** 2379-2384.

Husebye, H., Chadchawan, S., Winge, P., Thangstad, O.P., and Bones, A.M. (2002). Guard cell- and phloem idioblast-specific expression of thioglucoside glucohydrolase 1 (myrosinase) in Arabidopsis. Plant Physiol. **128:** 1180-1188.

Keurentjes, J.J.B., Fu, J., de Vos, C.H., Lommen, A., Hall, R.D., Bino, R.J., van der Plas, L.H.W., Jansen, R.C., Vreugdenhil, D., and Koornneef, M. (2006). The genetics of plant metabolism. Nat. Genet. **38:** 842-849.

Kim, J.H., and Jander, G. (2007). *Myzus persicae* (green peach aphid) feeding on Arabidopsis induces the formation of a deterrent indole glucosinolate. Plant J. **49:** 1008-1019.

Kim, J.H., Lee, B.W., Schroeder, F.D., and Jander, G. (2008). Identification of indole glucosinolate breakdown products with antifeedant effects on *Myzus persicae* (green peach aphid). Plant J. **54:** 1015-1026.

King, J.J., Stimart, D.P., Fisher, R.H., and Bleecker, A.B. (1995). A mutation altering auxin homeostasis and plant morphology in Arabidopsis. Plant Cell **7:** 2023–37.

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

Kliebenstein, D.J., Gershenzon, J., and Mitchell-Olds, T. (2001b). Comparative quantitative trait loci mapping of aliphatic, indolic and benzylic glucosinolate production in *Arabidopsis thaliana* leaves and seeds. Genetics **159:** 359-370.

Kliebenstein, D.J., Lambrix, V.M, Reichelt, M., Gershenzon, J., and Mitchell-Olds, T. (2001c). Gene duplication in the diversification of secondary metabolism: Tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in Arabidopsis. Plant Cell **13:** 681-693.

Kliebenstein, D.J., Pedersen, D., Barker, B., and Mitchell-Olds, T. (2002). Comparative analysis of quantitative trait loci controlling glucosinolates, myrosinase and insect resistance in *Arabidopsis thaliana*. Genetics **161:** 325-332.

Kliebenstein, D.J., Kroymann, J., and Mitchell-Olds, T. (2005). The glucosinolate-myrosinase system in an ecological and evolutionary context. Curr. Opin. Plant Biol. **8:** 264-271.

Koroleva, O.A., Davies. A., Deeken, R., Thorpe, M.R., Tomos, A.D., and Hedrich, R. (2000). Identification of a new glucosinolate-rich cell type in Arabidopsis flower stalk. Plant Physiol. **124:** 599-608.

Kroymann, J., Textor, S., Tokuhisa, J.G., Falk, K.L., Bartram, S., Gershenzon, J., and Mitchell-Olds, T. (2001). A gene controlling variation in Arabidopsis glucosinolate composition is part of the methionine chain elongation pathway. Plant Physiol. **127:** 1077-1088.

Kroymann, J., Donnerhacke, S., Schnabelrauch, D., and Mitchell-Olds, T. (2003). Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait locus. Proc. Natl. Acad. Sci. USA **100:** 14587-14592.

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

Lambrix, V., Reichelt, M., Mitchell-Olds, T., Kliebenstein, D.J., and Gershenzon, J. (2001). The Arabidopsis epithiospecifier protein promotes the hydrolysis of glucosinolates to nitriles and influence *Trichoplusia ni* herbivory. Plant Cell **13:** 2793-2807.

Lehman, A., Black, R., and Ecker, J.R. (1996). *HOOKLESS1*, an ethylene response gene, is required for differential cell elongation in the *Arabidopsis* hypocotyl. Cell **85:** 183–94.

Levy, M., Wang, Q., Kaspi, R., Parrella, M.P., and Abel, S. (2005). Arabidopsis IQD1, a novel calmodulin-binding nuclear protein, stimulates glucosinolate accumulation and plant defense. Plant J. **43:** 79-96.

Li, J., Chen, S., Zhu, L., and Last, R.L. (1995). Isolation of cDNAs encoding the tryptophan pathway enzyme indole-3-glycerol phosphate synthase from *Arabidopsis thaliana*. Plant Physiol. **108:** 877-878.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods **25:** 402-408.

Ljung, K., Hull, A.K., Celenza, J.L., Yamada, M., Estelle, M., Normaly, J., and Sandberg, G. (2005). Sites and regulation of auxin biosynthesis in Arabidopsis roots. Plant Cell **17:** 1090–104.

Lüthy, B., Matile, P.H. (1984). The mustard oil bomb - rectified analysis of the subcellular organization of the myrosinase system. Biochem. Physiol. Pfl. **179:** 5-12.

Matile, P.H. (1980). "Die Senfölbombe": Zur Kompartmentierung des Myrosinasesystems. Biochem. Physiol. Pflanz. **175:** 722-731.

Mewis, I., Appel, H.M., Hom, A., Raina, R., and Schultz, J.C. (2005). Major signaling pathways modulate Arabidopsis glucosinolate accumulation and response to both phloem-feeding and chewing insects. Plant Physiol. **138:** 1149-1162.

Mewis, I., Tokuhisa, J.G., Schultz, J.C., Appel, H.M., Ulrichs, C., and Gershenzon, J. (2006) Gene expression and glucosinolate accumulation in *Arabidopsis thaliana* in response to generalist and specialist herbivores of different feeding guilds and the role of defense signaling pathways. Phytochemistry **67:** 2450-2462.

Mikkelsen, M.D., Hansen, C.H., Wittstock, U., and Halkier, B.A. (2000). Cytochrome P450 CYP79B2 from *Arabidopsis* catalyzes the conversion of tryptophan to indole-3-acetaldoxime, a precursor of glucosinolates and indole-3-acetic acid. J. Biol. Chem. **43:** 33712-33717.

Mikkelsen, M.D., Naur, P., and Halkier, B.A. (2004). *Arabidopsis* mutants in the *C-S* lyase of glucosinolate biosynthesis establish a critical role for indole-3-acetaldoxime in auxin homeostasis. Plant J. **37:** 770–77.

Mohn T., Cutting B., Ernst B., and Hamburger M. (2007). Extraction and analysis of intact glucosinolates – A validated pressurized liquid extraction/liquid

chromatography – mass spectrometry protocol for *Isatis tinctoria*, and qualitative analysis of other cruciferous plants. J. Chromatogr. A **1166:** 142-151.

Naur, P., Petersen, B.L., Mikkelsen, M.D., Bak, S., Rasmussen, H., Olsen, C.E., and Halkier, B.A. (2003). CYP83A1 and CYP83B1, two nonredundant cytochrome P450 enzymes metabolizing oximes in the biosynthesis of glucosinolates in Arabidopsis. Plant Physiol. **133:** 63-72.

Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. **10:**1-6.

Paquette, S., Bak, S., and Feyereisen, R. (2000). Intron-exon organization and phylogeny in a large superfamily, the paralogous cytochrome P450 genes of *Arabidopsis thaliana*. DNA Cell Biol. **19:** 307-317.

Pfalz, M., Vogel, H., Mitchell-Olds, T., and Kroymann, J. (2007). Mapping of QTL for resistance against the crucifer specialist insect herbivore *Pieris brassicae* in a new Arabidopsis inbred line population, Da(1)-12 × Ei-2. PLoS ONE **2:** e578.

Piotrowski, M., Schemenewitz, A., Lopukhinat, A., Müller, A., Janowitz, T., Weiler, E.W., and Oecking, C. (2004). Desulfoglucosinolate sulfotransferases from *Arabidopsis thaliana* catalyze the final step in the biosynthesis of the glucosinolate core structure. J. Biol. Chem. **279:** 50717–25.

Pollard, D.G. (1972). Plant penetration by feeding aphids (Hemiptera, Aphidoidea): a review. Bull. Entomol. Res. **62:** 631-714.

Raybould, A.F, and Moyes, C.L. (2001). The ecological genetics of aliphatic glucosinolates. Heredity **87:** 383-391.

Reichelt, M., Brown, P.D., Schneider, B., Oldham, N.J., Stauber, E., Tokuhisa, J., Kliebenstein, D.J., Mitchell-Olds, T., and Gershenzon, J. (2002). Benzoic acid glucosinolate esters and other glucosinolates from *Arabidopsis thaliana*. Phytochemistry **59:** 663-671.

Shelton, A.M., Cooley, R.J., Kroening, M.K., Wilsey, W.T., and Eigenbrode, S.D. (1991). Comparative analysis of two rearing procedures for diamond-back moth, *Plutella xylostella* (Lepidoptera: Plutellidae*)*. J. Entomol. Sci. **26:** 17-26.

Skirycz, A., Reichelt, M., Burow, M., Birkemeyer, C., Rolcik, J., Kopka, J., Zanor, M.I., Gershenzon, J., Strnad, M., Szopa, J., Mueller-Roeber, B., and Witt, I. (2006). DOF transcription factor AtDof1.1 (OBP2) is part of a regulatory network controlling glucosinolate biosynthesis in Arabidopsis. Plant J. **47:** 10-24.

Smolen, G., and Bender, J. (2002). Arabidopsis cytochrome P450 *cyp83B1* mutations activate the tryptophan biosynthetic pathway. Genetics **160:** 323–32.

Stotz, H.U., Pittendrigh, B.R, Kroymann, J., Weniger, K., Fritsche, J., Bauke, A., and Mitchell-Olds, T. (2000). Induced plant defense responses against chewing insects. Ethylene signaling reduces resistance of Arabidopsis against Egyptian cotton worm but not diamondback moth. Plant Physiol. **124:** 1007-1017.

Thangstad, O.P., Gilde, B., Chadchawan, S., Seem, M., Husebye, H. Bradley, D., and Bones, A.M. (2004). Cell specific cross-species expression of myrosinases in *Brassica napus*, *Arabidopsis thaliana* and *Nicotiana tabacum*. Plant Mol. Biol. **54:** 597-611.

Thies, W. (1988). Isolation of sinigrin and glucotropaeolin from cruciferous seeds. Fat.Sci. Technol. **90:** 311-314.

Wang, S., Basten, C.J., and Zeng, Z.B. (2001 – 2004). Windows QTL cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC (http://statgen.ncsu.edu.qtlcart/WQTLCart.htm).

Wentzell, A.M, Rowe, H.C, Hansen, B.G., Ticconi, C., Halkier, B.A., and Kliebenstein, D.J. (2007). Linking metabolic QTLs with network and *cis*-eQTLs controlling biosynthetic pathways. PLoS Genetics **3:** e162.

Werck-Reichhart, D., and Feyereisen, R. (2000). Cytochromes P450: a success story. Genome Biol. **1:** doi:10.1186/gb-2000-1-6-reviews3003 (http://genomebiology.com/2000/1/6/reviews/3003).

Werck-Reichhart, D., Bak, S., and Paquette, S. (2002). Cytochromes P450. *In* Somerville, C.R., and Meyerowitz, E.M. (eds.) The Arabidopsis Book. American Society of Plant Biologists, Rockville, MD, doi: 10.1199/tab.0028 (http://www.aspb.org/publications/arabidopsis/).

Wetter, L.R., and Chisholm, M.D. (1968). Sources of sulfur in thioglucosides of various higher plants. Can. J. Biochem. **46:** 931-935.

Zhang, Z., Ober, J.A., and Kliebenstein, D.J. (2006). The gene controlling the quantitative trait locus *EPITHIOSPECIFIER MODIFIER1* alters glucosinolate hydrolysis and insect resistance in *Arabidopsis*. Plant Cell **18:** 1524-1536.

Zhao, J., Last, R.L. (1995). Immunological characterization and chloroplast localization of the tryptophan biosynthetic enzymes of the flowering plant *Arabidopsis thaliana*. J. Biol. Chem. **270:** 6081-6087.

Zhao, Y., Hull A.K., Gupta, N.R., Goss, K.A., Alonso, J., Ecker, J.A., Normanly, J., Chory, J., and Celenza, J.L. (2002). Trp-dependent auxin biosynthesis in *Arabidopsis*: involvement of cytochrome P450s CYB79B2 and CYP79B3. Genes Dev. **16:** 3100-3112.

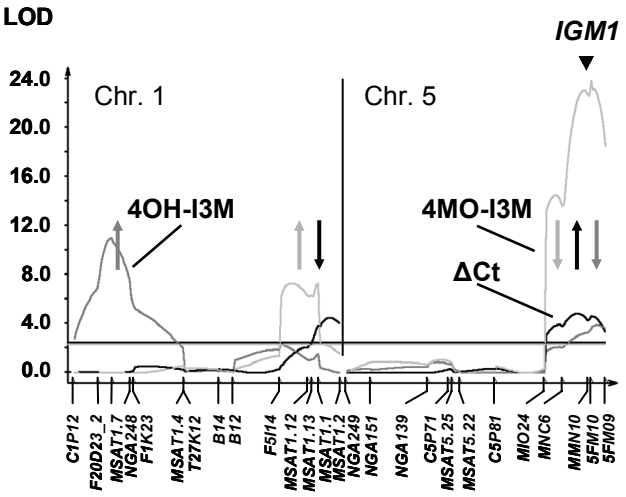

Figure 1. Metabolic and Expression QTL for Modified Indole Glucosinolates.

Shown are chromosomes (1 and 5) with significant QTL for 4OH-I3M (dark grey), 4MO-I3M (light grey) and eQTL for *CYP81F2* expression (black). Markers from Pfalz et al. (2007) are indicated on the horizontal axis, and the vertical axis quantifies statistical significance as LOD scores. Horizontal lines represent significance thresholds (~ 2.5 LOD), based on 1,000 permutations of the data. Vertical arrows indicate QTL effect directions; upwards arrows correspond to higher values for the Da(1)-12 allele. Note that higher ΔCt values indicate lower expression.

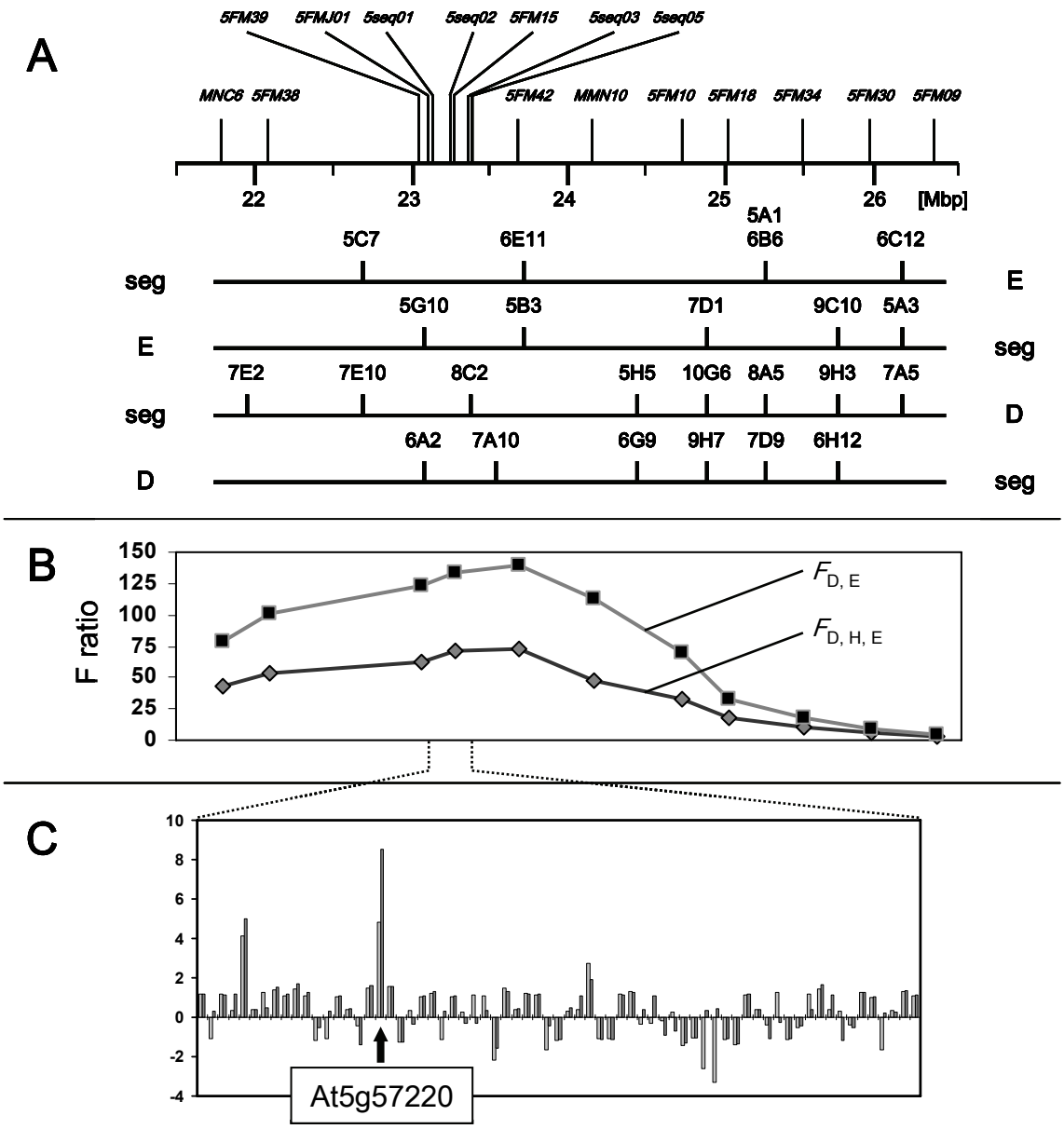

Figure 2. Identification of Candidate Genes for the *IGM1* QTL.

(A) Near isogenic lines used for fine-mapping of *IGM1*. Markers are indicated on top, with distances in Mbp corresponding to the marker position in Col-0. NIL families have fixed genotypes (either D = Da(1)-12 or E = Ei-2) on one side but segregate (seg) at the other side. As an example, family 5C7 is Ei-2 at markers *5FM39* to *5FM09* and segregates at markers *MNC6* and *5FM38*. One NIL family turned out to have only Ei-2 alleles and is not included in the panel.
(B) Statistical support for the 4MO-I3M QTL, given as *F* ratios for the comparison between Da(1)-12 and Ei-2 genotypes ($F_{D, E}$) and when heterozygotes are included ($F_{D, H, E}$).
(C) Differences in transcript abundance between DE089 and Ei-2 in the QTL candidate region under two different experimental conditions. Values are given as average fold changes for Ei-2 *versus* DE089 from three biological replicates each for non-manipulated plants (light grey bars) and herbivory-treated plants (dark grey).

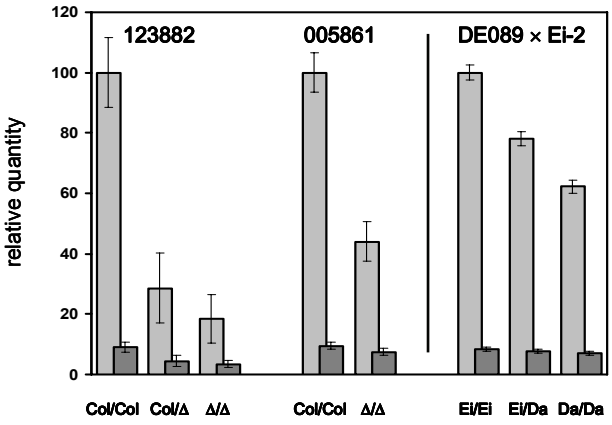

Figure 3. Modified Indole Glucosinolate Content in SALK T-DNA Insertion Lines and in DE089 × Ei-2 NILs Contingent on *CYP81F2* Genotype.

4MO-I3M (light grey) and 4OH-I3M (dark grey) quantity in SALK_123882 and SALK_005861 and in DE089 × Ei-2 NILs contingent on the genotype at *CYP81F2*. Vertical lines indicate standard errors. Note that 4MO-I3M and 4OH-I3M content are significantly reduced in plants with mutant alleles but are not completely absent. Note also that heterozygotes from SALK_123882 contain only a little more 4MO-I3M and 4OH-I3M than homozygous mutants, in contrast to heterozygous NILs which display a phenotype intermediate between parental lines.

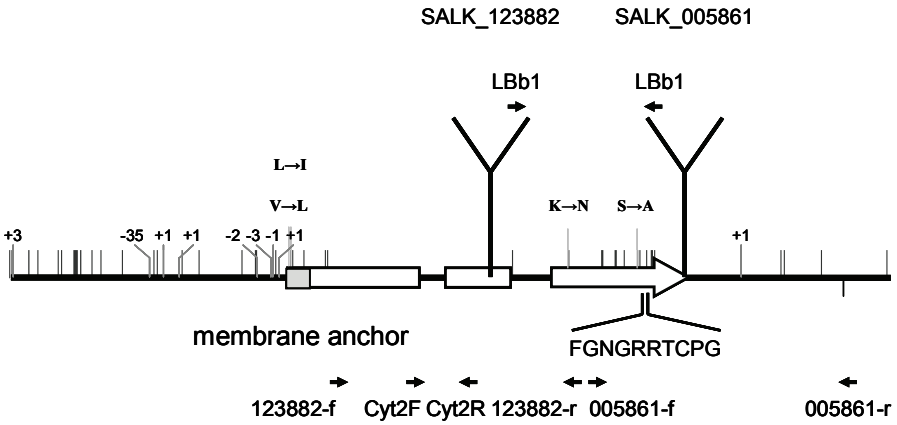

Figure 4. *CYP81F2* in Different Arabidopsis Accessions.

Small black vertical lines show nucleotide and insertion/deletion polymorphisms and grey vertical lines show amino acid polymorphisms between Ei-2 and Col-0 (top) and between Da(1)-12 and Col-0 (bottom). Positive numbers correspond to an insertion and negative numbers to a deletion relative to the Col-0 reference. Positions of N-terminal membrane anchor and cysteine heme-iron ligand structure are indicated, as well as location and direction of T-DNA insertions in SALK lines. Positions of selected primers for RT-PCR are also shown.

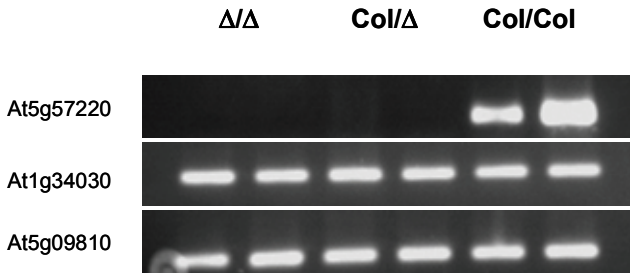

Figure 5. RT-PCR of SALK_lines.

SALK_123882 lines with a T-DNA insertion in At5g57220 lack *CYP81F2* transcript. Shown are results for primers 123882-f and Cyt2R, in comparison to control genes. Each genotype is represented by two biological replicates. We obtained the same result with primer combinations 123882-f/123882-r and Cyt2F/123882-r (Figure 4). By comparison, we found more or less identical *CYP81F2* cDNA quantities for different *CYP81F2* genotypes from SALK_008561.

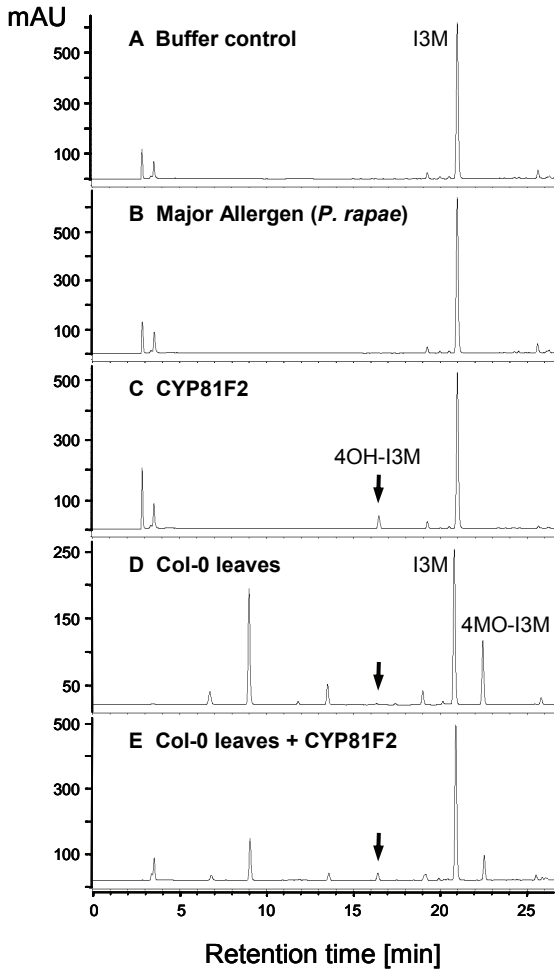

Figure 6. CYP81F2 Catalyzes the Conversion of I3M to 4OH-I3M.

Compared to buffer controls **(A)** or heterologously expressed major allergen **(B)**, HPLC runs of I3M incubated with heterologously expressed *CYP81F2* **(C)** have an additional peak, corresponding to hydroxylated I3M and having the same retention time as 4OH-I3M in an extract of leaf glucosinolates from Col-0 **(D)**, as shown with a mix of I3M incubated with CYP81F2 and Col-0 leaf glucosinolates **(E)**. All HPLC runs were done with desulfo-glucosinolates. The Figure shows representative examples from several independent experimental replicates. Note that Col-0 leaves contain much more 4MO-I3M than 4OH-I3M. mAU = milli Absorption Units at 226 nm wavelength.

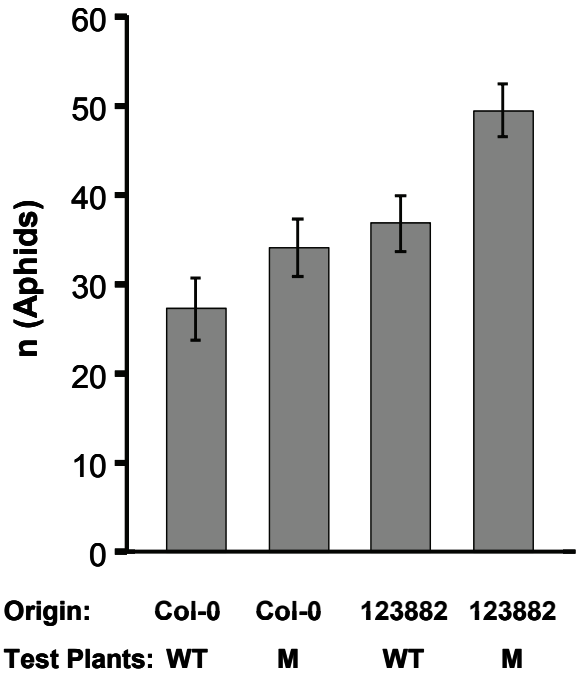

Figure 7. Aphid Proliferation Contingent on *CYP81F2* Genotype.

Shown are aphid numbers (± standard error) with different combinations of pre-experimental rearing ('Origin', either on Col-0 or on homozygous SALK_123882 mutants) and experimental test plants ('Test Plants', either with *CYP81F2* wildtype alleles (WT) or with mutant alleles (M)). We conducted these assays three times with the same results.

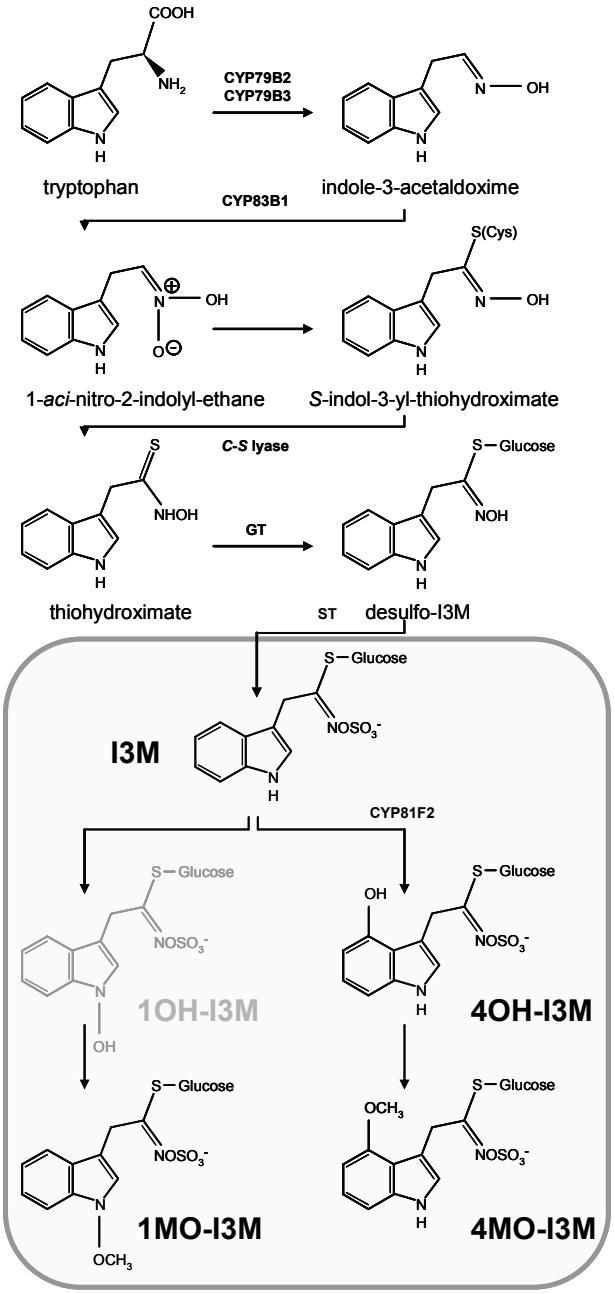

Figure 8. Biosynthesis of Indole Glucosinolates from Tryptophan.

Shown are intermediates and known biosynthesis enzymes. GT: *S*-glucosyltransferase, ST: sulfotransferase. Glucosinolates are boxed. I3M = indol-3-yl-methyl, 4OH-I3M = 4-hydroxy-indol-3-yl-methyl, 4MO-I3M = 4-methoxy-indol-3-yl-methyl, 1OH-I3M = 1-hydroxy-indol-3-yl-methyl, and 1MO-I3M = 1-methoxy-indol-3-yl-methyl glucosinolate. Note that we postulate a two-step synthesis of 1MO-I3M from I3M with 1OH-I3M (grey) as an intermediate.

Supplementary Results:

QTL Fine-mapping with DE089 × Ei-2 F_3 Progeny

The 4MO-I3M QTL was segregating in families 6A2, 5G10, 8C2, and 5H5, but not in families 5C7, 7E10, 7A10, and 6G9 (Figure 2B). Among the families not segregating for the glucosinolate phenotype, 5C7 had an Ei-2 genotype at *5FM39* and family 7E10 had a Da(1)-12 genotype at *5FM39* but both families segregated at *5FM38*, pointing to a QTL location downstream of *5FM38*. Family 6G9 was Da(1)-12 at *MMN10* but segregated at *5FM10*. Likewise, 7A10 did not segregate at *5FM15* but segregated at *5FM42*. These data allowed us to position the 4MO-I3M QTL upstream of *5FM42*. Among the families segregating for the glucosinolate phenotype, 6A2 was Da(1)-12 at *5FM39* and family 5G10 had an Ei-2 genotype at this locus, but both families segregated at *5FM15*, indicating that the 4MO-I3M QTL was downstream of *5FM39*. Family 8C2 carried Da(1)-12 alleles at *5FM42*, but segregated at *5FM15*, thus providing support for a QTL position upstream of *5FM42*.

Comparison	Replicate A	$\Delta Ct_{probe-control}$(A)	Replicate B	$\Delta Ct_{probe-control}$(B)	$\Delta\Delta Ct_{(A-B)}$	$2^{-(\Delta\Delta Ct)}$	Average	SE
DE089 *vs.* Da(1)-12	DE089.1	5.24	Da(1)-12.1	3.87	1.37	0.39	**0.308**	**0.041**
	DE089.1	5.24	Da(1)-12.2	3.40	1.84	0.28		
	DE089.1	5.24	Da(1)-12.3	4.30	0.94	0.52		
	DE089.2	5.50	Da(1)-12.1	3.87	1.63	0.32		
	DE089.2	5.50	Da(1)-12.2	3.40	2.10	0.23		
	DE089.2	5.50	Da(1)-12.3	4.30	1.20	0.44		
	DE089.3	6.24	Da(1)-12.1	3.87	2.37	0.19		
	DE089.3	6.24	Da(1)-12.2	3.40	2.84	0.14		
	DE089.3	6.24	Da(1)-12.3	4.30	1.94	0.26		
DE089 *vs.* Ei-2	DE089.1	5.24	Ei-2.1	4.18	1.06	0.48	**0.309**	**0.057**
	DE089.1	5.24	Ei-2.2	2.57	2.67	0.16		
	DE089.1	5.24	Ei-2.3	4.39	0.85	0.55		
	DE089.2	5.50	Ei-2.1	4.18	1.32	0.40		
	DE089.2	5.50	Ei-2.2	2.57	2.93	0.13		
	DE089.2	5.50	Ei-2.3	4.39	1.11	0.46		
	DE089.3	6.24	Ei-2.1	4.18	2.06	0.24		
	DE089.3	6.24	Ei-2.2	2.57	3.67	0.08		
	DE089.3	6.24	Ei-2.3	4.39	1.85	0.28		
Da(1)-12 *vs.* Ei-2	Da(1)-12.1	3.87	Ei-2.1	4.18	-0.31	1.24	**1.070**	**0.195**
	Da(1)-12.1	3.87	Ei-2.2	2.57	1.30	0.41		
	Da(1)-12.1	3.87	Ei-2.3	4.39	-0.52	1.43		
	Da(1)-12.2	3.40	Ei-2.1	4.18	-0.78	1.72		
	Da(1)-12.2	3.40	Ei-2.2	2.57	0.83	0.56		
	Da(1)-12.2	3.40	Ei-2.3	4.39	-0.99	1.99		
	Da(1)-12.3	4.30	Ei-2.1	4.18	0.12	0.92		
	Da(1)-12.3	4.30	Ei-2.2	2.57	1.73	0.30		
	Da(1)-12.3	4.30	Ei-2.3	4.39	-0.09	1.06		

Supplementary Table 1. Calculation of *CYP81F2* Transcript Abundance in Da(1)-12, Ei-2, and DE089 from qRT-PCR Data.

Name	Sequence (5'→3')	Purpose
MNC6f	GTTTGGGTCCAATGATAAAATC	QTL fine mapping
MNC6r	GCCTATTGGGCTGAGTTTTC	
5FM38f	CTCTTTCCAGGCTGGTTTGA	
5FM38r	CCTTACAAACATACTCAAATTAAC	
5FM39f	CATTAGCCATTCTCGCACTTTC	
5FM39r	TGTATATTCTGGGCAGGAAAC	
5FMJ01f	GAGTGGGGAATTATACAACAAT	
5FMJ01r	CGCAAAATTATTTGATGAACTC	
5seq01f	TGAAAACCACTAAAACTGAAGC	
5seq01r	CAACGCAGAGAAGTGTTAGTC	
5seq02f	CATTCCAACCATAGCAGCATC	
5seq02r	TCCTCTTTGCATTAGCATTTAC	
5FM15f	GTTGGTCTGTCCTAATTCATAC	
5FM15r	CCTATTCCCAAAGATGAAGAAC	
5seq03f	CTATCTTGTTACCATTCCTTAG	
5seq03r	GAATCGCTCTTACAGTTTAGC	
5seq05f	TCGGTGAAGTTGGGAATTTAG	
5seq05r	GGACTCATAGGCTTCAAGTG	
5FM42f	GAGTTTGAGATTGCGCATTTTG	
5FM42r	TCTCAATCACACTTCATCGTC	
MMN10f	CAGTGTCGGCTAATTTCGAC	
MMN10r	CAGTCGACATTTCAAAGGTTC	
5FM10f	GATTTGACGACTGATTACATAAC	
5FM10r	GCTTGAAATTTGTGTGTATTGTC	
5FM18f	CAAGTAGGAGCGTTTCACATC	
5FM18r	TGTTATACATCCGCGCCAAC	
5FM34f	GAATTAATCTCTATTTGCCAAAG	
5FM34r	GGATAACGAAGGAGATAACAAG	
5FM30f	CAATTTACACACTCTGTAATTGG	
5FM30r	CAACCTTCACTTGAGCATTTG	
5FM09f	CAATTTCTTGTTATCTGCTTATG	
5FM09r	CCATTGCCATATGTTTCCCTC	
LBb1	AGTTGCAGCAAGCGGTCCACGC	Verification of T-DNA insertions in SALK lines
005861-f	GACTTTTCGGTTGTGTCCAG	
005861-r	GTTCTTTCAGGCATTTCATCG	
123882-f	TCGCCAAGTCGTAGTGATCTC	Verification of T-DNA insertions in SALK lines; RT-PCR
123882-r	TGCCTTTTTCAACACTTCAGG	
T-DNA_L1F	TCCAGTTTGGAACAAGAGTCCA	Amplification of a T-DNA fragment as a hybridization probe
T-DNA-L1R	CCTGGCGTTACCCAACTTAATC	
5-CYP	CGTCGTCACCTTCGATGAAC	Amplification of genomic DNA containing At5g57220
3-CYP	GAGAAATGTTGAATCTCCCATC	
23204480F	GCATCTATTTATCGGACGGGAG	Sequencing of the At5g57220 region
23205533R	GACGGTTAGAGGAGAGGATTTC	
23206263R	CACACACAAGTTGTTAGTTCATC	
23206821R	CATAGCCATTCCAGGATTCTC	
23207360R	GCTTCTTCCGATTTGCTTTCAC	
Seq57720F	GTGAAAGCACTCGGCGAAG	
Seq57720R	GGCAAACGGCTCGTAATTG	
Cyt2F	GTCACAGGGAGACGCTACTAC	qRT-PCR and RT-PCR, test gene: CYP81F2 (At5g57220)
Cyt2R	CACCACTGTTGTCATTGATGTC	
RPS18b-F	ATTGTTGCAAACCCAAGACAG	qRT-PCR and RT-PCR, control gene: At5g09810 (RPS18)
RPS18b-R	CAGACCACGATGGTTTCTGAT	
EIF4-F	AGCTTGACTATGCCCTTCTCC	qRT-PCR, control gene: At1g54270 (EIF-4A2)
EIF4-R	AACACAGGCATGAACCTTGAC	
At2g18110_F	ATCGACACCATGATCGAAGAG	qRT-PCR, control gene: At2g18110 (EF 1-beta)
At2g18110_R	AGCCCTCCCACAGAGTACAAT	
At2g29550_F	CCATGTCTGGTGTAACCTGCT	qRT-PCR, control gene: At2g29550 (Tubulin beta-7 chain)
At2g29550_R	CTGAGATCCACGAGAGGTGAG	
At5g09810_F	ATTCCGTTGTCCTGAGGTTCT	qRT-PCR and RT-PCR, control gene: At5g09810 (Actin 7)
At5g09810_R	CCACCACTGAGAACGATGTTT	
IGM1f	GTGATGGATTACGTTTTGATTGTTTTGC	Heterologous expression of *CYP81F2*
IGM1r	AGCCAAGAGATTAGTCATAATGG	
MA-ExF	ATGAAGGCAGCCGTAGTTC	Heterologous expression of a major allergen gene from *Pieris rapae*
MA-ExR	TTGTCCCCAGAGGGTTGG	

Supplementary Table 2. Primer Sequences Used in this Work.

4. Discussion

4.1 Mapping of QTL for resistance against crucifer specialists

Insect herbivores specialized on glucosinolate-producing host plants have evolved counteradaptations that render the glucosinolate-myrosinase system ineffective. *Plutella xylostella* larvae benefit from expressing a sulfatase in their gut which blocks the myrosinase-catalyzed glucosinolate hydrolysis by desulfating glucosinolates (Ratzka et al., 2002). *Pieris rapae* and other Pieridae utilize a nitrile-specifier protein that avoids the formation of the more toxic isothiocyanates (Wittstock et al., 2004; Wheat et al., 2007). This leads to the question whether and how cruciferous plants defend themselves against specialist insects.

Several studies have indicated that cruciferous plants are not defenseless against specialists. However, it is not clear whether the glucosinolate-myrosinase system might be involved. Herbivory by the cabbage stem flea beetle *Psylliodes chrysocephala* differed significantly among various oilseed rape lines, but damage was not explained by variation in the glucosinolate content (Bartlet et al., 1996). Similarly, variation in resistance to the larvae of the turnip root fly *Delia floralis* was not linked to total or individual glucosinolate content (Birch et al., 1992). Other studies, however, reported herbivory by lepidopteran specialists to be affected by myrosinase activity or to be positively correlated with plant glucosinolate content (Giamoustaris & Mithen, 1995; Li et al., 2000; Kliebenstein et al., 2005). By contrast, damage by the specialist flea beetles, *Psylliodes convexior* and *Phyllotreta zimmermani*, decreased on *A. thaliana* lines with high glucosinolate levels (Mauricio, 1998).

Consistent with other studies, *Manuscript I* showed that resistance to herbivory by the specialist lepidopterans *Pieris brassicae* and *P. xylostella* varied significantly among Arabidopsis accessions, reinforcing the idea that cruciferous plants can defend themselves against specialists. Furthermore, plant resistance against both herbivores was positively correlated, indicating that some defense-related factors influence herbivory by both specialists similarly.

The genetic basis for variation in resistance against *P. brassicae* and *P. xylostella* larvae was investigated in a Da(1)-12 × Ei-2, a new Arabidopsis recombinant inbred line population (*Manuscript I*). QTL mapping revealed several QTL for resistance against *P. brassicae*, and one weak QTL for *Plutella*. Known defense-related traits such as glucosinolate accumulation or myrosinase activity were also analyzed but did not explain the

herbivory QTL convincingly, indicating that resistance against these specialist insect herbivores has a different genetic basis than resistance against generalist insect herbivores. Similarly, trichome density, a defense trait found to be negatively correlated with generalist and specialist insect herbivores in several studies (Agren & Schemske, 1993; Mauricio, 1998; Handley et al., 2005) had no detectable impact on *P. brassicae* or *P. xylostella* herbivory. Of course, glucosinolates or their hydrolysis products can be important at other levels in the interaction between plants and specialist insects. Indeed, adult females of several specialist insect herbivores, including *P. rapae* and *P. xylostella*, use glucosinolates or their hydrolysis products as cues for host localization and oviposition (Huang & Renwick, 1994; Pivnick et al., 1994; De Jong & Städler, 1999; Renwick et al., 2006; De Vos et al., 2008).

Ultimately, the specialist herbivory QTL detected in *Manuscript I* need to be fine mapped and cloned to identify the genes that underlie resistance against specialist herbivores. However, these analyses require large sample sizes and are very time-consuming, which precluded such an approach during this thesis.

4.2 *Plutella*-induced changes in the Arabidopsis transcriptome

Gene expression microarrays provide a valuable tool for studying herbivore-elicited transcriptional regulation of many genes simultaneously and have been successfully applied to investigate transcriptome changes in response to herbivory in different plants (Reymond et al., 2004; Schmidt et al., 2004; Voelckel & Baldwin, 2004; Schmidt et al., 2005; Vogel et al., 2007; Ehlting et al., 2008). These studies typically revealed that hundreds or even thousands of genes are differentially regulated upon herbivory. For example, Schmidt et al. (2004) reported 155 genes to be significantly regulated upon flea beetle attack in *Solanum nigrum*, when using a microarray representing 558 genes from different plant sources. In another study, Schmidt et al. (2005) used a microarray representing 10,000 cDNAs from *Solanum tuberosum* to investigate transcriptional responses of *Nicotiana attenuata* and *S. nigrum* upon *Manduca sexta* herbivory, and found 409 and 280 significantly regulated genes, respectively. There was little overlap in the sets of responding genes in both plant species, indicating that the same insect species can cause profound differences in the transcriptional response of closely related plants. In *Arabidopsis thaliana*, Reymond et al. (2004) used a microarray representing approximately 7,200 unique genes to compare the plant transcriptional response to two different lepidopteran insect herbivores, *P. rapae* and *Spodoptera littoralis*. More than hundred genes were insect-responsive, and the sets of differentially expressed genes were

quite similar for both insects. Other studies with Arabidopsis or close relatives, however, found that there was little overlap in the plant transcriptional signature in response to different insect herbivores (De Vos et al., 2005; Vogel et al., 2007; Ehlting et al., 2008).

In summary, these studies demonstrate massive changes in the plant transcriptome in response to herbivory, both for generalist and for specialist insect herbivores. However, the impact of these changes for plant performance is not at all clear. Fitness consequences have been investigated only for a very limited set of genes. Transgenic *N. attenuata* plants silenced in a gene that catalyzes the key step in nicotine biosynthesis were found to be more frequently attacked by herbivores when planted in their natural environment (Steppuhn et al., 2004). Manipulation of trypsin proteinase inhibitor levels in *N. attenuata* affected larval developmental time and survivorship (Zavala et al., 2004). Likewise, manipulation of WRKY3 and WRKY6, plant-specific transcription factors, generated *N. attenuata* plants which experienced extensive damage in their native habitat (Skibbe et al., 2008).

In Arabidopsis, similar studies have mostly focussed on mutants impaired in phytohormone biosynthesis or perception. Several phytohormone-mediated signal cascades are involved in the elicitation of induced defenses, including the jasmonate acid, ethylene and salicylic acid pathways (De Vos et al., 2005; Bodenhausen & Reymond, 2007). The *coi1* mutant, insensitive to jasmonic acid, was found to be more susceptible to *S. littoralis* than wildtype plants, whereas *npr1* mutants, impaired in the salicylic acid pathway, had increased resistance to this insect herbivore (Stotz et al., 2002; Reymond et al., 2004). However, alterations in signalling pathways cause pleiotropic effects, and affect the expression of a large number of genes (Reymond et al., 2000; Reymond et al., 2004; De Vos et al., 2005; Vogel et al., 2007). Thus, despite that these studies are useful in understanding the importance of phytohormone signalling in defense responses, the distal genes that ultimately affect the herbivore remain undetected.

The goal of *Manuscript IV* was to analyze potential fitness effects of a set of more than 30 genes that are differentially regulated upon herbivory by the specialist insect herbivore *P. xylostella*. This work compared feeding damage caused by *P. xylostella* larvae on T-DNA insertion mutants *versus* wildtype plants. When up-regulation of a *Plutella*-responsive gene contributes to defense against this insect, then an insertion of T-DNA into this gene should negatively affect plant performance. *Vice versa*, an insertion into a down-regulated gene should increase plant fitness. However, a surprisingly low number of insertions showed an impact on plant resistance against *P. xylostella* larvae. Thus, a large proportion of *Plutella*-

responsive genes do not improve plant fitness upon *Plutella* herbivory. Rather, differential regulation may serve other purposes such as minimizing secondary harmful effects caused by potential pathogen infection or water loss through the wound that is inflicted by herbivory. These possibilities remain to be tested. However, in the few cases where a T-DNA insertion resulted in a phenotype, fitness effects were contrary to expectations. Two mutants with T-DNA insertions in upregulated genes were less susceptible and one mutant with an insertion in a downregulated gene was more susceptible to herbivory than wildtype plants. This suggested that *Plutella* larvae may be able to manipulate the Arabidopsis transcriptome. Indeed, several other studies provide evidence for herbivore manipulation of plant defenses. In *Helicoverpa zea* larvae a glucose oxidase was identified as a principle salivary enzyme which suppressed induced defenses in *N. attenuata* and prevented the induction of nicotine (Musser et al., 2002). The silverleaf whitefly was found to be able to repress JA-regulated defenses that deter nymph development (Zarate et al., 2007). Finally, jasmonic acid-mediated upregulation of defense genes was suppressed in response to *P. rapae* (Bodenhausen & Reymond, 2007). These findings indicate that the adaptation of specialist herbivores to their host plants may be more substantial than previously thought. Apparently, specialist insect herbivores have not only evolved counteradaptations against preformed activated defenses such as the glucosinolate-myrosinase system but also appear to possess counteradaptations against induced defenses, thus adding a further level of complexity to interactions between specialist insect herbivores and their host plants.

4.3 Natural variation in Arabidopsis glucosinolate structures

Glucosinolates are a structurally diverse group of plant secondary compounds, and glucosinolate composition varies between and within species (Fahey et al., 2001; Kliebenstein et al., 2001a; Windsor et al., 2005; Arany et al., 2008; *Manuscript I*). In *A. thaliana*, more than 30 different glucosinolates have been identified, derived from methionine, tryptophan or phenylalanine (Kliebenstein et al., 2001a; Reichelt et al., 2002). Most of the structural variability in aliphatic glucosinolates, the largest class of glucosinolates in Arabidopsis, can be attributed to four polymorphic genetic loci, *MAM, GSL-AOP, GSL-OX* and *GSL-OH* (Magrath et al., 1994; Mithen et al., 1995; Campos de Quiros et al., 2000; Kliebenstein et al., 2001b). Enzymes encoded at *GSL-AOP*, *GSL-OX* and *GSL-OH* catalyze modifications of the glucosinolate side chain (Magrath et al., 1994; Mithen et al., 1995; Giamoustaris & Mithen; 1996; Kliebenstein et al., 2001a; Kliebenstein et al., 2001b; Hansen et al., 2007), while *MAM*

gene products determine the length of the carbon chain of aliphatic glucosinolates (Kroymann et al., 2001; Kroymann et al., 2003; Textor et al., 2004; Benderoth et al., 2006; Textor et al., 2007; *Manuscript III*). *MAM* gene products act during methionine chain elongation, the initial stage in aliphatic glucosinolate biosynthesis. Therefore, they play a central role for the diversification of aliphatic glucosinolate structures. *Manuscript III* reviews the genetic and biochemical basis for *MAM*-mediated variation in carbon chain lengths. It also provides an overview of the evolutionary forces that are responsible for the generation of chain length variation and for maintaining this diversity.

Compared to aliphatic glucosinolates, much less attention has been paid to tryptophan-derived indole glucosinolates, although this class of glucosinolates is also structurally variable in Arabidopsis and related species. Furthermore, recent studies suggest that indole glucosinolates may contribute to aphid resistance (Kim & Jander, 2007; Kim et al., 2008). Most Arabidopsis accessions contain four different indole glucosinolates, indole-3-yl-methyl (I3M), 1-methoxy-indole-3-yl-methyl (1MO-I3M), 4-hydroxy-indole-3-yl-methyl (4OH-I3M) and 4-methoxy-indole-3-yl-methyl (4MO-I3M) glucosinolate (Kliebenstein et al., 2001a; Brown et al., 2003). Mapping studies with Arabidopsis RIL populations have detected numerous QTL for qualitative and quantitative variation in indole glucosinolate profiles (Kliebenstein et al., 2001c; *Manuscript I*). *Manuscript V* is the first study to report the cloning of a QTL for indole glucosinolates, *Indole Glucosinolate Modifier 1* (*IGM1*). This QTL influenced 4OH-I3M and 4MO-I3M accumulation in the Da(1)-12 × Ei-2 RIL population (*Manuscript I*), caused by differential regulation of the underlying structural gene, *CYP81F2*. This gene encodes a cytochrome P450 monooxygenase which converts I3M to 4OH-I3M. 4OH-I3M serves, in turn, as a substrate for the generation of 4MO-I3M. Therefore, natural variation of *CYP81F2* among Arabidopsis accessions or artificial manipulation of this gene in mutant lines affects both 4OH-I3M and 4MO-I3M simultaneously.

Manuscript V proposes that the remaining modified indole glucosinolate in Arabidopsis, 1MO-I3M, is also generated via a hydroxy-intermediate, in analogy to the pathway for generation of 4MO-I3M, and candidate genes have already been identified. Furthermore, the work suggests that the reactions from 4OH-I3M to 4MO-I3M and from 1OH-I3M to 1MO-I3M, respectively, are catalyzed by methyltransferases. Therefore, the identification of the first gene involved in the modification of indole glucosinolate structures represents a major advance in one of the last unresolved areas of Arabidopsis glucosinolate biosynthesis research (Halkier & Gershenzon, 2006), and promises that the entire gene

repertoire for the generation of modified indole glucosinolates will be uncovered in the near future.

Finally, reduction in 4OH-I3M and 4MO-I3M levels in *CYP81F2* mutants benefited aphids but not herbivory by larvae from several lepidopteran species, irrespective of their host range. Aliphatic glucosinolates, on the other hand, appear to be particularly important for defense against generalist lepidopteran larvae (Kroymann et al., 2003; Kliebenstein et al., 2005; Beekwilder et al., 2008, *Manuscript III*). Thus, different elements of the glucosinolate-myrosinase system have different enemies as their target, reinforcing the idea that the complexity of the glucosinolate-myrosinase system has evolved, at least in part, in response to biotic challenges inflicted by a diverse community of herbivores, pathogens and other enemies.

5.1 Summary

This work investigates plant-insect interactions in the model plant *Arabidopsis thaliana* at the molecular and genetic level.

Among Arabidopsis accessions, there is natural genetic variation for resistance against herbivory by the crucifer specialist insects *Plutella xylostella* and *Pieris brassicae*. Six QTL for resistance against *P. brassicae* and one QTL for resistance against *P. xylostella* were detected in the Arabidopsis Da(1)-12 × Ei-2 recombinant inbred line population. Although Da(1)-12 × Ei-2 harbours numerous QTL for glucosinolates, myrosinase activity and trichome density, none of these traits explained the resistance QTL convincingly. This indicates that resistance against specialist lepidopterans has a different genetic basis in Arabidopsis than resistance against generalist lepidopterans.

Insect herbivory triggers massive changes in the plant transcriptome. The impact of *Plutella*-responsive genes on plant fitness was investigated with Arabidopsis T-DNA insertion lines. Most genes that are differentially regulated upon *Plutella* herbivory do not alter plant resistance against *Plutella* larvae. Some genes exhibit fitness consequences contrary to expectations based on transcript patterns, suggesting that herbivores may be able to manipulate plant gene expression to their own benefit.

One of the indole glucosinolate QTL that segregated in Da(1)-12 × Ei-2 was cloned by a combination of QTL fine mapping and transcript profiling. This QTL, termed *Indole Glucosinolate Modifier 1* (*IGM1*), influences the accumulation of 4-hydroxy-indole-3-yl-methyl (4OH-I3M) and 4-methoxy-indole-3-yl-methyl (4MO-I3M) glucosinolates. The function of the underlying gene, *CYP81F2*, was confirmed with Arabidopsis T-DNA insertion lines and enzyme assays with heterologously expressed *CYP81F2*. CYP81F2 catalyzes the conversion of indole-3-yl-methyl (I3M) to 4-methoxy-indole-3-yl-methyl. *CYP81F2* is the first identified gene involved in the biosynthesis of modified indole glucosinolates. A rectified model for the biosynthesis of indole glucosinolates was proposed. The *IGM1* QTL is largely explained by variation in *CYP81F2* expression, controlled by a combination of *cis* and *trans* acting eQTL. Differences in the accumulation of modified indole glucosinolates influence proliferation of the aphid *Myzus persicae* but not herbivory by larvae from four lepidopteran species.

5.2 Zusammenfassung

Die vorliegende Arbeit befaßt sich mit der Verteidigung der Modelpflanze *Arabidopsis thaliana* (Brassicales) gegenüber Insekten. Sekundäre Pflanzeninhaltsstoffe spielen bei der Interaktion zwischen Pflanzen und Insekten eine entscheidende Rolle. Während einige Insekten polyphag sind und als Generalisten bezeichnet werden, hat sich ein Großteil der herbivoren Insekten im Laufe ihrer Evolution auf bestimmte Pflanzengruppen spezialisiert. Diese Spezialisierung beruht unter anderem auf der Überwindung von Verteidigungsmechanismen der Pflanzen, die der Abwehr von Fraßfeinden dienen sollen. Das Glucosinolat-Myrosinase-System, das in *A. thaliana* und anderen Vertretern der Ordnung der Brassicales zu finden ist, stellt ein Beispiel für effektives aktivierbares Abwehrsystem gegenüber pflanzenfressenden Insekten dar. Insekten, die auf Kreuzblütler spezialisiert sind, haben jedoch Anpassungsmechanismen entwickelt, die dieses System außer Kraft setzen.

Ein Vergleich mehrerer Arabidopsis-Akzessionen zeigte, daß es zwischen diesen Akzessionen natürliche genetische Variabilität in der Suszeptibilität gegenüber den Spezialisten *Plutella xylostella* (Kohlmotte) und *Pieris brassicae* (Großer Kohlweissling) gibt: Das Ausmaß des durch die Raupen der beiden Spezies hervorgerufenen Fraßschadens unterschied sich signifikant zwischen den Akzessionen. Die QTL-Kartierung (QTL = Quantitative Trait Locus/Loci) einer neuen Population rekombinanter Inzuchtlinien (RIL), hervorgegangen aus einer Kreuzung der Akzessionen Da(1)-12 und Ei-2, deckte sechs genetische Loci auf, die zur Resistenz gegenüber *P. brassicae*-Larven beitragen. Zudem wurde ein QTL für die Verteidigung gegen *P. xylostella* detektiert. Untersuchungen der Glucosinolatprofile, der Myrosinaseaktivität und der Trichomdichte der Blätter innerhalb dieser Population führten zur Identifikation einer Vielzahl von QTL für diese Merkmale. Jedoch erklärten diese QTL nicht in überzeugender Weise die Resistenz-QTL gegen die Spezialisten. Dies deutet darauf hin, dass die Resistenz gegenüber spezialisierten Lepidopterenlarven durch genetische Faktoren bestimmt wird, die sich von denen unterscheiden, die für Generalisten wichtig sind.

Neben der konstitutiven Abwehr gegen Fraßfeinde können induzierte Verteidigungsmechanismen eine Rolle spielen. Herbivorie durch Insekten löst üblicherweise massive Änderungen im pflanzlichen Transkriptom aus. Allerdings ist im wesentlichen unbekannt, ob und inwieweit die differentielle Regulierung dieser Vielzahl von Genen der Verteidigung gegen herbivore Insekten dient. Daher wurde die potentielle Bedeutung einer

Auswahl an Genen, die nach Fraß von Raupen des Spezialisten *P. xylostella* mit einer Änderung ihrer Transkriptmenge reagieren, für die pflanzliche Fitneß untersucht. Hierzu wurden Arabidopsispflanzen mit Insertionen von T-DNA oder Transposons in den Kandidatengenen mit Wildtyplinien im Hinblick auf den durch *Plutella*-Larven verursachten Fraßschaden verglichen. Die Insertion von Fremd-DNA führte in den meisten der betrachteten Gene zu einer deutlichen oder sogar vollständigen Disruption der Genfunktion. Während etwa ein Viertel der untersuchten T-DNA-Insertionsmutanten im Vergleich zu Wildtyppflanzen statistisch signifikante Änderungen in der Pflanzengröße aufwies, gab es nur in drei von mehr als 30 betrachteten Linien eine Auswirkung auf den Resistenzphänotyp. Dies läßt darauf schließen, daß die meisten der nach *Plutella*-Fraß auftretenden Änderungen im pflanzlichen Transkriptom nicht zur Verteidigung gegenüber diesem Insekt beitragen. Zudem standen die Fitnesseffekte im Widerspruch zu den auf Transkriptmustern basierenden Erwartungen. Für zwei Gene, deren Transkriptmenge sich nach Insektenherbivorie erhöhte, zeigten sich die Insertionsmutanten resistenter als der Wildtyp. Für ein weiteres Gen mit verringerter Transkriptmenge war der an der Mutantenlinie verursachte Fraßschaden höher als an den Kontrollpflanzen. Dies bedeutet, daß die differentielle Regulierung dieser Gene weniger der Pflanze als vielmehr dem Insekt nützt, was vermuten läßt, daß spezialisierte Insekten die Fähigkeit entwickelt haben könnten, die Genexpression der Pflanzen zu ihren eigenen Gunsten zu manipulieren.

Die Analyse der Glucosinolatprofile der Da(1)-12 × Ei-2 RIL-Population ergab neben vielen anderen, zum Teil bislang unbekannten QTL, einen QTL für die Akkumulation der modifizierten Indolglucosinolate 4-Hydroxy-3-indolylmethyl (4OH-I3M) und 4-Methoxy-3-indolylmethyl (4MO-I3M). Dieser QTL wurde als *Indole Glucosinolate Modifier 1* (*IGM1*) bezeichnet. Die Kombination zweier Methoden, QTL-Feinkartierung und Transkriptprofilierung mittels Mikroarrays, ermöglichte die Identifikation des Gens, das *IGM1* zugrunde lag. Die Funktion dieses Genes, welches für das Cytochrom P450 CYP81F2 codiert, wurde mittels T-DNA Insertionslinien und einem enzymatischen Assay mit heterolog exprimiertem *CYP81F2* verifiziert. Das Genprodukt war in der Lage, die Reaktion von Indol-3-yl-methyl-glucosinolat (I3M) zu 4-Hydroxy-indol-3-yl-methyl-glucosinolat zu katalysieren. *CYP81F2* ist somit das erste identifizierte Gen, das eine Rolle in der Modifikation von Indolglucosinolaten spielt. Der *IGM1* QTL erklärt sich im wesentlichen durch Unterschiede in der Expression von *CYP81F2*, welche über eine Kombination von *cis* und *trans* eQTL kontrolliert wird.

Untersuchungen der Auswirkungen auf das Fraßverhalten verschiedener Insekten ergaben, daß Unterschiede in der Akkumulation modifizierter Indolglucosinolate keinen Einfluß auf den durch Raupen der Lepidopteren *P. xylostella* und *P. brassicae* verursachten Fraßschaden hatten. Ebenso wurden zwei verschiedene Generalisten unter den Lepidopteren, *Spodoptera exigua* und *Trichoplusia ni*, nicht beeinträchtigt. Im Gegensatz dazu hatten modifizierte Indolglucosinolate Konsequenzen für die Vermehrung der Blattlaus *Myzus persicae*, die sich bei erhöhter Konzentration an modifizierten Indolglucosinolaten deutlich langsamer vermehrte. Generalisten unter den Lepidopteren reagieren üblicherweise empfindlich auf aliphatische Glucosinolate. Dies wurde unter anderem am Beispiel des *MAM* QTL gezeigt, der einen wesentlichen Einfluß auf die Akkumulation aliphatischer Glucosinolate hat. Daher läßt sich schließen, daß die verschiedenen Komponenten des Glucosinolat-Myrosinase-Systems unterschiedliche Insekten zum Ziel haben könnten. Dies unterstreicht die Idee, daß sich die Komplexität im Glucosinolate-Myrosinase-System als Antwort auf die durch eine Vielzahl von Pflanzenfeinden gestellten Herausforderungen entwickelt hat.

6. References

Agrawal AA, Kurashige NS (2003). A role for isothiocyanates in plant resistance against the specialist herbivore *Pieris rapae*. *J. Chem. Ecol.* **29**, 1403–1415.

Agren J, Schemske DW (1993). The cost of defense against herbivores: an experimental study of tricome production in *Brassica rapa*. *Am. Nat.* **141**, 338-350.

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

Alonso-Blanco C, Blankenstijn-de Vries H, Hanhart CJ, Koornneef M (1999). Natural allelic variation at seed size loci in relation to other life history traits of *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **96**, 4710-4717.

Andreasson E, Jorgensen LB, Hoglund AS, Rask L, Meijer J (2001). Different myrosinase and idioblast distribution in Arabidopsis and *Brassica napus*. *Plant Physiol.* **127**, 1750-1763.

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

Arany AM, de Jong TJ, Kim HK, van Dam NM, Choi YH, Verpoorte R, Van der Meijden E (2008). 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.

Bak S, Feyereisen R (2001). The involvement of two P450 enzymes, *CYP83A1* and *CYP83B1*, in auxin homeostasis and glucosinolate biosynthesis. *Plant Physiol.* **127**, 108-218.

Baldwin IT (1999). Inducible nicotine production in native Nicotiana as an example of adaptive phenotypic plasticity. *J Chem. Ecol.* **25**, 3-30.

Barth C, Jander G (2006). Arabidopsis myrosinases TGG1 and TGG2 have redundant function in glucosinolate breakdown and insect defense. *Plant J.* **46**, 549-562.

Bartlet E, Mithen R, Clark SJ (1996). Feeding of the cabbage stem flea beetle *Psylliodes chrysocephala* on high and low glucosinolate cultivars of oilseed rape. *Entomol. Exp. Appl.* **80**, 87-89.

Beekwilder J, van Leeuwen W, van Dam NM, Bertossi M, Grandi V, Mizzi L, Soloviev M, Szabados L, Molthoff JW, Schipper B, Verbocht H, de Vos RC, Morandini P, Aarts MGM, Bovy A (2008). The impact of the absence of aliphatic glucosinolates on insect herbivory in Arabidopsis. *PLoS ONE* **3**, e2068.

Benderoth M, Textor S, Windsor AJ, Mitchell-Olds T, Gershenzon J, Kroymann J (2006). Positive selection driving diversification in plant secondary metabolism. *Proc. Natl. Acad. Sci. USA* **103**, 9118-9123.

Binder S, Knill T, Schuster J (2007). Branched-chain amino acid metabolism in higher plants. *Physiol. Plant.* **129**, 68-78.

Birch ANE, Griffiths DW, Hopkins RJ, Macfarlane Smith WH, McKinlay RG (1992). Glucosinolate responses of swede, kale, forage and oilseed rape to root damage by turnip root fly (*Delia floralis*) larvae. *J. Sci. Food Agric.* **1**, 1-9.

Bodenhausen N, Reymond P (2007). Signaling pathways controlling induced resistance to insect herbivores in Arabidopsis. *Mol. Plant Microbe Interact.* **20**, 1406-1420.

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.

Burow M, Müller R, Gershenzon J, Wittstock U (2006). Altered glucosinolate hydrolysis in genetically engineered *Arabidopsis thaliana* and its influence on the larval development of *Spodoptera littoralis*. *J. Chem. Ecol.* **32**, 2333-2349.

Burow M, Bergner A, Gershenzon J, Wittstock U (2007). Glucosinolate hydrolysis in *Lepidium sativum* – identification of the thiocyanate-forming protein. *Plant Mol. Biol.* **63**, 49-61.

Campos de Quiros H, Magrath HCR, McCallum D, Kroymann J, Schnabelrauch D, Mitchell-Olds T, Mithen R (2000). Quantitative trait loci and the biochemical genetics of glucosinolate biosynthesis in *Arabidopsis thaliana*. *Theor. Appl. Genet.* **101**, 429-437.

Chen S, Glaswischnig E, Jorgensen K, Naur P, Jorgensen B, Olsen CE, Hansen CH, Rasmussen H, Pickett JA, Halkier BA (2003). *CYP79F1* and *CYP79F2* have distinct functions in the biosynthesis of aliphatic glucosinolates in Arabidopsis. *Plant J.* **33**, 923-937.

Chen MS (2008). Inducible direct plant defense against insect herbivores: a review. *Insect Sci.* **15**, 101-114.

Chisholm MD, Wetter LR (1964). Biosynthesis of mustard oil glucosides: IV The administration of methionine-C^{14} and related compounds to horseradish. *Can. J. Biochem.* **42**, 1033-1040.

Clark RM, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P, Warthmann N, Hu TT, Fu G, Hinds DA, Chen H, Frazer KA, Huson DH, Schölkopf B, Nordborg M, Rätsch G, Ecker JR, Weigel D (2007). Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* **317**, 338-342.

Daxenbichler ME, Spencer GF, Carlson DG, Rose GB, Brinker AM, Powell RG (1991). Glucosinolate composition of seeds from 297 species of wild plants. *Phytochemistry* **30**, 2623-2638.

De Jong R, Städler E (1999). The influence of odour on the oviposition behaviour of the cabbage root fly. *Chemoecology* **9**, 151-154.

De Vos, Van Oosten VR, Van Poecke RMP, Van Pelt JA, Pozo MJ, Mueller MJ, Buchala AJ, Metraux JP, Van Loon LC, Dicke M, Pieterse CMJ (2005). Signal signature and transcriptome changes of Arabidopsis during pathogen and insect attack. *Mol. Plant Microbe Interact.* **18**, 923-937.

De Vos M, Kriksunov KL, Jander G (2008). Indole-3-acetonitrile production from indole glucosinolates deters oviposition by *Pieris rapae*. *Plant Physiol.* **146**, 916-926.

Dicke M (1999). Evolution of induced indirect defense of plants. In R Tollrian, CD Harvell eds, The Ecology and Evolution of Inducible Defenses. Princeton University Press, Princeton, pp. 62-88.

Eigenbrode SD, Espelie KE (1995). Effects of plant epicuticular lipids on insect herbivores. *Annu. Rev. Entomol.* **40**, 171-194.

Ehlting J, Chowrira SG, Mattheus N, Aeschliman DS, Arimura GI, Bohlmann J (2008). Comparative transcriptome analysis of *Arabidopsis thaliana* infested by diamond back moth (*Plutella xylostella*) larvae reveals signatures of stress response, secondary metabolism, and signalling. *BMC Genomics* **9**, 154.

Fahey JW, Zalcmann AT, Talalay P (2001). The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry* **56**, 5-51.

Giamoustaris A, Mithen R (1995). The effect of modifying the glucosinolate content of leaves of oilseed rape (*Brassica napus* ssp. *oleifera*) on its interaction with specialist and generalist pests. *Ann. Appl. Biol.* **126**, 347-363.

Giamoustaris A, Mithen R (1996). Genetics of aliphatic glucosinolates. IV. Side-chain modification in *Brassica oleracea*. *Theor. Appl. Genet.* **93**, 1006-1010.

Glawischnig E, Hansen BG, Olsen CE, Halkier BA (2004). Camalexin is synthesized from indole-3-actaldoxime, a key branching point between primary and secondary metabolism in *Arabidopsis*. *Proc. Natl. Acad. Sci USA* **101**, 8245-8250.

Graser G, Schneider B, Oldham NJ, Gershenzon J (2000). The methionine chain elongation pathway in the biosynthesis of glucosinolates in *Eruca sativa* (Brassicaceae). *Arch. Biochem. Biophys.* **378**, 411-419.

Grubb CD, Ludwig-Muller J, Masuno MN, Molinski TF, Abel S (2004). Arabidopsis glucosyltransferase *UGT74B1* functions in glucosinolate biosynthesis and auxin homeostasis. *Plant J.* **40**, 893-908.

Halkier BA, Gershenzon J (2006). Biology and biochemistry of glucosinolates. *Ann. Rev. Plant Biol.* **57**, 303-333.

Handley R, Ekbom B, Agren J (2005). Variation in trichome density and resistance against a specialist insect herbivore in natural populations of *Arabidopsis thaliana*. *Ecol. Entomol.* **30**, 284-292.

Hansen CH, Wittstock U, Olsen CE, Hick AJ, Pickett JA, Halkier BA (2001). Cytochrome P450 *CYP79F1* from Arabidopsis catalyzes the conversion of dihomomethionine and trihomomethionine to the corresponding aldoximes in the biosynthesis of aliphatic glucosinolates. *J. Biol. Chem.* **276**, 24790-24796.

Hansen BG, Kliebenstein DJ, Halkier BA (2007). Identification of a flavin-monooxygenase as the S-oxygenating enzyme in aliphatic glucosinolate biosynthesis in Arabidopsis. *Plant J.* **50**, 902-910.

Harvey JA, Witjes LMA, Benkirane M, Wagenaar R (2007). Nutritional suitability and ecological relevance of *Arabidopsis thaliana* and *Brassica oleracea* as foodplant for the cabbage butterfly, *Pieris rapae*. *Plant Ecol.* **189**, 117-126.

Howe GA, Jander G (2008). Plant immunity to insect herbivores. *Annu. Rev. Plant Biol.* **59**, 41-66.

Huang X, Renwick JAA (1994). Relative activities of glucosinolates as oviposition stimulants for *Pieris rapae* and *P. napi oleracea*. *J. Chem. Ecol.* **20**, 1025-1037.

Hull AK, Vij R, Celenza JL (2000). Arabidopsis cytochrome P450s that catalyze the first step of tryptophan-dependent indole-3-acetic acid biosynthesis. *Proc. Natl. Acad. Sci. USA* **97**, 2379-2384.

Kim JH, Jander G (2007). *Myzus persicae* (green peach aphid) feeding on Arabidopsis induces the formation of a deterrent indole glucosinolate. *Plant J.* **49**, 1008-1019.

Kim JH, Lee BW, Schroeder FC, Jander G (2008). Identification of indole glucosinolate breakdown products with antifeedant effects on *Myzus persicae* (green peach aphid). *Plant J.* **54**, 1015-1026.

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

Kliebenstein DJ, Lambrix VM, Reichelt M, Gershenzon J, Mitchell-Olds T (2001b). Gene duplication in the diversification of secondary metabolism: Tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in Arabidopsis. *Plant Cell* **13**, 681-693.

Kliebenstein DJ, Gershenzon J, Mitchell-Olds T (2001c). Comparative quantitative trait loci mapping of aliphatic, indolic and benzylic glucosinolate production in *Arabidopsis thaliana* leaves and seeds. *Genetics* **159**, 359-370.

Kliebenstein DJ, Pedersen D, Barker B, Mitchell-Olds T (2002). Comparative Analysis of quantitative trait loci controlling glucosinolates, myrosinase and insect resistance in *Arabidopsis thaliana*. *Genetics* **161**, 325-332.

Kliebenstein DJ, Kroymann J, Mitchell-Olds T (2005). The glucosinolate-myrosinase system in an ecological and evolutionary context. *Curr. Opin. Plant Biol.* **8**, 264-271.

Kliebenstein D (2008). A quantitative genetics and ecological model system: understanding the aliphatic glucosinolate biosynthetic network via QTLs. *Phytochem. Rev.* DOI 10.1007/s11101-008-09102-8.

Knill T, Schuster J, Reichelt M, Gershenzon J, Binder S (2008). Arabidopsis branched-chain aminotransferase 3 functions in both amino acid and glucosinolate biosynthesis. *Plant Physiol.* **146**, 1028-1039.

Kobayashi Y, Koyama H (2002). QTL analysis of Al tolerance in recombinant inbred lines of *Arabidopsis thaliana*. *Plant Cell Physiol.* **43**, 1526-1533.

Koornneef M, Alonso-Blanco C, Peeters AJM, Soppe W (1998). Genetic control of flowering time in Arabidopsis. *Annu. Rev. Plant Physiol.* **49**, 345-70.

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

Koroleva OA, Davies A, Deeken R, Thorpe MR, Tomos AD, Hedrich R (2000). Identification of a new glucosinolate-rich cell type in Arabidopsis flower stalk. *Plant Physiol.* **124**, 599-608.

Kroymann J, Textor S, Tokuhisa JG, Falk KL, Bartram S, Gershenzon J, Mitchell-Olds T (2001). A gene controlling variation in *Arabidopsis thaliana* glucosinolate composition is part of the methionine chain elongation pathway. *Plant Physiol.* **127**, 1077-1088.

Kroymann J, Donnerhacke S, Schnabelrauch D, Mitchell-Olds T (2003). Evolutionary dynamics of an Arabidopsis insect resistance quantitative trait locus. *Proc. Natl. Acad. Sci. USA* **100**, 14587-14592.

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

Lambrix V, Reichelt M, Mitchell-Olds T, Kliebenstein D, Gershenzon J (2001). The *Arabidopsis* epithiospecifier protein promotes the hydrolysis of glucosinolates to nitriles and influences *Trichoplusia ni* herbivory. *Plant Cell* **13**, 2793-2807.

Li Q, Eigenbrode SD, Stringam GR, Thiagarajah MR (2000). Feeding and growth of *Plutella xylostella* and *Spodoptera eridania* on *Brassica juncea* with varying glucosinolate concentrations and myrosinase activities. *J. Chem. Ecol.* **26**, 2401-2419.

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

Lister C, Dean C (1993). Recombinant inbred lines for mapping RFLP and phenotypic markers. *Plant J.* **4**, 745-750.

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

Magrath R, Bano F, Morgner M, Parkin I, Sharpe A, Lister C, Dean C, Turner J, Ludiate D, Mithen R (1994). Genetics of aliphatic glucosinolates: I. Side chain elongation in *Brassica napus* and *Arabidopsis thaliana*. *Heredity* **72**, 290-299.

Matsuo M, Yamazaki M (1964). Biosynthesis of sinigrin. *Chem. Pharmacol. Bull.* **12**, 1388-1389.

Mauricio R (1998). Costs of resistance to natural enemies in field populations of the annual plant *Arabidopsis thaliana*. *Am. Nat.* **151**, 20-28.

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

Mikkelsen MD, Hansen CH, Wittstock U, Halkier BA (2000). Cytochrome P450 CYP79B2 from Arabidopsis catalyzes the conversion of tryptophan to indole-3-acetaldoxime, a precursor of indole glucosinolates and indole-3-acetic acid. *J. Biol. Chem.* **275**, 33712-33717.

Mikkelsen MD, Naur P, Halkier BA (2004). Arabidopsis mutants in the C-S lyase of glucosinolate biosynthesis establish a critical role for indole-3-acetaldoxime in auxin homeostasis. *Plant J.* **37**, 770-777.

Mitchell-Olds T (2001). *Arabidopsis thaliana* and its wild relatives: a model system for ecology and evolution. *Trends Ecol. Evol.* **16**, 693-700.

Mithen R, Clarke J, Lister C, Dean C (1995). Genetics of aliphatic glucosinolates. III. Side chain structure of aliphatic glucosinolates in *Arabidopsis thaliana*. *Heredity* **74**, 210-215.

Musser RO, Hum-Musser SM, Eichenseer H, Peiffer M, Ervin G, Murphy JB, Felton GW (2002). Herbivory: caterpillar saliva beats plant defenses. *Nature* **416**, 599-600.

Naur P, Petersen BL, Mikkelsen MD, Bak S, Rasmussen H, Olsen CE, Halkier BA (2003). *CYP83A1* and *CYP83B1*, two nonredundant cytochrome P450 enzymes metabolizing oximes in the biosynthesis of glucosinolates in Arabidopsis. *Plant Physiol.* **133**, 67-72.

O'Neill CM, Morgan C, Kirby J, Tschoep H, Deng PX, Brennan M, Rosas U, Fraser F, Hall C, Gill S, Bancroft I (2008). Six new recombinant inbred populations for the study of quantitative traits in *Arabidopsis thaliana*. *Theor. Appl. Genet.* **116**, 623-634.

Piotrowski M, Schemenewitz A, Lopukhina A, Müller A, Janowitz T, Weiler EW, Oecking C (2004). Desulfoglucosinolate sulfotransferases from *Arabidopsis thaliana* catalyze the final step in the biosynthesis of the glucosinolate core structure. *J. Biol. Chem.* **279**, 50717-50725.

Pivnick K, Jarvis BJ, Slater GP (1994). Identification of olfactory cues used in host-plant finding by diamondback moth, *Plutella xylostella* (Lepidoptera: Plutellidae). *J. Chem. Ecol.* **20**, 1407-1427.

Ratzka A, Vogel H, Kliebenstein D, Mitchell-Olds T, Kroymann J (2002). Disarming the mustard oil bomb. *Proc. Natl. Acad. Sci. USA* **99**, 11223-11228.

Reichelt M, Brown PD, Schneider B, Oldham NJ, Stauber EJ, Tokuhisa J, Kliebenstein DJ, Mitchell-Olds T, Gershenzon J (2002). Benzoic acid glucosinolate esters and other glucosinolates from *Arabidopsis thaliana*. *Phytochemistry* **59**, 663-671

Renwick JAA, Haribal M, Gouinguené S, Städler E (2006). Isothiocyanates stimulating oviposition by the diamondback moth, *Plutella xylostella*. *J. Chem. Ecol.* **32**, 755-766.

Reymond P, Weber H, Damond M, Farmer EE (2000). Differential gene expression in response to mechanical wounding and insect feeding in Arabidopsis. *Plant Cell* **12**, 707-719.

Reymond P, Bodenhausen N, Van Poecke RMP, Krishnamurthy V, Dicke M, Farmer EE (2004). A conserved transcript pattern in response to a specialist and a generalist herbivore. *Plant Cell* **16**, 3132-3147.

Riddle NC, Richards EJ (2002). The control of natural variation in cytosine methylation in Arabidopsis. *Genetics* **162**, 355-363.

Rodman JE (1991a). A taxonomic analysis of glucosinolate-producing plants, part 1: phenetics. *Syst. Bot.* **16**, 598-618.

Rodman JE (1991b). A taxonomic analysis of glucosinolate-producing plants, part 2: cladistics. *Syst. Bot.* **16**, 619-629.

Schmidt D, Kessler A, Kessler D, Schmidt S, Lim M, Gase K, Baldwin IT (2004). *Solanum nigrum*: A model ecological expression system and its tools. *Mol. Ecol.* **13**, 981-995.

Schmidt D, Voelckel C, Hartl M, Schmidt S, Baldwin IT (2005). Specificity in ecological interactions. Attack from the same lepidopteran herbivore results in species-specific transcriptional responses in two solanaceous host plants. *Plant Physiol.* **138**, 1763-1773.

Schuster J, Knill T, Reichelt M, Gershenzon J, Binder S (2006). BRANCHED-CHAIN AMINOTRANSFERASE4 is part of the chain elongation pathway in the biosynthesis of methionine-derived glucosinolates in Arabidopsis. *Plant Cell* **18**, 2664-2679.

Shroff R, Vergara F, Muck A, Svatoš A, Gershenzon J (2008). Nonuniform distribution of glucosinolates in *Arabidopsis thaliana* leaves has important consequences for plant defense. *Proc. Natl. Acad. Sci. USA* **105**, 6196-6201.

Skibbe M, Qu N, Galis I, Baldwin IT (2008). Induced plant defenses in the natural environment: *Nicotiana attenuata* WRKY3 and WRKY6 coordinate responses to herbivory. *Plant Cell* **20**, 1984-2000.

Smallegange RC, Van Loon JJA, Blatt SE, Harvey JA, Agerbirk N, Dicke M (2007). Flower vs. leaf feeding by *Pieris brassicae*: glucosinolate-rich flower tissues are preferred and sustain higher growth rate. *J. Chem. Ecol.* **33**, 1831-1844.

Steppuhn A, Gase K, Krock B, Halitschke R, Baldwin IT (2004). Nicotine's defensive function in nature. *PLoS Biol.* **2**, 1074-1080.

Stotz HU, Kroymann J, Mitchell-Olds T (1999). Plant-insect interactions. *Curr. Opin. Plant Biol.* **2**, 268-72.

Stotz HU, Koch T, Biedermann A, Weniger K, Boland W, Mitchell-Olds T (2002). Evidence for regulation of resistance in Arabidopsis to Egyptian cotton worm by salicylic and jasmonic acid signaling pathways. *Planta* **214**, 648-652.

Talekar NS, Shelton AM (1993). Biology, ecology and management of diamondback moth. *Annu. Rev. Entomol.* **38**, 275-301.

Thangstad OP, Gilde B, Chadchawan S, Seem M, Husebye H, Bradley D, Bones AM (2004). Cell specific, cross-species expression of myrosinases in *Brassica napus*, *Arabidopsis thaliana* and *Nicotiana tabacum*. *Plant Mol. Biol.* **54**, 597-611.

Textor S, Bartram S, Kroymann J, Falk KL, Hick AJ, Pickett JA, Gershenzon J (2004). Biosynthesis of methionine-derived glucosinolates in *Arabidopsis thaliana*: recombinant expression and characterization of methylthioalkylmalate synthase, the condensing enzyme of the chain-elongation cycle. *Planta* **218**, 1026-1035.

Textor S, De Kraker JW, Hause B, Gershenzon J, Tokuhisa J (2007). MAM3 catalyzes the formation of all aliphatic glucosinolate chain lengths in Arabidopsis. *Plant Physiol.* **144**, 60-71.

Traw MB, Dawson TE (2002). Differential induction of trichomes by three herbivores of black mustard. *Oecologia* **131**, 526–532.

Underhill EW (1968). Biosynthesis of mustard oil glucosides: 3-benzylmalic acid, a precursor of 2-amino-4-phenylbutyric acid and of gluconasturtiin. *Can. J. Biochem.* **46**, 401-405.

Van Poecke RMP, Dicke M (2004). Indirect defence of plant against herbivores: using *Arabidopsis thaliana* as a model plant. *Plant Biol.* **6**, 387-401.

Van Poecke RMP (2007). Arabidopsis-insect interactions. In CR Somerville, EM Meyerowitz eds, The Arabidopsis Book. American Society of Plant Biologists, Rockville, MD, doi:10.1199/tab.0107, www.aspb.org/publications/arabidopsis/

Voelckel C, Baldwin IT (2004). Generalist and specialist lepidopteran larvae elicit different transcriptional responses in *Nicotiana attenuata*, which correlate with larval FAC profiles. *Ecol. Letters* **7**, 770-775.

Vogel H, Kroymann J, Mitchell-Olds T (2007). Different transcript patterns in response to specialist and generalist herbivores in the wild Arabidopsis relative *Boechera divaricarpa*. *PLoS ONE* **2**, e1081.

Walling LL (2000). The myriad plant responses to herbivores. *J. Plant Growth Regul.* **19**, 195-216.

Wheat C, Vogel H, Wittstock U, Braby MF, Underwood D, Mitchell-Olds T (2007). The genetic basis of a plant-insect coevolutionary key innovation. *Proc. Natl. Acad. Sci. USA* **104**, 20427-20431.

Wilson IW, Schiff CL, Hughes DE, Sommerville SC (2001). Quantitative trait loci analysis of powdery mildew disease resistance in the *Arabidopsis thaliana* accession Kashmir-1. *Genetics* **158**, 1301-1309.

Windsor A, Reichelt M, Figuth A, Svatos A, Kroymann J, Kliebenstein DJ, Gershenzon J, Mitchell-Olds T (2005). Geographic and evolutionary diversification of glucosinolates among near relatives of *Arabidopsis thaliana* (Brassicaceae). *Phytochemistry* **66**, 1321-1333.

Wittstock U, Agerbirk N, Stauber EJ, Olsen CE, Hippler M, Mitchell-Olds T, Gershenzon J, Vogel H (2004). Successful herbivore attack due to metabolic diversion of a plant chemical defense. *Proc. Natl. Acad. Sci. USA* **101**, 4859-4864.

Zarate SI, Kempema LA, Walling LL (2007). Silverleaf whitefly induces salicylic acid defenses and suppresses effectual jasmonic acid defenses. *Plant Physiol.* **143**, 866-875.

Zavala JA, Patankar A, Gase K, Hui D, Baldwin IT (2004). Manipulation of endogenous trypsin proteinase inhibitor production in *Nicotiana attenuata* demonstrates their function as antiherbivore defenses. *Plant Physiol.* **134**, 1181-1190.

Zeller G, Clark RM, Schneeberger K, Bohlen A, Weigel D, Rätsch G (2008). Detecting polymorphic regions in *Arabidopsis thaliana* with resequencing mircoarrays. *Genome Res.* **18**, 918-929.

Zhang Z, Ober JA, Kliebenstein DJ (2006). The gene controlling the quantitative trait locus *EPITHIOSPECIFIER MODIFIER1* alters glucosinolate hydrolysis and insect resistance in *Arabidopsis*. *Plant Cell* **18**, 1524-1536.

Zhao Y, Hull AK, Gupta NR, Goss KA, Alonso J, Ecker JR, Normanly J, Chory J, Celenza JL (2002). Trp-dependent auxin biosynthesis in Arabidopsis: involvement of cytochrome P450s CYP79B2 and CYP79B3. *Gen. Dev.* **16**, 3100-3112.

Zhou N, Tootle TL, Glazebrook J (1999). Arabidopsis *PAD3*, a gene required for camalexin biosynthesis, encodes a putative cytochrome P450 monooxygenase. *Plant Cell* **11**, 2419-2428.

7. Acknowledgements

I would like to thank:

Juergen Kroymann for supervision and, together with **Heiko Vogel**, for their teaching me many molecular and genetic methods, and for their constructive comments and discussions,

Tom Mitchell-Olds for providing me the opportunity to work in his lab,

Michael Reichelt for sharing his knowledge and for constructive advice on glucosinolate isolation and analytical assistance in enzyme assays,

John D'Auria for his help with plant transformation and for proofreading of manuscripts,

Wilhelm Boland, **Jonathan Gershenzon**, **David Heckel** and **Ian T. Baldwin** for continuing support after the department of Genetic & Evolution was closed down,

Markus Benderoth for pleasant colleagueship in the lab and office,

Choon Wei Wee for sharing his knowledge on cytochrome P450s,

Yannick Pauchet for providing the basis for insect cell transfection and microsome isolation,

G**rit Kuhnert** and **Sybille Unsicker** for kindly providing aphids and sharing their knowledge in aphid rearing and handling,

Domenica Schnabelrauch for sequencing,

Petra Eismann and **Katrin Eberhardt** for helping with plant work and data collection,

Jutta Steffen for helping with insect rearing,

Andrea Hirsch for lots of HiWi help,

the **gardeners**, the **IT** and **library** teams, and the **administration** of the MPICE for friendly support,

my family for their permanent support during the many years of study,

Maria Soledade C. Pedras for providing a glucosinolate intermediate that is not commercially available,

the **referees** for their efforts,

the Max Planck Society, the European Union (NATURAL), the BMBF (GABI-Gènoplante) and the DFG for funding,

"Die Herde" for their friendship.

8. Eigenständigkeitserklärung

Ich erkläre hiermit, daß mir die geltende Promotionsordnung der Biologisch-Pharmazeutischen Fakultät der Friedrich-Schiller-Universität Jena bekannt ist.

Die vorliegende Dissertation habe ich selbstständig angefertigt. Alle von mir benutzten Hilfsmittel, persönliche Mitteilungen und Quellen sind angegeben.

Personen, die mich bei der Auswahl und Auswertung des Materials sowie bei der Erstellung der Manuskripte unterstützt haben, sind in der Auflistung der Manuskripte genannt.

Ich habe weder die Hilfe eines Promotionsberaters in Anspruch genommen noch haben Dritte für Arbeiten, die im Zusammenhang mit dem Inhalt der vorliegenden Dissertation stehen, geldwerte Dienstleistungen erhalten.

Die vorgelegte Dissertation wurde weder als Prüfungsarbeit für eine Staatliche oder andere Prüfung noch als Dissertation an einer anderen Hochschule eingereicht.

Marina Pfalz

Jena, den 23. Oktober 2008

9. Curriculum vitae

Personal data

Name:	Marina Pfalz
Date of birth:	03.01.1973
Place of birth:	Dresden, Germany
Address:	Gartenstr. 7, D-07743 Jena

Education

2004 – present	PhD student; Max Planck Institute for Chemical Ecology, Jena
2001	Diploma in Biology; Friedrich Schiller University, Jena
1993 – 1997, 1998 – 2001	Studies of Biology; Friedrich Schiller University, Jena
1990 – 1992	High School Exam (Abitur); Polytechnische Oberschule Friedrich Engels, Dresden
1980 – 1990	Polytechnische Schule Gottfried Semper, Dresden

Professional and Scientific Career

2004 – 2008	Ph.D thesis; Max Planck Institute for Chemical Ecology, Jena (Supervisor Dr. habil. (Paris-Orsay) Juergen Kroymann): Analysis of Plant Resistance against Herbivorous Insects with Molecular and Quantitative Genetics
2002 – 2003	Research associate at the Clinical Research Center Lobeda/Jena
1999 – 2001	Diploma thesis; Friedrich Schiller University, Jena (Supervisor HDoz. Dr. Helga Dietrich): Molekularsystematische und palynologische Untersuchungen innerhalb der Laeliinae (Orchidaceae)
1997 – 1998	Participation in a rainforest reforestation project; Fundacion Jatun Sacha, Ecuador

10. Scientific publications & Conference contributions

Publications:

Pfalz M, Vogel H, Kroymann J. The gene controlling the *Indole Glucosinolate Modifier 1* QTL alters indole glucosinolate structures and aphid resistance in Arabidopsis. *Plant Cell*, *in review*.

Pfalz M, Vogel H, Kroymann J. Analysis of fitness consequences of *Plutella*-regulated genes in Arabidopsis. *PLoS ONE*, *in review*.

Benderoth M, **Pfalz M**, Kroymann J (2008). Methylthioalkylmalate synthases: genetics, ecology and evolution. *Phytochem. Rev.* DOI.101007/s11101-008-9097-1.

Böttcher C, Centeno D, Freitag J, Höfgen R, Köhl K, Kopka J, Kroymann J, Matros A, Mock HP, Neumann S, **Pfalz M**, von Roepenack-Lahaye E, Schauer N, Trenkamp S, Zubriggen M, Fernie AR (2008). Teaching (and learning from) metabolomics: The 2006 PlantMetaNet ETNA metabolomics research school. *Physiol. Plant.* **132**, 136-149.

Pfalz M, Vogel H, Mitchell-Olds T, Kroymann J (2007). Mapping of QTL for resistance against the crucifer specialist insect herbivore *Pieris brassicae* in a new Arabidopsis recombinant inbred line population, Da(1)-12 × Ei-2. *PLoS ONE* **2**, e578.

Wölfl S, Dummer A, Pusch L, **Pfalz M**, Wang L, Clement JH, Leube I, Ehricht R (2004). Analyzing proteins and protein modifications with ArrayTube™ antibody microarrays. *In:* Mark Schena (ed.) *Protein Microarrays*. Jones & Bartlett, Boston, MA.

Verbal presentations:

Kurt Mothes Workshop, Jena, Germany 09/2008
Pfalz M, Vogel H, Kroymann J: The genetic basis of variation in Arabidopsis indole glucosinolate biosynthesis

MPICE Institute Symposium, Jena, Germany 09/2007
Pfalz M: *IGM1* alters Indole Glucosinolate Structures and Aphid Resistance in Arabidopsis

Institute for Plant Physiology, University of Jena, Germany 06/2006
Pfalz M, Vogel H, Kroymann J: Analysis of Arabidopsis Resistance against Crucifer Specialist Insects

GABI-Génoplante Status Meeting, Potsdam, Germany 02/2005
Pfalz M, Vogel H, Kroymann J: Insect Resistance in *Arabidopsis thaliana*

NATURAL Workshop in Golm, Germany 11/2004
Pfalz M, Eberhardt K, Mitchell-Olds T, Kroymann J: New Arabidopsis RILs & Insect Resistance QTL

Posters:

MPICE Institute Symposium, Jena, Germany 09/2008
Pfalz M, Vogel H, Kroymann J: *Indole Glucosinolate Modifier 1* alters Indole Glucosinolate Structures and Aphid Resistance in Arabidopsis

18th International Conference on Arabidopsis Research, Beijing, China 06/2007
Pfalz M, Vogel H, Kroymann J: Analysis of Arabidopsis Resistance against Crucifer Specialist Insects

1st International Glucosinolate Conference, Jena, Germany 09/2006
Pfalz M, Vogel H, Kroymann J: Mapping of QTL for *Pieris* Resistance and for Defense-related Traits in Arabidopsis Da(1)-12 x Ei-2 Recombinant Inbred Lines

MPICE Institute Symposium, Jena, Germany 06/2006
Pfalz M, Vogel H, Kroymann J: Analysis of Plant Resistance against Crucifer Specialist Insects

15th International Conference on Arabidopsis Research, Berlin, Germany 07/2004
Pfalz M, Vogel H, Mitchell-Olds T, Kroymann J: Da(1)-12 × Ei-2 Recombinant Inbred Lines: A Tool for Mapping Genes that Control Resistance to Specialist Insect Herbivores