the plant journal

SEB Society for Experimental Biology

The Plant Journal (2017) **89,** 204–220 doi: 10.1111/tpj.13377

PYK10 myrosinase reveals a functional coordination between endoplasmic reticulum bodies and glucosinolates in *Arabidopsis thaliana*

Ryohei T. Nakano[1,2,3], Mariola Piślewska-Bednarek[4], Kenji Yamada[5,†], Patrick P. Edger[6,‡], Mado Miyahara[3,§], Maki Kondo[5], Christoph Böttcher[7,¶], Masashi Mori[8], Mikio Nishimura[5], Paul Schulze-Lefert[1,2,*], Ikuko Hara-Nishimura[3,*,#,‖] and Paweł Bednarek[4,*,#]

[1]*Department of Plant Microbe Interactions, Max Planck Institute for Plant Breeding Research, Carl-von-Linné-Weg 10, D-50829 Köln, Germany,*

[2]*Cluster of Excellence on Plant Sciences (CEPLAS), Max Planck Institute for Plant Breeding Research, Carl-von-Linné-Weg 10, D-50829 Köln, Germany,*

[3]*Department of Botany, Graduate School of Science, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan,*

[4]*Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznań, Poland,*

[5]*Department of Cell Biology, National Institute of Basic Biology, Okazaki 444-8585, Japan,*

[6]*Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA,*

[7]*Department of Stress and Developmental Biology, Leibniz Institute of Plant Biochemistry, D-06120 Halle (Saale), Germany, and*

[8]*Ishikawa Prefectural University, Nonoichi, Ishikawa 834-1213, Japan*

Received 29 March 2016; revised 30 August 2016; accepted 5 September 2016; published online 19 December 2016.

*For correspondence (e-mails bednarek@ibch.poznan.pl; ihnishi@gr.bot.kyoto-u.ac.jp; schlef@mpipz.mpg.de).

#These authors contributed equally to this work.

†Present address: Malopolska Centre of Biotechnology, Jagiellonian University, 30-387 Krakow, Poland.

‡Present address: Department of Horticulture, Michigan State University, East Lansing, MI, USA.

§Present address: Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo 113-0033, Japan.

¶Present address: Julius Kühn-Institute, Federal Research Centre for Cultivated Plants, Institute for Ecological Chemistry, Plant Analysis and Stored Product Protection, Königin-Luise-Strasse 19, 14195 Berlin, Germany.

‖Present address: Faculty of Science and Engineering, Konan University, Kobe 658-8501, Japan.

SUMMARY

The endoplasmic reticulum body (ER body) is an organelle derived from the ER that occurs in only three families of the order Brassicales and is suggested to be involved in plant defense. ER bodies in *Arabidopsis thaliana* contain large amounts of β-glucosidases, but the physiological functions of ER bodies and these enzymes remain largely unclear. Here we show that PYK10, the most abundant β-glucosidase in *A. thaliana* root ER bodies, hydrolyzes indole glucosinolates (IGs) in addition to the previously reported *in vitro* substrate scopolin. We found a striking co-expression between ER body-related genes (including *PYK10*), glucosinolate biosynthetic genes and the genes for so-called specifier proteins affecting the terminal products of myrosinase-mediated glucosinolate metabolism, indicating that these systems have been integrated into a common transcriptional network. Consistent with this, comparative metabolite profiling utilizing a number of *A. thaliana* relatives within Brassicaceae identified a clear phylogenetic co-occurrence between ER bodies and IGs, but not between ER bodies and scopolin. Collectively, our findings suggest a functional link between ER bodies and glucosinolate metabolism *in planta*. In addition, *in silico* three-dimensional modeling, combined with phylogenomic analysis, suggests that PYK10 represents a clade of 16 myrosinases that arose independently from the other well-documented class of six thioglucoside glucohydrolases. These findings provide deeper insights into how glucosinolates are metabolized in cruciferous plants and reveal variation of the myrosinase–glucosinolate system within individual plants.

Keywords: ER body, glucosinolate, myrosinase, *Arabidopsis thaliana*, Brassicaceae, co-expression, molecular phylogeny.

INTRODUCTION

Plants have evolved sophisticated mechanisms to respond to adverse biotic and abiotic stresses in a given ecological niche. One major strategy is dependent on the biosynthesis of low-molecular-weight compounds known as secondary metabolites, characterized by a huge structural diversity and high phylogenetic specificity (Bednarek and Osbourn, 2009; Weng *et al.*, 2012). Sugar conjugation is a way to control the bioactivities of these compounds, in which enzymatic cleavage of the sugar moiety can trigger immediate responses to external stimuli. The enzymes required for this cleavage are often physically separated from their substrates at a cellular or subcellular level. It has been proposed that the endoplasmic reticulum body (ER body) provides such compartmentalization for the PYK10/BGLU23 β-glucosidase (Hara-Nishimura and Matsushima, 2003; Yamada *et al.*, 2011; Nakano *et al.*, 2014).

The ER body is a rod-shaped and membrane-bound organelle that is derived from, and continuous with, the ER (Figure S1a), and has been observed only in a monophyletic taxonomic lineage within the order Brassicales, namely the families Brassicaceae, Caparaceae and Cleomaceae (Figure S1b) (Bonnett and Newcomb, 1965; Iversen and Flood, 1969; Iversen, 1970b; Cresti *et al.*, 1974; Hoefert, 1975; Endress and Sjolund, 1976; Jørgensen *et al.*, 1977; Behnke and Eschlbeck, 1978; Gailhofer *et al.*, 1979; Jørgensen, 1981). ER bodies that are present in *Arabidopsis thaliana* roots accumulate large amounts of PYK10/BGLU23 β-glucosidase together with its closest homologs BGLU21 and BGLU22 (Matsushima *et al.*, 2003; Nagano *et al.*, 2008). ER bodies in leaves, whose formation is stimulated by wounding or methyl jasmonate (MeJA) treatment, contain another β-glucosidase designated as BGLU18 (Matsushima *et al.*, 2002; Ogasawara *et al.*, 2009).

In line with the typical role of plant β-glucosidases in plant immunity (Morant *et al.*, 2008), a role of PYK10 and ER bodies in plant defense has been proposed (Yamada *et al.*, 2011; Nakano *et al.*, 2014). Consistent with this, *PYK10* expression is increased in roots parasitized by the nematode *Heterodera schachtii* (Nitz *et al.*, 2001) and in leaves during induced systemic resistance triggered upon root colonization by the growth-promoting rhizobacterium *Pseudomonas fluorescens* (Pozo *et al.*, 2008). In addition, genetic depletion of either *PYK10* or the whole ER body system by knocking out the *NAI1* transcription factor results in slightly increased colonization with the mutualistic root-endophytic fungus *Piriformospora indica*, which caused a loss of growth-promotion activity by the fungus (Sherameti *et al.*, 2008).

The narrow taxonomic distribution of the ER body to few lineages within the order Brassicales raises the possibility that the substrate(s) of PYK10 is also restricted to these evolutionary clades. Based on this idea, early studies suggested that ER bodies (formerly referred to as 'dilated cisternae') could serve as a repository for β-thioglucoside glucosyl hydrolases (referred to as myrosinases), which are involved in the metabolism of Brassicales-specific β-*S*-glucosides, known as glucosinolates (Figures 1a and S1) (Iversen, 1970a, b; Behnke and Eschlbeck, 1978; Jørgensen, 1981; Mithen *et al.*, 2010). Myrosinase-mediated hydrolysis of glucosinolates leads to various end products, some of which are important components of plant defense against insects and microbial pathogens (Barth and Jander, 2006; Bednarek *et al.*, 2009; Clay *et al.*, 2009). The processes that determine the end products following glucosinolate hydrolysis are controlled by the functions of epithiospecifier protein, epithiospecifier modifier protein and nitrile specifier proteins, which in turn determine the target of the whole metabolic/catabolic cascade (Wittstock and Burow, 2010; Iven *et al.*, 2012). Sequence-based prediction did not assign ER body-accumulated β-glucosidases as myrosinases (Burmeister *et al.*, 1997; Rask *et al.*, 2000). This was further supported by a study showing undetectable activity of PYK10, BGLU21 and BGLU22 towards sinigrin, a methionine-derived aliphatic glucosinolate (AG) (Ahn *et al.*, 2010), and a study showing that the bulk of root-associated myrosinase activity toward sinigrin is dependent on the root-specific myrosinases THIOGLUCOSIDE GLUCOHYDROLASE 4 (TGG4) and TGG5 (Fu *et al.*, 2016). Instead, these ER-body β-glucosidases hydrolyzed a series of coumarinyl *O*-glucosides (scopolin, esculin and 4-methylumbelliferyl-β-D-glucopyranoside; Figure 1a), with highest activity towards scopolin (Ahn *et al.*, 2010). However, there has so far been no functional evidence connecting ER bodies and scopolin metabolism *in planta*. Of note, studies on the PENETRATION2 (PEN2) β-glucosidase, which belongs to a sister clade of PYK10, revealed that TGG-specific sequence signatures (Burmeister *et al.*, 1997) are not necessarily required for myrosinase activity toward tryptophan-derived indole glucosinolates (IGs) (Bednarek *et al.*, 2009; Clay *et al.*, 2009). This led us to hypothesize that PYK10 is also able to hydrolyze IGs (Nakano *et al.*, 2014), but experimental validation for this is lacking.

Here we show that PYK10 indeed has *in vitro* myrosinase activity toward IGs and that the co-expressed gene cluster of *PYK10* is enriched in genes required for the production of glucosinolates. We thereby propose that glucosinolates are *in planta* substrates for PYK10 that are tightly linked to the physiological functions of ER bodies. Our phylogenomic and three-dimensional (3D) structural modeling analysis identified the presence of two independent classes of myrosinases, represented by PYK10/PEN2 and thioglucoside glucohydrolases (TGGs). Overall, our results point to a functional link between ER bodies and glucosinolate metabolism and provide an insight into the diversity and evolutionary processes of glucosinolate/myrosinase systems in the order Brassicales.

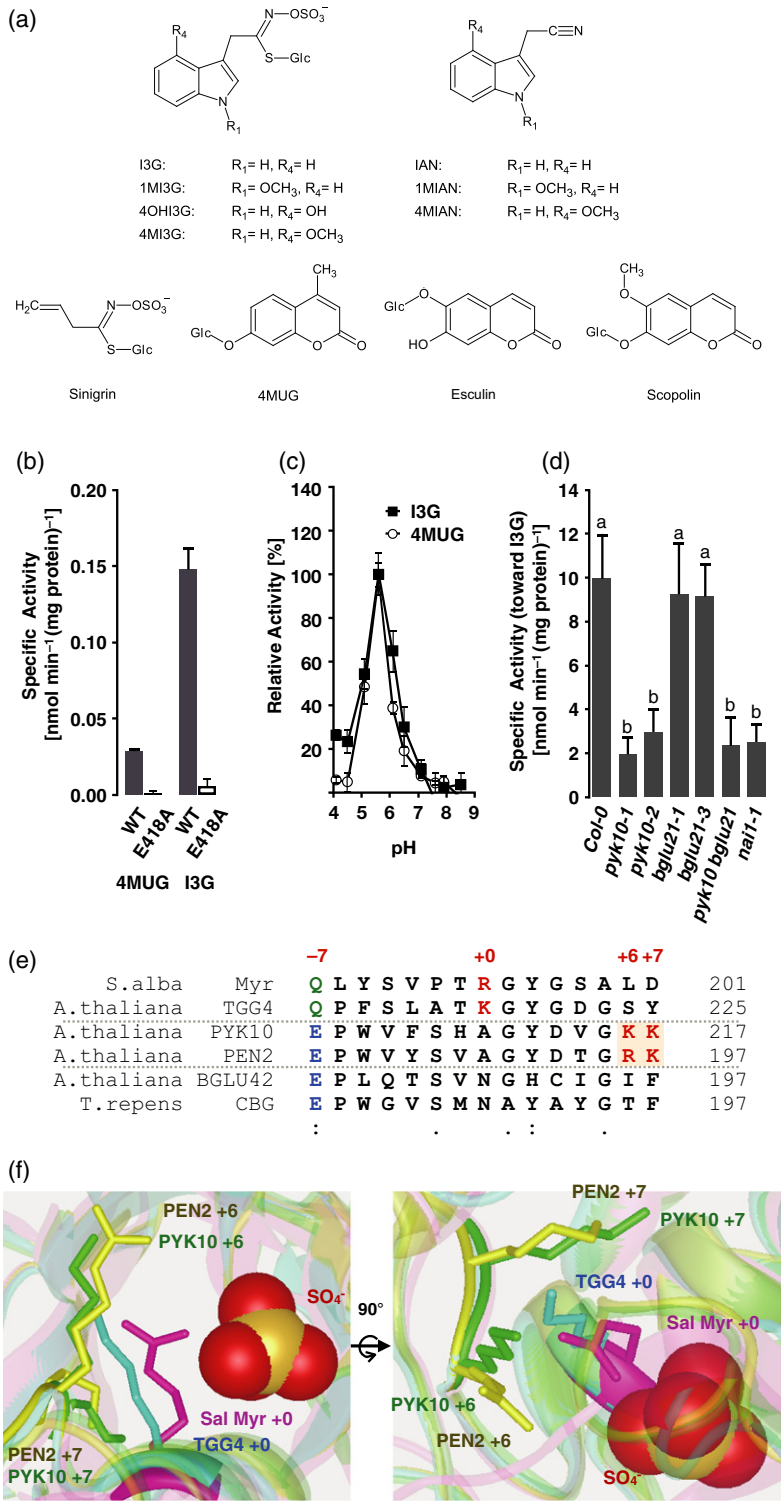


Figure 1. PYK10 is a novel myrosinase whose activity is conferred by distinct basic residues.(a) Chemical structures of the compounds described in this study. I3G, indol-3-ylmethyl glucosinolate; 1MI3G, 1-methoxyindol-3-ylmethylglucosinolate; 4OHI3G, 4-hydroxyindol-3-ylmethylglucosinolate; 4MI3G, 4-methoxyindol-3-ylmethylglucosinolate; IAN, indol-3-acetonitrile; 1MIAN, 1-methoxyindol-3-acetonitrile; 4MIAN, 4-methoxyindol-3-acetonitrile; 4MUG, 4-methylumbelliferyl-β-D-glucoside. (b) PYK10 exhibits both *O*-glucosidase and *S*-glucosidase activity. Activity of purified PYK10 protein with the wild-type (WT) sequence and with alanine substitution at the deduced catalytic center (E418A) was examined using either 4MUG or I3G *in vitro*. (c) Both *O*- and *S*-glucosidase activities displayed sharp pH optima under acidic conditions. Activity is shown relative to the activity at pH 5.5. (d) Total myrosinase activity toward I3G in root protein extract from Col-0, *pyk10-1*, *pyk10-2*, *bglu21-1*, *bglu21-3*, *pyk10-1 bglu21-1* and *nai1-1* measured *in vitro*. Letters indicate statistically significant differences (Tukey's honestly significant difference, $P < 0.05$). (e) Sequence comparison of myrosinases (TGG4 of *Arabidopsis thaliana* and Myr of *Sinapis alba*), PEN10 and PEN2 atypical myrosinases, and β-glucosidase not annotated as myrosinases [BGLU42 of *A. thaliana* and CYANOGENIC BETA-GLUCOSIDASE (CBG) of *Trifolium repens*]. PYK10 and PEN2 possess substrate-binding basic residues at different positions (+6 and +7, yellow shaded) from those previously identified in myrosinases (+0). Substitution of proton donor glutamate residue at the −7 position by glutamine was seen in annotated myrosinases but not the others. Deduced glucosinolate-binding residues are indicated by red letters, and glutamate and glutamine residues at position −7 are indicated by blue and green letters, respectively. (f) 3D structures at the substrate pocket of TGG4 (cyan), PEN2 (yellow) and PYK10 (green) are modeled based on sequence alignment with *S. alba* myrosinase (magenta). Side chains of +0 residues of TGG4 and Myr and of +6/+7 residues of PEN2 and PYK10 are shown with wire models. Two panels depict the same region from different directions as indicated. All basic residues identified in (e) oriented their side chains toward the negative charge of glucosinolates (red and yellow spheres, marked as SO_4^-). Backbone positions of respective residues are highlighted.

RESULTS

PYK10 exhibits *in vitro* myrosinase activity against indol-3-ylmethyl glucosinolate

PYK10 was shown to be unable to hydrolyze sinigrin (Ahn *et al.*, 2010), a short-chain AG, but its activity toward IGs was not investigated. We assayed *in vitro* activity of recombinant PYK10 protein (Figure S2a) toward indol-3-ylmethyl glucosinolate (I3G) and 4-methylumbelliferyl-β-D-glucopyranoside (4MUG), an artificial coumarinyl glucoside (Figure 1a). The results showed that PYK10 exhibited a five times higher activity toward I3G than

4MUG (Figure 1b, WT), both activities having a sharp acidic pH optimum at 5.5 (Figure 1c). The activities were abolished when the glutamate residue at the deduced catalytic center was substituted by alanine (Figure 1b, E418A). The kinetics of the enzyme revealed that PYK10 preferred I3G by 5.6 times over 4MUG, as estimated by the catalytic efficiency (V_{Max}/K_M; Table S1 and Figure S2b). Consistently, hydrolytic activity toward I3G in root crude protein extracts was reduced by approximately 80% in two independent *pyk10* alleles (Figure 1d). Comparable levels of activity reduction were observed in *nai1-1* and *pyk10 bglu21* root extracts, but not in *bglu21* roots. These results show that PYK10 hydrolyzes both coumarin glucosides and glucosinolates, and accounts for the bulk of total myrosinase activity against I3G in *A. thaliana* roots. Analogous results were reported in a recent study, in which *tgg4 tgg5* double-mutant plants exhibited a major reduction of myrosinase activity toward sinigrin in root protein extracts (Fu *et al.*, 2016), suggesting that the major myrosinases in *A. thaliana* hydrolyzing this short-chain AG are TGG4 and TGG5.

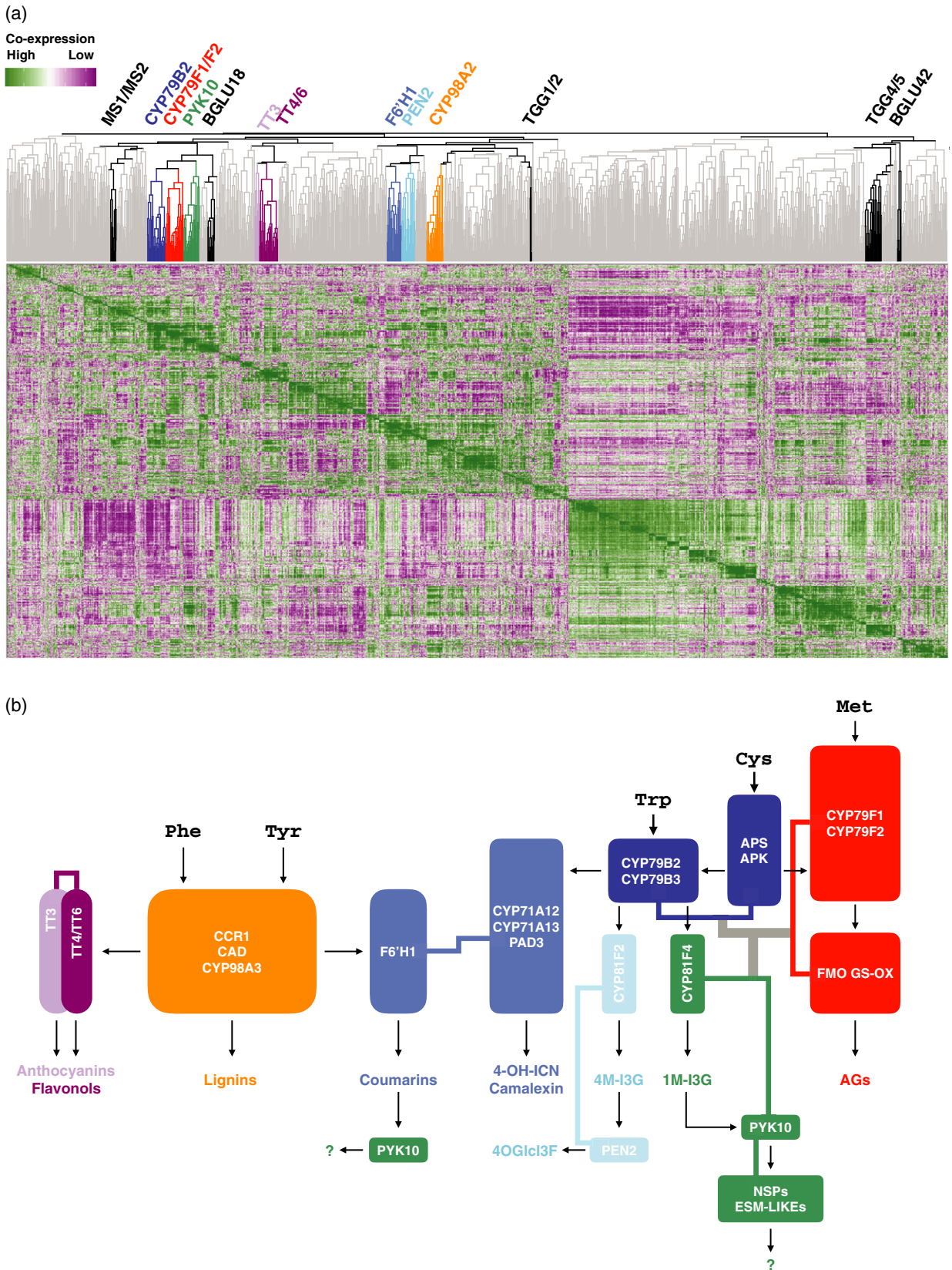
In contrast, we did not detect any significant differences in the accumulation of IG between the roots of a similar set of wild type (WT) and mutants, even upon MeJA treatment, which is known to stimulate the biosynthesis of IG in leaves (Figure S3) (Mikkelsen *et al.*, 2003). Untargeted comparative metabolomic analyses also failed to find clear evidence that PYK10 affects steady-state levels of secondary metabolites (Figure S4, Appendix S1). These results suggest that PYK10 is metabolically sequestered and inactivated in intact root cells.

Previous crystal structure analyses of a TGG from *Sinapis alba* identified the amino acid residues that are exclusively conserved across myrosinases and therefore thought to be required for their activity (Burmeister *et al.*, 1997; Zhou *et al.*, 2012). These signatures include a glutamine residue at one catalytic position (serves as a proton donor, −7 position; Figure 1e) and a basic arginine residue at the surface of the substrate-binding pocket (+0 position; Figure 1e), the latter provides positive charge to the binding pocket and enables electrostatic interaction with negatively charged glucosinolates. Although neither signature was found in PYK10 or PEN2, our 3D structural modeling identified two basic residues shared between PYK10 and PEN2 orienting their side chains toward the sulfate group of the glucosinolate, as the +0 signature residue does, which explains the unexpected activity of these β-glucosidases (Figures 1e,f). These amino acids were located only six and seven residues away (+6 and +7) from the typical +0 basic residue. We found no evidence for an indel polymorphism between the +0 and +6/+7 positions that could cause a residue shift (Figure 1e), suggesting that these substrate-binding sites arose independently.

The ER body formation/functions and IG biosynthesis/metabolism form a tightly connected transcriptional network

Our results, together with a previous report that identified PYK10 activity toward scopolin (Ahn *et al.*, 2010), prompted the question of which class of substrates, namely coumarinyl glucosides or IGs, is more tightly linked to the physiological function of ER bodies *in planta*. It has been reported that *PYK10* is significantly co-expressed with genes encoding NAI2, jacalin-related lectins (JALs), a GDSL lipase-like protein (GLL) and Meprin and TRAF homology (MATH) domain-containing proteins (MATHs), which are either required for ER body formation (Yamada *et al.*, 2008) or define the components of heteromeric PYK10 protein complexes (Figure S5) (Nagano *et al.*, 2008; Nakano *et al.*, 2012). Two genes encoding enzymes involved in IG modification pathways [a cytochrome P450 monooxygenase (*CYP81F4*) and *INDOLE GLUCOSINOLATE O-METHYLTRANSFERASE 5* (*IGMT5*)] were also found to be co-expressed with *PYK10*, which is similar to the co-expression among *PEN2*, *CYP81F2* and *IGMT1/-2/-3/-4* (see below), suggesting a link between the ER body system and IG metabolism.

We next employed a comprehensive, targeted approach to test for co-expression among the genes involved in various primary/secondary metabolic pathways, including glucosinolate, flavonoid, lignin and coumarin pathways (Shirley *et al.*, 1995; Vanholme *et al.*, 2010; Chan *et al.*, 2011; Siwinska *et al.*, 2014), as well as other *BGLU*s and genes involved in the ER body system. Sequence homologs of these genes with a low similarity threshold were included in order to ensure sufficient coverage, and the cross-hybridized genes on the ATH1 gene chip were incorporated by utilizing a probe-wise co-expression dataset. Probe-wise and genome-wide mutual ranks (MRs) (Obayashi and Kinoshita, 2009) were calculated from 15 275 public microarray data (ATTED-II, v. Ath-m6.0) for each combination of the 1474 probes corresponding to 1549 genes and a $\log_2$-transformed MR matrix was used for hierarchical clustering. The co-expressed gene clusters responsible for each metabolic pathway were objectively determined by silhouette widths analysis (Rousseeuw, 1987) (Figure S6, Table S2; see Appendix S1 for details). The heatmap and the dendrogram shown in Figure 2 indicate that the genes indeed formed a number of co-expressed gene clusters, as represented by the green boxes aligned on the diagonal axis. Clusters optimized for *CYP79B2*, *CYP79F1/F2* and *PYK10* contained almost all the genes required for biosynthesis of Trp and IGs, AGs and ER bodies, respectively (Data S1), which is consistent with previous studies (Hirai *et al.*, 2007; Chan *et al.*, 2011). The *PYK10* cluster also contained nitrile specifier proteins (*NSP1/-3/-4*) that affect the terminal products of



glucosinolate metabolism. Notably, *CYP79B2*, *CYP79F1/F2* and *PYK10* clusters collectively formed a monophyletic branch in the clustering dendrogram, suggesting that the ER body system and glucosinolate biosynthetic pathway are transcriptionally linked to each other. In contrast, the cluster optimized for *F6′H1*, and therefore likely to be coumarin-related, was far apart from all three aforementioned clusters, even from the *CYP98A3* cluster that corresponded to the common pathway of phenylpropanoid biosynthesis. The *F6′H1* cluster also contained genes required for camalexin and 4-hydroxyindole-3-carbonyl nitrile (4-OH-ICN) biosynthesis, namely *CYP71A12/13*, *PAD3* (Nafisi *et al.*, 2007; Muller *et al.*, 2015) and *CYP82C2* (Rajniak *et al.*, 2015). Furthermore, the *F6′H1* cluster was located close to the *PEN2* cluster, composed of *PEN2*, *MYB51* and IG modification genes (*IGMT1-4* and *CYP81F2*). This suggests a metabolic/functional link between the Phe/Tyr-derived coumarin biosynthetic pathway and Trp-derived branch pathways. Interestingly, the *BGLU18* cluster composed of *BGLU18*, *TSA1* (the only sequence homolog of *NAI2* in *A. thaliana*; Figure 5b) and other *GLL*s, *JAL*s and *MATH*s was closely located to the AG/IG/ER body clusters. This suggests that the inducible leaf ER body accumulating BGLU18 (Ogasawara *et al.*, 2009) is also linked with glucosinolate metabolism.

A condition-wise co-expression analysis using a limited number of microarrays (Obayashi *et al.*, 2011) failed to detect an increased degree of co-expression between ER bodies and glucosinolate biosynthesis when developmental patterns were used to calculate correlations (Figures S7–S9, Tables S3 and S4, Data S2 and Appendix S1), rejecting the possibility that the correlation was merely due to a high expression in specific tissues/organs. This is consistent with the fact that the spatiotemporal distributions of AGs and IGs largely differ from each other (Brown *et al.*, 2003). Instead, the analysis depicted a co-expression facilitated by hormone treatments between *BGLU18*, *MYB34*, *NAI1* and IG biosynthetic genes, supporting the link between inducible ER bodies and IG metabolism (Appendix S1).

Overall, this analysis identified a tight link between transcriptional regulation of genes required for AG production, IG production and ER body formation. This finding, combined with a previous study that suggested a co-occurrence of ER bodies and IG in root cortex cells (Moussaieff *et al.*, 2013), supports our hypothesis that ER bodies serve in physical and biochemical sequestration that is indispensable for concomitant production of glucosinolates and PYK10 myrosinase within the same cell.

Phylogenetic co-occurrence of ER bodies and IGs

Assuming that the functions of ER bodies are evolutionarily conserved, we tested the hypothesis that the taxonomic occurrence of the physiological substrates of PYK10 is associated with the occurrence of ER bodies. We investigated the distribution of root-associated ER bodies, IGs and scopolin among eight selected Brassicaceae species (Figure 3a), which covered the broad phylogenetic diversity of Brassicaceae (Couvreur *et al.*, 2010). Among these species, only *Cardamine hirsuta* and *Arabis alpina* have previously been reported to develop ER bodies in their roots (Iversen, 1970b), whereas no detailed analysis has been reported for the other species. All species were previously analyzed for glucosinolate accumulation in leaves (Bednarek *et al.*, 2011), but not in roots. Scopolin accumulation has only been analyzed for *A. thaliana* (Bednarek *et al.*, 2005).

Transient expression of an ER-localized GFP (pro35S::*SP-GFP-HDEL* construct) (Hayashi *et al.*, 2001) in root epidermis cells of *Arabis alpina*, *Crucihimalaya lasiocarpa*, *Capsella rubella* and *Olimarabidopsis pumila* labeled rod-shaped structures that were similar to *A. thaliana* ER bodies, in addition to the typical ER network (Figure 3b). The ER bodies in *Cardamine hirsuta* and *A. thaliana* roots were clearly delineated when the ER membranes were stained with the ER-specific fluorescent dye (ER-tracker Green; Figure 3c). We also subjected *Arabis alpina*, *Thellugiella salsuginea*, *Sisymbrium irio, Cardamine hirsuta* and *Capsella rubella* roots to transmission electron microscopy (TEM) for ultrastructural analysis. TEM images of *A. thaliana* roots show that ER bodies have a relatively high electron density and are often surrounded by ribosomes (Hayashi *et al.*, 2001; Matsushima *et al.*, 2002). For all five species tested with ultrastructural analysis, TEM identified similar rod-shaped structures that resembled ER bodies in *A. thaliana* and were dissimilar to any other organelles reported previously (Figure 3d). Taken together, all eight species tested formed ER body-like structures in their root cells, as indicated by at least one of the three methods used. This was also supported by the identification of *NAI2*

Figure 2. *PYK10* and other ER body-related genes are tightly co-expressed with genes required for glucosinolate biosynthesis.
(a) Mutual ranks (MRs) amongst 1474 probes (see Appendix S1) were calculated and illustrated as a heatmap. Green and magenta indicate high and low co-expression, respectively. Probes are aligned in the same order in rows and columns, according to hierarchical clustering using $\log_2$-transformed MRs, as shown as a dendrogram above the heatmap. Green rectangles aligned on the diagonal axis indicate gene clusters with high co-expression. Clusters responsible for different metabolic pathways are objectively determined (see Appendix S1 and Figure S6) and the corresponding branches in the dendrogram are colored according to the clusters.
(b) Schematic representation of the pathways analyzed for co-expression. Biosynthesis of phenylalanine (Phe), tyrosine (Tyr), tryptophan (Trp), cysteine (Cys) and methionine (Met) are not shown. Boxes and letters outside the boxes indicate (sub-)pathways and compounds produced out of the pathways with color coding identical to (a). For each pathway, representative genes are indicated in the boxes. Co-expressed gene modules are indicated with the same colors and connected with thick lines.

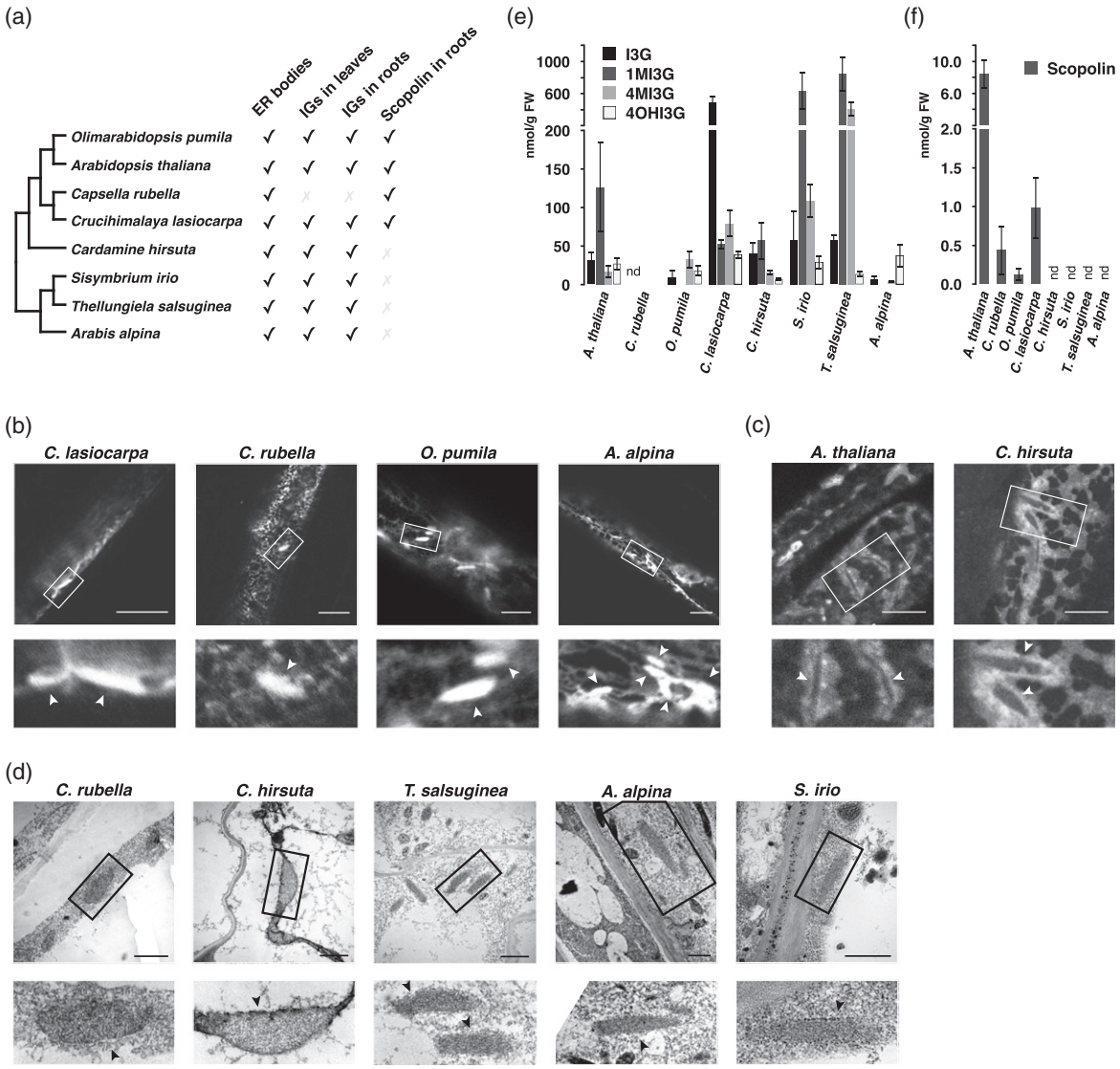


Figure 3. Endoplasmic reticulum (ER) bodies and indole glucosinolates (IGs) taxonomically co-occur in Brassicaceae.
(a) Detection of ER bodies (c, d, e), IGs (e), and scopolin (f) summarized with a phylogenetic tree according to a previous study (Couvreur *et al.*, 2010; not shown to scale). The presence of IGs in leaves is shown according to Bednarek *et al.* (2011). Black checks and gray crosses indicate presence and absence, respectively. (b) ER-localized GFP (SP-GFP-HDEL) was transiently expressed in *Crucihimalaya lasiocarpa*, *Capsella rubella*, *Olimarabidopsis pumila* and *Arabis alpina*, which visualized rod-shaped structures that resembled ER bodies in *Arabidopsis thaliana* (arrowheads in bottom panels). (c) The ER membrane was chemically stained in living roots of *A. thaliana* and *Cardamine hirsuta*, which outlined ER bodies (arrowheads). (d) Roots of *C. rubella*, *C. hirsuta*, *Thellugiella salsuginea*, *A. alpina* and *Sisymbrium irio* were analyzed by transmission electron microscopy, which identified electron-dense and ribosome-attached spindle structures (arrowheads). The bottom panels (b, c, d) are enlarged view of insets in the top panels. (e) Amounts of IGs in the Brassicaceae species quantified by HPLC/UV. All tested species except for *C. rubella* accumulated substantial amount of IGs. (f) Scopolin abundance in the same set of species was assessed with HPLC/fluorescence detectors; nd, not detected. Bars are shown with standard deviations.

in all genome sequences from Brassicaceae species deposited in the Phytozome database (version 11, released on 12 January 2016), but not in a basal Brassicales species, *Carica papaya* (Figures 4a,b).

Almost all species accumulated the four IGs analyzed (Figure 3e). The only exception was *Capsella rubella*, which has lost the ability to accumulate IGs in leaves (Bednarek *et al.*, 2011). A series of CYP enzymes required for IG biosynthesis was also conserved in, and was in most cases specific to, Brassicaceae (Figures 4c–f and S10a–c). In contrast, we detected scopolin in roots of only four out of the eight tested species at levels clearly lower than those of IGs (Figure 3f). The putative *F6′H1* oxygenase orthologs, on the other hand, are distributed throughout dicotyledonous species (Figures 4g and S10d) (Schmid *et al.*, 2014), suggesting high variability in the products of this pathway. This also raised a possibility that the tested species accumulate discrete sets of coumarin glucosides that

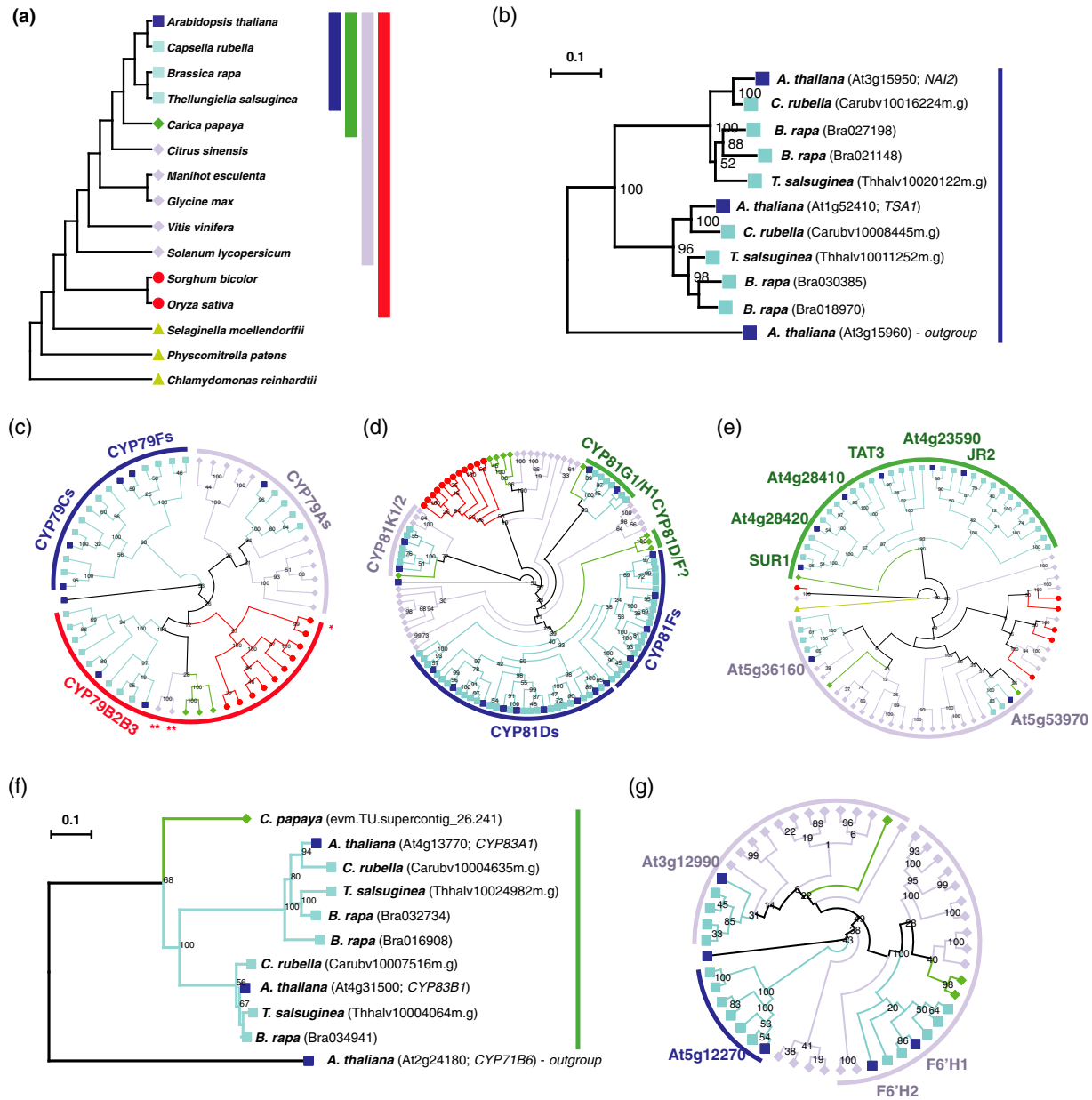


Figure 4. All Brassicaceae species retain genes required for endoplasmic reticulum (ER) body formation and indole glucosinolate (IG) biosynthesis.
(a) Schematic representation of the species used for phylogenomic analysis, according to the Phytozome database (not shown to scale). Tree leaves colored with blue, cyan, green, light purple, red and yellow indicate *Arabidopsis thaliana*, other Brassicaceae plant, *Carica papaya*, other dicotyledonous plants, monocotyledonous plants, and other basal plants, respectively. Color strips beside the tree correspond to the colors of strips next to neighbor-joining (NJ) tree shown in this figure (blue, green, purple and red for Brassicaceae-specific, Brassicales-specific, dicotyledonous-specific and angiosperm-specific clades, respectively). (b)–(g) The NJ trees of NAI2s, CYP79s, CYP81s, SUR1s, CYP83s and F6'H1s (see Appendix S1 for details). (b) NAI2 homologs were only present in Brassicaceae, not in *C. papaya* or other tested species. The scale bar indicates substitutions per site. (c) Sequence homologs of CYP79B2/B3 were commonly found in angiosperms (indicated by red bars), including *Manihot esculenta* (cassava) CYP79D1/D2 and CYP79A1 (asterisks). CYP79Cs and CYP79Fs appear to be specific to Brassicaceae. (d) The closest homolog of CYP81Fs in *C. papaya* was found to be also ancestral to CYP81Ds, the function of which is unknown. (e) SUR1 homologs contained a series of genes with function unknown, including *TAT3* and *JR2*. (f) CYP83s formed an exclusive monophyletic clade composed of Brassicales genes, amongst a huge diversity of CYP71s. *Carica papaya* CYP83 is ancestral to both CYP83As and CYP83Bs. The scale bar indicates substitutions per site. (g) F6'H1 and F6'H2 were widely conserved in dicotyledonous plants but not in monocotyledonous plants. Bootstrap values with 100 replications are indicated for each node. Trees in (c, d, e and g) are not shown to scale, and respective trees with branch length and gene IDs can be found in Figure S10.

were not detected in our analysis, but this requires a more complex experimental design using an identical set of species/isolates in both metabolomics and phylogenetic analysis. A detailed phylogenetic analysis of coumarins also requires identification of further enzymes specifically involved in this pathway (Siwinska *et al.*, 2014). For instance, genes encoding UGT76B1 and UGT73D1 were tightly co-expressed with *F6'H1* (Data S1), and therefore

these two UDP-glucosyltransferases may be good candidates for coumarin glucosylating enzymes.

Our results from both phylogenetic co-occurrence patterns (ER body formation and IG metabolites in roots) and correlated transcriptional regulation in *A. thaliana* point to a tight link between ER bodies and glucosinolate biosynthesis. Thus, we propose that the physiological function of ER bodies *in planta* is at least partially associated with glucosinolate metabolism.

Evolutionary origin of myrosinases

Finally, we asked whether other β-glucosidases stored in ER bodies (BGLU18, BGLU21 and BGLU22) could also hydrolyze glucosinolates. As described above, we identified a set of potential glucosinolate-binding residues at positions +6 and +7 that explained the 'atypical' myrosinase activity of PYK10 and PEN2 (Figures 1e,f). Notably, the presence and structural location of at least one of these basic residues is conserved in 16 (BGLU18–BGLU33) out of 47 annotated *A. thaliana* β-glucosidases (Figures S11 and S12a). This suggests that almost half of the *A. thaliana* β-glucosidases (22 out of 47) possess myrosinase activity. Hereafter, we refer to the 'atypical' versus 'classic' classes of myrosinases as EE versus QE myrosinases, respectively, based on their sequence signatures.

To address whether EE- and QE-type β-glucosidase enzymes share a common ancestral myrosinase, we conducted a phylogenomic analysis using 52 public plant genome sequences in the Phytozome database (see Appendix S1 for details). The neighbor-joining (NJ) tree shown in Figure 5(a) illustrates that TGG1–TGG6 (QE myrosinases) and BGLU18–BGLU33 (EE myrosinases) formed monophyletic clades with high bootstrap values (96 and 100, respectively) that were specific to Brassicales, as reported previously (Xu *et al.*, 2004; Nakano *et al.*, 2014). The distribution of glucosinolate-binding signatures exhibited an irregular pattern, except for those amongst EE and QE myrosinases, indicating that these signatures are under purifying selective pressure (Figure 5a, outer rings). In contrast, the other signature of myrosinases (i.e. a glutamine residue at the −7 position) was specific to QE but not to EE myrosinases. *Carica papaya* TGG1 and TGG2 were reported to have no basic residues for glucosinolate binding (Nong *et al.*, 2010). However, we identified yet another basic residue at position +4, the side chain of which was also oriented to the deduced substrate-binding

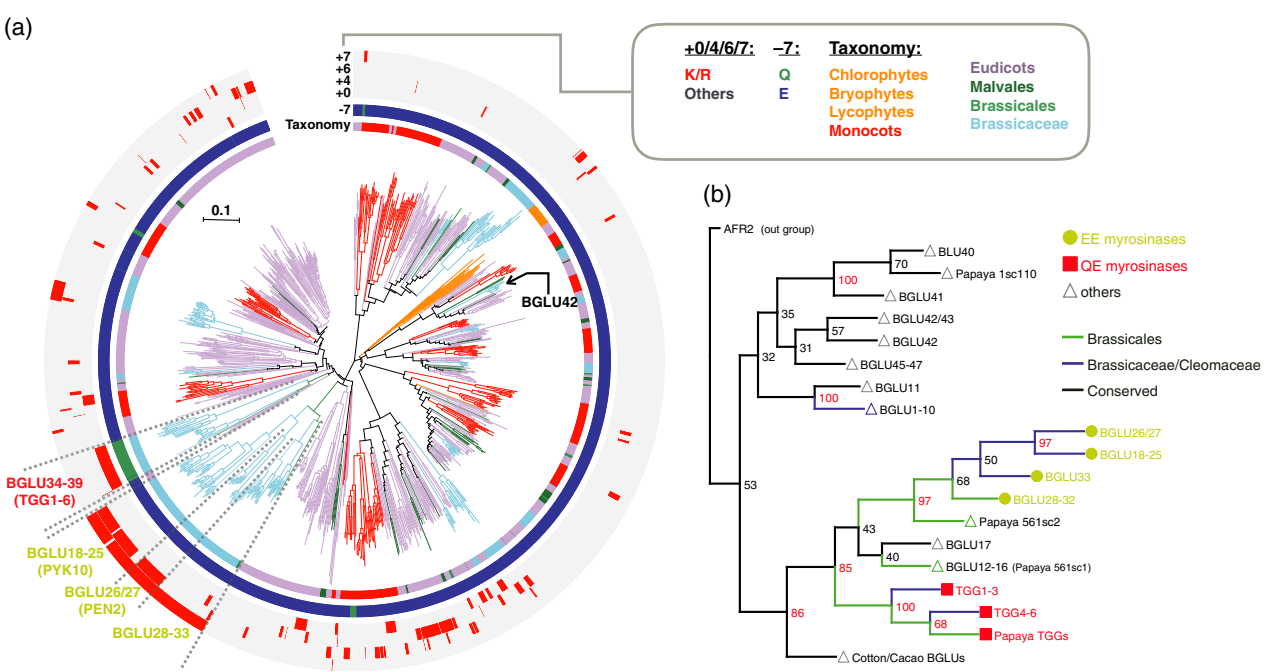


Figure 5. Lineage specific distribution of β-glucosidases and glucosinolate-binding residues across the plant kingdom.
(a) Neighbor-joining tree of 962 β-glucosidases from 52 plant species with genome sequences deposited to the Phytozome database (details can be found at https://phytozome.jgi.doe.gov). Scale bar indicated substitutions per site. Branches are colored according to the taxonomy. Rings surrounding the tree indicate taxonomy and the residues at positions −7, +0, +4, +6, and +7. Color coding is shown in the inset. BGLU34–BGLU39 (TGG1–TGG6) and BGLU18–BGLU33, corresponding to QE and EE myrosinases, form monophyletic clades that are specific to Brassicales. The glutamine residue at −7 and the basic residue at +0 are conserved within QE myrosinases, and basic residues at +4/+6/+7 are conserved within EE myrosinases. Other BGLUs have basic residues at any of the +0/+4/+6/+7 positions and a glutamine residue at −7, but in a irregular pattern.
(b) β-Glucosidases from Brassicales and Malvales were subjected to maximum likelihood analysis. BGLU subfamilies that were grouped with high bootstrap values (>97%) were collapsed. EE and QE myrosinases are indicated by yellow and red letters. No footprint for common acquisition of glucosinolate-binding capacity between EE and QE myrosinases was found.

pocket (Figure S12b, cyan and green side chains). A subset of EE myrosinases, namely BGLU33 orthologs, also possessed +4 basic residues in addition to +6 and +7 (Figure S12c), which was in contrast to QE myrosinases among which we did not detect any TGGs that possessed both +0 and +4 signatures (Figure 5a). One of the *Carica papaya* EE myrosinases (evm.TU.supercontig_444.5) that also possessed three basic residues at positions +4/+6/+7 was used to confirm that the direction of the side chain spanned from the +4 position is conserved within *Carica papaya* myrosinases (Figure S12b) and between *Carica papaya* and *A. thaliana* EE myrosinases (Figure S12c).

Interestingly, the two clades of myrosinases constituted independent groups in the NJ tree. This was also supported by a maximum likelihood (ML) analysis using β-glucosidases only from Brassicales species, with those from Malvales species as out-groups, namely cotton (*Gossypium raimondii*) and cocoa bean (*Theobroma cacao*) (Figure 5b). The *Carica papaya* TGGs formed an independent clade that was separated from both subclasses of TGGs (TGG1–TGG3 and TGG4–TGG6) (Nong *et al.*, 2010), suggesting that the divergence of TGG subclasses was obtained after *Carica papaya* was separated from the other Brassicales species. The *Carica papaya* gene evm.TU.supercontig_561.2 encoded a β-glucosidase that belongs to the EE myrosinase clade supported by a high bootstrap value (97%); however, none of the glucosinolate-binding sites (+0, +4, +6, and +7) was a basic amino acid residue. This might suggest that this gene is the common ancestor for all EE myrosinases, which is also supported by the finding that the closest sequence homolog of this gene (evm.TU.supercontig_561.1) was found in the BGLU12–BGLU16 clade.

Overall, our analysis strongly suggests that all ER-resident type β-glucosidases, including ER body-stored BGLU18, BGLU21 and BGLU22, are potent to act as myrosinases, further supporting the idea that ER bodies are functionally linked with glucosinolate metabolism. Independent evolutionary processes that provided the ER body–EE myrosinase system and the myrosin cell–QE myrosinase system point to the complex and highly diversified machineries of glucosinolate–myrosinase systems in the Brassicaceae.

DISCUSSION

PYK10 is a novel myrosinase that requires external signals for activation

Here we have shown that PYK10 hydrolyzes both an *S*-glucoside (I3G) and an *O*-glucoside (4MUG) with a narrow pH optimum range (Figure 1b). The pH in the ER body is presumably identical to the canonical ER, i.e. neutral around 7 (Shen *et al.*, 2013), which would render PYK10 enzyme inactive inside intact cells. This is consistent with the lack of significant differences in steady-state accumulation of secondary metabolites between wild-type and *pyk10* mutant plants (Figures S3 and S4). Hence we propose that ER bodies provide a dual-layer safeguard system that physically segregates PYK10 from its substrates, and biochemically inactivates the enzyme. This provides a clear contrast to the foliar myrosin cells, where myrosinases and glucosinolates are stored in discrete cells and therefore the system does not require biochemical inactivation of the enzymes (Jørgensen, 1981; Koroleva *et al.*, 2000, 2010; Ueda *et al.*, 2006). Moreover, as accumulation in separate cells does not require stringently co-regulated transcriptional machineries, expression of *TGG1/-2* is not linked to glucosinolate biosynthesis (Figure 2). Activity of the other reported EE-myrosinase, PEN2, is also safeguarded by two independent machineries (Lipka *et al.*, 2005; Bednarek *et al.*, 2009; Fuchs *et al.*, 2016). This similarity may suggest a common ancestral innovation that enabled *intra*-cellular mustard bombs that are linked with EE myrosinases, which is independent from *inter*-cellular mustard bombs. Unlike feeding damage by insects, abiotic and microbial stresses are not necessarily accompanied by cell disruption, which might have facilitated innovation of such dual-layer safeguarding systems within a single cell.

ER bodies are potentially engaged in plant–microbe interactions via IG metabolism and in abiotic stress responses via coumarin metabolism

The IG metabolic pathway comprises Brassicales-specific defense strategies against microbial pathogens such as powdery mildew fungi (Bednarek *et al.*, 2009; Clay *et al.*, 2009) and a nectrotrophic fungus *Plectosphaerella cucumerina* (Frerigmann *et al.*, 2016). Recent studies also assigned that the *A. thaliana* IG metabolic pathway an essential role in the establishment of beneficial interactions with fungal root endophytes, namely *Piriformospora indica*, *Sebacina vermifera* (Lahrmann *et al.*, 2015) and *Colletotrichum tofieldiae* (Hiruma *et al.*, 2016). Similarly, mutants lacking either NAI1 or PYK10 showed hypercolonization by *P. indica* that resulted in loss of the beneficial interaction (Sherameti *et al.*, 2008). This phenotypic similarity, combined with the transcriptional link between ER bodies and IG biosynthesis described in this study, suggests that ER bodies are engaged in interactions with soil-borne microbes, including both pathogenic and beneficial fungi.

Scopolin metabolism, on the other hand, appears to have a role in iron uptake in *A. thaliana* roots (Fourcroy *et al.*, 2014; Schmid *et al.*, 2014). Recent work has demonstrated that another β-glucosidase, BGLU42, is crucial for the secretion of phenolic compounds under iron deficiency, most likely including scopolin derivatives (Zamioudis *et al.*, 2014). BGLU42 belongs to one of the most conserved clades of BGLUs (Figure 5a), implying that the ability to hydrolyze coumarin glucosides is an ancient

feature of conserved BGLUs. This is consistent with the broader distribution of coumarins and *F6'H1* orthologs in dicotyledonous plants (Figure 4g) (Gutierrez-Mellado *et al.*, 1996; Churngchow and Rattarasarn, 2001; Chong *et al.*, 2002). However, this does not exclude the possibility that PYK10 is involved in coumarin glucoside metabolism in *A. thaliana*, for example under specific adverse environmental conditions. Similarly, it also remains conceivable that AG metabolism is linked to the physiological functions of ER bodies. To determine which class of substrates is linked to the physiological function(s) of ER bodies, single and multiple mutants that lack both ER bodies and either class of glucosides (e.g. *myb34/51/122* for IGs, *myb28/29* for AGs and *f6'h1* for coumarins) need to be tested for common phenotypic differences from the wild type.

Co-expression analysis is a powerful tool but requires careful interpretation

The ability of co-expression analysis to infer metabolic pathways has been evident from a number of past studies (e.g. Obayashi *et al.*, 2007; Saito *et al.*, 2008; Saito and Matsuda, 2010; Yonekura-Sakakibara and Saito, 2013), including identification of glucosinolate biosynthetic/modification enzymes (Hirai *et al.*, 2007). In our analysis, we confirmed a functional link between PEN2 and IG metabolism, a metabolic link known to play an important role in plant immunity (Bednarek *et al.*, 2009; Clay *et al.*, 2009). We also detected co-expression of *BGLU42* with *MYB72* (Figure 2 and Data S1), which is consistent with previous reports that identified a shared role for BGLU42 and MYB72 in the secretion of fluorescent coumarin derivatives (Fourcroy *et al.*, 2014; Schmid *et al.*, 2014; Zamioudis *et al.*, 2014). Remarkably, a potential role for TGG4 and TGG5 in auxin production suggested by a very recent study (Fu *et al.*, 2016) is supported by our result that *YUCCA* genes, which are involved in CYP79B2/B3-mediated auxin biosynthesis (Zhao, 2014), are indeed located next to the *TGG4/-5* cluster (Data S1). These results validate the relevance of our co-expression analysis and allowed us to look for further functional links between other genes.

On the other hand, despite its potential to identify functional gene modules, co-expression analysis fails to identify functional segregation. This is mainly because functional coordination is not always accompanied by transcriptional co-regulation, as indeed is demonstrated by the lack of co-expression between *TGG1/-2* and glucosinolate pathways despite a clear functional relationship. Therefore, our analysis does not suggest any functional separation between PYK10/ER bodies and coumarin glucoside metabolism. Nonetheless, the tight co-expression between ER bodies and glucosinolate biosynthetic pathways, combined with the biochemical properties of PYK10 as a myrosinase, suggest a functional link between these systems.

Gene modules composed of IG modification, hydrolysis and catabolism genes may facilitate a functional differentiation among myrosinases

The IG modification genes, such as *IGMT*s and *CYP81F*s (Pfalz *et al.*, 2011), were included in *PEN2* and *PYK10* clusters but not in the *CYP79B2* cluster, suggesting that IG modification is linked with metabolic activity of the respective myrosinases. In fact, the production of 4MI3G by CYP81F2 is indispensable for PEN2-mediated immunity against microbial pathogens (Bednarek *et al.*, 2009; Clay *et al.*, 2009; Pfalz *et al.*, 2009; Sanchez-Vallet *et al.*, 2010; Fuchs *et al.*, 2016; Lu *et al.*, 2015). This in turn suggests that 1-methoxylation of I3G by CYP81F4 and IGMT5 is important for the specific function of PYK10.

The end products of glucosinolate hydrolysis are determined by the conditions and proteins co-occurring with myrosinases during this metabolic process (Wittstock and Burow, 2010). These proteins include epithiospecifier protein (ESP), epithiospecifier modifier 1 (ESM1) and nitrile specifier proteins (NSPs). In our analysis *ESP*, encoding a protein promoting epithionitrile and nitrile formation (Lambrix *et al.*, 2001; Burow *et al.*, 2009), was found to be tightly co-expressed with *TGG1* and *TGG2*, whereas *NSP1/-3/-4*, encoding proteins promoting simple nitrile formation (Burow *et al.*, 2009; Kissen and Bones, 2009), were found in the *PYK10* cluster (Data S1). This finding is consistent with the preferential end products in shoots and roots (Wentzell and Kliebenstein, 2008). In the case of IGs, it has been shown that ESP and ESM1 promote production of indole acetonitrile (IANs; Appendix S1) (Burow *et al.*, 2008).The type of end product of glucosinolate metabolism appears to be crucial for the physiological output (Jander *et al.*, 2001; Lambrix *et al.*, 2001; Buxdorf *et al.*, 2013). Hence, a stringent transcriptional control to coordinately express ER body structural proteins, myrosinases, substrate modification enzymes and specifier proteins may be crucial for orderly function of the myrosinase–glucosinolate system.

Convergent evolution of myrosinases in plants, insects and bacteria

Our results indicate that PYK10 belongs to the class of EE myrosinases with a greater number of members than the classic QE myrosinases (Figure 5a). Our NJ and ML analyses did not find any clear evidence that EE myrosinases arose from QE myrosinases, or vice versa. We also found no evidence for insertion or deletion between positions +0 and +4/6/7, making it less likely that short-length structural variation caused a shift of binding residues. Considering the frequent but irregular emergence of basic residues at these positions across the plant kingdom (Figure 5a), we conclude that the two classes of myrosinases arose independently during differentiation of the order Brassicales. In

the animal kingdom some specialist herbivores [*Brevicoryne brassicae* (cabbage aphid), *Lipaphis erysimi* (turnip aphid) and *Phyllotreta striolata* (a flea beetle)] have been reported to be myrosinase innovators (Bridges *et al.*, 2002; Beran *et al.*, 2014). These insect enzymes possess a glutamate residue as a proton donor at position −7 and a basic residue at position +0, a combination distinct from both classes of plant myrosinases. Interestingly, a bacterial myrosinase was recently isolated from *Citrobacter* strain WYE1 (Albaser *et al.*, 2016). This enzyme belongs to another class of glycosidases (family 3 glycosyl hydrolase) and depicts convergent evolution of myrosinase activity across three different kingdoms.

Diversification of glucosinolate–myrosinase systems accompanied with multiple transcriptional co-options

Our findings on the huge diversity of myrosinases and on transcriptional co-regulations, combined with phylogenomic analysis, have revealed that the glucosinolate–myrosinase system has been remarkably diversified during the differentiation of Brassicales (Figure 6). The ancestral Brassicales species evolved only benzyl glucosinolates, together with one subclass of QE myrosinases stored in myrosin cells and one subclass of EE myrosinases localized in vacuoles (Figure 6a–c). Subsequently, both EE and QE myrosinases have undergone diversification into three subclasses of EE myrosinases (BGLU18–BGLU25, PEN2/BGLU27 and BGLU28–BGLU33) and two of QE myrosinases (TGG1/-2 and TGG4/-5) (Figure 6d,g). Diversification of glucosinolates resulted in three subclasses of these metabolites (BGs, IGs and AGs) (Figure 6e,f). In addition, a limited number of species have acquired specific subcellular compartments (i.e. ER bodies) to store EE myrosinases that acquired ER retention signals (BGLU18–BGLU25) (Figure 6h). Root constitutive ER bodies and leaf-inducible ER bodies are transcriptionally linked with the core glucosinolate biosynthetic pathway (Figure 6j), as well as IG modification pathways (Figure 6i). The co-option between glucosinolate biosynthesis and JA response pathways (Schweizer *et al.*, 2013; Frerigmann and Gigolashvili, 2014; Frerigmann *et al.*, 2014) (Figure 6l) and between TGG1/-2-accumulating myrosin cells in shoots and stomatal development (Shirakawa *et al.*, 2014, 2016) (Figure 6m) require MYC2/-3/-4 and FAMA, respectively, which are widely conserved in land plants (Kazan and Manners, 2013; Ran *et al.*, 2013). This raises the question of how a lineage-specific innovation can become integrated with an evolutionarily common pathway. Similarly, further investigation is required for the putative transcriptional links between auxin biosynthesis and TGG4/-5 (Figure 6n) and between camalexin/4-OH-ICN pathways and coumarin biosynthesis (Figure 6o).

Currently, plants of the Brassicaceae family are equipped with a glucosinolate–myrosinase system that has much higher complexity than that of the ancestral Brassicales species. However, it remains unclear when the sub-divergence within QE myrosinases and the subcellular diversification of EE myrosinases occurred. Importantly, the evolutionary origin of ER bodies remains elusive, as we detected the orthologs of *NAI2* and *TSA1* only within Brassicaceae but not in *Carica papaya*, even with a low E-value threshold (Figure 4b, Appendix S1). Nevertheless, the core Brassicales species underwent multiple innovations comprising metabolic enzymes (myrosinases), secondary metabolites (glucosinolates) and a subcellular structure (ER bodies), which contributed to the functional diversity and sophistication of the overall system.

EXPERIMENTAL PROCEDURES

Plant materials

We used wild-type *A. thaliana* (Col-0) plants and the following mutants obtained as described previously: *nai1-1* (Matsushima *et al.*, 2004), *pyk10-1* (Nagano *et al.*, 2008), *bglu21-1*, *bglu21-3*, *pyk10-1 bglu21-1* (Nagano *et al.*, 2009) and *cyp79B2 cyp79B3* (Bednarek *et al.*, 2009). The *pyk10*-2 (SAIL_156_E11; N871638) line was obtained from The European Arabidopsis Stock Centre. Seeds of *O. pumila*, *Capsella rubella*, *Crucihimalaya lasiocarpa*, *Cardamine hirsuta*, *S. irio*, *T. salsuginea* and *Arabis alpina* were obtained as described previously (Bednarek *et al.*, 2011).

Protein preparation for enzymatic assays

The recombinant PYK10 was expressed in tobacco BY-2 cells using the tomato mosaic virus (ToMV)-based expression system (Dohi *et al.*, 2006, 2008). The PYK10 coding sequence was fused with the His_6 tag and purified with a Ni^{2+} affinity column. The eluted fraction was dialyzed in 0.1 M sodium phosphate buffer (pH 7.0) and centrifuged for 5 min at 20 000 ***g*** to avoid aggregations. For the preparation of total protein extracts, plants were grown on vertical agar (0.9% w/v) plates containing complete Murashige–Skoog (MS) salts at 22°C with a 12-h photoperiod for 4 weeks. One hundred milligrams of roots were ground in 20 mM HEPES-KOH buffer (pH 7.5) supplemented with 13% (w/v) sucrose, 1 mM dithiothreitol, 1 mM EDTA and complete protease inhibitor cocktail (Roche, http://www.roche.com/).

In vitro myrosinase assays

Enzymatic assays of purified PYK10 with 4MUG (Sigma-Aldrich, http://www.sigmaaldrich.com/) and I3G were performed essentially as described previously (Bednarek *et al.*, 2009). Prior to the enzymatic assays, the amount of protein in the eluted fractions was measured using the Bradford protein assay (Bio-Rad, http://www.bio-rad.com/). The 100-μl reaction mixture contained either 25 μg of recombinant protein or 3 μg of root-extracted proteins. Kinetic analysis was performed using various concentrations of substrate (10, 20, 50, 100, 150, 250, 500, 1000, 1500 and 2000 μM), and fitting to the Michaelis–Menten formula was done using the weighted-fitting function of gnuplot v.4.2 software (http://www.gnuplot.info/).

Extraction of secondary metabolites and targeted HPLC analysis

For mutant analysis plants were grown on vertical agar media as described above. Roots of 4-week-old plants were sprayed with

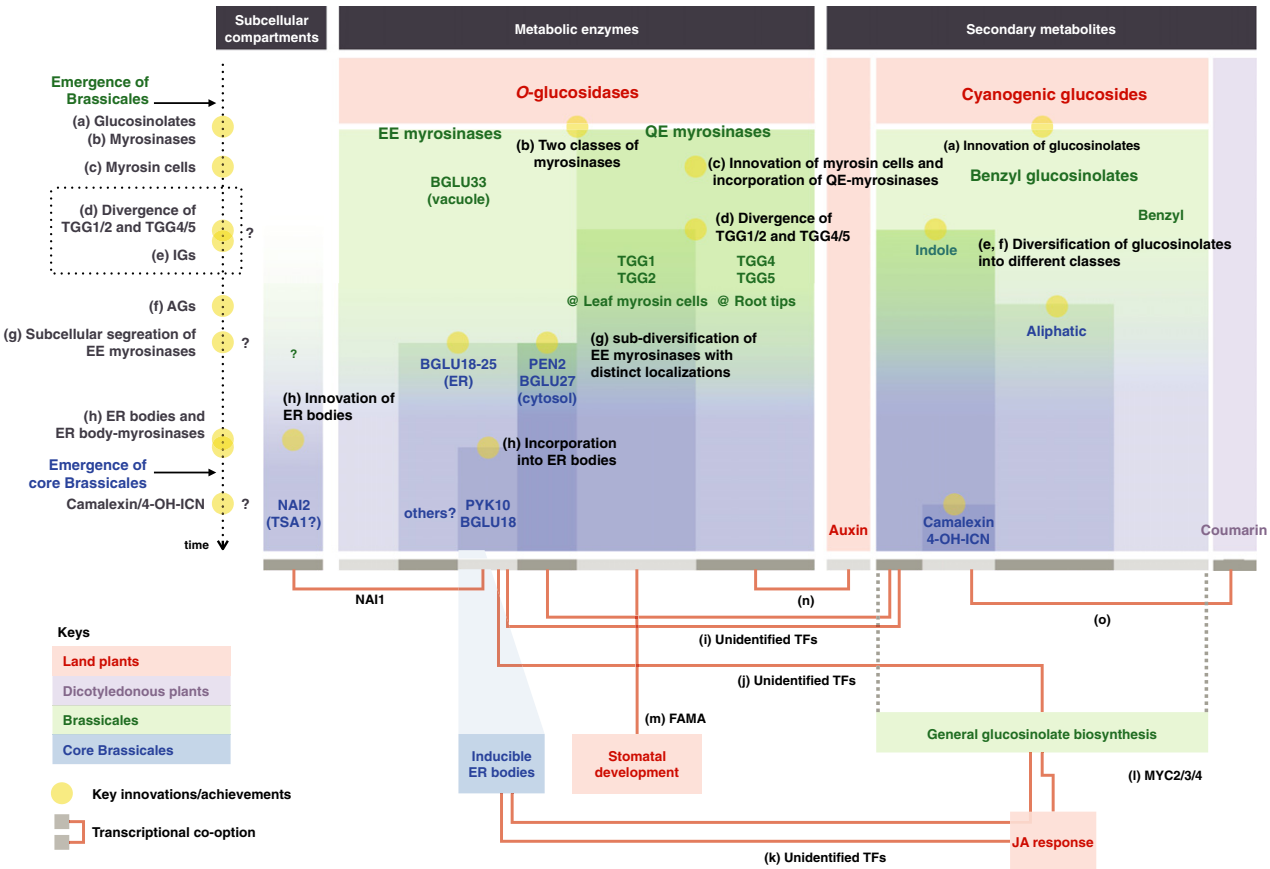


Figure 6. Proposed evolutionary trajectories of endoplasmic reticulum (ER) bodies, myrosinases, glucosinolates, and co-option among them with top-to-bottom axis of time.
Color keys are as indicated. Yellow circles indicate key innovations as follows. The ancestral Brassicales species initiated to produce benzyl glucosinolates (a). Concomitant with, or following this, two classes of myrosinases were acquired independently (b) and only QE myrosinases were incorporated into myrosin cells (c). QE myrosinases were subsequently diversified into TGG1/-2 and TGG4/-5 subclasses (d). During differentiation of the Brassicales, indole glucosinolates (IGs) and aliphatic glucosinolates (AGs) were sequentially acquired (e, f). The ancestral EE myrosinases localized to vacuoles, whereas BGLU18–BGLU27 acquired different subcellular localizations (BGLU18–BGLU25 in the ER and BGLU26/-27 in the cytosol) (g). The EE myrosinases with ER retention signals were incorporated into ER bodies when the ancestral core Brassicales species innovated ER bodies by acquiring NAI2 (h). Red connections indicate transcriptional co-option between innovations in *Arabidopsis thaliana* as follows. The ER body system and PEN2 pathway were transcriptionally integrated with IG modification pathways (i). The general glucosinolate biosynthetic pathway is further co-opted by the ER body system, including inducible ER bodies in leaves (j). Glucosinolate biosynthesis has also been integrated with the MYC2/-3/-4-mediated jasmonic acid (JA) responsive pathway (k) and the differentiation of foliar myrosin cells requires the FAMA transcription factor that is known to be a master regulator of stomatal development (l). TGG4/-5 and auxin biosynthesis (m), as well as camalexin/4-OH-ICN and coumarin pathways (n) appear to have co-opted each other. It remains unclear when sub-divergence within thioglucoside glucohydrolases (TGGs), within EE myrosinases and within glucosinolate occurred and when ER bodies and camalexin/4-OH-ICN were innovated.

250 μM MeJA (Sigma-Aldrich) or sterilized water as a mock control and incubated for 24 h. For phylogenetic analysis the plants were grown in Jiffy-7 Pellets (44 mm diameter, Jiffy, http://www.jiffy-group.com/) in a growth chamber at 20–23°C with a 10-h photoperiod. Adhering soil was washed with distilled water after being collected gently from the Jiffy Pellets. Collected root samples homogenized in DMSO (50 μl/20 mg fresh weight) were subjected to HPLC on an Agilent (http://www.agilent.com/) 1200 HPLC system equipped with diode array (DAD) and fluorescence (FLD) detectors. Samples were analyzed on an Atlantis T3 C18 column (150 × 2.1 mm, 3 μm; Waters, http://www.waters.com/) and a Zorbax SB-Aq column (150 × 2.1 mm, 3.5 μm; Agilent). Detection and quantification of IG was performed as described previously (Bednarek *et al.*, 2011). Scopolin was detected in comparison with a standard isolated from *A. thaliana* roots and was quantified based on fluorescence chromatograms (excitation 340 nm; emission 410 nm) (Bednarek *et al.*, 2005).

In silico co-expression analysis

The non-targeted co-expression analysis was performed with NetworkDrawer tool integrated in the ATTED-II database (v7.1; http://atted.jp/top_draw.shtml#NetworkDrawer) using the *PYK10* gene (*At3g09260*) as a query. The co-expression clustering of target genes were performed using R software v.3.1.2 (http://www.r-project.org/). The list of genes was prepared as described in Appendix S1. Hierarchical clustering was conducted via the average-linkage method based on $\log_2$-transformed MRs. The silhouette plotting method was applied to determine cluster number. Detailed protocols are available in Appendix S1.

Comparative phylogenomics

Detailed protocols for the protein search and phylogenetic analysis can be found in Appendix S1. Briefly, the protein search on the

Phytozome database (https://phytozome.jgi.doe.gov) was conducted either by protein Blast search or keyword search using Pfam IDs. Protein sequences were aligned by the MAFFT (multiple alignment using fast Fourier transform) algorithm (v.7.164; http://mafft.cbrc.jp/alignment/server/) (Katoh and Standley, 2013) with default parameters. Sequence sets were refined by the MaxAlign algorithm to exclude fragmentary sequences (Gouveia-Oliveira *et al.*, 2007). Phylogenetic relationships among β-glucosidases across Brassicales and Malvales species were estimated with a ML method (RaxML v.8.1.11) (Stamatakis, 2006), using the GTR+G substitution model following sequence alignment using MAFFT. Node support was estimated with bootstrap analysis with 1000 replicates. The phylogenetic tree was processed and modified by Dendroscope 3 software (v.3.2.9; http://dendroscope.org/) (Huson *et al.*, 2007; Huson and Scornavacca, 2012) and the Interactive Tree of Life (iTOL) web tool (http://itol.embl.de) (Letunic and Bork, 2016).

Molecular 3D modeling

Threading was carried out using HH-suite 2.0 (Soding, 2005; Remmert *et al.*, 2012) with default settings and the pdb70 template database. Homology models for three templates (Protein Data Bank entries 2dga, 1v02 and 1e4 m) were rendered by fragment assembly using Spanner (Lis *et al.*, 2011). The resulting 3D structures were aligned and processed with Protean 3D (v.11.0.0; DNASTAR).

Microscopic analysis

Transient expression assay using a particle bombardment system (Bio-Rad) was performed as described previously (Matsushima *et al.*, 2004) using 2- to 3-week-old seedlings of *Arabis alpina*, *Crucihimalaya lasiocarpa*, *Capsella rubella*, and *O. pumila* grown on vertical agar (1% w/v) plates containing complete MS salts with 1% (w/v) sucrose at 22°C under continuous light. Surface-sterilized seeds of *A. thaliana* and *Cardamine hirsuta* were sown into liquid media containing half-strength MS salts with 1% sucrose, and grown at 22°C under continuous light for 1–2 weeks without agitation, followed by staining with 1 μM ER-Tracker Green (Molecular Probes, Thermo Fisher Scientific, https://www.thermofisher.com/de/de/home/brands/molecular-probes.html) for 2 h in the dark. Root samples were observed with a laser scanning confocal microscope (Zeiss LSM780, Carl Zeiss, http://www.zeiss.com/) equipped with a 488-nm 40-mW Ar/Kr laser and a 100 × 1.45 numerical aperture oil-immersion objective (alpha Plan-Fluar, 000000-1084-514, Carl Zeiss). Images were analyzed and processed with ImageJ (https://imagej.nih.gov/ij/). For TEM, *T. salsuginea* was grown on vermiculite for 4–5 weeks, *Capsella rubella* and *S. irio* were grown in the liquid MS media for 2–3 weeks, and *Arabis alpina* and *Cardamine hirsuta* were grown on the vertical agar MS media for 3–4 weeks. The roots were collected and fixed for 2 h with 4% (w/v) paraformaldehyde and 1% (v/v) glutaraldehyde in 0.05 M cacodylate buffer, pH 7.4. After following the procedures described previously (Shimada *et al.*, 2003), the ultrathin sections were stained with 2% uranyl acetate and examined with a transmission electron microscope (JEM-1011; JEOL, http://www.jeol.co.jp/en/) at 100 kV. Images were taken by a CCD camera (VELETA; Olympus, http://www.olympus.com/) and processed with ImageJ.

Accession numbers

Sequence data used in Figure 1(e) can be found in the Arabidopsis Genome Initiative (for *A. thaliana*), EMBL/GenBank (for *T. repens*) or UniPlot (for *S. alba*) data libraries under the following accession numbers: *PYK10* (*At3g09260*), *PEN2* (*At2g44490*), *TGG4* (*At1g47600*), *BGLU42* (*At5g36890*), *S. alba* myrosinase (P29736), and *T. repens* CBG (ABV54720). The accession numbers for the genes used in co-expression analysis (Figures 2 and S6–S9) can be found in Data S1 and S2. The accession numbers for the proteins used in phylogenomic analysis (Figures 4 and 5) can be found in Appendix S1.

ACKNOWLEDGEMENTS

We are grateful to Dr Danon M. Standley (Osaka University, Japan) for helping with 3D modeling of BGLUs, to Dr Yasuko Koumoto (Kyoto University, Japan) for assisting with TEM analysis and to the Arabidopsis Biological Resource Center for providing the T-DNA tagged lines of *A. thaliana*. Probe-wise co-expression data were kindly provided by Dr Takeshi Obayashi (Tohoku University, Japan) and the ATTED-II management team. This work was supported by a 'Specially Promoted Research' Grant-in-Aid for Scientific Research to IH-N (no. 22000014) from Japan Society for the Promotion of Science (JSPS), by Grants-in-Aid for Scientific Research to I.H-.N. (no. 15H05776) from JSPS, by an EMBO Installation Grant and Polish Ministry of Science and Higher Education (KNOW program) to PB, by the funds to PS-L from the Max Planck Society and the 'Cluster of Excellence on Plant Sciences' program funded by the Deutsche Forschungsgemeinschaft, by a Grant-in-Aid for JSPS Fellows to RTN (no. 21001585) from JSPS, NSF PGRP 1202793 to PPE and in part by the Grant for Excellent Graduate Schools Program by MEXT, Japan, to IH-N. The authors declare no conflict of interest.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Glucosinolates and endoplasmic reticulum bodies are lineage-specific innovations within Brassicales.

Figure S2. Purification and kinetic analysis of PYK10.

Figure S3. Targeted HPLC analysis of indole glucosinolate accumulation in roots of *nai1*, *pyk10*, *bglu21*, and *pyk10 bglu21*.

Figure S4. Unbiased metabolite profiling of wild-type and *pyk10-1* seedlings.

Figure S5. Untargeted co-expressed gene network surrounding *PYK10*.

Figure S6. Silhouette plotting associated with the clustering shown in Figure 2.

Figure S7. Heatmaps of co-expression clustering in specific conditions.

Figure S8. Developmental pattern of gene expression is not the driving force of co-expression between the endoplasmic reticulum body system and glucosinolate metabolism.

Figure S9. Boxplots of mutual ranks calculated from condition-wise co-expression analysis.

Figure S10. Neighbor-joining trees of CYP79s, CYP81s, SUR1s and F6'H1s with branch length and gene IDs, associated with Figure 4.

Figure S11. Basic residues at positions +6/+7 are conserved among BGLU subfamilies 3–6.

Figure S12. The basic residues at positions +0/+4/+6/+7 are exposed to the substrate-binding pocket.

Table S1. Kinetics of PYK10 enzymatic activity in comparison with PEN2.

Table S2. A list of representative genes for each pathway used in co-expression analysis.

Table S3. List and details of experiments included in the ATTED-II dataset Ath-m c4.1.

Table S4. Genes used for category-wise mutual rank plotting in Figure S9.

Data S1. Co-expression clustering results of whole dataset as a gene table, associated with Figure 2.

Data S2. Co-expression clustering results of partial datasets as a gene table, associated with Figures S7–S9.

Appendix S1. Supplemental texts, methods and references.

REFERENCES

Ahn, Y.O., Shimizu, B.-I., Sakata, K., Gantulga, D., Zhou, Z., Bevan, D.R. and Esen, A. (2010) Scopolin-hydrolyzing β-glucosidases in roots of *Arabidopsis*. *Plant Cell Physiol.* **51**, 132–143.

Albaser, A., Kazana, E., Bennett, M., Cebeci, F., Luang-In, V., Spanu, P.D. and Rossiter, J.T. (2016) Discovery of a bacterial glycoside hydrolase family 3 (GH3) β-glucosidase with myrosinase activity from a Citrobacter strain isolated from soil. *J. Agric. Food Chem.* **64**, 20–27.

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

Bednarek, P. and Osbourn, A. (2009) Plant-microbe interactions: chemical diversity in plant defense. *Science*, **324**, 746–748.

Bednarek, P., Schneider, B., Svatos, A., Oldham, N.J. and Hahlbrock, K. (2005) Structural complexity, differential response to infection, and tissue specificity of indolic and phenylpropanoid secondary metabolism in *Arabidopsis* roots. *Plant Physiol.* **138**, 1058–1070.

Bednarek, P., Pislewska-Bednarek, M., Svatos, A. *et al.* (2009) A glucosinolate metabolism pathway in living plant cells mediates broad-spectrum antifungal defense. *Science*, **323**, 101–106.

Bednarek, P., Piślewska-Bednarek, M., Ver Loren van Themaat, E., Maddula, R.K., Svatoš, A. and Schulze-Lefert, P. (2011) Conservation and clade-specific diversification of pathogen-inducible tryptophan and indole glucosinolate metabolism in *Arabidopsis thaliana* relatives. *New Phytol.* **192**, 713–726.

Behnke, H.D. and Eschlbeck, G. (1978) Dilated cisternae in Capparales – an attempt towards the characterization of a specific endoplasmic reticulum. *Protoplasma*, **97**, 351–363.

Beran, F., Pauchet, Y., Kunert, G. *et al.* (2014) Phyllotreta striolata flea beetles use host plant defense compounds to create their own glucosinolate-myrosinase system. *Proc. Natl Acad. Sci. USA*, **111**, 7349–7354.

Bonnett, H.T. and Newcomb, E.H. (1965) Polyribosomes and cisternal accumulations in root cells of radish. *J. Cell Biol.* **27**, 423–432.

Bridges, M., Jones, A.M., Bones, A.M., Hodgson, C., Cole, R., Bartlet, E., Wallsgrove, R., Karapapa, V.K., Watts, N. and Rossiter, J.T. (2002) Spatial organization of the glucosinolate-myrosinase system in brassica specialist aphids is similar to that of the host plant. *Proc. Biol. Sci.* **269**, 187–191.

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

Burmeister, W.P., Cottaz, S., Driguez, H., Iori, R., Palmieri, S. and Henrissat, B. (1997) The crystal structures of *Sinapis alba* myrosinase and a covalent glycosyl enzyme intermediate provide insights into the substrate recognition and active-site machinery of an *S*-glycosidase. *Structure*, **5**, 663–676.

Burow, M., Zhang, Z.Y., Ober, J.A., Lambrix, V.M., Wittstock, U., Gershenzon, J. and Kliebenstein, D.J. (2008) ESP and ESM1 mediate indol-3-acetonitrile production from indol-3-ylmethyl glucosinolate in Arabidopsis. *Phytochemistry*, **69**, 663–671.

Burow, M., Losansky, A., Muller, R., Plock, A., Kliebenstein, D.J. and Wittstock, U. (2009) The genetic basis of constitutive and herbivore-induced ESP-independent nitrile formation in Arabidopsis. *Plant Physiol.* **149**, 561–574.

Buxdorf, K., Yaffe, H., Barda, O. and Levy, M. (2013) The effects of glucosinolates and their breakdown products on necrotrophic fungi. *PLoS ONE*, **8**, e70771.

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.

Chong, J., Baltz, R., Schmitt, C., Beffa, R., Fritig, B. and Saindrenan, P. (2002) Downregulation of a pathogen-responsive tobacco UDP-Glc:phenylpropanoid glucosyltransferase reduces scopoletin glucoside accumulation, enhances oxidative stress, and weakens virus resistance. *Plant Cell*, **14**, 1093–1107.

Churngchow, N. and Rattarasarn, M. (2001) Biosynthesis of scopoletin in *Hevea brasiliensis* leaves inoculated with *Phytophthora palmivora*. *J. Plant Physiol.* **158**, 875–882.

Clay, N.K., Adio, A.M., Denoux, C., Jander, G. and Ausubel, F.M. (2009) Glucosinolate metabolites required for an *Arabidopsis* innate immune response. *Science*, **323**, 95–101.

Couvreur, T.L.P., Franzke, A., Al-Shehbaz, I.A., Bakker, F.T., Koch, M.A. and Mummenhoff, K. (2010) Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae). *Mol. Biol. Evol.* **27**, 55–71.

Cresti, M., Pacini, E. and Simoncioli, C. (1974) Uncommon paracrystalline structures formed in the endoplasmic reticulum of the integumentary cells of *Diplotaxis erucoides* ovules. *J. Ultrastruct. Res.* **49**, 219–223.

Dohi, K., Nishikiori, M., Tamai, A., Ishikawa, M., Meshi, T. and Mori, M. (2006) Inducible virus-mediated expression of a foreign protein in suspension-cultured plant cells. *Arch. Virol.* **151**, 1075–1084.

Dohi, K., Tamai, A. and Mori, M. (2008) Insertion in the coding region of the movement protein improves stability of the plasmid encoding a tomato mosaic virus-based expression vector. *Arch. Virol.* **153**, 1667–1675.

Endress, A.G. and Sjolund, R.D. (1976) Ultrastructural cytology of callus cultures of *Streptanthus tortuosus* as affected by temperature. *Am. J. Bot.* **63**, 1213–1224.

Fourcroy, P., Siso-Terraza, P., Sudre, D., Saviron, M., Reyt, G., Gaymard, F., Abadia, A., Abadia, J., Alvarez-Fernandez, A. and Briat, J.F. (2014) Involvement of the ABCG37 transporter in secretion of scopoletin and derivatives by Arabidopsis roots in response to iron deficiency. *New Phytol.* **201**, 155–167.

Frerigmann, H. and Gigolashvili, T. (2014) MYB34, MYB51, and MYB122 distinctly regulate indolic glucosinolate biosynthesis in *Arabidopsis thaliana*. *Mol. Plant*, **7**, 814–828.

Frerigmann, H., Berger, B. and Gigolashvili, T. (2014) bHLH05 is an interaction partner of MYB51 and a novel regulator of glucosinolate biosynthesis in Arabidopsis. *Plant Physiol.* **166**, 349–369.

Frerigmann, H., Pislewska-Bednarek, M., Sanchez-Vallet, A., Molina, A., Glawischnig, E., Gigolashvili, T. and Bednarek, P. (2016) Regulation of pathogen-triggered tryptophan metabolism in *Arabidopsis thaliana* by MYB transcription factors and indole glucosinolate conversion products. *Mol. Plant*, **9**, 682–695.

Fu, L., Wang, M., Han, B., Tan, D., Sun, X. and Zhang, J. (2016) Arabidopsis myrosinase genes AtTGG4 and AtTGG5 Are root-tip specific and contribute to auxin biosynthesis and root-growth regulation. *Int. J. Mol. Sci.* **17**, 892.

Fuchs, R., Kopischke, M., Klapprodt, C., Hause, G., Meyer, A.J., Schwarzlander, M., Fricker, M.D. and Lipka, V. (2016) Immobilized subpopulations of leaf epidermal mitochondria mediate PEN2-dependent pathogen entry control in Arabidopsis. *Plant Cell.* **28**, 130–145.

Gailhofer, M., Thaler, I. and Rücker, W. (1979) Dilated ER in callus cells and in cells from Armoracia plants cultivated in vitro. *Protoplasma*, **98**, 263–274.

Gouveia-Oliveira, R., Sackett, P.W. and Pedersen, A.G. (2007) MaxAlign: maximizing usable data in an alignment. *BMC Bioinformatics*, **8**, 312, doi: 10.1186/1471-2105-8-312.

Gutierrez-Mellado, M.C., Edwards, R., Tena, M., Cabello, F., Serghini, K. and Jorrin, J. (1996) The production of coumarin phytoalexins in different plant organs of sunflower (*Helianthus annuus* L.). *J. Plant Physiol.* **149**, 261–266.

Hara-Nishimura, I. and Matsushima, R. (2003) A wound-inducible organelle derived from endoplasmic reticulum: a plant strategy against environmental stresses? *Curr. Opin. Plant Biol.* **6**, 583–588.

Hayashi, Y., Yamada, K., Shimada, T., Matsushima, R., Nishizawa, N., Nishimura, M. and Hara-Nishimura, I. (2001) A proteinase-storing body that prepares for cell death or stresses in the epidermal cells of Arabidopsis. *Plant Cell Physiol.* **42**, 894–899.

Hirai, M.Y., Sugiyama, K., Sawada, Y. *et al.* (2007) Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc. Natl Acad. Sci. USA*, **104**, 6478–6483.

Hiruma, K., Gerlach, N., Sacristan, S. *et al.* (2016) Root endophyte colletotrichum tofieldiae confers plant fitness benefits that are phosphate status dependent. *Cell*, **165**, 464–474.

Hoefert, L.L. (1975) Tubules in dilated cisternae of endoplasmic reticulum of *Thlaspi arvense* (Cruciferae). *Am. J. Bot.* **62**, 756–760.

Huson, D.H. and Scornavacca, C. (2012) Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst. Biol.* **61**, 1061–1067.

Huson, D.H., Richter, D.C., Rausch, C., Dezulian, T., Franz, M. and Rupp, R. (2007) Dendroscope: an interactive viewer for large phylogenetic trees. *BMC Bioinformatics*, **8**, 460.

Iven, T., Konig, S., Singh, S., Braus-Stromeyer, S.A., Bischoff, M., Tietze, L.F., Braus, G.H., Lipka, V., Feussner, I. and Droge-Laser, W. (2012) Transcriptional activation and production of tryptophan-derived secondary metabolites in arabidopsis roots contributes to the defense against the fungal vascular pathogen Verticillium longisporum. *Mol. Plant*, **5**, 1389–1402.

Iversen, T.-H. (1970a) Cytochemical localization of myrosinase (β-thioglucosidase) in root tips of *Sinapis alba*. *Protoplasma*, **71**, 451–466.

Iversen, T.-H. (1970b) The morphology, occurrence, and distribution of dilated cisternae of the endoplasmic reticulum in tissues of plants of the Cruciferae. *Protoplasma*, **71**, 467–477.

Iversen, T.-H. and Flood, P.R. (1969) Rod-shaped accumulations in cisternae of the endoplasmic reticulum in root cells of *Lepidium sativum* seedlings. *Planta*, **86**, 295–298.

Jander, G., Cui, J., Nhan, B., Pierce, N.E. and Ausubel, F.M. (2001) The TASTY locus on chromosome 1 of Arabidopsis affects feeding of the insect herbivore Trichoplusia ni. *Plant Physiol.* **126**, 890–898.

Jørgensen, L.B. (1981) Myrosin cells and dilated cisternae of the endoplasmic reticulum in the order Capparales. *Nord. J. Bot.* **1**, 433–445.

Jørgensen, L.B., Behnke, H.D. and Mabry, T.J. (1977) Protein-accumulating cells and dilated cisternae of the endoplasmic reticulum in three glucosinolate-containing genera: Armoracia, Capparis, Drypetes. *Planta*, **137**, 215–224.

Katoh, K. and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780.

Kazan, K. and Manners, J.M. (2013) MYC2: the master in action. *Mol. Plant*, **6**, 686–703.

Kissen, R. and Bones, A.M. (2009) Nitrile-specifier proteins involved in glucosinolate hydrolysis in *Arabidopsis thaliana*. *J. Biol. Chem.* **284**, 12057–12070.

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

Koroleva, O.A., Gibson, T.M., Cramer, R. and Stain, C. (2010) Glucosinolate-accumulating S-cells in Arabidopsis leaves and flower stalks undergo programmed cell death at early stages of differentiation. *Plant J.* **64**, 456–469.

Lahrmann, U., Strehmel, N., Langen, G., Frerigmann, H., Leson, L., Ding, Y., Scheel, D., Herklotz, S., Hilbert, M. and Zuccaro, A. (2015) Mutualistic root endophytism is not associated with the reduction of saprotrophic traits and requires a noncompromised plant innate immunity. *New Phytol.* **207**, 841–857.

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

Letunic, I. and Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245.

Lipka, V., Dittgen, J., Bednarek, P. *et al.* (2005) Pre- and postinvasion defenses both contribute to nonhost resistance in *Arabidopsis*. *Science*, **310**, 1180–1183.

Lis, M., Kim, T., Sarmiento, J., Kuroda, D., Dinh, H., Kinjo, A.R., Amada, K., Devadas, S., Nakamura, H. and Standley, D.M. (2011) Bridging the gap between single-template and fragment based protein structure modeling using Spanner. *Immunome Res.* **7**, 1–8.

Lu, X., Dittgen, J., Pislewska-Bednarek, M. *et al.* (2015) Mutant allele-specific uncoupling of PENETRATION3 functions reveals engagement of the ATP-binding cassette transporter in distinct tryptophan metabolic pathways. *Plant Physiol.* **168**, 814–827.

Matsushima, R., Hayashi, Y., Kondo, M., Shimada, T., Nishimura, M. and Hara-Nishimura, I. (2002) An endoplasmic reticulum-derived structure that is induced under stress conditions in Arabidopsis. *Plant Physiol.* **130**, 1807–1814.

Matsushima, R., Kondo, M., Nishimura, M. and Hara-Nishimura, I. (2003) A novel ER-derived compartment, the ER body, selectively accumulates a β-glucosidase with an ER-retention signal in Arabidopsis. *Plant J.* **33**, 493–502.

Matsushima, R., Fukao, Y., Nishimura, M. and Hara-Nishimura, I. (2004) NAI1 gene encodes a basic-helix-loop-helix–type putative transcription factor that regulates the formation of an endoplasmic reticulum-derived structure, the ER body. *Plant Cell*, **16**, 1536–1549.

Mikkelsen, M.D., Petersen, B.L., Glawischnig, E., Jensen, A.B., Andreasson, E. and Halkier, B.A. (2003) Modulation of CYP79 genes and glucosinolate profiles in Arabidopsis by defense signaling pathways. *Plant Physiol.* **131**, 298–308.

Mithen, R., Bennett, R. and Marquez, J. (2010) Glucosinolate biochemical diversity and innovation in the Brassicales. *Phytochemistry*, **71**, 2074–2086.

Morant, A.V., Jørgensen, K., Jørgensen, C., Paquette, S.M., Sánchez-Pérez, R., Møller, B.L. and Bak, S. (2008) β-Glucosidases as detonators of plant chemical defense. *Phytochemistry*, **69**, 1795–1813.

Moussaieff, A., Rogachev, I., Brodsky, L., Malitsky, S., Toal, T.W., Belcher, H., Yativ, M., Brady, S.M., Benfey, P.N. and Aharoni, A. (2013) High-resolution metabolic mapping of cell types in plant roots. *Proc. Natl Acad. Sci. USA*, **110**, E1232–E1241.

Muller, T.M., Bottcher, C., Morbitzer, R., Gotz, C.C., Lehmann, J., Lahaye, T. and Glawischnig, E. (2015) TRANSCRIPTION ACTIVATOR-LIKE EFFECTOR NUCLEASE-mediated generation and metabolic analysis of camalexin-deficient cyp71a12 cyp71a13 double knockout lines. *Plant Physiol.* **168**, 849–858.

Nafisi, M., Goregaoker, S., Botanga, C.J., Glawischnig, E., Olsen, C.E., Halkier, B.A. and Glazebrook, J. (2007) Arabidopsis cytochrome P450 monooxygenase 71A13 catalyzes the conversion of indole-3-acetaldoxime in camalexin synthesis. *Plant Cell*, **19**, 2039–2052.

Nagano, A.J., Fukao, Y., Fujiwara, M., Nishimura, M. and Hara-Nishimura, I. (2008) Antagonistic jacalin-related lectins regulate the size of ER body-type β-glucosidase complexes in *Arabidopsis thaliana*. *Plant Cell Physiol.* **49**, 969–980.

Nagano, A.J., Maekawa, A., Nakano, R.T., Miyahara, M., Higaki, T., Kutsuna, N., Hasezawa, S. and Hara-Nishimura, I. (2009) Quantitative analysis of ER body morphology in an *Arabidopsis* mutant. *Plant Cell Physiol.* **50**, 2015–2022.

Nakano, R.T., Matsushima, R., Nagano, A.J., Fukao, Y., Fujiwara, M., Kondo, M., Nishimura, M. and Hara-Nishimura, I. (2012) ERMO3/MVP1/GOLD36 is involved in a cell type-specific mechanism for maintaining ER morphology in *Arabidopsis thaliana*. *PLoS ONE*, **7**, e49103.

Nakano, R.T., Yamada, K., Bednarek, P., Nishimura, M. and Hara-Nishimura, I. (2014) ER bodies in plants of the Brassicales order: biogenesis and association with innate immunity. *Front. Plant Sci.* **5**, 73, doi: 10.3389/fpls.2014.00073.

Nitz, I., Berkefeld, H., Puzio, P.S. and Grundler, F.M.W. (2001) *Pyk10*, a seedling and root specific gene and promoter from *Arabidopsis thaliana*. *Plant Sci.* **161**, 337–346.

Nong, H., Zhang, J.M., Li, D.Q., Wang, M., Sun, X.P., Zhu, Y.J., Meijer, J. and Wang, Q.H. (2010) Characterization of a novel beta-thioglucosidase CpTGG1 in Carica papaya and its substrate-dependent and ascorbic acid-independent O-beta-glucosidase activity. *J. Integr. Plant Biol.* **52**, 879–890.

Obayashi, T. and Kinoshita, K. (2009) Rank of correlation coefficient as a comparable measure for biological significance of gene coexpression. *DNA Res.* **16**, 249–260.

Obayashi, T., Kinoshita, K., Nakai, K., Shibaoka, M., Hayashi, S., Saeki, M., Shibata, D., Saito, K. and 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.

Obayashi, T., Nishida, K., Kasahara, K. and Kinoshita, K. (2011) ATTED-II updates: condition-specific gene coexpression to extend coexpression

analyses and applications to a broad range of flowering plants. *Plant Cell Physiol.* **52**, 213–219.

Ogasawara, K., Yamada, K., Christeller, J.T., Kondo, M., Hatsugai, N., Hara-Nishimura, I. and Nishimura, M. (2009) Constitutive and inducible ER bodies of *Arabidopsis thaliana* accumulate distinct β-glucosidases. *Plant Cell Physiol.* **50**, 480–488.

Pfalz, M., Vogel, H. and 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.

Pfalz, M., Mikkelsen, M.D., Bednarek, P., Olsen, C.E., Halkier, B.A. and Kroymann, J. (2011) Metabolic engineering in Nicotiana benthamiana reveals key enzyme functions in Arabidopsis indole glucosinolate modification. *Plant Cell*, **23**, 716–729.

Pozo, M.J., Van Der Ent, S., Van Loon, L.C. and Pieterse, C.M.J. (2008) Transcription factor MYC2 is involved in priming for enhanced defense during rhizobacteria-induced systemic resistance in *Arabidopsis thaliana*. *New Phytol.* **180**, 511–523.

Rajniak, J., Barco, B., Clay, N.K. and Sattely, E.S. (2015) A new cyanogenic metabolite in Arabidopsis required for inducible pathogen defence. *Nature*, **525**, 376–379.

Ran, J.H., Shen, T.T., Liu, W.J. and Wang, X.Q. (2013) Evolution of the bHLH genes involved in stomatal development: implications for the expansion of developmental complexity of stomata in land plants. *PLoS ONE*, **8**, e78997.

Rask, L., Andreasson, E., Ekbom, B., Eriksson, S., Pontoppidan, B. and Meijer, J. (2000) Myrosinase: gene family evolution and herbivore defense in Brassicaceae. *Plant Mol. Biol.* **42**, 93–113.

Remmert, M., Biegert, A., Hauser, A. and Soding, J. (2012) HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods*, **9**, 173–175.

Rousseeuw, P.J. (1987) Silhouettes – a graphical aid to the interpretation and validation of cluster-analysis. *J. Comput. Appl. Math.* **20**, 53–65.

Saito, K. and Matsuda, F. (2010) Metabolomics for functional genomics, systems biology, and biotechnology. *Annu. Rev. Plant Biol.* **61**, 463–489.

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

Sanchez-Vallet, A., Ramos, B., Bednarek, P., Lopez, G., Pislewska-Bednarek, M., Schulze-Lefert, P. and Molina, A. (2010) Tryptophan-derived secondary metabolites in *Arabidopsis thaliana* confer non-host resistance to necrotrophic Plectosphaerella cucumerina fungi. *Plant J.* **63**, 115–127.

Schmid, N.B., Giehl, R.F., Doll, S., Mock, H.P., Strehmel, N., Scheel, D., Kong, X., Hider, R.C. and von Wiren, N. (2014) Feruloyl-CoA 6′-Hydroxylase1-dependent coumarins mediate iron acquisition from alkaline substrates in arabidopsis. *Plant Physiol.* **164**, 160–172.

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.

Shen, J., Zeng, Y., Zhuang, X., Sun, L., Yao, X., Pimpl, P. and Jiang, L. (2013) Organelle pH in the Arabidopsis endomembrane system. *Mol. Plant*, **6**, 1419–1437.

Sherameti, I., Venus, Y., Drzewiecki, C., Tripathi, S., Dan, V.M., Nitz, I., Varma, A., Grundler, F.M. and Oelmüller, R. (2008) PYK10, a β-glucosidase located in the endoplasmatic reticulum, is crucial for the beneficial interaction between *Arabidopsis thaliana* and the endophytic fungus *Piriformospora indica*. *Plant J.* **54**, 428–439.

Shimada, T., Yamada, K., Kataoka, M. *et al.* (2003) Vacuolar processing enzymes are essential for proper processing of seed storage proteins in *Arabidopsis thaliana*. *J. Biol. Chem.* **278**, 32292–32299.

Shirakawa, M., Ueda, H., Nagano, A.J., Shimada, T., Kohchi, T. and Hara-Nishimura, I. (2014) FAMA is an essential component for the differentiation of two distinct cell types, myrosin cells and guard cells, in Arabidopsis. *Plant Cell*, **26**, 4039–4052.

Shirakawa, M., Ueda, M., Shimada, T. and Hara-Nishimura, I. (2016) FAMA: a molecular link between stomata and myrosin cells. *Trends Plant Sci.* **21**, 861–871.

Shirley, B.W., Kubasek, W.L., Storz, G., Bruggemann, E., Koornneef, M., Ausubel, F.M. and Goodman, H.M. (1995) Analysis of Arabidopsis mutants deficient in flavonoid biosynthesis. *Plant J.* **8**, 659–671.

Siwinska, J., Kadzinski, L., Banasiuk, R., Gwizdek-Wisniewska, A., Olry, A., Banecki, B., Lojkowska, E. and Ihnatowicz, A. (2014) Identification of QTLs affecting scopolin and scopoletin biosynthesis in *Arabidopsis thaliana*. *BMC Plant Biol.* **14**, 280.

Soding, J. (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics*, **21**, 951–960.

Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–2690.

Ueda, H., Nishiyama, C., Shimada, T., Koumoto, Y., Hayashi, Y., Kondo, M., Takahashi, T., Ohtomo, I., Nishimura, M. and Hara-Nishimura, I. (2006) AtVAM3 is required for normal specification of idioblasts, myrosin cells. *Plant Cell Physiol.* **47**, 164–175.

Vanholme, R., Demedts, B., Morreel, K., Ralph, J. and Boerjan, W. (2010) Lignin biosynthesis and structure. *Plant Physiol.* **153**, 895–905.

Weng, J.-K., Philippe, R.N. and Noel, J.P. (2012) The rise of chemodiversity in plants. *Science*, **336**, 1667–1670.

Wentzell, A.M. and Kliebenstein, D.J. (2008) Genotype, age, tissue, and environment regulate the structural outcome of glucosinolate activation. *Plant Physiol.* **147**, 415–428.

Wittstock, U. and Burow, M. (2010) Glucosinolate breakdown in Arabidopsis: mechanism, regulation and biological significance. *Arabidopsis Book*, **8**, e0134.

Xu, Z., Escamilla-Trevino, L., Zeng, L. *et al.* (2004) Functional genomic analysis of *Arabidopsis thaliana* glycoside hydrolase family 1. *Plant Mol. Biol.* **55**, 343–367.

Yamada, K., Nagano, A.J., Nishina, M., Hara-Nishimura, I. and Nishimura, M. (2008) NAI2 Is an endoplasmic reticulum body component that enables ER body formation in *Arabidopsis thaliana*. *Plant Cell*, **20**, 2529–2540.

Yamada, K., Hara-Nishimura, I. and Nishimura, M. (2011) Unique defense strategy by the endoplasmic reticulum body in plants. *Plant Cell Physiol.* **52**, 2039–2049.

Yonekura-Sakakibara, K. and Saito, K. (2013) Transcriptome coexpression analysis using ATTED-II for integrated transcriptomic/metabolomic analysis. *Methods Mol. Biol.* **1011**, 317–326.

Zamioudis, C., Hanson, J. and Pieterse, C.M.J. (2014) β-Glucosidase BGLU42 is a MYB72-dependent key regulator of rhizobacteria-induced systemic resistance and modulates iron deficiency responses in Arabidopsis roots. *New Phytol.* **204**, 368–379.

Zhao, Y. (2014) Auxin biosynthesis. *Arabidopsis Book*, **12**, e0173.

Zhou, C., Tokuhisa, J.G., Bevan, D.R. and Esen, A. (2012) Properties of beta-thioglucoside hydrolases (TGG1 and TGG2) from leaves of *Arabidopsis thaliana*. *Plant Sci.* **191–192**, 82–92.