Amino Acids (2010) 39:1067–1075
DOI 10.1007/s00726-010-0681-5

ORIGINAL ARTICLE

Widely targeted metabolomics and coexpression analysis as tools to identify genes involved in the side-chain elongation steps of aliphatic glucosinolate biosynthesis

Doris Albinsky · Yuji Sawada · Ayuko Kuwahara · Mutsumi Nagano · Akiko Hirai · Kazuki Saito · Masami Yokota Hirai

Received: 21 May 2010 / Accepted: 25 June 2010 / Published online: 10 July 2010

Abstract Amino acid and glucosinolate biosynthesis are two interdependent pathways; amino acid synthesis as a part of primary metabolism provides the precursors for glucosinolate biosynthesis in secondary metabolism. In our previous studies, the combination of coexpression analysis and metabolite profiling led to the identification of genes and key regulators involved in glucosinolate biosynthesis. Moreover, the integration of transcriptome and metabolome data of sulphur-deprived *Arabidopsis* plants revealed coordinate changes in the expression profiles of genes involved in glucosinolate and amino acid metabolism.

This review provides an overview of our recent studies involving *Arabidopsis* mutant plants that exhibit impairment in the side-chain elongation process occurring during aliphatic glucosinolate biosynthesis by means of coexpression analysis and a novel metabolite profiling approach based on ultra-performance liquid chromatography coupled with tandem quadrupole mass spectrometry (UPLC-TQMS) (Sawada et al. 2009a). Thus, this review highlights the advantages of the omics-based approach in identifying genes involved in glucosinolate biosynthesis.

Keywords Glucosinolate biosynthesis · Widely targeted metabolomics · Batch-learning self-organising map (BL-SOM) · Side-chain elongation · Leucine biosynthesis

D. Albinsky, Y. Sawada contributed equally to this work.

D. Albinsky · Y. Sawada · A. Kuwahara · M. Nagano · A. Hirai · K. Saito · M. Y. Hirai
RIKEN Plant Science Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

D. Albinsky · Y. Sawada · A. Kuwahara · M. Nagano · M. Y. Hirai
JST, CREST, 4-1-8 Hon-chou, Kawaguchi, Saitama 332-0012, Japan

K. Saito
Graduate School of Pharmaceutical Sciences, Chiba University, 1-33 Yayoi-cho, Inage-ku, Chiba 263-8522, Japan

M. Y. Hirai (✉)
Graduate School of Bioagricultural Sciences, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan
e-mail: myhirai@psc.riken.jp

Introduction

Glucosinolates (GSLs) represent a class of secondary metabolites that predominantly occur in the *Brassicaceae* family. They are sulphur- and nitrogen-rich compounds (thioglucosides), and their breakdown products confer resistance to herbivores and pathogens or exhibit anticarcinogenic properties in humans.

GSL biosynthesis is closely associated with the synthesis of the aliphatic amino acids alanine, leucine, isoleucine, valine, and methionine, and the aromatic amino acids phenylalanine, tyrosine, and tryptophan, in primary metabolism. These amino acids form the GSL skeleton. Side-chain elongation, core GSL biosynthesis, and side-chain modifications lead to the final products, namely, aliphatic, indolic, and aromatic GSLs (Fig. 1) (for overview, see Halkier and Gershenzon 2006).

The biosynthesis of aliphatic and indolic GSLs is controlled by several R2R3-MYB transcription factors (Schneider et al. 2005; Hirai et al. 2007; Sønderby et al. 2007; Gigolashvili et al. 2007a, b, 2008; Beekwilder et al. 2008; Malitsky et al. 2008). *Myb28* and *Myb51* are the main positive regulators of the aliphatic and indolic

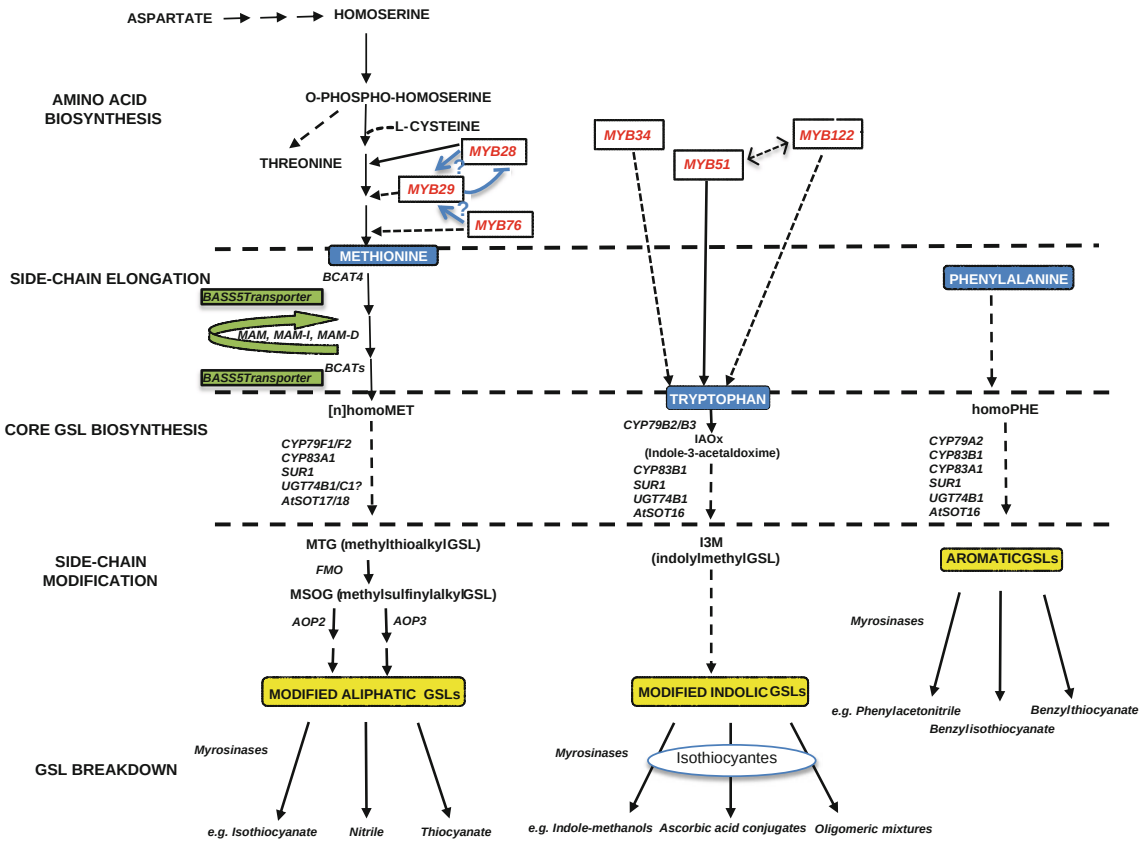

Fig. 1 Overview of the biosynthesis of aliphatic, indolic, and aromatic glucosinolates in *Arabidopsis thaliana* from the main amino acid precursors methionine, tryptophan, and phenylalanine, respectively. Sønderby et al. (2010) recently reported a positive regulation of *Myb29* by *Myb28* and *Myb76* via a putative interaction partner and a putative function of *Myb29* as repressor of *Myb28*, in *Arabidopsis* plants (Col-0 accession); these are indicated in *grey*

biosynthetic pathways, respectively. *Myb29* and *Myb76* control the biosynthesis of aliphatic GSLs, and *Myb34* and *Myb122* control that of indolic GSLs. These regulators respond to different environmental cues such as wounding, methyl jasmonates (MeJAs), and sulphur deficiency; therefore, the GSL species synthesised vary depending upon the specific environmental stimuli, developmental phase or tissue. Moreover, several recent studies (Hemm et al. 2003; Gigolashvili et al. 2008) have shown that the aliphatic and indolic GSL biosynthesis pathways are reciprocally regulated probably in order to balance the GSL and sulphur content of cells, as described in an overview of the regulation of indolic and aliphatic GSL biosynthesis (Gigolashvili et al. 2009a).

Leucine biosynthesis and aliphatic side-chain elongation

GSL biosynthesis depends on amino acid biosynthesis, and the regulation of these synthesis pathways is well balanced in order to adapt to varying developmental and environmental stimuli and to sustain a regular metabolic flow. The linkage between GSL and amino acid biosynthesis is evident from the association between leucine (Leu) biosynthesis and the side-chain elongation process in aliphatic GSL biosynthesis. The enzymatic steps involved in both processes are analogous, and the dysregulation of one pathway can affect the other pathway (Knill et al. 2008). The similarities between both pathways and the identification of genes involved in the bacterial Leu biosynthesis pathway were the premise for the identification and characterisation of genes involved in the side-chain elongation process of aliphatic GSL biosynthesis.

Methionine (Met)-derived side-chain elongation begins with the deamination of Met to 4-methylthio-2-oxobutyrate (MTOB) that is catalysed by the branched-chain aminotransferase 4 (BCAT4) in the cytosol (Schuster et al. 2006). The next three steps occur in the plastid; these steps are common to the Leu biosynthesis pathway and comprise in *Arabidopsis* up to six cycles of: (1) acetyl-CoA condensation, (2) isomerisation and (3) oxidative decarboxylation. These processes form different GSL end products and hence contribute to the diversity of GSLs. Step (1), i.e. the condensation of acetyl-CoA and 2-oxo-acid is catalysed by

the methylthioalkylmalate synthases MAM1 and MAM3 and results in aliphatic GSLs with different chain lengths (Kroymann et al. 2001, 2003; Textor et al. 2007).

The genes involved in the isomerisation (2) and oxidative decarboxylation (3) steps have been recently characterised, mainly on the basis of transcriptome coexpression analysis and different metabolic profiling approaches (Sawada et al. 2009b, c; He et al. 2009; Knill et al. 2009; Gigolashvili et al. 2009b). In the present review, we will focus on our work related to the elucidation of genes involved in the side-chain elongation of aliphatic GSLs by coexpression analysis and widely targeted metabolomics.

Transcriptome coexpression analysis

Coexpression analysis is based on the assumption that genes, which are simultaneously expressed, are probably involved in the same metabolic pathway. Thus, it enables the identification of unknown genes and their functions (Wei et al. 2006; for overview, see Obayashi and Kinoshita 2010; Aoki et al. 2007; Usadel et al. 2009; Hirai 2009). The number of plant gene characterisation studies, based on coexpression analysis, has increased in the last 5 years, thereby highlighting the usefulness of this approach (Gachon et al. 2005; Tohge et al. 2005; Vanderauwera et al. 2005; Ehlting et al. 2006, 2008; Hirai et al. 2007; Yonekura-Sakakibara et al. 2007; Saito et al. 2008).

In our previous coexpression studies, we investigated the temporal changes occurring in the transcriptome and metabolome of sulphur-starved *Arabidopsis* plants (Hirai et al. 2005). Data analysis using batch-learning self-organising map (BL-SOM) analysis—a type of cluster analysis—revealed clustering (coexpression) of many genes and/or metabolites involved in the same pathway, e.g. GSL biosynthesis. Homologs of bacterial Leu biosynthesis genes, i.e. *AtLeuC1* (At4g13430), *AtLeuD1* (At2g43100), *AtLeuD2* (At3g58990) and *AtIMD1* (At5g14200), were found to be coexpressed with genes known to be involved in GSL biosynthesis, e.g. *MAM*s, *CYP79*s, *CYP83*s and *SUR1*; therefore, we focussed on these homologs in our further studies. The branched-chain aminotransferases 3 and 4 (*BCAT3* and *BCAT4*) were also coexpressed with GSL biosynthesis genes and later confirmed to be involved in GSL biosynthesis (Knill et al. 2008; Schuster et al. 2006); cytosolic BCAT4 deaminates Met and [n]homoMet to their corresponding 2-oxo acids, and BCAT3 in the plastid catalyses the final transamination of short-chain keto-acids to homomethionine and dihomomethionine, respectively. Transcriptome coexpression analysis for the genes *AtLeuC1*, *AtLeuD1*, *AtLeuD2* and *AtIMD1* has been performed using the publicly available databases *Arabidopsis thaliana trans*-factor and *cis*-element prediction database (ATTED-II; http://atted.jp/) (Obayashi et al. 2007) and the *Arabidopsis* coexpression data-mining tool (ACT; http://www.arabidopsis.leeds.ac.uk/act/coexpanalyser.php) (Manfield et al. 2006). All the four genes were strongly coexpressed with known Met-derived GSL biosynthesis genes (Table 1 for *AtLeuC1* and *AtIMD1*). Moreover, our analysis has revealed that GSL biosynthesis genes and *BASS5*, a bile acid: sodium symporter like family protein, are coexpressed (Table 1).

On the basis of data from coexpression analysis and from a study investigating the relationship between *Myb28* and downstream biosynthetic genes (Hirai et al. 2007), we assumed that the above-mentioned genes that are homologs of the Leu biosynthesis genes are probably involved in the biosynthesis of aliphatic GSLs, especially in the steps of the side-chain elongation pathway. Based on homology of the nucleotide sequences, *AtLeuC1*, *AtLeuD*s and *AtIMD1* were thought to encode the methylthioalkylmalate isomerase (MAM-I) large subunit, the MAM-I small subunits, and the methylthioalkylmalate dehydrogenase (MAM-D), respectively. Because *BASS5* was also regulated by Myb28 and its protein was located in the chloroplast (Sawada et al. 2009b; Gigolashvili et al. 2009b), BASS5 was thought to function as a transporter of intermediates of the Met-derived GSLs across the chloroplast membrane.

Widely targeted metabolomics and its implication on the identification of genes involved in aliphatic side-chain elongation

We analysed the leaves of *AtLeuC1* (At4g13430) and *AtIMD1* (At5g14200) knockout *Arabidopsis* plants by widely targeted metabolomics and identified changes in the synthesis profiles of Met-derived GSLs and related metabolites (Table 2) (Sawada et al. 2009c), which indicates that both genes are definitely involved in Met-derived GSL biosynthesis. Therefore, *AtLeuC1* and *AtIMD1* were termed *MAM-IL1* (MAM-I large subunit1) and *MAM-D1*, respectively. Knill et al. (2009) also supported the conclusion.

Arabidopsis AtLeuD1 knockouts did not exhibit remarkable changes in the GSL content, suggesting that the function of *AtLeuD1* and *AtLeuD2* may be redundant because both genes encode the small subunits of MAM-I in Leu biosynthesis (Sawada et al. 2009c). Indeed, indications that *AtLeuD1* and *AtLeuD2* perform redundant functions in the isomerisation step (2) during chain elongation also came from studies undertaken by Knill et al. (2009). The authors analysed the metabolite profiles of leaves and seeds of the two single knockout plants for the two small subunits of the MAM isomerases (*MAM-I1/2* and *AtLeuD1/D2*). In accordance with our data, they did not observe remarkable changes in the metabolite profiles; only in seeds of the *AtLeuD1* knockout line, the content of short-chain GSLs

Table 1 Pearson's correlation coefficient (PCC) for candidate chain elongation genes and genes involved in glucosinolate (GSL) biosynthesis

AGI	GSL biosynthesis genes	*AtLeuC1 (MAM-IL1)*	*AtIMD1 (MAM-D1)*[a]	*BASS5*	References
At5g61420	*MYB28*	0.62	0.76	0.77	Hirai et al. 2007; Gigolashvili et al. 2007b; Sønderby et al. 2007
At5g07690	*MYB29*	0.42	0.58	0.68	Hirai et al. 2007; Gigolashvili et al. 2008; Sønderby et al. 2007
At3g49680	*Branched-chain aminotransferase 3 (BCAT3)*	0.72	0.78	0.54	Knill et al. 2008
At3g19710	*BCAT4*	0.66	0.78	0.88	Schuster et al. 2006
At5g23010	*Methylthioalkylmalate synthase 1 (MAM1)*	0.61	0.69	0.84	Kroymann et al. 2001
At5g23020	*2-Isopropylmalate synthase 2 (MAM3; IMS2)*	0.49	0.58	0.51	Textor et al. 2007
At1g16410[b]	*CYP79F1*	0.68	0.80	0.85	Reintanz et al. 2001; Chen et al. 2003
At1g16400[b]	*CYP79F2*	0.68	0.80	0.85	Chen et al. 2003
At2g22330	*CYP79B3*	n.d.	n.d.	0.46	Hull et al. 2000
At4g13770	*CYP83A1*	0.71	0.78	0.88	Bak and Feyereisen 2001; Naur et al. 2003
At4g31500	*CYP83B1*	n.d.	n.d.	0.36	Bak et al. 2001; Barlier et al. 2000; Hansen et al. 2001; Smolen and Bender 2002
At2g20610	*Superroot 1 (SUR1)*	0.63	0.78	0.63	Mikkelsen et al. 2004
At1g24100	*UDP-glucosyl transferase 74B1 (UGT74B1)*	0.53	0.56	0.53	Grubb et al. 2004; Gachon et al. 2005
At2g31790	*UGT74C1*	0.71	0.78	0.74	Gachon et al. 2005
At1g74100	*Sulphotransferase 16 (SOT16)*	n.d.	n.d.	0.45	Piotrowski et al. 2004; Hirai et al. 2005
At1g18590	*SOT17*	0.61	0.71	0.72	Piotrowski et al. 2004; Hirai et al. 2005
At1g74090	*SOT18*	0.71	0.82	0.84	Piotrowski et al. 2004; Hirai et al. 2005
At1g65860	*Flavin-monooxygenase glucosinolate S-oxygenase 1 ($FMO_{GS\text{-}OX1}$)*	0.45	0.74	0.73	Hansen et al. 2007
At1g62560	$FMO_{GS\text{-}OX3}$	0.56	0.65	0.72	Hansen et al. 2007; Li et al. 2008
At4g03060	*Alkenyl hydroxalkyl producing 2 (AOP2)*	0.38	n.d.	0.59	Kliebenstein et al. 2001

[a] Determined from the *Arabidopsis* Coexpression Data-mining Tool (ACT) database

[b] Affymetrix probe sets hybridise to transcripts of both homologs of cytochrome P450 79F1/79F2

n.d. not determined

was slightly enhanced. Additionally, liquid chromatography mass spectrometry (LC–MS) analysis revealed that *MAM-IL1* plays a role in both Leu and GSL biosynthesis because the intermediates of Leu biosynthesis and Met chain elongation were accumulated in the respective *Arabidopsis mam-il1* lines (Knill et al. 2009).

Widely targeted metabolomics of *Arabidopsis* knockout plants for the plastidic bile acid: sodium symporter-like family protein *BASS5* confirmed a predicted function of this protein in the transport of MTOB from the cytosol to the plastid for side-chain elongation and/or in the transport of 2-keto-acids with elongated chains ([n]homoMet), from the plastid to the cytosol for conversion into GSLs (Sawada et al. 2009b) (Fig. 2). Gigolashvili et al. (2009b) reported similar conclusions by using a different research strategy.

We observed two trends by comparing the metabolite profiles of leaves of *MAM-IL1, MAM-D1* and *BASS5* knockout plants with those of leaves of *MAM1*, *MAM3* knockout plants, which are genes known to be involved in side-chain elongation (Table 2).

First, short- and long-chain elongation in GSL biosynthesis appeared to be distinct processes and dependent on particular gene functions. The *mam-il1*, *mam-d1* and *bass5-2* plants showed significantly reduced long-chained GSLs, whereas *mam1* and *mam3* plants did not (Table 2). However, GSLs with C3 chains tended to increase in all the knockout lines except *bass5-2* (Table 2). BCAT3 catalyses the final transamination of only short-chain keto-acids to homomethionine and dihomomethionine (Knill et al. 2008). However, decrease in the content of

Table 2 Fold changes of selected Met-derived metabolites and GSLs in *mam-il1, mam-d1, bass5, mam1*, and *mam3 Arabidopsis* mutants (Sawada et al. 2009b, c)

Compound name	*mam-il1* (*atleuc1-1*)[a]		*mam-d1* (*atimd1-1*)[a]		*bass5-2*[b]		*mam1*[b]		*mam3*[b]	
	Fold change	*p* value	Fold change	*p* value	Fold change	*p* value	Fold change	*p* value	Fold change	*p* value
Methionine sulphoxide	**5.3**	0.00	**3.0**	0.00	**9.9**	0.00	6.3	0.08	**5.3**	0.04
Methionine	**2.6**	0.00	1.6	0.08	**7.5**	0.00	2.1	0.08	2.0	0.05
S-methylmethionine	**2.7**	0.00	2.3	0.06	**3.4**	0.00	1.7	0.24	1.5	0.22
S-adenosyllmethionine	**2.1**	0.04	1.5	0.33	**3.3**	0.00	0.9	0.52	1.1	0.65
5′-deoxy-5′-methylthioadenosine	**2.2**	0.00	**1.4**	0.03	**2.9**	0.00	**2.0**	0.04	2.2	0.08
MTc3	**4.2**	0.01	**3.5**	0.01	0.2	0.08	3.4	0.14	3.2	0.12
MSc3	**7.8**	0.00	**5.0**	0.01	**0.3**	0.00	1.8	0.19	1.7	0.16
OHc3	3.0	0.06	2.1	0.07	1.4	0.16	3.2	0.26	4.7	0.15
BOc3	**3.2**	0.01	**2.6**	0.01	0.8	0.63	1.0	0.88	0.9	0.78
MTc4	**0.2**	0.00	**0.5**	0.00	0.5	0.05	**0.3**	0.00	**0.4**	0.03
MSc4	**0.5**	0.01	**0.7**	0.04	**0.5**	0.02	**0.3**	0.00	**0.4**	0.04
OHc4	1.8	0.14	0.8	0.64	1.5	0.36	**0.3**	0.03	**0.4**	0.05
BOc4	**0.2**	0.01	**0.3**	0.01	**2.0**	0.03	1.5	0.38	1.5	0.22
MTc5	**0.2**	0.00	**0.3**	0.00	**0.5**	0.03	**0.4**	0.01	0.5	0.05
MSc5	**0.4**	0.00	**0.4**	0.00	**0.6**	0.02	**0.5**	0.03	0.7	0.35
MTc6	1.2	0.59	**0.4**	0.01	0.8	0.57	1.0	0.82	0.8	0.44
MSc6	**0.1**	0.00	0.1	–	**0.4**	0.00	0.7	0.37	0.6	0.08
MTc7	**0.2**	0.00	0.0	–	**0.2**	0.00	0.6	0.06	**0.6**	0.02
MSc7	**4.7**	0.01	2.4	–	2.1	0.51	3.7	–	8.6	0.21
MTc8	1.1	0.72	0.6	–	**0.3**	0.01	**2.5**	0.05	1.7	0.06
MSc8	**0.02**	0.01	**0.001**	0.01	**0.3**	0.00	1.4	0.55	0.8	0.64

MT methylthioalkyl GSL, *MS* methylsulfinylalkyl GSL, *OH* hydroxyalkyl GSL, *BO* benzoyloxyalkyl GSL, *c[n]* indicates the number of carbon chains

Fold changes with p values <0.05 are represented in bold letters

– p values could not be determined because of the small sample number

[a] 4 weeks after germination

[b] 3 weeks after germination

aliphatic GSLs of all chain-length classes (i.e. C3–C8) was observed in *bass5–2* (Table 2) and *bcat4* plants (Sawada et al. 2009c). Taken together, these data suggest that short- and long-chain formation processes are biosynthetically separated in terms of condensation, isomerisation, oxidative decarboxylation and transamination. This hypothesis implies that several genes involved in the side-chain elongation process of GSL biosynthesis, which are not regulated by the known transcription factor *Myb28*, are hitherto unidentified and that they may respond to specific environmental stimuli. Different GSLs with varying side-chain lengths were found to be synthesised in response to developmental and environmental changes (Brown et al. 2003; Mewis et al. 2006). Another explanation for the partial reduction in the GSL content of the knockout plants analysed may be a partial redundancy in the function of genes from other pathways. In fact, several members of the gene family encoding genes involved in Leu biosynthesis were shown to have similar or even dual functions with genes involved in the side-chain elongation of aliphatic GSLs in *Arabidopsis* (Knill et al. 2008; He et al. 2009; Knill et al. 2009) (Fig. 2).

Further, the knockout plants of all genes involved in the side-chain elongation of aliphatic GSLs, i.e. *MAM-IL1*, *MAM-D1*, *MAM1*, *MAM3* and *BASS5*, exhibited partially significant induction in the synthesis of Met-derived metabolites (Table 2). This finding indicates that the biosynthetic precursors of GSL are channelled towards Met biosynthesis and the synthesis of related compounds when the GSL biosynthetic pathway is stalled or at least partially blocked.

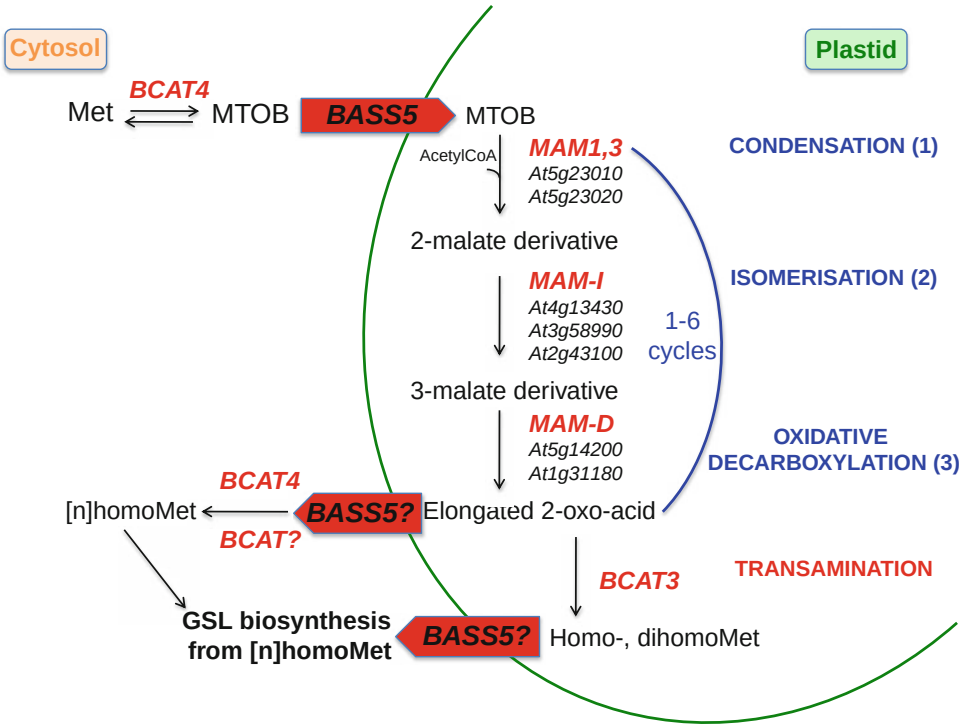

Fig. 2 Overview of side-chain elongation during the biosynthesis of aliphatic GSLs showing the steps in which *BASS5* is probably involved

The impact of the omics-based approach on the identification of genes involved in GSL biosynthesis

The combination of coexpression analysis and metabolite profiling has been invaluable in the identification of genes involved in regulatory and biosynthetic pathways (Tohge et al. 2005; Hirai et al. 2007; Cho et al. 2008). The R2R3-MYB transcription factors involved in aliphatic GSL biosynthesis (Hirai et al. 2007) and genes involved in the three steps of Met-derived GSL biosynthesis could be identified using the above-mentioned method (Hirai et al. 2005; Knill et al. 2008; He et al. 2009). With respect to the number of GSL biosynthesis genes successfully decoded, this approach is equivalent to the method for identifying genes on the basis of the metabolic changes occurring in recombinant inbred lines (RILs) by quantitative trait loci (QTL)-mapping (Textor et al. 2004; Kliebenstein et al. 2001; Kroymann et al. 2001, 2003; Hansen et al. 2008; Pfalz et al. 2009). However, both approaches have their specific advantages and limitations that are contradictory. Therefore, the choice of method depends on the goal of the investigation.

The systems biology approach using QTL mapping can identify genes on the basis of their effects on natural metabolic variation, although not all GSL biosynthetic gene products may cause metabolic variations in different RILs. Moreover, a combination of coexpression and QTL analysis may be required in order to reduce the number of putative candidate genes suggested to be involved in GSL biosynthesis.

The main advantage of coexpression analysis is that it is rapid, can be conducted *in silico* without wet experiments and provides abundant information on gene relations. However, a 'targeted approach' requires the use of genes already characterised to be involved in the same process as 'bait'. Moreover, in cases where gene regulation is dependent on environmental or developmental factors, coexpression analysis by using publicly available transcriptome databases may not provide conclusive results. Because these genes are only expressed under specific conditions, data sets may not adequately represent these particular conditions.

Noteworthy, different GSL classes can accumulate in various organs or cell types (Brown et al. 2003; Shroff et al. 2008; Matsuda et al. 2009) in response to developmental or environmental cues (Brown et al. 2003; Mikkelsen et al. 2003; Matsuda et al. 2009). Considering this tissue-specificity of GSLs (Wentzell and Kliebenstein 2008), another very promising way to identify novel genes involved in GSL biosynthesis could be the performance of coexpression analysis with transcriptomic data derived from specific tissues or even individual cell types.

With respect to the aliphatic side-chain elongation, *mam-il1*, *mam-d1*, *mam1* and *mam3* mutants revealed only partial reduction in their GSL content, thereby suggesting hitherto unidentified missing links in this process of GSL biosynthesis, which are most probably under different regulation from *Myb28*.

A recent study examined the specific effects of the transcription factors *Myb28*, *Myb29* and *Myb76* on aliphatic

GSL biosynthesis, GSL distribution and transcription of genes involved in GSL biosynthesis (Sønderby et al. 2010). The authors unveiled a complex interactive network of the three transcription factors involved in GSL biosynthesis and suggested that additional regulators are involved in GSL biosynthesis. Especially, the synthesis of long-chain GSLs cannot be exclusively linked to the expression of *Myb28* because *Myb29* or *Myb76* alone could induce long-chain GSL synthesis in the absence of the other two transcription factors involved in aliphatic GSL biosynthesis.

Altogether, the dissection of short- and long-chain synthesis in GSL formation remains the challenge in future studies. The omics-based approach is expected to be indispensable once the mode of regulation of side-chain synthesis has been identified.

Acknowledgments This work was supported in part by the Core Research for Evolutional Science and Technology (CREST) of the Japan Science and Technology Agency (Project name: 'Elucidation of Amino Acid Metabolism in Plants Based on Integrated Omics Analyses').

References

Aoki K, Ogata Y, Shibata D (2007) Approaches for extracting practical information from gene co-expression networks in plant biology. Plant Cell Physiol 48:381–390

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

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

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 MG, Bovy A (2008) The impact of the absence of aliphatic glucosinolates on insect herbivory in *Arabidopsis*. PLoS One 3:e2068

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

Chen S, Glawischnig 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

Cho K, Shibato J, Agrawal GK, Jung YH, Kubo A, Jwa NS, Tamogami S, Satoh K, Kikuchi S, Higashi T, Kimura S, Saji H, Tanaka Y, Iwahashi H, Masuo Y, Rakwal R (2008) Integrated transcriptomics, proteomics, and metabolomics analyses to survey ozone responses in the leaves of rice seedling. J Proteome Res 7:2980–2998

Ehlting J, Provart NJ, Werck-Reichhart D (2006) Functional annotation of the *Arabidopsis* P450 superfamily based on large-scale co-expression analysis. Biochem Soc Trans 34:1192–1198

Ehlting J, Sauveplane V, Olry A, Ginglinger JF, Provart NJ, Werck-Reichhart D (2008) An extensive (co-)expression analysis tool for the cytochrome P450 superfamily in *Arabidopsis thaliana*. BMC Plant Biol 8:47

Gachon CM, Langlois-Meurinne M, Henry Y, Saindrenan P (2005) Transcriptional co-regulation of secondary metabolism enzymes in *Arabidopsis*: functional and evolutionary implications. Plant Mol Biol 58:229–245

Gigolashvili T, Berger B, Mock HP, Müller C, Weisshaar B, Flügge UI (2007a) The transcription factor HIG1/MYB51 regulates indolic glucosinolate biosynthesis in *Arabidopsis thaliana*. Plant J 50:886–901

Gigolashvili T, Yatusevich R, Berger B, Müller C, Flügge UI (2007b) 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, Yatusevich R, Müller C, Flügge UI (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, Berger B, Flügge UI (2009a) Specific and coordinated control of indolic and aliphatic glucosinolate biosynthesis by R2R3-MYB transcription factors in *Arabidopsis thaliana*. Phytochem Rev 8:3–13

Gigolashvili T, Yatusevich R, Rollwitz I, Humphry M, Gershenzon J, Flügge UI (2009b) The plastidic bile acid transporter 5 is required for the biosynthesis of methionine-derived glucosinolates in *Arabidopsis thaliana*. Plant Cell 21:1813–1829

Grubb CD, Zipp BJ, Ludwig-Müller 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. Annu Rev Plant Biol 57:303–333

Hansen CH, Du L, Naur P, Olsen CE, Axelsen KB, Hick AJ, Pickett JA, Halkier BA (2001) CYP83b1 is the oxime-metabolizing enzyme in the glucosinolate pathway in *Arabidopsis*. 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

Hansen BG, Kerwin RE, Ober JA, Lambrix VM, Mitchell-Olds T, Gershenzon J, Halkier BA, Kliebenstein DJ (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

He Y, Mawhinney TP, Preuss ML, Schroeder AC, Chen B, Abraham L, Jez JM, Chen S (2009) A redox-active isopropylmalate dehydrogenase functions in the biosynthesis of glucosinolates and leucine in *Arabidopsis*. Plant J 60:679–690

Hemm MR, Ruegger MO, 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 MY (2009) A robust omics-based approach for the identification of glucosinolate biosynthetic genes. Phytochem Rev 8:15–23

Hirai M, Klein M, Fujikawa Y, Yano M, Goodenowe D, 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

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

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

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

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

Knill T, Reichelt M, Paetz C, Gershenzon J, 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

Kroymann J, Textor S, Tokuhisa JG, Falk KL, Bartram S, Gershenzon J, 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, Mitchell-Olds T (2003) Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait locus. Proc Natl Acad Sci USA 100: 14587–14592

Li J, Hansen BG, Ober JA, Kliebenstein DJ, Halkier BA (2008) Subclade of flavin-monooxygenases involved in aliphatic glucosinolate biosynthesis. Plant Physiol 148:1721–1733

Malitsky S, Blum E, Less H, Venger I, Elbaz M, Morin S, Eshed Y, Aharoni A (2008) The transcript and metabolite networks affected by the two clades of *Arabidopsis* glucosinolate biosynthesis regulators. Plant Physiol 148:2021–2049

Manfield IW, Jen CH, Pinney JW, Michalopoulos I, Bradford JR, Gilmartin PM, Westhead DR (2006) *Arabidopsis* Co-expression Tool (ACT): web server tools for microarray-based gene expression analysis. Nucleic Acids Res 34:W504–W509

Matsuda F, Hirai MY, Sasaki E, Akiyama K, Yonekura-Sakakibara K, Provart NJ, Sakurai T, Shimada Y, Saito K (2009) AtMetExpress development: a phytochemical atlas of *Arabidopsis thaliana* development. Plant Physiol 152:566–578

Mewis I, Tokuhisa JG, Schultz JC, Appel HM, Ulrichs C, 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 MD, Petersen BL, Glawischnig E, Jensen AB, Andreasson E, Halkier BA (2003) Modulation of CYP79 genes and glucosinolate profiles in *Arabidopsis* by defense signaling pathways. Plant Physiol 131:298–308

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

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

Obayashi T, Kinoshita K (2010) Coexpression landscape in ATTED-II: usage of gene list and gene network for various types of pathways. J Plant Res 123:311–319

Obayashi T, Kinoshita K, Nakai K, Shibaoka M, Hayashi S, Saeki M, Shibata D, Saito K, Ohta H (2007) ATTED-II: a database of co-expressed genes and cis elements for identifying co-regulated gene groups in *Arabidopsis*. Nucleic Acids Res 35:D863–D869

Pfalz M, Vogel H, Kroymann J (2009) The gene controlling the indole glucosinolate modifier1 quantitative trait locus alters indole glucosinolate structures and aphid resistance in *Arabidopsis*. Plant Cell 21:985–999

Piotrowski M, Schemenewitz A, Lopukhina A, Muller 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

Reintanz B, Lehnen M, Reichelt M, Gershenzon J, Kowalczyk M, Sandberg G, Godde M, Uhl R, Palme K (2001) *Bus*, a bushy *Arabidopsis CYP79F1* knockout mutant with abolished synthesis of short-chain aliphatic glucosinolates. Plant Cell 13:351–367

Saito K, Hirai MY, Yonekura-Sakakibara K (2008) Decoding genes with coexpression networks and metabolomics—'majority report by precogs'. Trends Plant Sci 13:36–43

Sawada Y, Akiyama K, Sakata A, Kuwahara A, Otsuki H, Sakurai T, Saito K, Hirai MY (2009a) Widely targeted metabolomics based on large-scale MS/MS data for elucidating metabolite accumulation patterns in plants. Plant Cell Physiol 50:37–47

Sawada Y, Kuwahara A, Nagano M, Narisawa T, Sakata A, Saito K, Hirai MY (2009b) 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, Hirai MY (2009c) *Arabidopsis* bile acid:sodium symporter family protein 5 is involved in methionine-derived glucosinolate biosynthesis. Plant Cell Physiol 50:1579–1586

Schneider A, Kirch T, Gigolashvili T, Mock HP, Sonnewald U, Simon R, Flügge UI, Werr W (2005) A transposon-based activation-tagging population in *Arabidopsis thaliana* (TAMARA) and its application in the identification of dominant developmental and metabolic mutations. FEBS Lett 579:4622–4628

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, Svatos 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

Smolen G, Bender J (2002) *Arabidopsis* cytochrome P450 cyp83B1 mutations activate the tryptophan biosynthetic pathway. Genetics 160:323–332

Sønderby IE, Hansen BG, Bjarnholt N, Ticconi C, Halkier BA, Kliebenstein DJ (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

Sønderby IE, Burow M, Rowe HC, Kliebenstein DJ, Halkier BA (2010) A complex interplay of three R2R3 MYB transcription factors determines the profile of aliphatic glucosinolates in *Arabidopsis*. Plant Physiol 153:348–363

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 JG (2007) MAM3 catalyzes the formation of all aliphatic glucosinolate chain lengths in *Arabidopsis*. Plant Physiol 144:60–71

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

Usadel B, Obayashi T, Mutwil M, Giorgi FM, Bassel GW, Tanimoto M, Chow A, Steinhauser D, Persson S, Provart NJ (2009) Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. Plant Cell Environ 32:1633–1651

Vanderauwera S, Zimmermann P, Rombauts S, Vandenabeele S, Langebartels C, Gruissem W, Inze D, Van Breusegem F (2005) Genome-wide analysis of hydrogen peroxide-regulated gene expression in *Arabidopsis* reveals a high light-induced transcriptional cluster involved in anthocyanin biosynthesis. Plant Physiol 139:806–821

Wei H, Persson S, Mehta T, Srinivasasainagendra V, Chen L, Page GP, Somerville C, Loraine A (2006) Transcriptional coordination of the metabolic network in *Arabidopsis*. Plant Physiol 142:762–774

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

Yonekura-Sakakibara K, Tohge T, Niida R, Saito K (2007) Identification of a flavonol 7-O-rhamnosyltransferase gene determining flavonoid pattern in *Arabidopsis* by transcriptome coexpression analysis and reverse genetics. J Biol Chem 282:14932–14941