

UC Davis

UC Davis Previously Published Works

Title

The ease and complexity of identifying and using specialized metabolites for crop engineering

Permalink

<https://escholarship.org/uc/item/8dk454kq>

Journal

Emerging Topics in Life Sciences, 6(2)

ISSN

2397-8554

Authors

Muhich, Anna Jo
Agosto-Ramos, Amanda
Kliebenstein, Daniel J

Publication Date

2022-04-15

DOI

10.1042/etls20210248

Peer reviewed

Review Article

The ease and complexity of identifying and using specialized metabolites for crop engineering

Anna Jo Muhich^{1,2}, Amanda Agosto-Ramos^{1,2} and  Daniel J. Kliebenstein^{1,2,3}

¹Department of Plant Science, University of California, Davis, CA, U.S.A.; ²Plant Biology Graduate Group, University of California, Davis, CA, U.S.A.; ³DynaMo Center of Excellence, University of Copenhagen, Thorvaldsensvej 40, DK-1871 Frederiksberg C, Denmark

Correspondence: Daniel J. Kliebenstein (kliebenstein@ucdavis.edu)



Plants produce a broad variety of specialized metabolites with distinct biological activities and potential applications. Despite this potential, most biosynthetic pathways governing specialized metabolite production remain largely unresolved across the plant kingdom. The rapid advancement of genetics and biochemical tools has enhanced our ability to identify plant specialized metabolic pathways. Further advancements in transgenic technology and synthetic biology approaches have extended this to a desire to design new pathways or move existing pathways into new systems to address long-running difficulties in crop systems. This includes improving abiotic and biotic stress resistance, boosting nutritional content, etc. In this review, we assess the potential and limitations for (1) identifying specialized metabolic pathways in plants with multi-omics tools and (2) using these enzymes in synthetic biology or crop engineering. The goal of these topics is to highlight areas of research that may need further investment to enhance the successful application of synthetic biology for exploiting the myriad of specialized metabolic pathways.

Introduction

Plant specialized metabolites are highly diverse, lineage-specific compounds that help plants combat a wide variety of abiotic and biotic stresses. This broad range of activities means that many specialized metabolites are highly useful for medicine, agriculture, bioenergy, and beyond [1]. Transitioning these compounds to successful application is a particularly exciting goal for synthetic biology and bioengineering research. These synthetic/engineering efforts involve a conceptually simple process of finding the metabolite-producing genes and moving them into another plant to convey the metabolites and associated biological activities to the new plant [2]. A boost to this potential is the fact that specialized metabolites are often under changing selective pressures from changing environments drives their diversification [3]. This diversifying evolution of specialized metabolites provides immense raw material to find and create novel compounds. However, this rapid evolution can also complicate the application efforts. In this article, we are going to focus on two particular complications to synthetic/engineering created from specialized metabolite rapid evolution.

The first complication to synthetic/engineering biology created by rapid evolution is that specialized metabolites have narrow taxonomic distributions confounding the ability to find them [4]. Individual specialized metabolites and pathways can be limited to plant families, individual species, or even specific genotypes within a species. Correspondingly, the genes controlling the production of specialized metabolites are also limited in their distribution and not easy or impossible to find via simple homology searches in part due to pangenomic variation both within and between species. In the past decade, simple alternatives to homology-based identification like co-expression and gene clusters have been developed to solve this conundrum. However, as we will cover, there are potential overlooked limitations to these methods that could be key to finding the genes needed to precisely modify a specialized metabolite's function and/or structure.

Received: 22 November 2021
Revised: 20 January 2022
Accepted: 24 January 2022

Version of Record published:
18 March 2022

The second complication that we will focus on is the potential for specialized metabolism to be interconnected with more conserved plant processes. Connections with core plant processes could complicate the ability to efficiently move specialized pathways between species. The idea of specialized metabolite rapid evolution has created a general theoretical assumption that specialized metabolism is largely disconnected from a plant's physiology [5]. This assumption arises from the idea that if specialized metabolism is highly integrated into the plant's core metabolism and physiology, the ability of specialized metabolism to rapidly evolve or change would be constrained [6]. If this hypothesis is true, it is appealing in a synthetic biology/engineering premise, as it suggests that these pathways should be easier to move between plants without requiring significant modification of the plant into which the pathway is being moved. If a pathway functions well in the native plant with minimal connections, then it should function well when transferred to a naive plant. However, recent studies are beginning to show that specialized metabolite pathways intersect with primary metabolism and physiology in unexpected ways [7]. We will cover some of this evidence and the implications it has for the engineering of metabolic pathways into new species.

Integrated transcriptomics and metabolomics for identification of specialized metabolic biosynthetic pathways in plants

The rapid and extensive diversification of specialized metabolites means that most have not been identified and an even larger fraction of the genes required for their synthesis are unidentified [2]. Classical approaches to finding new pathways relied on biochemical and/or genetic characterization of single genes/enzymes at a time to build up the pathway. This one-by-one approach is highly time consuming and will obviously not make much progress in cloning the entirety of plant specialized metabolism. New approaches have been developed to address this bottleneck including the use of machine learning to identify specialized metabolic genes with publicly available omics datasets [8], but this may be limited by the scale of available data [9]. Another approach is looking for gene/enzyme clusters under the assumption that the physical collocation of genes on chromosomes is a predictor for co-expression and/or shared function in a pathway. The cluster approach has identified important specialized metabolic pathways in plants, including DIBOA biosynthesis in maize [10], avenacin biosynthesis in oat [11], steroidal glycoalkaloids in tomato and potato [12], and more. However, it is becoming apparent that most plant specialized metabolite pathways may not be clustered, and the pathway genes are dispersed widely across the genome [13,14]. Expanding our catalogue of cloned specialized metabolites requires powerful tools to find these pathways and their associated compounds.

Easy pathway identification

One solution to find pathways is through transcriptomics, the analysis of the near-entire set of RNA transcripts in an organism at a given time and tissue. Transcriptomics is showing that specialized metabolic pathways can be found via simple co-expression analysis, wherein most genes for a pathway correlate in their expression across a set of samples. This has been especially streamlined with the advent of the next-gen sequencing-based technology, RNA-seq. Gene co-expression analysis for the discovery of metabolic pathways harnesses a key feature of the nature of biosynthetic enzymes in that they have to be co-ordinated for optimal efficiency. This is often driven by co-expression due to cell- or tissue-type specificity shaping the gene's expression patterns [15]. Another possibility is that co-expression may also map to metabolon formation. These enzymes are often organized into spatially localized macromolecular complexes, or metabolons, to channel metabolites through a sequence of functional modifications [16–18]. This ensures that unstable intermediates are not broken down before being converted to the final product, or that intermediates which may be toxic do not damage the plant. The hypothesis is that to efficiently form a metabolon or at least a co-ordinated pathway, the genes encoding enzymes in a metabolon should be co-expressed. Co-expression analysis has identified numerous different biosynthetic pathways producing specialized metabolites in plants, including aliphatic glucosinolates in *Arabidopsis* [19], terpenoids in *Arabidopsis* and rice and benzoxazinoids in maize [13], podophyllotoxin in mayapple [20], and saponins in legumes [21]. More importantly, co-expression can be used in any species with or without an available genome making this a key tool for cataloging plant specialized metabolism.

Applications of multi-omics

An interesting consequence of the technical ease of global co-expression is that it is now possible, if not common, to find putative biosynthetic pathways without having an associated metabolite. This limits our ability to understand what the putative pathway is doing. One solution to this problem is multi-omics, the specific integration of transcriptomics and metabolomics. Multi-omics is a promising avenue for uncovering the metabolites synthesized by newly identified biosynthetic pathways that depend on a tight interconnection in gene-to-metabolite networks. Global gene co-expression data can be paired with metabolomics to determine which genes are co-expressing and which metabolites are accumulating in a correlated fashion under specific conditions (Figure 1). This provides sets of genes with a hypothesized function in formation of a given metabolite. Functional confirmation of candidate genes can be further conducted with reverse genetics or protein functional analysis.

Multi-omics is also easily applied across non-model plants. For example, this approach found a key enzyme for the initial reaction in forming amine moieties in alkaloids in *Echinacea* tissues and organs [22]. Multi-omics also found the key enzymes for kavalactone and flavokavain biosynthesis in the medicinal plant kava [23]. In addition to finding individual enzymes, multi-omics approaches are beginning to be validated by their ability to recover known biosynthetic pathways. Integrated transcriptomics and metabolomics were used in *Sindora glabra* to identify sesquiterpenes in the stems of mature trees before and after treatment with methyl jasmonate. Combining this with the transcriptomics identified known terpenoid backbone biosynthesis genes, two new terpene synthase genes, and several transcription factors [24]. Similarly, UPLC-QTOF-MS and corresponding RNA-seq on different tissues of *Narcissus pseudonarcissus* confirmed all previously proposed genes encoding alkaloid biosynthetic enzymes [25].

Multi-omics analysis also allows for the identification of unknown pathways for biosynthesis of different metabolic classes in previously poorly understood systems. Perez de Souza et al. [26] measured the metabolome of several common bean accessions and assigned chemical structures and classes for ~39% of measured metabolites. Integrating this dataset with transcriptomics generated whole pathway predictions for a wide array of specialized metabolites, including flavonoids, triterpenoid saponins, and hydroxycinnamates. Combined RNA-seq and GC-MS headspace analysis generated a proposed terpenoid biosynthetic pathway in *Curcuma weyujin* resulting in the formation of 87 main terpenoids [27]. Other studies have similarly developed near-complete biosynthetic pathways for colchicine and salidroside [28,29].

Beyond metabolic pathways, multi-omics can also identify regulatory genes in various pathways. Combined metabolomics and transcriptomics in the Chinese medicinal herb *Salvia miltiorrhiza* provided evidence for a specialized metabolic stress response involving 70 transcription factors and eight cytochrome P450s as candidate genes for the production of terpenoid biosynthesis [30]. Metabolomics and transcriptomics in tea leaves across different harvest times showed differential terpenoid volatile accumulation and enabled the construction of a cross-talk regulatory network. This network involved 13 transcription factors possibly regulating terpenoid biosynthesis, while a key terpenoid biosynthetic gene was functionally characterized [31].

Extending multi-omics to include phenotypic data can help to determine the larger systemic consequences of specialized metabolite accumulation patterns and assess their biological roles. For example, combined phenotyping, RNA-seq and volatile and nonvolatile untargeted metabolite analysis was used to predict the genes and metabolites contributing to pepper susceptibility to spider mite infestation, including flavonoid and diterpene glycoside production [32].

Complications of pathway identification

Co-expression and multi-omics approaches to specialized metabolite analysis do have limitations and caveats that are not always fully apparent. The first caveat is that gene co-expression does not typically find an entire biosynthetic pathway. For example, enzymes controlling side-chain modifying processes like hydroxylation or benzoylation in Arabidopsis glucosinolate biosynthesis are encoded by genes that do not co-express with other genes in the pathway [13,33–35] (Figure 1). This lack of co-expression may arise from the possibility that these genes appear to be newly evolved and may have not had the time to coalesce to the core expression pattern. Alternatively, the structural modification may be critical for a new biological activity that is controlled by signaling pathways different from the central pathway [36,37].

Additionally, these studies are showing that co-expression extends beyond the central specialized metabolic pathway to peripheral supporting genes. For example, in glucosinolate co-expression analysis, the sulfur assimilation pathway is frequently identified in addition to the core glucosinolate pathway [13]. Because sulfur is a

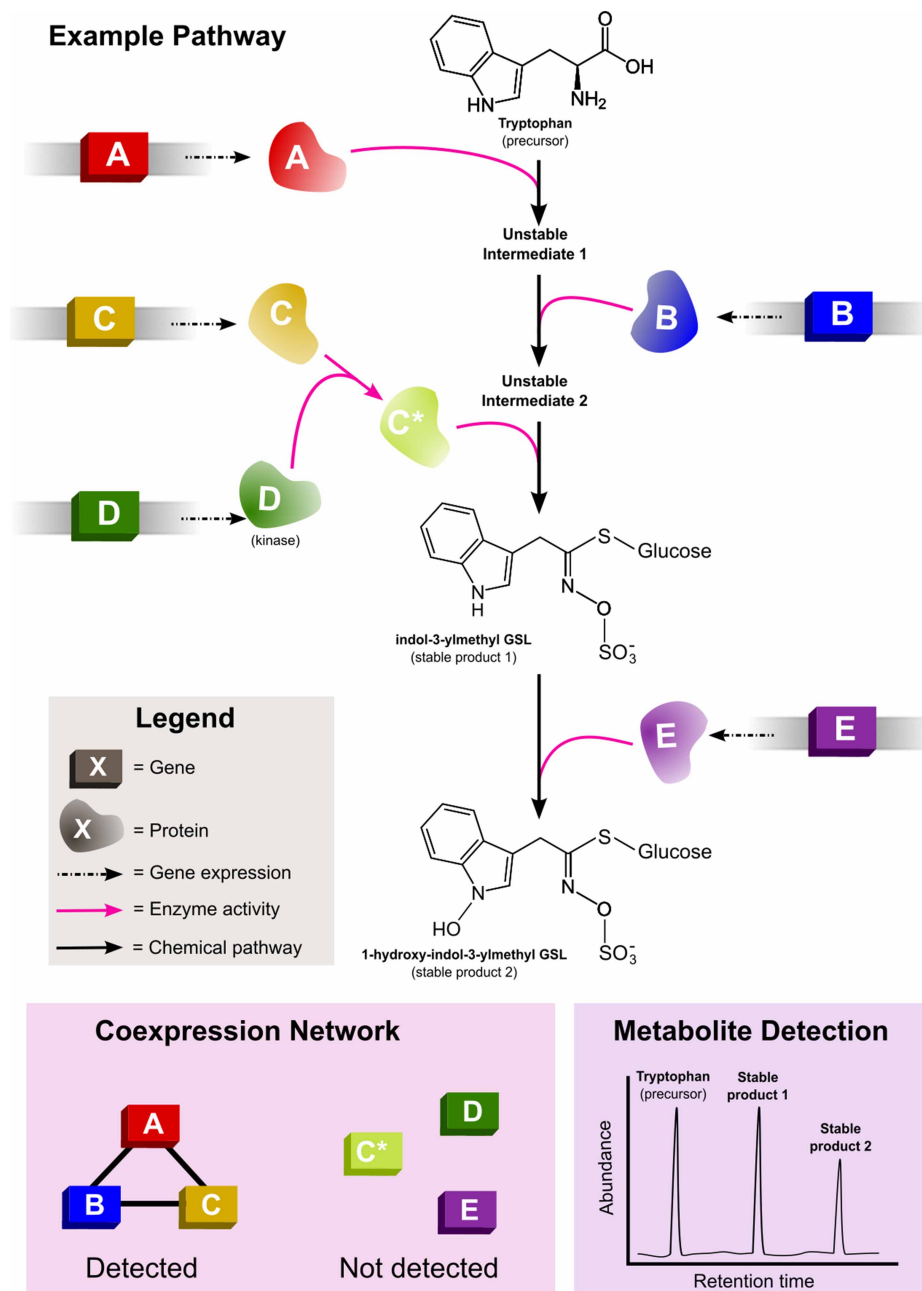


Figure 1. A simplified glucosinolate biosynthetic pathway illustrating the power and limitations of integrated transcriptomics and metabolomics for identifying specialized metabolic pathways.

The simplified pathway uses tryptophan as a precursor and generates two stable glucosinolate products using several genes and enzymes. Enzymes A and B are directly involved in the production of the first stable product. Enzyme C must be activated by Enzyme D (a kinase) before it can also contribute to the pathway. As a result, Genes A, B, and C are detected in the co-expression network, but Gene C* is not detected because the enzyme isn't generated directly from a transcript. Gene D encoding the kinase may also not be reliably detected in the co-expression network. Enzyme E is directly involved in a final functional modification of stable product 1 to form stable product 2, so it also may not reliably co-express at the same time as the other biosynthetic genes. Metabolic profiles reveal both stable glucosinolate products in addition to the precursor molecule.

critical component required to produce a glucosinolate, this suggests that specialized metabolic pathways can become integrated with core metabolism. This indicates that co-expression will find genes required for a pathway's function which are not part of the formal pathway. These peripheral support genes could confuse the analysis if one assumes all co-expressed genes are a direct part of the pathway. Other confounding aspects of peripheral gene co-expression will be discussed later.

Another complication is that these methods are highly dependent on the specific samples and the analytical technologies used. Since many specialized metabolites are stress inducible, gene co-expression for a particular metabolite may require very specific environmental conditions to be induced. As such, care should be taken to identify a broad swath of environmental conditions to maximize the potential to find a pathway [38]. In addition to the samples, a focus on transcriptome and metabolome without integrated proteomics will overlook the effects of post-translational modification, which can have diverse and dynamic consequences for protein localization and function [39]. Post-translational modifications have been shown to have major roles in regulating specialized metabolic pathways (Figure 1). For example, phenylalanine ammonia lyase (PAL), a pivotal enzyme in the phenylpropanoid biosynthetic pathway, is regulated by ubiquitination from Kelch repeat F-box proteins [40,41]. Solving this ascertainment bias is difficult as the cost of proteomics and even metabolomics, due to their specialized equipment, means that these experiments are expensive and throughput is also decreased. Future technological solutions would be required to resolve post-translational aspects key to regulating a specialized metabolite pathway.

Pathway engineering: are specialized metabolic enzymes enough?

Once the genes of a pathway are identified and characterized, we have ample tools available to directly engineer metabolic pathways. These include diverse expression constructs and systems to enable heterologous expression of biosynthetic pathways. For example, these have been applied to model plants such as *Nicotiana benthamiana* to express terpenes [42], indolic glucosinolates [37], and flavonols [43]. These direct pathway transfer efforts highlight a complication in extending pathway identification to successful synthetic biology or engineering applications. In most cases, the introduction of a biosynthetic pathway for a specialized metabolite core structure (here refers to the central or shared chemical structure of a class of specialized metabolites) is not sufficient for the pathway to be successfully expressed and the desired product achieved [44]. For example, introducing the indolic glucosinolate pathway into *N. benthamiana* leads to the accumulation of a large number of spurious and unwanted side products [37]. This can be due to gene silencing, sequestration and degradation of the engineered metabolites, unexpected effects in fitness, spurious side reactions from endogenous enzymes and a host of other causes. In microbes, this problem of moving from initial pathway introduction to optimization has been partially circumvented by the design/build/test/learn approach (DBLT), where pathways are tested in microbial systems [45]. In plants, DBLT is more difficult because generation times are significantly longer, desired organisms often are polyploids, and transformation protocols can be intensive [46,47]. An additional complication is that the critical random mutagenesis and selection step key to microbial DBLT approaches is only possible because of their quick generation times and smaller genome size. This random mutagenesis step is largely impossible in plants with slow generation times and constrained population sizes. Thus, to maximize our ability to engineer the newly identified specialized metabolite pathways into novel plant systems, we need a comprehensive understanding of plant metabolism and its intersection with gene regulation, physiology, and development to work towards predictive models. Examples of these interconnection studies drawn from the glucosinolate and phenolic literature are described below. There are further examples that we are unable to go into due to space limitations [48–53].

Role of peripheral genes

Critical to a successful predictive model to engineer metabolism is developing a better understanding of which and by how peripheral genes (any gene not considered part of the core structure metabolic pathway, but that supports the core structure pathway's function) influence the pathway. These peripheral factors could be from a range of processes including transcription factors and transporter proteins to general development, physiology, etc. For example, benzylisoquinoline alkaloid biosynthesis in opium poppy spans three different cell types indicating the involvement of developmental regulators and transport processes as peripheral components necessary for pathway function [54]. While this idea has a long history, it was recently formalized into the proposed omnigenic model that states gene regulatory networks are connected in such a way that genes expressed in trait-relevant cells can affect the core structure genes associated with the trait in question [55]. By extension,

any process affecting the trait-relevant cell including metabolism, physiology, transport, or ontology would equally apply. Of more immediate importance for predictive metabolic models, the cumulative effect of peripheral genes is expected to potentially exceed that of core structure genes. Identifying the full suite of developmental, transporter, regulator, and primary metabolism genes that allow a specialized metabolite pathway to function fully would facilitate synthetic biology, engineering, and crop improvement efforts.

Pathway connectedness

To illustrate how peripheral genes may integrate a specialized metabolite into a plant, we will focus on studies from the aliphatic glucosinolate pathway. Glucosinolates are composed of three main parts: a glucose molecule, a sulfur group, and an amino acid-derived R group that provides chemical diversity. In the case of aliphatic glucosinolates, this side chain is methionine derived, synthesized by the formation of cysteine from 3-PGA with serine as an intermediate [56,57]. The methionine cycle also produces S-adenosylmethionine which serves as a precursor to both ethylene and polyamines [58] (Figure 2). This established a flux connection between these pathways that extends to co-ordinated transcriptional regulation. For example, overexpressing the aliphatic glucosinolate transcription factor MYB28 led to genes in cysteine and methionine being up-regulated, as well as sulfur-deficient-inducible genes [19]. Thus, aliphatic glucosinolate transcription factors also control processes necessary for flux into the pathway. An additional flux connection is presented by Sugiyama *et al.* [59] where a sulfur reallocation pathway occurs in low sulfur concentrations. In this instance, aliphatic glucosinolates are catabolized into the sulfur-containing amino acid cysteine (Figure 2). Thus, changes in specialized metabolism likely involve evolutionary shifts in primary metabolism. This is supported by observations that betalain production in the Caryophyllales required alterations in the production of the precursor amino acid tyrosine [60].

In addition to the specialized metabolism-to-central metabolism connections, there can also be cross-talk between specialized metabolite pathways. This was recently shown when it was observed that accumulation of Tyr and Phe derived glucosinolate intermediates limits phenylpropanoid production [61,62]. Studying the cross-talk between the two biosynthetic pathways showed that this is mediated by a group of KFBs that degrade the phenylpropanoid entry point enzyme PAL [63]. The KFBs respond to alterations in glucosinolate production by an unknown mechanism allowing them to regulate PAL accumulation.

The connections between specialized metabolic pathways also extend to evolutionary observations that the gain of a particular pathway might result in loss or conversion of endogenous pathways. In the Capparales, glucosinolates appear to have evolved from cyanogenic glucosides with a concomitant loss of the cyanogenic pathway [64]. Similarly, in the order Caryophyllales, betalains replace the role usually filled by anthocyanins. However, some Caryophyllales families revert to anthocyanin pigmentation and the betalains are simultaneously lost; with no coexistence being reported [60]. For the glucosinolate/cyanogenic relationship, it is possible to engineer a plant containing both pathways. This suggests that these mutual antagonisms may not be chemical or mechanistic but may reflect selective/ecological tradeoffs [65]. Thus, when attempting to introduce a pathway into a new organism, it is necessary to consider how the existing pathways might be affected both mechanistically at the level of cross-talk and at potential ecological or evolutionary levels.

Effect of exogenous metabolites

In addition to pathway-level flux connections, specialized metabolites themselves can also influence the plants as potential regulatory compounds. In *Arabidopsis*, 3-hydroxypropylglucosinolate influences root growth through the target of rapamycin (TOR) pathway [66]. Similarly, the glucosinolate breakdown product indole-3-carbinol can rescue auxin-induced phenotypes by allosterically perturbing TIR1 and Aux/IAAs interactions, both central to the canonical auxin signaling [67]. An interesting observation from these studies is that both specialized metabolites are specific to either *Arabidopsis* or the Brassicas, yet they influence highly conserved genes (TOR and TIR1). Supporting this hypothesis was the finding that the 3-hydroxypropylglucosinolate signaling function affected plants without glucosinolates and even the fungus, *Saccharomyces cerevisiae*. Beyond glucosinolates, phenylpropanoid homeostasis can influence lignin biosynthesis and growth through the conserved Mediator complex [68]. The potential for specialized metabolites to interact with conserved mechanisms suggests that it is possible for plant species that have never produced a compound to respond to that compound. Thus, there is a need to further study the breadth of specialized metabolites that may interact with existing signaling pathways and developmental processes to better predict how moving a compound into a new plant may affect that plant.

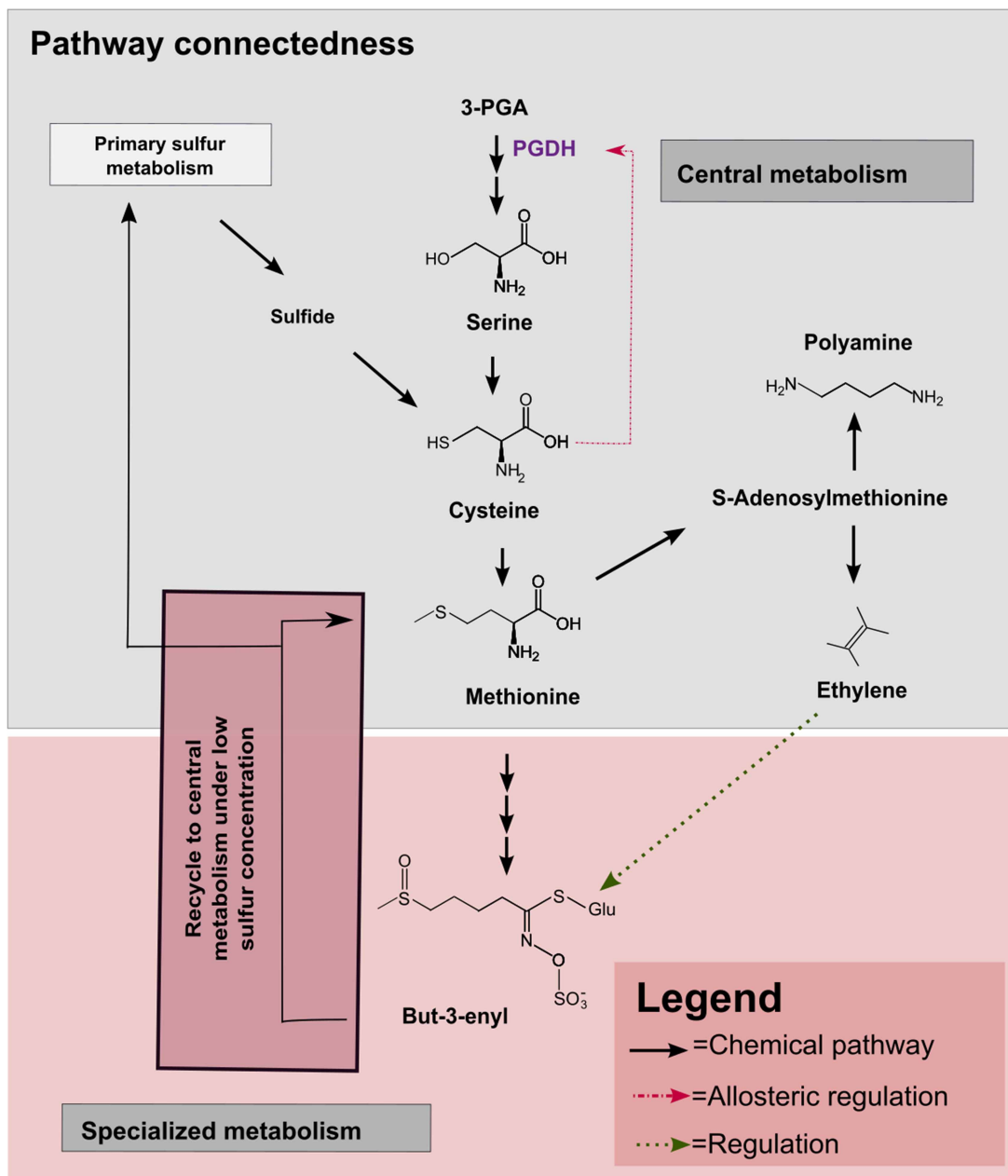


Figure 2. Simplified aliphatic glucosinolate pathway connectedness to other biochemical and signaling pathways.

Aliphatic glucosinolates' side chain is methionine derived which is synthesized by the formation of cysteine from 3-PGA with serine as an intermediate. Within this pathway, the intermediate methionine can allosterically regulate the enzyme PGDH which controls the entry point to the phosphorylated pathway of serine biosynthesis. Methionine also produces S-adenosylmethionine which serves as precursor to both ethylene and polyamines biosynthesis. Methionine gives rise to various glucosinolates, of which some have been observed to be catabolized to cysteine in low sulfur concentrations.

Conclusion

We are at the beginning of an exciting stage in synthetic biology and engineering where it is rapidly becoming feasible to identify new metabolic pathways in nearly any species and engineer them into naive systems. However, the experimental potential to conduct these studies is out-running our fundamental knowledge that is required to fully engineer metabolic systems. This review gives hints of just a few areas in which missing fundamental knowledge about specialized metabolite pathways could hinder future application efforts while there

are many others that we were unable to highlight for lack of space or newness of approach (e.g. scRNAseq and cell type specificity in metabolism). While there are numerous avenues to address these gaps like more funding, interdisciplinary groups, etc., probably the biggest need is to step back and reassess the basic assumptions we make about any system. At every level, we should ask if this assumption is supported by actual experiments that test the assumption, or is this assumption largely built on historical contingencies that have not really been tested. Then using these self-evaluations to direct where fundamental studies should focus to maximize the future ability of engineering or synthetic biology efforts.

Summary

- The identification of new plant specialized metabolic pathways is progressing at an accelerating rate with multi-omics approaches.
- The compounds from biosynthetic pathways have diverse biological activities that are highly desirable for synthetic biology or engineering applications.
- Current co-expression approaches for pathway identification may miss key enzymes and regulators needed for these pathways requiring new methods.
- These pathways are potentially highly connected to the rest of the host plant's physiology and metabolism.
- Understanding the depth and identity of potential connections between the processes occurring in a plant is key to successfully applying these pathways in synthetic biology or engineering methods.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Funding

This work was supported by the USDA National Institute of Food and Agriculture (grant nos. CA-D-PLS-7033-H and 2019-05709, to D.J.K.); and the US National Science Foundation (grant nos. IOS 2020754, MCB 1906486, and IOS 1655810, to DJK).

Open Access

Open access for this article was enabled through a transformative open access agreement between Portland Press and the University of California.

Author Contributions

All authors contributed to the writing of this article.

Abbreviations

DBLT, design/build/test/learn approach; PAL, phenylalanine ammonia lyase; TOR, target of rapamycin.

References

- 1 Owen, C., Patron, N.J., Huang, A. and Osbourn, A. (2017) Harnessing plant metabolic diversity. *Curr. Opin. Chem. Biol.* **40**, 24–30 <https://doi.org/10.1016/j.cbpa.2017.04.015>
- 2 Wurtzel, E.T. and Kutchan, T.M. (2016) Plant metabolism, the diverse chemistry set of the future. *Science* **353**, 1232–1236 <https://doi.org/10.1126/science.aad2062>
- 3 Weng, J.-K., Philippe, R.N. and Noel, J.P. (2012) The rise of chemodiversity in plants. *Science* **336**, 1667–1670 <https://doi.org/10.1126/science.1217411>
- 4 Weng, J. (2014) The evolutionary paths towards complexity: a metabolic perspective. *New Phytol.* **201**, 1141–1149 <https://doi.org/10.1111/nph.12416>

- 5 Lu, X., Tang, K. and Li, P. (2016) Plant metabolic engineering strategies for the production of pharmaceutical terpenoids. *Front. Plant Sci.* **7**, 1647 <https://doi.org/10.3389/fpls.2016.01647>
- 6 Moghe, G. and Last, R.L. (2015) Something old, something new: conserved enzymes and the evolution of novelty in plant specialized metabolism. *Plant Physiol.* **169**, 1512–1523. <https://doi.org/10.1104/pp.15.00994>
- 7 Lynch, J.H., Huang, X.-Q. and Dudareva, N. (2021) Silent constraints: the hidden challenges faced in plant metabolic engineering. *Curr. Opin. Biotechnol.* **69**, 112–117 <https://doi.org/10.1016/j.copbio.2020.12.014>
- 8 Moore, B.M., Wang, P., Fan, P., Leong, B., Schenck, C.A., Lloyd, J.P. et al. (2019) Robust predictions of specialized metabolism genes through machine learning. *Proc. Natl Acad. Sci. U.S.A.* **116**, 2344–2353 <https://doi.org/10.1073/pnas.1817074116>
- 9 Mahood, E.H., Kruse, L.H. and Moghe, G.D. (2020) Machine learning: a powerful tool for gene function prediction in plants. *Appl. Plant Sci.* **8**, e11376. <https://doi.org/10.1002/aps3.11376>
- 10 Frey, M., Chomet, P., Glawischnig, E., Stettner, C., Grün, S., Winklmair, A. et al. (1997) Analysis of a chemical plant defense mechanism in grasses. *Science* **277**, 696–699 <https://doi.org/10.1126/science.277.5326.696>
- 11 Qi, X., Bakht, S., Leggett, M., Maxwell, C., Melton, R. and Osbourn, A. (2004) A gene cluster for secondary metabolism in oat: implications for the evolution of metabolic diversity in plants. *Proc. Natl Acad. Sci. U.S.A.* **101**, 8233–8238 <https://doi.org/10.1073/pnas.0401301101>
- 12 Itkin, M., Heinig, U., Tzfadia, O., Bhide, A.J., Shinde, B., Cardenas, P.D. et al. (2013) Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. *Science* **341**, 175–179 <https://doi.org/10.1126/science.1240230>
- 13 Wisecaver, J.H., Borowsky, A.T., Tzin, V., Jander, G., Kliebenstein, D.J. and Rokas, A. (2017) A global coexpression network approach for connecting genes to specialized metabolic pathways in plants. *Plant Cell* **29**, 944–959 <https://doi.org/10.1105/tpc.17.00009>
- 14 Rai, A., Saito, K. and Yamazaki, M. (2017) Integrated omics analysis of specialized metabolism in medicinal plants. *Plant J.* **90**, 764–787 <https://doi.org/10.1111/tpj.13485>
- 15 Tholl, D. and Lee, S. (2011) Terpene specialized metabolism in *Arabidopsis thaliana*. *Arab Book* **9**, e0143 <https://doi.org/10.1199/tab.0143>
- 16 Winkel, B.S.J. (2004) Metabolic channeling in plants. *Annu. Rev. Plant Biol.* **55**, 85–107 <https://doi.org/10.1146/annurev.arplant.55.031903.141714>
- 17 Laursen, T., Møller, B.L. and Bassard, J.-E. (2015) Plasticity of specialized metabolism as mediated by dynamic metabolons. *Trends Plant Sci.* **20**, 20–32 <https://doi.org/10.1016/j.tplants.2014.11.002>
- 18 Robinson, J.B. and Srere, P.A. (1985) Organization of Krebs tricarboxylic-acid cycle enzymes in mitochondria. *J. Biol. Chem.* **260**, 800–805.
- 19 Hirai, M.Y., Sugiyama, K., Sawada, Y., Tohge, T., Obayashi, T., Suzuki, A. et al. (2007) Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc. Natl Acad. Sci. U.S.A.* **104**, 6478–6483 <https://doi.org/10.1073/pnas.0611629104>
- 20 Lau, W. and Sattely, E.S. (2015) Six enzymes from mayapple that complete the biosynthetic pathway to the topoisomerase aglycone. *Science* **349**, 1224–1228 <https://doi.org/10.1126/science.aac7202>
- 21 Naoumkina, M.A., Modolo, L.V., Huhman, D.V., Urbanczyk-Wochniak, E., Tang, Y., Sumner, L.W. et al. (2010) Genomic and coexpression analyses predict multiple genes involved in triterpene saponin biosynthesis in *Medicago truncatula*. *Plant Cell* **22**, 850–866 <https://doi.org/10.1105/tpc.109.073270>
- 22 Rizhsky, L., Jin, H., Shepard, M.R., Scott, H.W., Teitgen, A.M., Perera, M.A. et al. (2016) Integrating metabolomics and transcriptomics data to discover a biocatalyst that can generate the amine precursors for alkaloid biosynthesis. *Plant J.* **88**, 775–793 <https://doi.org/10.1111/tpj.13295>
- 23 Pluskal, T., Torrens-Spence, M.P., Fallon, T.R., De Abreu, A., Shi, C.H. and Weng, J.-K. (2019) The biosynthetic origin of psychoactive kavalactones in kava. *Nat. Plants* **5**, 867–878 <https://doi.org/10.1038/s41477-019-0474-0>
- 24 Yu, N., Chen, Z., Yang, J., Li, R. and Zou, W. (2021) Integrated transcriptomic and metabolomic analyses reveal regulation of terpene biosynthesis in the stems of *Sindora glabra*. *Tree Physiol.* **41**, 1087–1102 <https://doi.org/10.1093/treephys/tpaa168>
- 25 Singh, A. and Desgagné-Penix, I. (2017) Transcriptome and metabolome profiling of *Narcissus pseudonarcissus* ‘King Alfred’ reveal components of Amaryllidaceae alkaloid metabolism. *Sci. Rep.* **7**, 17356 <https://doi.org/10.1038/s41598-017-17724-0>
- 26 Perez de Souza, L., Scossa, F., Proost, S., Bitocchi, E., Papa, R., Tohge, T. et al. (2019) Multi-tissue integration of transcriptomic and specialized metabolite profiling provides tools for assessing the common bean (*Phaseolus vulgaris*) metabolome. *Plant J.* **97**, 1132–1153 <https://doi.org/10.1111/tpj.14178>
- 27 Jiang, C., Fei, X., Pan, X., Huang, H., Qi, Y., Wang, X. et al. (2021) Tissue-specific transcriptome and metabolome analyses reveal a gene module regulating the terpenoid biosynthesis in *Curcuma wenyujin*. *Ind. Crops Prod.* **170**, 113758 <https://doi.org/10.1016/j.indcrop.2021.113758>
- 28 Nett, R.S., Lau, W. and Sattely, E.S. (2020) Discovery and engineering of colchicine alkaloid biosynthesis. *Nature* **584**, 148–153 <https://doi.org/10.1038/s41586-020-2546-8>
- 29 Torrens-Spence, M.P., Pluskal, T., Li, F.-S., Carballo, V. and Weng, J.-K. (2018) Complete pathway elucidation and heterologous reconstitution of rhodiola salidroside biosynthesis. *Mol. Plant* **11**, 205–217 <https://doi.org/10.1016/j.molp.2017.12.007>
- 30 Gao, W., Sun, H.-X., Xiao, H., Cui, G., Hillwig, M.L., Jackson, A. et al. (2014) Combining metabolomics and transcriptomics to characterize tanshinone biosynthesis in *Salvia miltiorrhiza*. *BMC Genomics* **15**, 73 <https://doi.org/10.1186/1471-2164-15-73>
- 31 Xu, Q., He, Y., Yan, X., Zhao, S., Zhu, J. and Wei, C. (2018) Unraveling a crosstalk regulatory network of temporal aroma accumulation in tea plant (*Camellia sinensis*) leaves by integration of metabolomics and transcriptomics. *Environ. Exp. Bot.* **149**, 81–94 <https://doi.org/10.1016/j.envexpbot.2018.02.005>
- 32 Zhang, Y., Bouwmeester, H.J. and Kappers, I.F. (2020) Combined transcriptome and metabolome analysis identifies defence responses in spider mite-infested pepper (*Capsicum annuum*). *J. Exp. Bot.* **71**, 330–343 <https://doi.org/10.1093/jxb/erz422>
- 33 Hansen, B.G., Kerwin, R.E., Ober, J.A., Lambrix, V.M., Mitchell-Olds, T., Gershenzon, J. et al. (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 <https://doi.org/10.1104/pp.108.129981>
- 34 Hansen, B.G., Kliebenstein, D.J. and Halkier, B.A. (2007) Identification of a flavin-monoxygenase as the S-oxygenating enzyme in aliphatic glucosinolate biosynthesis in Arabidopsis: identification of a flavin monoxygenase. *Plant J.* **50**, 902–910 <https://doi.org/10.1111/j.1365-3113X.2007.03101.x>
- 35 Zhang, Y., Li, B., Huai, D., Zhou, Y. and Kliebenstein, D.J. (2015) The conserved transcription factors, MYB115 and MYB118, control expression of the newly evolved benzoxyl glucosinolate pathway in *Arabidopsis thaliana*. *Front. Plant Sci.* **6**, 343. <https://doi.org/10.3389/fpls.2015.00343>
- 36 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 <https://doi.org/10.1105/tpc.108.063115>

- 37 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 <https://doi.org/10.1105/tpc.110.081711>
- 38 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 <https://doi.org/10.1038/nature14907>
- 39 Vu, L.D., Gevaert, K. and De Smet, I. (2018) Protein language: post-translational modifications talking to each other. *Trends Plant Sci.* **23**, 1068–1080 <https://doi.org/10.1016/j.tplants.2018.09.004>
- 40 Zhang, X., Gou, M. and Liu, C.-J. (2014) *Arabidopsis* kelch repeat F-box proteins regulate phenylpropanoid biosynthesis via controlling the turnover of phenylalanine ammonia-lyase. *Plant Cell* **25**, 4994–5010 <https://doi.org/10.1105/tpc.113.119644>
- 41 Yu, S., Kim, H., Yun, D.-J., Suh, M.C. and Lee, B. (2019) Post-translational and transcriptional regulation of phenylpropanoid biosynthesis pathway by Kelch repeat F-box protein SAGL1. *Plant Mol. Biol.* **99**, 135–148 <https://doi.org/10.1007/s11103-018-0808-8>
- 42 Reed, J. and Osbourn, A. (2018) Engineering terpenoid production through transient expression in *Nicotiana benthamiana*. *Plant Cell Rep.* **37**, 1431–1441 <https://doi.org/10.1007/s00299-018-2296-3>
- 43 Irmisch, S., Jancsik, S., Yuen, M.M.S., Madilao, L.L. and Bohlmann, J. (2019) Biosynthesis of the anti-diabetic metabolite montbretin A: glucosylation of the central intermediate mini-MbA. *Plant J.* **100**, 879–891 <https://doi.org/10.1111/tpj.14493>
- 44 Lynch, J.H., Orlova, I., Zhao, C., Guo, L., Jaini, R., Maeda, H. et al. (2017) Multifaceted plant responses to circumvent Phe hyperaccumulation by downregulation of flux through the shikimate pathway and by vacuolar Phe sequestration. *Plant J.* **92**, 939–950 <https://doi.org/10.1111/tpj.13730>
- 45 Pouvreau, B., Vanhercke, T. and Singh, S. (2018) From plant metabolic engineering to plant synthetic biology: the evolution of the design/build/test/learn cycle. *Plant Sci.* **273**, 3–12 <https://doi.org/10.1016/j.plantsci.2018.03.035>
- 46 Birch, R.G. (1997) Plant transformation: problems and strategies for practical application. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **48**, 297–326 <https://doi.org/10.1146/annurev.arplant.48.1.297>
- 47 Soltis, D.E., Visger, C.J. and Soltis, P.S. (2014) The polyploidy revolution then...and now: stebbins revisited. *Am. J. Bot.* **101**, 1057–1078 <https://doi.org/10.3732/ajb.1400178>
- 48 Gaudinier, A., Tang, M. and Kliebenstein, D.J. (2015) Transcriptional networks governing plant metabolism. *Curr. Plant Biol.* **3–4**, 56–64 <https://doi.org/10.1016/j.cpb.2015.07.002>
- 49 Li, D., Heiling, S., Baldwin, I.T. and Gaquerel, E. (2016) Illuminating a plant's tissue-specific metabolic diversity using computational metabolomics and information theory. *Proc. Natl Acad. Sci. U.S.A.* **113**, E7610–8 <https://doi.org/10.1073/pnas.1610218113>
- 50 Tan, Q.W. and Mutwil, M. (2020) Inferring biosynthetic and gene regulatory networks from *Artemisia annua* RNA sequencing data on a credit card-sized ARM computer. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* **1863**, 194429 <https://doi.org/10.1016/j.bbagr.2019.194429>
- 51 Li, Y., Van den Ende, W. and Rolland, F. (2014) Sucrose induction of anthocyanin biosynthesis is mediated by DELLA. *Mol. Plant* **7**, 570–572. <https://doi.org/10.1093/mp/sst161>
- 52 Meihls, L.N., Handrick, V., Glauser, G., Barbier, H., Kaur, H., Haribal, M.M. et al. (2013) Natural variation in maize aphid resistance is associated with 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one glucoside methyltransferase activity. *Plant Cell* **25**, 2341–2355 <https://doi.org/10.1105/tpc.113.112409>
- 53 Ahmad, S., Veyrat, N., Gordon-Weeks, R., Zhang, Y., Martin, J., Smart, L. et al. (2011) Benzoxazinoid metabolites regulate innate immunity against aphids and fungi in maize. *Plant Physiol.* **157**, 317–327 <https://doi.org/10.1104/pp.111.180224>
- 54 Beaudoin, G.A.W. and Facchini, P.J. (2014) Benzylisoquinoline alkaloid biosynthesis in opium poppy. *Planta* **240**, 19–32 <https://doi.org/10.1007/s00425-014-2056-8>
- 55 Boyle, E.A., Li, Y.I. and Pritchard, J.K. (2017) An expanded view of complex traits: from polygenic to omnigenic. *Cell* **169**, 1177–1186 <https://doi.org/10.1016/j.cell.2017.05.038>
- 56 Li, J., Hansen, B.G., Ober, J.A., Kliebenstein, D.J. and Halkier, B.A. (2008) Subclade of flavin-monooxygenases involved in aliphatic glucosinolate biosynthesis. *Plant Physiol.* **148**, 1721–1733 <https://doi.org/10.1104/pp.108.125757>
- 57 Okamura, E., Ohtaka, K., Nishihama, R., Uchida, K., Kuwahara, A., Mochida, K. et al. (2021) Diversified amino acid-mediated allosteric regulation of phosphoglycerate dehydrogenase for serine biosynthesis in land plants. *Biochem. J.* **478**, 2217–2232 <https://doi.org/10.1042/BCJ20210191>
- 58 Sauter, M., Moffatt, B., Saechao, M.C., Hell, R. and Wirtz, M. (2013) Methionine salvage and S-adenosylmethionine: essential links between sulfur, ethylene and polyamine biosynthesis. *Biochem. J.* **451**, 145–154 <https://doi.org/10.1042/BJ20121744>
- 59 Sugiyama, R., Li, R., Kuwahara, A., Nakabayashi, R., Sotta, N., Mori, T. et al. (2021) Retrograde sulfur flow from glucosinolates to cysteine in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. U.S.A.* **118**, e2017890118 <https://doi.org/10.1073/pnas.2017890118>
- 60 Lopez-Nieves, S., Yang, Y., Timoneda, A., Wang, M., Feng, T., Smith, S.A. et al. (2018) Relaxation of tyrosine pathway regulation underlies the evolution of betalain pigmentation in Caryophyllales. *New Phytol.* **217**, 896–908 <https://doi.org/10.1111/nph.14822>
- 61 Hemm, M.R., Ruegger, M.O. and Chapple, C. (2003) The *Arabidopsis* *ref2* mutant is defective in the gene encoding CYP83A1 and shows both phenylpropanoid and glucosinolate phenotypes. *Plant Cell* **15**, 179–194 <https://doi.org/10.1105/tpc.006544>
- 62 Kim, J.I., Dolan, W.L., Anderson, N.A. and Chapple, C. (2015) Indole glucosinolate biosynthesis limits phenylpropanoid accumulation in *Arabidopsis thaliana*. *Plant Cell* **27**, 1529–1546 <https://doi.org/10.1105/tpc.15.00127>
- 63 Kim, J.I., Zhang, X., Pascuzzi, P.E., Liu, C. and Chapple, C. (2020) Glucosinolate and phenylpropanoid biosynthesis are linked by proteasome-dependent degradation of PAL. *New Phytol.* **225**, 154–168 <https://doi.org/10.1111/nph.16108>
- 64 Sønderby, I.E., Geu-Flores, F. and Halkier, B.A. (2010) Biosynthesis of glucosinolates – gene discovery and beyond. *Trends Plant Sci.* **15**, 283–290 <https://doi.org/10.1016/j.tplants.2010.02.005>
- 65 Bak, S., Olsen, C.E., Petersen, B.L., Møller, B.L. and Halkier, B.A. (1999) Metabolic engineering of p-hydroxybenzylglucosinolate in *Arabidopsis* by expression of the cyanogenic CYP79A1 from *Sorghum bicolor*. *Plant J.* **20**, 663–671 <https://doi.org/10.1046/j.1365-313X.1999.00642.x>
- 66 Malinovsky, F.G., Thomsen, M.-L.F., Nintemann, S.J., Jagd, L.M., Bourguine, B., Burrow, M. et al. (2017) An evolutionarily young defense metabolite influences the root growth of plants via the ancient TOR signaling pathway. *eLife* **6**, e29353 <https://doi.org/10.7554/eLife.29353>
- 67 Katz, E., Nisani, S., Yadav, B.S., Woldemariam, M.G., Shai, B., Obolski, U. et al. (2015) The glucosinolate breakdown product indole-3-carbinol acts as an auxin antagonist in roots of *Arabidopsis thaliana*. *Plant J.* **82**, 547–555 <https://doi.org/10.1111/tpj.12824>
- 68 Mao, X., Weake, V.M. and Chapple, C. (2019) Mediator function in plant metabolism revealed by large-scale biology. *J. Exp. Bot.* **70**, 5995–6003 <https://doi.org/10.1093/jxb/erz372>