

## Review

Lea Møller Jensen, Barbara Ann Halkier and Meike Burow\*

# How to discover a metabolic pathway? An update on gene identification in aliphatic glucosinolate biosynthesis, regulation and transport

**Abstract:** Identification of enzymes, regulators and transporters involved in different metabolic processes is the foundation to understand how organisms function. There are, however, many difficulties in identifying candidate genes as well as in proving their *in vivo* roles. In this review, we describe different approaches utilized in *Arabidopsis thaliana* to identify gene candidates and experiments required to prove the function of a given candidate. For example, we use the production of methionine-derived aliphatic glucosinolates that represent major defence compounds in *A. thaliana*. Nearly all biosynthetic genes, as well as the first sets of regulators and transporters, have been identified. An array of approaches, i.e. classical mapping, quantitative trait loci (QTL) mapping, eQTL mapping, co-expression, genome wide association studies (GWAS), mutant screens and phylogenetic analyses, has been exploited to increase the number of identified genes. Here we summarize the lessons learned from the different approaches used over the years with the aim to help designing and combining new approaches in the future.

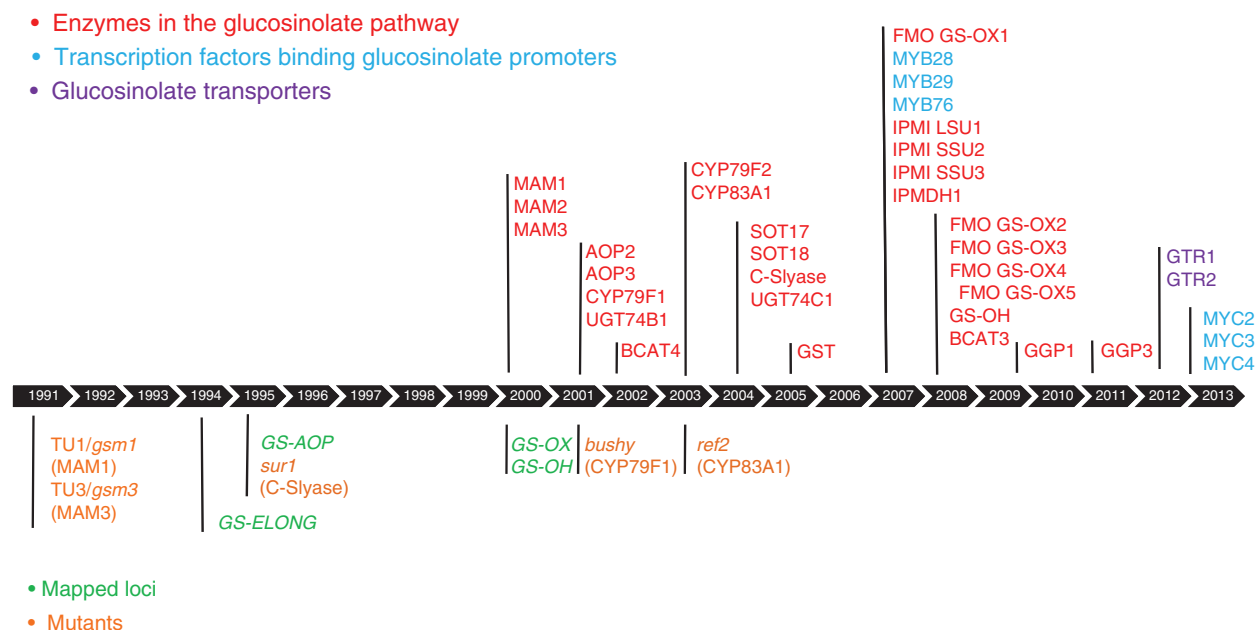
**Keywords:** aliphatic glucosinolates; *Arabidopsis thaliana*; gene identification; mapping; mutant screen.

## Introduction

The production of metabolites is essential for any organism. Primary metabolism has for many years been of main interest, but increasing attention has been paid to the synthesis and function of specialized metabolites. Chemically diverse classes of secondary metabolites are found in different plant species and serve important biological functions, e.g., as defence compounds against herbivores and pathogens, as attractants of pollinators and as storage and signaling compounds. Various approaches have been used to identify enzymes catalyzing steps in biosynthetic pathways, regulators controlling production and transport proteins moving intermediates and end products to where they are needed. In this review, we describe and discuss different approaches for gene identification in *Arabidopsis thaliana*. This well-studied model organism is characterized by a short generation time, small size, small genome and sequence information on several hundred naturally varying accessions adapted to different natural environments. As *A. thaliana* is inbreeding, these different accessions have a high degree of homozygosity. Differences among accessions provide a basis for identification of genes underlying a variety of developmental, physiological, biochemical and environmental traits (Alonso-Blanco and Koornneef, 2000). Gene identification has become increasingly feasible in *A. thaliana* as it is the first plant sequenced. On top of the growing amount of sequence information for a large number of different accessions, several mapping populations and mutant collections are available that can be exploited to assign gene functions (Weigel, 2012). The combination of tools in *A. thaliana* enables the use of many different strategies to find candidate genes for a given trait. In this review, we focus on the identification of biosynthetic genes, regulators and transporters of a group of specialized metabolites, i.e., aliphatic glucosinolates (Figure 1). A broad array of different approaches has been used in order

\*Corresponding author: Meike Burow, DynaMo Center of Excellence, Department of Plant and Environmental Sciences, Faculty of Science, University of Copenhagen, Thorvaldsensvej 40, DK-1871 Frederiksberg, C., Denmark, e-mail: mbu@plen.ku.dk  
Lea Møller Jensen and Barbara Ann Halkier: DynaMo Center of Excellence, Department of Plant and Environmental Sciences, Faculty of Science, University of Copenhagen, Thorvaldsensvej 40, DK-1871 Frederiksberg, C., Denmark

Brought to you by | University of California at Los Angeles  
Authenticated | 128.97.90.221  
Download Date | 5/5/14 2:06 AM



**Figure 2** Time line of identification of genes related to the aliphatic glucosinolate pathway.

The time line gives an overview of when mutants (orange) and mapped loci (green) for aliphatic glucosinolate production were identified. Included are biosynthetic enzymes in the pathway (red), transcription factors binding to promoters of glucosinolate biosynthetic genes (blue) and glucosinolate transporters (purple).

locus (Magrath et al., 1994). Six years later, fine-scale mapping in Ler × Col-0 RILs using cosmids, YACs and BACs identified a 1.2 kb region without recombination events. This fragment was sequenced and two 2-isopropylmalate synthases (IPMSs, later named MAM1/2) were identified as top candidates for the condensation of the 2-oxo acid with acetyl-CoA based on their homology to the respective enzymes in leucine biosynthesis (Figure 1) (Chisholm and Wetter, 1964; de Quiros et al., 2000). While the mapping populations were able to point out the locus responsible for the ratio of C3 to C4 glucosinolates, gene identification was based on the enzymatic reaction proposed for chain elongation of methionine. As early as 1991, an EMS mutant screen with glucosinolates as phenotype (see section on mutant screens) had revealed the mutant TU1/*gsm1* with nearly no C4 glucosinolates (Haughn et al., 1991), but the underlying *MAM1* gene was not identified.

The two *MAM* genes in the *GS-ELONG* locus were further investigated as the identification of the genes could not explain if the elongated 2-oxo acid was to be further chain-elongated or enter core biosynthesis. The *MAM1* in Col-0 was shown to be responsible for two rounds of chain elongation, whereas C3 profiles were only found in plants homozygous for *MAM2* or in plants expressing a non-functional *MAM1*. Enzyme assays after heterologous expression in *Escherichia coli* showed that *MAM1* was able to catalyze

the condensation reaction of 4-methylthio-2-oxo-butanoic acid and acetyl-CoA required for production of C4 glucosinolates, but neither a mutated *MAM1* nor *MAM2* could catalyze this reaction (Kroymann et al., 2001). However, *MAM2* was later re-characterized by additional enzyme assays, showing that the enzyme does indeed show activity towards 4-methylthio-2-oxo-butanoic acid, but only *MAM1* converts the 5-methylthio-2-oxo-pentanoic acid, which leads to the production of C4 glucosinolates (Benderoth et al., 2006). Furthermore, 27 *A. thaliana* accessions were sequenced for *GS-ELONG*, which showed variance in *MAM1* and *MAM2* sequences and frequency, whereas an additional gene in the same locus, *MAM3*, was found in all accessions analyzed (Kroymann et al., 2003). Enzyme kinetics show that *MAM3* can catalyze condensation of 4-methylthio-2-oxo-butanoic acids, leading to formation of C3 to C8 glucosinolates (Textor et al., 2007). The variation in the ratio of C3 to C4 glucosinolates can therefore be explained by variation in *MAM1* and *MAM2* alone. Mathematical modeling has been used in order to understand how *MAM1* and *MAM3* play together in Col-0 (Knoke et al., 2009), but it is still not entirely understood how *MAMs* control the glucosinolate chain length *in planta*. For *GS-ELONG*, the classical mapping approach identified the major locus containing the genes responsible for the chain length variation. The discrete C4 phenotype is linked to the presence or absence of *MAM1* expression. Both *MAM2*

and *MAM3* seem to contribute to the C3 production, which could theoretically have resulted in mapping both *MAM2* and *MAM3*. These genes are, however, tightly linked and *MAM3* shows little variation, which made it impossible to identify them as two loci.

Distinct variation in leaf glucosinolate profiles exists for side chain modifications of C3 and C4 glucosinolates (Figure 1). Absence or presence of hydroxyalkyl and alkenyl glucosinolates produced from methylsulfinylalkyl glucosinolates are observed as discrete profiles. Genes controlling the three different glucosinolate profiles methylsulfinylalkyl, hydroxyalkyl, and alkenyl were mapped in two RIL populations (Limburg-5 × H51 and Ler × Col-0), which each segregated for accumulation of two of the profiles. Both identified a locus in a similar position on chromosome 4 (later named *GS-AOP*) (Mithen et al., 1995). For the fine-scale mapping in Ler × Col-0 and Ler × Cvi RILs, three different BACs were found to cover the region (Kliebenstein et al., 2001c). Little fine-scale mapping was required as the sequences of the BACs became available in 1998 and 1999 (The Arabidopsis Information Resource, TAIR, [www.arabidopsis.org](http://www.arabidopsis.org)), which made gene identification less labor-intensive.

Even with the sequence information at hand, it is still important to choose the correct gene and prove its contribution to the phenotype. The sequence of the *GS-AOP* locus revealed two 2-oxoglutarate-dependent dioxygenases (2-ODDs), *AOP2* and *AOP3*, as candidate enzymes for the side chain modification reactions resulting in alkenyl and hydroxyalkyl glucosinolates, respectively. Additionally, an *AOP2* allele with a frame shift mutation was found in Col-0, giving rise to a non-functional enzyme. In any given tissue, only one allele from the *GS-AOP* locus is expressed. The enzymatical non-functional *AOP2* allele was therefore proposed to cause accumulation of the methylsulfinylalkyl glucosinolate substrates (Figure 1). *AOP2* and *AOP3* were further investigated for *in planta* gene expression and association with glucosinolate profiles as well as *in vitro* enzyme assays (Kliebenstein et al., 2001c). Due to the three discrete phenotypes observed, two different mapping populations had to be used, which unexpectedly led to identification of three different allelic states in the same locus.

Qualitative mapping is a powerful approach to investigate genes controlling phenotypes that are indeed discrete and expected to be explained by either one or two genes. A limitation occurs if the phenotype depends on many genes varying in the background, as this would mask the segregation pattern of the major locus. In these situations a quantitative mapping approach (see later) is required, which is possibly followed up by fine-scale

mapping in introgression lines, where only the locus of interest varies, while the rest of the genome is stable.

## Quantitative trait loci (QTL) mapping can identify loci controlling levels and ratios

Most phenotypes are multigenic and thus not discrete but continuous. The continuous distribution varies among accessions because of polymorphisms in the many genes controlling this kind of phenotype. Segregating mapping populations have different combinations of segregating loci, which makes it possible to associate markers with the variation in the phenotypes (Alonso-Blanco and Koornneef, 2000). This approach relies on having markers for the entire genome for each plant, which can then be associated with the phenotypes to identify QTLs that are above the significant threshold and are therefore better linked than by random chance. In contrast to classical mapping approaches, QTL mapping can be used to identify loci that are highly dependent on the rest of the genome. This could be phenotypes where the outcome is not a new feature, but for instance a shift in size, timing or metabolite levels, where several genes are expected to be involved and contribute to the observed variation.

The total levels of aliphatic glucosinolates vary among accessions, which makes it possible to map loci controlling the levels of glucosinolates produced. A QTL mapping in Ler × Cvi RILs showing transgressive segregation with a 39-fold difference in glucosinolate levels with a heritability of 0.8 pointed out that several genes are involved in controlling the levels of glucosinolates (Kliebenstein et al., 2001a). Three major QTLs were found to control total levels of glucosinolates in seeds and three were found for leaves. Two of the leaf QTLs were positioned close to *GS-ELONG* and *GS-AOP*. Even though a tightly linked QTL cannot be excluded based on mapping, it was suggested that the allelic status of *GS-ELONG* and *GS-AOP* in addition to controlling the chemical diversity of the glucosinolates produced, also control the levels (Kliebenstein et al., 2001a). However, mapping approaches cannot give any information on how this could be possible. Ectopic expression of *AOP2* in Col-0 has nevertheless shown that *AOP2* is able to increase total levels of glucosinolates as well as to increase transcript levels of the biosynthetic genes (Wentzell et al., 2007) by an unknown molecular mechanism. Thus, additional studies are needed to understand the regulatory function of *GS-AOP* and *GS-ELONG* as



well as to identify the underlying genes in the other loci identified.

Another naturally varying quantitative trait seen for aliphatic glucosinolates is the ratio of methylthioalkyl to methylsulfinylalkyl glucosinolates (Figure 1). Variation in the ratio of substrate to product could be a consequence of differential regulation or variation in the enzyme catalyzing this step. Glucosinolate profile based QTL mapping enabled identification of the *GS-OX* locus as responsible for the ratio (Kliebenstein et al., 2001b). Fine-scale mapping in a Wei-0 × Ler F2 population and expression quantitative trait locus (eQTL, see next section) data identified five flavin-dependent monooxygenases (FMO), *GS-OX1-5*, suggesting that the difference in ratios are mainly caused by different enzymes (Li et al., 2008). FMO *GS-OX1* had been identified earlier based on co-expression analysis and the proposed biochemical enzymatic reaction (Hansen et al., 2007). Even though *GS-OX1* was shown to catalyze the oxygenation of 4-methylthiobutyl *in vitro*, the knock out still contained the product of the reaction, i.e. 4-methylsulfinylbutyl (Hansen et al., 2007) pointing to redundant genes being expressed as confirmed by Li et al. (2008). *In vitro* studies revealed broad specificity of *GS-OX1-4*, whereas *GS-OX5* showed preference for the C8 substrate. Furthermore, the enzymes are capable of catalyzing the conversion of the desulfo-methylthioalkyl glucosinolates *in vitro* (Figure 1), suggesting that this modification might take place before the sulfation *in planta*.

*Arabidopsis thaliana* accessions also show variation in 2-hydroxybut-3-enyl glucosinolate production (Kliebenstein et al., 2001b). A QTL mapping approach in Ler × Cvi RILs identified a region with two enzymes with the potential to catalyze the hydroxylation of but-3-enyl glucosinolate (Kliebenstein et al., 2001a; Hansen et al., 2008). To limit the two candidate genes to one, further mapping was initiated in a Tac × Cvi F2 population that was chosen based on keeping the genes producing the but-3-enyl substrate constant. This made it possible to map a smaller region only containing one 2-ODD (At2g25450), hereafter *GS-OH* (Hansen et al., 2008). Exploiting natural variation combined with feeding experiments with but-3-enyl glucosinolate showed that the conversion to (*R/S*)-2-hydroxybut-3-enyl glucosinolate correlated with 2-ODD expression levels (Hansen et al., 2008).

QTL mapping is a very useful approach to identify loci controlling complex continuous phenotypes. Furthermore, loci controlling qualitative traits can also be found by QTL mapping. Thus, a QTL approach might reveal more players, but a qualitative approach offers advantages for fine-scale mapping. The mapping resolution is indeed one of the major issues as there is a bias for choosing genes

encoding known proteins once a genomic region has been associated with the phenotype. Thus, there might be a risk to overlook genes encoding proteins with unknown function or non-coding RNAs.

## eQTL mapping and omics approaches facilitate identification of direct regulators

Variation in gene expression among accessions is caused by polymorphisms either within the genes or in the genes controlling their expression. This variation can be mapped as *cis*- or *trans*-expression quantitative trait loci (e-QTLs). *Cis*-eQTLs have the potential to identify genes that are differentially expressed among accessions and thereby control a specific phenotype. This can for instance be used to identify the gene of interest in a previously mapped QTL, as an expression polymorphism instead of an enzymatic polymorphism may be responsible for the given trait. *Trans*-eQTLs potentially represent strong candidates for new regulatory genes, as they control expression of other genes. A global eQTL mapping study can identify from 5000 to 37 000 eQTLs (Keurentjes et al., 2007; West et al., 2007; Cubillos et al., 2012).

As mentioned above, QTL mapping of genes controlling glucosinolate accumulation identified three loci to control seed accumulation and three loci for leaf accumulation (Kliebenstein et al., 2001a). The seed QTL found at the bottom of chromosome V was also found by mapping leaf glucosinolate accumulation in a RIL population of Sha × Bay (Sonderby et al., 2007). This QTL was compared to an eQTL study that identified both network eQTLs and eQTL of the individual glucosinolates biosynthetic genes (Kliebenstein et al., 2006). To identify potential regulators within the QTL region, 63 genes with *cis*-eQTLs were identified and checked for being *trans*-eQTLs for the pathway. These were analyzed for co-expression with the aliphatic glucosinolate biosynthetic pathway, which revealed one gene encoding a R2R3 MYB transcription factor, *MYB28* (Sonderby et al., 2007). A large amount of data was included in order to find the gene underlying this QTL. Phylogenetic analysis based on *MYB28* led to the identification of two close homologues, *MYB29* and *MYB76*, which showed co-expression with the aliphatic glucosinolate biosynthetic genes (Sonderby et al., 2007).

Two other groups simultaneously identified the MYBs as regulators of aliphatic glucosinolate biosynthesis. Using co-expression data (see next section), *MYB28* and

MYB29 were found due to conditional co-expression with the aliphatic glucosinolate biosynthetic genes (Hirai et al., 2007). In parallel, transactivation assay-based screening of a collection of MYB transcription factors was used to test for the proteins' ability to activate aliphatic biosynthetic genes (Gigolashvili et al., 2007). This approach did indeed identify MYB28, MYB29, and MYB76 as direct activators of the pathway. The direct activation of the biosynthetic genes is not accounted for by the other approaches and therefore provides additional information on the identified candidates. Conversely, the screen was limited to MYB transcription factors that directly activate a defined subset of biosynthetic genes.

In order to confirm that the identified MYB transcription factors are regulators of the pathway *in planta*, several approaches have been utilized. The strongest evidence was obtained by analyses of different kinds of MYB knock outs and overexpression lines that showed effects on glucosinolate levels as well as expression levels of biosynthetic transcripts (Gigolashvili et al., 2007; Hirai et al., 2007; Sonderby et al., 2007; Beekwilder et al., 2008; Gigolashvili et al., 2008). These studies did also reveal that the MYB transcription levels are not sufficient to predict the outcome of glucosinolate production and that the transcription factor proteins are post-transcriptionally regulated.

Yeast-two-hybrid and pull-down experiments have shown that MYB28, MYB29, and MYB76 interact with the bHLH transcription factors MYC2, MYC3, and MYC4 (Schweizer et al., 2013). The identification of MYCs as regulators of glucosinolate biosynthesis was based on transcriptome data of the *myc234* triple knock out that showed reduced expression levels of glucosinolate biosynthetic genes and produced nearly no glucosinolates. These three MYCs were not found as *trans*-eQTLs for the pathway, which indicates that their individual transcript levels are not the determining factor for regulating transcript levels of the pathway. As MYC2, 3 and 4 interact with both MYBs and JAZs (Qi et al., 2011; Schweizer et al., 2013) further studies are needed to understand, how they regulate glucosinolate biosynthesis.

As described above, eQTL studies are able to predict the involvement of already known regulatory networks as well as new players. However, there are some strict limitations to what kind of new players that can be identified. Transcript levels need to reflect the network to a large enough degree, as post-transcriptional regulation cannot be accounted for. Keeping this in mind, eQTL studies have the capacity to identify new candidates for regulatory networks of pathways, where little is known about the regulation. A requirement is that at least some genes

in a pathway are regulated by transcription factors that increase transcription levels of the pathway.

## Co-expression analysis enables identification of biosynthetic genes based on their expected enzymatic functions

Several databases are available for *A. thaliana*, especially for transcriptome data. The information found in these databases is useful for finding genes that co-express globally, in a given tissue, at a certain developmental stage, under specific conditions, or in a mutant without the need to conduct wet lab experiments. In order to find a new gene in a pathway by co-expression, it is a prerequisite to know some genes in the biosynthetic pathway and to know that these genes co-express. The co-expression-based strategy has – however – limitations, for instance if a gene is constitutively expressed or involved in several pathways. If co-expression occurs, the guilt-by-association principle is so significant that there is a good chance of identifying a gene involved in the pathway (Gachon et al., 2005). The assigned gene function needs to be validated *in planta* to reduce the potential risk to find a gene that co-expresses because of a specific stimulus that induces several pathways.

The formation of chain-elongated methionine derivatives in the biosynthesis of aliphatic glucosinolates resembles the conversion of valine to leucine in primary metabolism (Chisholm and Wetter, 1964). It was therefore hypothesized that similar enzymes would take part in the chain elongation of methionine (Sawada et al., 2009a). Thus, when genes encoding an isomerase and a dehydrogenase were found to co-express with glucosinolate biosynthetic genes, they were immediately proposed as candidate genes (Hirai et al., 2007). The knock out mutants of these candidate genes (*IPMI LSU1*, *IPMI SSU2*, *IPMI SSU3* and *IPMDH1*) showed higher accumulation of C3 glucosinolates and a decrease in C4–C8 glucosinolates. Furthermore, in some of the mutants, the level of other methionine-derived metabolites was increased, which suggested that the candidate genes are in fact involved in methionine chain elongation (Knill et al., 2009; Sawada et al., 2009a). In spite of this, aliphatic glucosinolates were still found in the knock out mutants in nearly as high quantities as the wild type indicating compensation by other enzymes. The relatively small changes in overall glucosinolate production possibly explain why

these genes were not identified as major QTLs in mapping populations.

Additionally, two glutathione-S-transferases (GSTs) showed co-expression with the aliphatic glucosinolate pathway (Hirai et al., 2005). To date, little is known about the two GSTs identified, probably because of the difficulties of working with GST. GST have been assigned many different functions along with catalyzing the glutathione transfer to an unstable *aci*-nitro intermediate (Dixon et al., 2010). Thus, involvement of a GST in glucosinolate biosynthesis seems likely for the production of a glutathione conjugate in glucosinolate biosynthesis (Figure 1). The latter was suggested based on the observation that glutathione-deficient mutants produce less glucosinolates (Schlaeppi et al., 2008; Geu-Flores et al., 2009).

Co-expression analysis with *SUR1* and *CYP83B1* of the glucosinolate core structure biosynthesis identified a gamma-glutamyl peptidase (hereafter GGP1) that was shown to be able to cleave glutathione conjugate intermediates in the glucosinolate pathway (Geu-Flores et al., 2009). Analysis of a *ggp1* knock down mutant did, however, not show any significant reduction in glucosinolate levels and the constitutively expressed GGP3 enzyme was shown to be able to catalyze the same reaction *in vitro* (Geu-Flores et al., 2011). As *GGP3* does not co-express with glucosinolate biosynthetic genes, this example of functional redundancy illustrates the importance of checking homologous genes of the proposed candidate genes. The double knock down of *GGP1* and *GGP3* was generated and showed a reduction in aliphatic glucosinolate (Geu-Flores et al., 2011) further supporting that both GGP1 and GGP3 are capable of catalyzing the respective step *in planta*.

Many biosynthetic pathways can be stimulated by external factors. It can therefore be useful to search for co-expression under different conditions. A study on sulfotransferases induced by the phytotoxin coronatine led to the identification of a sulfotransferase (*SOT16*), which together with two closely related SOTs (*SOT17* and *SOT18*), identified by phylogeny search, catalyzes the last step in the core biosynthesis of aliphatic glucosinolates (Figure 1) (Piotrowski et al., 2004). A global metabolomics study carried out under sulfur deficiency indicated the involvement of the three SOTs as well as the GSTs (see above) (Hirai et al., 2005). The identified SOTs have substrate specificity towards desulfoglucosinolates, and by comparing the substrate specificity for individual desulfoglucosinolates, *SOT17* and *18* were found to prefer different methionine-derived desulfoglucosinolates, while *SOT16* converted the tryptophan-derived desulfoglucosinolate more efficiently (Piotrowski et al., 2004; Hirai et al., 2005). The SOTs have also been shown to vary among natural accessions, but

investigations of a potential impact of their biochemical properties on glucosinolate profiles allowed the conclusion that the natural variation in glucosinolate profiles is not caused by variation in SOTs (Luczak et al., 2013). This conclusion is in agreement with the fact that the SOTs have not shown up as a major QTL in mapping studies.

Co-expression analysis is an easily accessible tool for model species. For *A. thaliana*, co-expression analysis has proven its potential to identify genes that either are part of an inducible pathway or regulators of the pathway. Several pathways in primary metabolism are more or less constitutively expressed. In such cases an approach based on co-expression analysis may not be as powerful. As for all other approaches, it is crucial to confirm the *in planta* function and involvement of the candidate genes and their closest homologues. Furthermore, one needs to keep an open mind towards the possibility that other types of genes than those expected could be part of a given pathway. This is exemplified by identification of GGP1, as this was not annotated as a member of the expected class of enzymes. Thus, biochemical bias in gene discovery may increase the risk of not identifying the correct enzyme.

## Mutant screens for identification of both biosynthetic and regulatory genes

One of the limitations of mapping approaches is the requirement for variation in the gene responsible for the phenotype of interest either by functionality or expression level. In a mutant collection, mutations have been introduced randomly across the genome, thereby enabling identification of genes with no polymorphic variation across accessions. Using mutant screens, potential difficulties in identifying genes include requirement for a gain-of-function mutation for the phenotype to be seen, redundancy with other genes, and expression only under certain conditions.

In order to set up a screen, a stable phenotype that is easy to score in a high-throughput manner is desirable. Two screens based on glucosinolate analysis have been reported. In 1991, 1200 EMS *Arabidopsis* mutants were screened and six mutants with altered glucosinolate phenotypes were identified (Haughn et al., 1991). The mutants TU1 and TU3 were later found to carry mutations in *MAM1* and *MAM3*, but at that time, the underlying genes were not identified. In another screen, glucosinolate analysis of seeds of 5500 EMS mutants revealed an enzyme required for the seed-specific production of benzoyloxy

glucosinolates (Kliebenstein et al., 2007). The screen was done after most of the other biosynthetic genes had already been identified (Figure 2). Generally, mutant screens *A. thaliana* genome sequence such a screen would have the potential to identify important genes controlling glucosinolate phenotypes.

Other mutant screens based on phenotypes not directly related to glucosinolate profiles have identified glucosinolate biosynthetic genes even though this was not the purpose. In a screen of *En-1* transposon mutants, a biosynthetic gene in the aliphatic glucosinolate pathway was identified based on its bushy phenotype suggesting a link to auxin-related genes (Reintanz et al., 2001). The mutant, *bushy*, showed a mutation in *CYP79F1*, a close homologue of genes that had already been shown to be involved in production of indole and aromatic glucosinolates. *Bushy* showed decreased levels of glucosinolates, which suggested *CYP79F1* as an enzyme in glucosinolate biosynthesis. As a side-effect of the mutation, auxin levels were increased, which caused the phenotype (Reintanz et al., 2001). Further evidence came from a targeted approach simultaneously identifying *CYP79F1* based on previously obtained knowledge of the well-studied group of CYP79s, which convert amino acids into the respective aldoximes (Nelson et al., 1993; Paquette et al., 2000; Hansen et al., 2001). Enzyme assays with a recombinant CYP79F1 protein showed specificity of the enzyme towards di- and trihomomethionine (Hansen et al., 2001). A later study demonstrated specificity of CYP79F2 towards long-chained glucosinolates (Chen et al., 2003). Knock out plants in *CYP79F1* or *CYP79F2* did indeed show drastic changes in glucosinolate profiles (Hansen et al., 2001; Reintanz et al., 2001; Chen et al., 2003). Accordingly, these mutants could have been identified in a more labor-intensive screen for mutants with altered glucosinolate profile as the phenotype.

*CYP83A1* was identified in a screen of 100 000 M2 seedlings for reduced fluorescence under UV light aiming at identifying sinapoylmalate-deficient mutants (Hemm et al., 2003). One of the mutants, *ref2*, was investigated by positional cloning, leading to the identification of 35 cosmids from the identified region. All of these were elegantly transformed into the *ref2* knock out to identify the region that complemented the phenotype. *CYP83A1* was found in all those cosmids complementing the phenotype, and analysis of the mRNA levels of *CYP83A1* in the *ref2* plant confirmed decreased *CYP83A1* expression. Furthermore, all knock out mutants were sequenced to show mutations, leading to either stop codons or amino acid changes in the *CYP83A1* gene (Hemm et al., 2003). This in-depth investigation was a powerful approach in

confirming that mutations in *CYP83A1* caused the phenotype and made it very unlikely that another gene was involved. *CYP83A1* was then related to glucosinolates as a closely related CYP83 had previously been reported to convert the tryptophan-derived aldoxime in indole glucosinolate biosynthesis (Bak and Feyereisen, 2001; Hemm et al., 2003). The analysis of a *CYP83A1* knock out mutant showed decreased levels of aliphatic glucosinolates along with the expected decrease in sinapoylmalate, which was suggested to be a side-effect of the over-accumulation of aldoximes (Hemm et al., 2003). The suggested function in aliphatic glucosinolate production was further confirmed by a concurrent study on the enzymatic activity of CYP83A1, showing substrate specificity towards intermediates of aliphatic glucosinolate biosynthetic pathway (Naur et al., 2003).

Other glucosinolate biosynthetic genes were discovered in screens for high auxin mutants. *CYP83B1* involved in indole glucosinolate biosynthesis was found in a screen as the *superroot2* (*sur2*) mutant (Delarue et al., 1998) and in an earlier screen, the *sur1* mutant had been isolated from an EMS population (Boerjan et al., 1995). It took 9 years before the *sur1* mutant was connected to glucosinolate biosynthesis. A C-S lyase was expected to be required for the biosynthesis of glucosinolates and a consensus sequence was made based on sequences from human, rat and zebrafish and blasted against the *A. thaliana* genome to identify *SUR1* that was originally annotated as an aminotransferase (Mikkelsen et al., 2004). The *sur1* mutant was completely incapable of glucosinolate production and *in vivo* feeding along with enzyme assays confirmed the enzymatic function of SUR1 (Mikkelsen et al., 2004).

Screening of mutant populations is a useful approach to identify genes leading to a specific phenotype. It is striking that several of the aliphatic glucosinolate biosynthetic genes have been identified in screens for phenotypes not directly associated with glucosinolate production. This illustrates how a phenotype can be caused by several other factors than those anticipated. Thus, the identified gene does not necessarily have the expected molecular function. The identification of *CYP79F1* (*bushy*), *CYP83A1* (*ref2*) and C-S-lyase (*sur1*) emphasizes this as the respective screens were performed to identify genes involved in auxin or sinapoylmalate metabolism. In all cases, knock out mutations in glucosinolate biosynthetic genes did indeed cause the phenotypes of interest, but due to pleiotropic effects most likely caused by intermediates of the glucosinolate pathway.

A screen for regulators of glucosinolate biosynthesis was performed with a high-throughput bioassay for leaf glucosinolate content of the progeny of 5000 T-DNA



activation tagged lines. Activation of a positive regulator of glucosinolate biosynthesis was assumed to induce glucosinolate accumulation. This screen identified a mutant of the transcription factor *IQD1*, which accumulated decreased levels of glucosinolates (Levy et al., 2005). Similarly, a screen based on an EMS-treated reporter line responsive to sulfur limitation identified the *slim1* mutant carrying a mutation in an EIF transcription factor. Along with other phenotypes, *slim1* was found to control glucosinolate levels (Maruyama-Nakashita et al., 2006). A reverse genetics approach showed that also the DOF1.1 transcription factor affects glucosinolate production (Skirycz et al., 2006). In all three cases, it remains to be shown how the transcription factors are involved in regulation of glucosinolate levels in order to place these transcription factors in the regulatory network controlling glucosinolate biosynthesis. This is in contrast to MYB28, 29 and 76 as well as MYC2, 3 and 4, as they have been shown to bind promoters of glucosinolate biosynthetic genes (Gigolashvili et al., 2007, 2008; Schweizer et al., 2013). For the identification of *SLIM1* it is not surprising that regulators of primary sulfur metabolism affect glucosinolate production, as two to three sulfur atoms are required for one aliphatic glucosinolate molecule. The identification of *SLIM1* shows how pathways cross-talk and thereby influence each other without necessarily directly affecting the transcript levels of the pathway.

## Genome wide association studies (GWAS) identify phenotype-associated genes

GWAS identify associations between genotype and phenotype in unrelated individuals. The use of unrelated individuals makes it possible to include many more different genotypes than in a QTL mapping population and offers therefore the possibility to identify other associations. GWAS have the risk of overlooking associations due to epistasis and rare genotypes. Further, false positives are frequently found, e.g. because genes are tightly linked. A large GWAS has been performed on 107 phenotypes using 96 *Arabidopsis* lines (Atwell et al., 2010). This GWAS has pinpointed associations between phenotypes and genotypes, and was able to identify some of the genes known to underlie specific phenotypes illustrating the power of the approach.

A GWAS including 96 accessions with focus on glucosinolate profiles under different conditions as the phenotype identified several hundred genes with significant associations and showed the importance of setting a

suitable threshold as well as filtering the identified candidates (Chan et al., 2011). For downstream analysis, co-expression network data was used to identify networks constituted of GWAS candidates. This led to the identification of several networks. A number of genes from these networks were chosen for knock out mutant analysis for glucosinolate profiles. The knock outs of several genes otherwise known to be involved in defense, response to blue light, chromatin remodeling and genes of unknown function all displayed altered glucosinolate production (Chan et al., 2011). However, these genes are unlikely to be direct regulators of the pathway, which makes it difficult to group the genes into regulators of the pathway, parts of other pathways that effect glucosinolate production and regulators of other pathways thereby affecting glucosinolate phenotypes.

GWAS has the potential to point out novel associations. The potential high number of false and negative associations has to be taken into account, as these can make it difficult to find the correct candidate gene. Furthermore, the associations give an array of candidate genes depending on the phenotype, which requires substantial downstream analyses to get an idea of the function. Nevertheless, GWAS have a big potential for identification of associations and in the future, it might be possible to make eGWAS, which have a higher chance of identifying regulators.

## Phylogenetic analyses identify genes belonging to well-characterized gene families

Completion of the *Arabidopsis* genome made it easier to find genes based on sequence similarity to known genes and investigate their phylogenetic relationship to other genes in the same family. The sequence information is very useful as many genes *in planta* arise from duplications that through evolution and neo-functionalization obtained specialized functions, albeit still catalyzing the same kind of reaction.

The first step in aliphatic glucosinolate biosynthesis represents the transamination of methionine. Based on their sequence similarity to yeast, human and bacterial genes, a family of amino acid transaminases was identified as candidates for this step (Diebold et al., 2002). One member of this family, BCAT4, did not complement a yeast mutant lacking endogenous BCAT activity, which made BCAT4 an interesting candidate for the aminotransferase in the glucosinolate pathway (Schuster et al., 2006).

Enzyme assays and knock out mutants showing reduced glucosinolate accumulation matched this hypothesis that was further supported by co-expression of *BCAT4* with the downstream *MAM1* gene. As another transamination step is required in glucosinolate biosynthesis, *BCAT3* was selected as additional candidate gene based on its expression pattern. Investigation of substrate specificity suggested that *BCAT3* takes part in both production of glucosinolates from methionine and amino acid biosynthesis (Knill et al., 2008). Nevertheless, no clear glucosinolate phenotype was seen in the *bcat3* knock out possibly because of compensation by *BCAT4* or other enzymes.

Another enzyme identified based on sequence similarity is UGT74B1. Glucosylation of the thiohydroximate intermediate had been proposed as a step in glucosinolate core structure biosynthesis. Based on an *S*-glycosyl transferase cDNA sequence from *Brassica napus*, a gene with 85% identity at the nucleotide level was identified and shown to be up-regulated in transgenic plants with induced glucosinolate biosynthesis (Petersen et al., 2001). Enzyme assays with the recombinant protein showed the ability of UGT74B1 to convert phenylacetothiohydroximate, an intermediate in benzyl glucosinolate biosynthesis (Grubb et al., 2004). Knock out plants only showed a 40–50% reduction in glucosinolate levels indicating that other enzymes can compensate in the knock out or redundant enzymes exist in the biosynthesis. Several phylogenetic analyses of *UGTs* have been carried out after the *A. thaliana* genome sequence became available (Li et al., 2001; Paquette et al., 2003). Nevertheless, it was a co-expression study that suggested *UGT74C1* as a likely candidate for an additional UGT involved in glucosinolate biosynthesis (Gachon et al., 2005). It is not fully understood how the *UGTs* play together in aliphatic glucosinolate biosynthesis. Until now, only UGT74B1 has been investigated in detail, with focus on the kinetic characteristics (Kopycki et al., 2013).

Identifying candidates for biosynthetic genes based on phylogenetic relationships is useful whenever proposed pathway intermediates allow prediction of the type of enzyme needed. As for any other approach to identify gene functions, it can nevertheless be difficult to prove that the chosen candidate is indeed the gene responsible for this activity *in planta*. To provide evidence for the *in planta* function of a candidate, it is important to look at expression and localization patterns along with knock out and over expression phenotypes, i.e. at levels of both end product and pathway intermediates. Phylogeny can furthermore be used across species and is thus one of the approaches that offer the possibility to speed up pathway identification in non-model species. It is important to

keep in mind that pathways may have evolved differently, and therefore can be specific for a given species and with unpredictable types of proteins involved.

The MYB transcription factors involved in aliphatic glucosinolate biosynthesis are an example of proteins identified based on similarity to other transcription factors (Gigolashvili et al., 2007) illustrating that phylogenetic analyses does also allow for identification of regulatory proteins. Used across species, it will probably be one of the ways to get the first insight into regulatory networks in new species being investigated.

## Identification of transporters requires diverse approaches

Many different approaches have been used to identify genes in the aliphatic glucosinolate pathway in *A. thaliana*. Until recently, only biosynthetic genes and regulators had been identified, although both transport of glucosinolates and intermediates is known to occur as the biosynthesis is localized partly in the cytosol and partly in the chloroplasts and as glucosinolates are subjected to long distance transport in *A. thaliana* (Nour-Eldin et al., 2012; Andersen et al., 2013). Transport of intermediates is required when and where the pathway is expressed and glucosinolates produced. The bile acid transporter 5 (*BAT5*) reported to move 2-oxo acids into the chloroplast was identified based on co-expression and activation by MYB28 (Gigolashvili et al., 2009; Sawada et al., 2009b). In support of the *in planta* function of *BAT5*, knock out mutants showed reduced levels of aliphatic glucosinolates and the protein localized to the chloroplast, but to date, there is no biochemical evidence to confirm its activity. *BAT5* had been annotated as a bile acid transporter based on homology to mammalian transporters. Thus, when identifying new genes, it needs to be taken into account that many annotations are based on sequence homology and might be misleading.

Sulfation of desulfoglucosinolates by SOTs requires PAPS, which functions as a sulfate donor (Figure 1) (Piotrowski et al., 2004). PAPS is produced in the plastids and transported into the cytosol where the sulfation reaction takes place (Klein and Papenbrock, 2004; Takahashi et al., 2011). Similarly to *BAT5*, an ADP/ATP antiporter (*TAAC*) transporter was identified based on co-expression and activation by the MYB transcription factors as a PAPS transporter candidate. Biochemical studies of the candidate protein showed its ability to transport PAP/PAPS and it was renamed to PAPST1 (Gigolashvili et al., 2012). This

illustrates that experimental evidence for one biochemical activity does not rule out other functions. An enzyme may catalyze several reactions or at least have the capability to do so because of low substrate specificity.

Transporters of pathway intermediates have to be expressed together with the biosynthetic genes unless they are involved in several metabolic processes. Conversely, transporters of glucosinolates do not need to strictly co-express, as glucosinolates do not need to be transported out of the cell directly after synthesis. However, storage within the cell might require intracellular transporters. These can be expected to co-express with the biosynthetic pathway unless they show broad substrate specificity. So far, no glucosinolate transporter has been identified based on co-expression analysis.

Using a functional genomic approach, a collection of *A. thaliana* genes predicted to encode importers was screened for uptake of 4-methylthiobutyl glucosinolate in *Xenopus oocytes* as expression system. The candidate identified in this screen and its closest homologues were investigated for localization and double knock out plants showed that the transporters were essential for import of glucosinolates into seeds (Nour-Eldin et al., 2012). Limitations of this functional genomic approach are the selection of transporters in the library, the prerequisite that the transporter consists of a single polypeptide, and the assay conditions (pH, ions, co-substrates) that may not mimic the native conditions. The transporters have not been discovered by other approaches, partially because they do not vary among accessions. A mutant screen could have revealed one of the glucosinolate transporters based on its seed glucosinolate knock out phenotype. Identification of additional transporters in mutant screens may require focus on glucosinolate distribution or uptake phenotypes.

## Concluding remarks

Many different approaches or combinations thereof can be used to identify genes in a metabolic pathway, i.e. biosynthetic enzymes, transcription factors and transporters of intermediates and end products. The use of complementary methods does not only enhance the chances of finding a gene underlying a phenotype, but also generates a stronger link between phenotype and the gene identified. The discovery of aliphatic glucosinolate biosynthetic genes illustrates that enzymes catalyzing the conversion of intermediates are most likely to be discovered by co-expression, phylogenetic analyses or in screens. Conversely, enzymes modifying end products can be found by mapping approaches if they vary among accessions,

which, does, however, create the need for identifying the correct gene within a locus.

In the case of the aliphatic glucosinolate biosynthetic pathway, the first direct regulators have all been discovered relatively recently (Figure 2), which is due to both the approaches used and the fact that the elucidation of regulatory networks is facilitated by knowledge about the biosynthetic genes in the pathway. The MYBs and MYCs are direct regulators, while *IQD1*, *SLIM1*, *DOF1.1* and genes identified in GWAS control glucosinolates by affecting connected networks. QTL mapping approaches have aimed at identifying regulators, but surprisingly identified the *GS-ELONG* and *GS-AOP* loci. These loci do not encode the expected transcription factors, but they still seem able to regulate the pathway. This emphasizes the need for detailed downstream investigation of identified candidates, as a gene might have an unexpected function or multiple functions. Furthermore, not only transcription factors but also RNAs and metabolites are involved in regulation, which suggests that it might not always be the protein encoded by the identified gene that is important for the phenotype of interest.

Identification of transporters moving pathway intermediates was guided by co-expression analysis, whereas the glucosinolate transporters known to date were found in a functional genomics screen. These approaches relied on prior knowledge of the pathway and the availability of the *A. thaliana* genome, respectively. In general, completion of the genomic sequence of *A. thaliana* around 2000 boosted the identification rate (Figure 2) because the increasing amount of sequence information made it possible to assign gene functions based on sequence similarity and to go more easily from locus to gene. Advances in microarray and RNA-sequencing technology have moreover made gene identification based on transcriptomic data more feasible. New techniques keep arising for high-throughput methods for genomics, transcriptomics, proteomics, and metabolomics allowing for quantification of different phenotypes and for new strategies to assign gene functions. An increasing number of studies combine different approaches, as the biological questions driving plant research change from identifying the enzymes and direct regulators towards upstream regulators, transporters, and cross-talk between pathways and regulatory networks. In addition to the expected enzymes, transcription factors or transporters, future approaches should be designed to also elucidate novel functions of proteins, RNAs and metabolites as players in networks.

Received November 27, 2013; accepted February 27, 2014; previously published online March 1, 2014

## References

- Alonso-Blanco, C. and Koornneef, M. (2000). Naturally occurring variation in *Arabidopsis*: an underexploited resource for plant genetics. *Trends Plant Sci.* 5, 22–29.
- Andersen, T.G., Nour-Eldin, H.H., Fuller, V.L., Olsen, C.E., Burow, M., and Halkier, B.A. (2013). Integration of biosynthesis and long-distance transport establish organ-specific glucosinolate profiles in vegetative *Arabidopsis*. *Plant Cell* 25, 3133–3145.
- Atwell, S., Huang, Y.S., Vilhjalmsdottir, B.J., Willems, G., Horton, M., Li, Y., Meng, D., Platt, A., Tarone, A.M., Hu, T.T., et al. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465, 627–631.
- Bak, S. and Feyerisen, R. (2001). The involvement of two P450 enzymes, CYP83B1 and CYP83A1, in auxin homeostasis and glucosinolate biosynthesis. *Plant Physiol.* 127, 108–118.
- Beekwilder, J., van Leeuwen, W., van Dam, N.M., Bertossi, M., Grandi, V., Mizzi, L., Soloviev, M., Szabados, L., Molthoff, J.W., Schipper, B., et al. (2008). The impact of the absence of aliphatic glucosinolates on insect herbivory in *Arabidopsis*. *PLoS One* 3, e2068.
- 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.
- Boerjan, W., Cervera, M.T., Delarue, M., Beeckman, T., Dewitte, W., Bellini, C., Caboche, M., Vanonckelen, H., Vanmontagu, M., and Inze, D. (1995). Superroot, a recessive mutation in *Arabidopsis*, confers auxin overproduction. *Plant Cell* 7, 1405–1419.
- Chan, E.K., Rowe, H.C., Corwin, J.A., Joseph, B., and Kliebenstein, D.J. (2011). Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in *Arabidopsis thaliana*. *PLoS Biol.* 9, e1001125.
- Chen, S.X., Glawischig, E., Jorgensen, K., Naur, P., Jorgensen, B., Olsen, C.E., Hansen, C.H., Rasmussen, H., Pickett, J.A., and Halkier, B.A. (2003). CYP79F1 and CYP79F2 have distinct functions in the biosynthesis of aliphatic glucosinolates in *Arabidopsis*. *Plant J.* 33, 923–937.
- Chisholm, M.D. and Wetter, L.R. (1964). Biosynthesis of mustard oil glucosides. 4. Administration of methionine-C14 + related compounds to horseradish. *Can. J. Biochem. Phys.* 42, 1033–1040.
- Cubillos, F.A., Yansouni, J., Khalili, H., Balzergue, S., Elftieh, S., Martin-Magniette, M.L., Serrand, Y., Lepiniec, L., Baud, S., Dubreucq, B., et al. (2012). Expression variation in connected recombinant populations of *Arabidopsis thaliana* highlights distinct transcriptome architectures. *Bmc Genomics* 13, 117.
- de Quiros, H.C., Magrath, R., McCallum, D., Kroymann, J., Schnabelrauch, D., Mitchell-Olds, T., and Mithen, R. (2000). Alpha-Keto acid elongation and glucosinolate biosynthesis in *Arabidopsis thaliana*. *Theor. Appl. Genet.* 101, 429–437.
- 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, 603–611.
- Diebold, R., Schuster, J., Daschner, K., and Binder, S. (2002). The branched-chain amino acid transaminase gene family in *Arabidopsis* encodes plastid and mitochondrial proteins. *Plant Physiol.* 129, 540–550.
- Dixon, D.P., Skipsey, M., and Edwards, R. (2010). Roles for glutathione transferases in plant secondary metabolism. *Phytochemistry* 71, 338–350.
- Gachon, C.M.M., Langlois-Meurinne, M., Henry, Y., and Saindrenan, P. (2005). Transcriptional co-regulation of secondary metabolism enzymes in *Arabidopsis*: functional and evolutionary implications. *Plant Mol. Biol.* 58, 229–245.
- Geu-Flores, F., Nielsen, M.T., Nafisi, M., Moldrup, M.E., Olsen, C.E., Motawia, M.S., and Halkier, B.A. (2009). Glucosinolate engineering identifies  $\gamma$ -glutamyl peptidase. *Nat. Chem. Biol.* 5, 575–577.
- Geu-Flores, F., Moldrup, M.E., Bottcher, C., Olsen, C.E., Scheel, D., and Halkier, B.A. (2011). Cytosolic  $\gamma$ -glutamyl peptidases process glutathione conjugates in the biosynthesis of glucosinolates and camalexin in *Arabidopsis*. *Plant Cell* 23, 2456–2469.
- Gigolashvili, T., Yatusovich, R., Berger, B., Muller, C., and Flugge, U.I. (2007). The R2R3-MYB transcription factor HAG1/MYB28 is a regulator of methionine-derived glucosinolate biosynthesis in *Arabidopsis thaliana*. *Plant J.* 51, 247–261.
- Gigolashvili, T., Engqvist, M., Yatusovich, R., Muller, C., and Flugge, U.I. (2008). HAG2/MYB76 and HAG3/MYB29 exert a specific and coordinated control on the regulation of aliphatic glucosinolate biosynthesis in *Arabidopsis thaliana*. *New Phytol.* 177, 627–642.
- Gigolashvili, T., Yatusovich, R., Rollwitz, I., Humphry, M., Gershenzon, J., and Flugge, U.I. (2009). The plastidic bile acid transporter 5 is required for the biosynthesis of methionine-derived glucosinolates in *Arabidopsis thaliana*. *Plant Cell* 21, 1813–1829.
- Gigolashvili, T., Geier, M., Ashykhmina, N., Frerigmann, H., Wulfert, S., Krueger, S., Mugford, S.G., Kopriva, S., Haferkamp, I., and Flugge, U.I. (2012). The *Arabidopsis* thylakoid ADP/ATP carrier TAAC Has an additional role in supplying plastidic phosphoadenosine 5'-phosphosulfate to the cytosol. *Plant Cell* 24, 4187–4204.
- Grubb, C.D., Zipp, B.J., Ludwig-Muller, 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.
- Hansen, C.H., Wittstock, U., Olsen, C.E., Hick, A.J., Pickett, J.A., and Halkier, B.A. (2001). Cytochrome P450CYP79F1 from *Arabidopsis* catalyzes the conversion of dihomomethionine and trihomomethionine to the corresponding aldoximes in the biosynthesis of aliphatic glucosinolates. *J. Biol. Chem.* 276, 11078–11085.
- Hansen, B.G., Kliebenstein, D.J., and Halkier, B.A. (2007). Identification of a flavin-monooxygenase as the S-oxygenating enzyme in aliphatic glucosinolate biosynthesis in *Arabidopsis*. *Plant J.* 50, 902–910.
- Hansen, B.G., Kerwin, R.E., Ober, J.A., Lambrix, V.M., Mitchell-Olds, T., Gershenzon, J., Halkier, B.A., and Kliebenstein, D.J. (2008). A novel 2-oxoacid-dependent dioxygenase involved in the formation of the goiterogenic 2-hydroxybut-3-enyl glucosinolate and generalist insect resistance in *Arabidopsis*. *Plant Physiol.* 148, 2096–2108.



- Haughn, G.W., Davin, L., Giblin, M., and Underhill, E.W. (1991). Biochemical genetics of plant secondary metabolites in *Arabidopsis thaliana* – the glucosinolates. *Plant Physiol.* 97, 217–226.
- Hemm, M.R., Ruegger, M.O., and Chapple, C. (2003). The *Arabidopsis* ref2 mutant is defective in the gene encoding CYP83A1 and shows both phenylpropanoid and glucosinolate phenotypes. *Plant Cell* 15, 179–194.
- Hirai, M.Y., Klein, M., Fujikawa, Y., Yano, M., Goodenowe, D.B., Yamazaki, Y., Kanaya, S., Nakamura, Y., Kitayama, M., Suzuki, H., et al. (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.
- Hirai, M.Y., Sugiyama, K., Sawada, Y., Tohge, T., Obayashi, T., Suzuki, A., Araki, R., Sakurai, N., Suzuki, H., Aoki, K., et al. (2007). Omics-based identification of *Arabidopsis* Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc. Natl. Acad. Sci. USA* 104, 6478–6483.
- Keurentjes, J.J.B., Fu, J.Y., Terpstra, I.R., Garcia, J.M., van den Ackerveken, G., Snoek, L.B., Peeters, A., J.M., Vreugdenhil, D., Koornneef, M., and Jansen, R.C. (2007). Regulatory network construction in *Arabidopsis* by using genome-wide gene expression quantitative trait loci. *Proc. Natl. Acad. Sci. USA* 104, 1708–1713.
- Klein, M. and Papenbrock, J. (2004). The multi-protein family of *Arabidopsis* sulphotransferases and their relatives in other plant species. *J. Exp. Bot.* 55, 1809–1820.
- Kliebenstein, D.J., Gershenzon, J., and Mitchell-Olds, T. (2001a). 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., Kroymann, J., Brown, P., Figuth, A., Pedersen, D., Gershenzon, J., and Mitchell-Olds, T. (2001b). Genetic control of natural variation in *Arabidopsis* glucosinolate accumulation. *Plant Physiol.* 126, 811–825.
- 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., West, M.A.L., van Leeuwen, H., Loudet, O., Doerge, R.W., and St Clair, D.A. (2006). Identification of QTLs controlling gene expression networks defined a priori. *BMC Bioinformatics* 7, 308.
- Kliebenstein, D.J., D'Auria, J.C., Behere, A.S., Kim, J.H., Gunderson, K.L., Breen, J.N., Lee, G., Gershenzon, J., Last, R.L., and Jander, G. (2007). Characterization of seed-specific benzoyloxyglucosinolate mutations in *Arabidopsis thaliana*. *Plant J.* 51, 1062–1076.
- Knill, T., Schuster, J., Reichelt, M., Gershenzon, J., and Binder, S. (2008). *Arabidopsis* branched-chain aminotransferase 3 functions in both amino acid and glucosinolate biosynthesis. *Plant Physiol.* 146, 1028–1039.
- Knill, T., Reichelt, M., Paetz, C., Gershenzon, J., and Binder, S. (2009). *Arabidopsis thaliana* encodes a bacterial-type heterodimeric isopropylmalate isomerase involved in both Leu biosynthesis and the Met chain elongation pathway of glucosinolate formation. *Plant Mol. Biol.* 71, 227–239.
- Knoke, B., Textor, S., Gershenzon, J., and Schuster, S. (2009). Mathematical modelling of aliphatic glucosinolate chain length distribution in *Arabidopsis thaliana* leaves. *Phytochem. Rev.* 8, 39–51.
- Kopycki, J., Wieduwild, E., Kohlschmidt, J., Brandt, W., Stepanova, A.N., Alonso, J.M., Pedras, M.S.C., Abel, S., and Grubb, C.D. (2013). Kinetic analysis of *Arabidopsis* glucosyltransferase UGT74B1 illustrates a general mechanism by which enzymes can escape product inhibition. *Biochem. J.* 450, 37–46.
- 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.
- Levy, M., Wang, Q.M., 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, Y., Baldauf, S., Lim, E.K., and Bowles, D.J. (2001). Phylogenetic analysis of the UDP-glycosyltransferase multigene family of *Arabidopsis thaliana*. *J. Biol. Chem.* 276, 4338–4343.
- Li, J., Hansen, B.G., Ober, J.A., Kliebenstein, D.J., and Halkier, B.A. (2008). Subclade of flavin-monooxygenases involved in aliphatic glucosinolate biosynthesis. *Plant Physiol.* 148, 1721–1733.
- Luczak, S., Forlani, F., and Papenbrock, J. (2013). Desulfo-glucosinolate sulfotransferases isolated from several *Arabidopsis thaliana* ecotypes differ in their sequence and enzyme kinetics. *Plant Physiol. Biochem.* 63, 15–23.
- Magrath, R., Bano, F., Morgner, M., Parkin, I., Sharpe, A., Lister, C., Dean, C., Turner, J., Lydiate, D., and Mithen, R. (1994). Genetics of aliphatic glucosinolates.1. Side-chain elongation in *Brassica napus* and *Arabidopsis thaliana*. *Heredity* 72, 290–299.
- Maruyama-Nakashita, A., Nakamura, Y., Tohge, T., Saito, K., and Takahashi, H. (2006). *Arabidopsis* SLIM1 is a central transcriptional regulator of plant sulfur response and metabolism. *Plant Cell* 18, 3235–3251.
- 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, 770–777.
- Mithen, R., Clarke, J., Lister, C., and Dean, C. (1995). Genetics of aliphatic glucosinolates.3. Side-chain structure of aliphatic glucosinolates in *Arabidopsis thaliana*. *Heredity* 74, 210–215.
- 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.
- Nelson, D.R., Kamataki, T., Waxman, D.J., Guengerich, F.P., Estabrook, R.W., Feyereisen, R., Gonzalez, F.J., Coon, M.J., Gunsalus, I.C., Gotoh, O., et al. (1993). The P450 superfamily – update on new sequences, gene-mapping, accession numbers, early trivial names of enzymes, and nomenclature. *DNA Cell Biol.* 12, 1–51.

- Nour-Eldin, H.H., Andersen, T.G., Burow, M., Madsen, S.R., Jorgensen, M.E., Olsen, C.E., Dreyer, I., Hedrich, R., Geiger, D., and Halkier, B.A. (2012). NRT/PTR transporters are essential for translocation of glucosinolate defence compounds to seeds. *Nature* 488, 531–534.
- Paquette, S.M., 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.
- Paquette, S., Moller, B.L., and Bak, S. (2003). On the origin of family 1 plant glycosyltransferases. *Phytochemistry* 62, 399–413.
- Petersen, B.L., Andreasson, E., Bak, S., Agerbirk, N., and Halkier, B.A. (2001). Characterization of transgenic *Arabidopsis thaliana* with metabolically engineered high levels of *p*-hydroxybenzylglucosinolate. *Planta* 212, 612–618.
- Piotrowski, M., Schemenewitz, A., Lopukhina, A., Muller, A., Janowitz, T., Weiler, E.W., and Oecking, C. (2004). Desulfo-glucosinolate sulfotransferases from *Arabidopsis thaliana* catalyze the final step in the biosynthesis of the glucosinolate core structure. *J. Biol. Chem.* 279, 50717–50725.
- Qi, T.C., Song, S.S., Ren, Q.C., Wu, D.W., Huang, H., Chen, Y., Fan, M., Peng, W., Ren, C.M., and Xie, D.X. (2011). The jasmonate-ZIM-domain proteins interact with the WD-Repeat/bHLH/MYB complexes to regulate jasmonate-mediated anthocyanin accumulation and trichome initiation in *Arabidopsis thaliana*. *Plant Cell* 23, 1795–1814.
- Reintanz, B., Lehnen, M., Reichelt, M., Gershenzon, J., Kowalczyk, M., Sandberg, G., Godde, M., Uhl, R., and Palme, K. (2001). bus, a bushy *Arabidopsis* CYP79F1 knockout mutant with abolished synthesis of short-chain aliphatic glucosinolates. *Plant Cell* 13, 351–367.
- Sawada, Y., Kuwahara, A., Nagano, M., Narisawa, T., Sakata, A., Saito, K., and Hirai, M.Y. (2009a). Omics-based approaches to methionine side chain elongation in *Arabidopsis*: characterization of the genes encoding methylthioalkylmalate isomerase and methylthioalkylmalate dehydrogenase. *Plant Cell Physiol.* 50, 1181–1190.
- Sawada, Y., Toyooka, K., Kuwahara, A., Sakata, A., Nagano, M., Saito, K., and Hirai, M.Y. (2009b). *Arabidopsis* bile acid:sodium symporter family protein 5 is involved in methionine-derived glucosinolate biosynthesis. *Plant Cell Physiol.* 50, 1579–1586.
- Schlaeppli, K., Bodenhausen, N., Buchala, A., Mauch, F., and Reymond, P. (2008). The glutathione-deficient mutant pad2-1 accumulates lower amounts of glucosinolates and is more susceptible to the insect herbivore *Spodoptera littoralis*. *Plant J.* 55, 774–786.
- Schuster, J., Knill, T., Reichelt, M., Gershenzon, J., and 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.
- Schweizer, F., Fernandez-Calvo, P., Zander, M., Diez-Diaz, M., Fonseca, S., Glauser, G., Lewsey, M.G., Ecker, J.R., Solano, R., and Reymond, P. (2013). *Arabidopsis* basic helix-loop-helix transcription factors MYC2, MYC3, and MYC4 regulate glucosinolate biosynthesis, insect performance, and feeding behavior. *Plant Cell* 25, 3117–3132.
- Skirycz, A., Reichelt, M., Burow, M., Birkemeyer, C., Rolcik, J., Kopka, J., Zanor, M.I., Gershenzon, J., Strnad, M., Szopa, J., et al. (2006). DOF transcription factor AtDof1.1 (OBP2) is part of a regulatory network controlling glucosinolate biosynthesis in *Arabidopsis*. *Plant J.* 47, 10–24.
- Sonderby, I.E., Hansen, B.G., Bjarnholt, N., Ticconi, C., Halkier, B.A., and Kliebenstein, D.J. (2007). A systems biology approach identifies a R2R3 MYB gene subfamily with distinct and overlapping functions in regulation of aliphatic glucosinolates. *PLoS One* 2, e1322.
- Takahashi, H., Kopriva, S., Giordano, M., Saito, K., and Hell, R. (2011). Sulfur Assimilation in Photosynthetic Organisms: Molecular Functions and Regulations of Transporters and Assimilatory Enzymes. In: *Annual Review of Plant Biology*, S.E. Fienberg, ed. (Palo Alto: Annual Reviews), vol. 62, pp. 157–184.
- Textor, S., de Kraker, J.W., Hause, B., Gershenzon, J., and Tokuhisa, J.G. (2007). MAM3 catalyzes the formation of all aliphatic glucosinolate chain lengths in *Arabidopsis*. *Plant Physiol.* 144, 60–71.
- Weigel, D. (2012). Natural variation in *Arabidopsis*: from molecular genetics to ecological genomics. *Plant Physiol.* 158, 2–22.
- 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, 1687–1701.
- West, M.A.L., Kim, K., Kliebenstein, D.J., van Leeuwen, H., Micheltore, R.W., Doerge, R.W., and Clair, D.A.S. (2007). Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in *Arabidopsis*. *Genetics* 175, 1441–1450.



Lea Møller Jensen holds an MSc in Biology-Biotechnology from 2010 received her PhD degree from the Faculty of Science at the University of Copenhagen in 2014. Her research interests lies within understanding the regulatory networks controlling glucosinolate biosynthesis and how these link to other phenotypes crucial for plant survival. The research aims at understanding how natural variation impacts the link from genotype-to-phenotype. Elucidation of the underlying molecular mechanisms is a necessity for understanding variation and changes in these regulatory networks.



Meike Burow is renowned for contributions to our understanding of the biosynthesis and activation of glucosinolates. In particular, her studies on plant specifier proteins gave new insights in the evolution and ecological functions the glucosinolate-myrosinase system. Her current research spans protein biochemistry, protein-protein interaction studies, regulatory networks and RNA-mediated regulation to exploit the well-studied biosynthetic pathway to glucosinolates as model system for studying complex regulatory mechanisms and their genetic bases in plants.



Barbara Ann Halkier, Head of DynaMo Center of Excellence at the University of Copenhagen, is leading in the field within pathway elucidation, identification of biosynthetic genes, and transporters of glucosinolates. She is actively pursuing pathway engineering in various host organisms, and has produced glucosinolates in yeast on a robust platform based on stable genomic expression. She has identified the NPF family as transporter family for secondary metabolites, exemplified by the glucosinolate transporters AtGTR1 and AtGTR2. She has generated highly valuable tool in the form of transporter cDNA libraries to be expressed in *Xenopus* oocytes and screened for uptake of small molecules. She has advanced ligase-independent USER cloning technologies for high-throughput transfer of cDNAs in systems biology.