the plant journal

SEB Society for Experimental Biology

The Plant Journal (2018) **94,** 933–942 doi: 10.1111/tpj.13950

FOCUSED REVIEW

Metabolomics 20 years on: what have we learned and what hurdles remain?

Saleh Alseekh[1,2] and Alisdair R. Fernie[1,2,*]

[1]*Max-Planck-Institute of Molecular Plant Physiology, Am Mühlenberg 1, Potsdam-Golm 14476, Germany, and*
[2]*Centre of Plant System Biology and Biotechnology, Plovdiv 4000, Bulgaria*

Received 1 March 2018; revised 20 April 2018; accepted 25 April 2018; published online 7 May 2018.
*For correspondence (e-mail fernie@mpimp-golm.mpg.de).

SUMMARY

The term metabolome was coined in 1998, by analogy to genome, transcriptome and proteome. The first research papers using the terms metabolomics, metabonomics, metabolic profiling or metabolite profiling were published shortly thereafter. In this short review we reflect on the major achievements brought about by the use of these approaches, and document the knowledge and technology gaps that are currently constraining its further development. Finally, we detail why we think that the time is ripe to refocus our efforts on the understanding of metabolic function.

Keywords: metabolomics, GC-MS, LC-MS, metabolite profiling, metabolite function.

INTRODUCTION

The plant kingdom is routinely stated to contain between 100 000 and 1 million metabolites (Dixon and Strack, 2003; Rai *et al.*, 2017), with any given species thought to contain upwards of 5000 metabolites (Fernie *et al.*, 2004). Although the core central metabolism of most plant species is largely comparable with that of non-plant species, plants and fungi contain a vast wealth of specialized compounds that account for the vast majority of the diversity within their metabolomes. These specialized (also known as secondary) metabolites collectively act as an effective arsenal against the myriad of biotic and abiotic stresses that they, as sessile organisms, may potentially be exposed to during their lifespan. Measuring such a vast number of entities is in its own right difficult; however, the problem is exacerbated by the fact that metabolites have highly diverse chemistry (D'Auria and Gershenzon, 2005), a massive dynamic range (Fernie *et al.*, 2004), and are often compartmented at both the cellular and subcellular level (Sweetlove and Fernie, 2013).

Given that this is a topical review, we will not dwell on historical aspects of (plant) metabolomics; however, it would be remiss not to provide a brief description of these. As mentioned above, the term metabolomics was coined by Steven Oliver in a review article on yeast functional genomics published in 1998 (Oliver *et al.*, 1998). Several research papers were subsequently published at the turn of the century that essentially defined the possibilities metabolite profiling afforded to studies of metabolic regulation and beyond (Katona *et al.*, 1999; Fiehn *et al.*, 2000; Fraser *et al.*, 2000; Roberts, 2000; Roessner *et al.*, 2001a; Sarry *et al.*, 2006; Duenas *et al.*, 2017). In essence, all of these papers describe the application of non-targeted profiling methods; however, although by nature largely descriptive, they heralded a widespread adoption of more sophisticated multivariate statistics in studies of plant and cellular function (Fiehn *et al.*, 2000; Roessner *et al.*, 2001a; Sumner *et al.*, 2003; Fraser *et al.*, 2007). Following on from these studies the tools of metabolomics have been brought to bear on an exhaustive range of biological questions. Indeed, they are now so plentiful that it is probably no longer possible to feature them all in a comprehensive review. Several recent reviews provide very good coverage on a topic-by-topic basis, however (Fridman and Pichersky, 2005; Tholl *et al.*, 2006; Gaquerel *et al.*, 2014; Feussner and Polle, 2015; Sumner *et al.*, 2015; Dong *et al.*, 2016; Tenenboim and Brotman, 2016; Wen *et al.*, 2016; Rai *et al.*, 2017; Showalter *et al.*, 2017). In the current article we plan mainly to review classical papers and provide a perspective of the challenges that remain both with regards to technical aspects but also in defining the *in vivo* function of metabolites (for highlight see Box 1).

METABOLOMICS TECHNOLOGIES

Current plant metabolomics strategies are reliant on either mass spectrometry- or nuclear magnetic resonance (NMR)-based approaches. A number of detailed protocols have been published (Lisec *et al.*, 2006; De Vos *et al.*, 2007; Kruger *et al.*, 2008; Tohge and Fernie, 2010), alongside several excellent technical reviews (Fiehn, 2002; Sumner *et al.*, 2003; Kopka *et al.*, 2004; Cajka and Fiehn, 2014; Fiehn *et al.*, 2015; Lu *et al.*, 2017). In response, we will restrict ourselves here to just a brief outline of the relative advantages and disadvantages of the three major methods currently in use, namely gas-chromatography mass-spectrometry (GC-MS), liquid-chromatography mass-spectrometry (LC-MS) and NMR. In brief, NMR is able to detect only highly abundant metabolites, or alternatively metabolites extracted from very large tissue volumes, because of its reliance on measuring atoms with non-zero magnetic moment. By contrast to MS-based approaches, which detect at the molecular level, NMR is an atomic-level approach providing high utility in both isotope tracing experiments and structural elucidation (Fernie and Tohge, 2017). Despite this advantage, coupled chromatographic and MS-based approaches, such as GC-MS and LC-MS, are far more frequently adopted. In GC-MS, polar metabolites are derivitized to render them volatile and then separated by GC. Electron impact ionization results in highly reproducible fragmentation patterns that are essential for large-scale experiments (Fernie *et al.*, 2004). GC-MS has the twin advantages of being relatively sensitive and highly robust, meaning that it can routinely measure hundreds of analytes (mass spectral features that are of either known or unknown chemical identity) in thousands of samples affording good coverage of primary metabolism. That said, it is LC-MS that currently provides the most comprehensive approach. Unlike GC-MS it does not require prior sample treatment, but the choice of columns, including reversed phase, ion exchange and hydrophobic interaction, provides metabolite separation on the basis of differential chemical properties. The twin developments of ultra-performance LC (UPLC) and high-resolution mass spectrometry rendered the technique even more powerful with regard to resolution, sensitivity and throughput (Fernie and Tohge, 2017), with data on over 1000 metabolites being accessible using either direct infusion or coupled UPLC high-resolution MS (Aharoni *et al.*, 2002; Kind and Fiehn, 2007; Giavalisco *et al.*, 2011). High-resolution MS provided a massive boost to metabolomics, with sufficient accuracy to determine the exact chemical composition, but not structure, of each analyte. Multiple rounds of MS can aid in structural identification, however, via the identification and assembly of the resultant metabolite fragments. It is important to note that although many analytes in GC-MS remain unannotated, relatively new novel metabolites are found by this method, which relies largely on co-elution with known standards, whereas in LC-MS a considerable number of novel compound annotations are currently being generated on a regular basis. Despite being the most comprehensive technique to hand, LC-MS by no means approaches the metabolic complement of a typical plant cell (Figure 1; Tables 1 and S2). Collectively, these data illustrate that despite considerable recent advances, improving the coverage of the plant metabolome

Box 1 Highlights

- The plant kingdom contains a vast repertoire of metabolites involved both in core essential functions and in plant interactions with their environment
- The coverage of the metabolic complement of this kingdom remains relatively sparse
- Both technical and computational efforts to improve coverage of the metabolome are under way, and our abilities are considerably beyond those of 20 years ago
- Of the few thousand metabolites we can currently assign either unambiguously or with a high level of confidence, the precise *in vivo* function of many is unknown, and even their spatial abundance is not always clearly understood
- Combining standard metabolomics approaches with studies of wide natural variance and the use of either genome-editing techniques or combinatorial transformation potentially offer powerful tools by which the function of an individual metabolite can be assigned

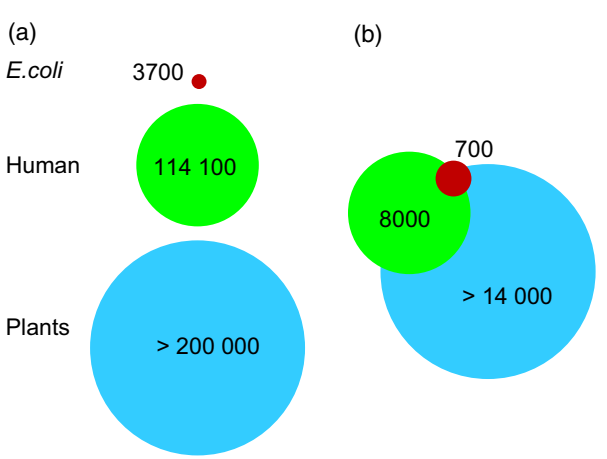

Figure 1. Metabolite diversity and coverage. (a) Approximate number of metabolites present in *Escherichia coli*, human and plants. (b) Approximate number of metabolites that we can measure. Metabolite numbers were taken from those reported in the following papers (*E. coli*, Guo *et al.*, 2013; Sajed *et al.*, 2016; human, Wishart *et al.*, 2018; and plants, Dixon and Strack, 2003; Saito and Matsuda, 2010; Afendi *et al.*, 2012; Wink, 2015; Rai *et al.*, 2017).

Table 1 Estimated number of specialized metabolites in plants per compound class

Alkaloids	
Amines	100
Cyanogenic glycosides	60
Glucosinolates	100
Alkamides	150
Lectins, peptides, polypeptide	2000
Non-protein amino acids	700
Phenolics	
Phenylpropanoids and flavonoids, anthocyanins, carotenoids	8000–9000
Polyacetylens	1000
Polyketides	750
Terpenoids	
Triterpenes, steroids, saponins	5000
Tetraterpenes	500
Diterpenes	2500
Sesquiterpenes	5000
Monoterpenes, iridoids	2500

Note that these are largely estimates based on the literature, and the true number is likely to be considerably higher, given that new metabolites are consistently being reported. Metabolite numbers were taken from the following studies: Tohge *et al.* (2014); Verpoorte and Alfermann (2000); Wink (2004); Wink (2015); Ziegler and Facchini (2008).

remains a major technical challenge for metabolomics. The wide range of numbers listed is illustrative of the difficulty in predicting the metabolite complement of the cell, which, unlike those of the transcript and protein, is genome-independent (Oliver *et al.*, 1998). We will return to the issue of the comprehensiveness of the plant metabolome after a brief historical overview of the major uses of metabolomics in research in the field of plant science. Before doing so, we should state that we use the terms metabolomics and metabolite profiling interchangeably here, using both to describe non-targeted metabolite analyses.

EARLY PLANT METABOLOMICS STUDIES

The earliest application of these tools in plants largely used GC-MS or LC-MS, or direct-injection MS, to profile transgenic plants (Fiehn *et al.*, 2000; Roessner *et al.*, 2001a; Aharoni *et al.*, 2002; Bovy *et al.*, 2002, Chen *et al.*, 2003), environmentally challenged plants (Hirai *et al.*, 2005; Nikiforova *et al.*, 2005; Suzuki *et al.*, 2005; Urbanczyk-Wochniak and Fernie, 2005; Allwood *et al.*, 2006; Lugan *et al.*, 2010) or both (Roessner *et al.*, 2001b). In addition, there were several studies that used metabolomics as screening approaches. For example, metabolomics was used in efforts to identify metabolic biomarkers that were able to predict future aspects of plant performance (Meyer *et al.*, 2007), to identify herbicide mode of action by comparing herbicides of known and unknown targets (Trenkamp *et al.*, 2009), and to assess whether genetically modified crops were substantially different from conventional cultivars (Catchpole *et al.*, 2005). These approaches, like those aimed at providing detailed and accurate descriptive information concerning the consequences of a biological intervention (detailed below), relied on the marriage of chemical analytics with statistical methods that were new to most people studying plant metabolism. In the early years of metabolomics both types of experiment were carried out at relatively low throughput, with any one publication reporting datasets with sample sizes below 100. There are three studies that we think are worthy of highlighting here. First, the study of Fiehn and co-workers used GC-MS to evaluate 326 analytes in comparing a developmental mutant (*stomatal density and distribution*, *sdd*) and a metabolic mutant (*digalactosyldiacylglycerol*, *dgd*) of Arabidopsis with their respective wild types (Fiehn *et al.*, 2000). Perhaps unsurprisingly, they observed that the metabolic mutant exhibited greater differences than the morphological mutant from their respective wild types. Secondly, the studies of Roessner *et al.* (2001a,b) carried out GC-MS profiling of 60 known and 27 unknown metabolites of a range of transgenic *Solanum tuberosum* (potato) lines expressing elevated sucrolytic activity in a tuber-specific manner alongside wild-type tuber material supplied with various exogenous concentrations of sugars. This transgenic material had been created with the aim of enhancing tuber starch content, but actually resulted in the opposite phenotype. The metabolic profiling studies were able to demonstrate that this was the result of a massive activation of respiration and amino acid biosynthesis (Roessner *et al.*, 2001a). Moreover, the combination of the feeding experiments and principal component analysis (PCA) indicated that this was induced by the high levels of glucose (and to a lesser extent fructose), thus providing a mechanistic explanation for the failed metabolic engineering strategies (Roessner *et al.*, 2001b). Thirdly, LC-MS-based metabolite profiles of Arabidopsis revealed 2000 different mass signals (analytes) from roots and leaves, with many of these representing specialized metabolites (von Roepenack-Lahaye *et al.*, 2004). This study used the chalcone synthase-deficient *tt4* mutant to demonstrate that subtle differences between samples can be observed; however, perhaps more importantly the demonstration of the sensitivity and resolving power of this technique greatly expanded the range of metabolic profiling at this time point.

These studies collectively broadened our depth of understanding of the metabolic responses to the respective perturbations, but in addition provided us with far greater direct understanding of metabolic network behavior than was available from previous studies. In this vein, correlation analysis was documented to be a powerful technique for assigning metabolites to the biochemical pathways to which they belong (Weckwerth and Fiehn, 2002). Such information was important in achieving a more holistic

understanding, both as a font of fundamental knowledge but also with regards to a more rational design of metabolic engineering strategies, aimed either at increasing the production of high-value metabolites or alternatively at increasing crop yield (Sweetlove *et al.*, 2017).

GENE FUNCTIONAL ANNOTATION AND METABOLIC QUANTITATIVE TRAIT LOCI ANALYSES

In the last decade, metabolomics has become an essential tool in the elucidation of functional annotation of genes associated with metabolism. For this purpose three major approaches have been used: (i) direct testing of candidate genes via the analysis of knock-out mutants; (ii) quantitative trait loci (QTL) mapping of the genes determining the abundance of specific metabolites; and more recently (iii) genome-wide association studies (GWASs) of metabolite abundance. Two groundbreaking papers using the first approach were published in 2005. The first of these used the combination of transcriptomics and metabolomics to identify genes and metabolites associated with the pigmentation of the PAP1 activation tagged line (Tohge *et al.*, 2005). The second used the inducibility of the triterpene saponoid pathway to identify glucosyl transferases in a similar manner (Achnine *et al.*, 2005). Follow-up research following further co-expressed genes related to these pathways have subsequently allowed the expansion of the phenylpropanoid pathway to include more than 40 genes and metabolites (Tohge and Fernie, 2010), whereas the saponin pathway has also been greatly further expanded (Thimmappa *et al.*, 2014). Other pathways wherein major improvements in gene annotation and even pathway structure have been elucidated following similar methods include cell wall and terpene biosynthesis, photorespiration, mitochondrial carrier proteins and monolignol transport (Persson *et al.*, 2005; Araujo *et al.*, 2011; Alejandro *et al.*, 2012; Geu-Flores *et al.*, 2012; Pick *et al.*, 2013). It is important to note that the use of the co-expression approach is of course limited to pathways that are mainly regulated at the level of expression, however. Neither QTL nor GWAS approaches suffer from this limitation.

The QTL approach was first taken in 2006 in parallel studies looking at the primary metabolism of *Solanum lycopersicum* (tomato; Schauer *et al.*, 2006) and the specialized metabolism of Arabidopsis (Keurentjes *et al.*, 2006). Keurentjes *et al.* quantified the levels of 2000 analytes representing the specialized metabolism of Arabidopsis in a population derived from the L*er* and Cvi ecotypes (Keurentjes *et al.*, 2006). Much of the follow-up work on Arabidopsis has been carried out by the laboratories of Dan Kliebenstein and Lothar Willmitzer (Lisec *et al.*, 2008; Rowe *et al.*, 2008; Joseph *et al.*, 2015a,b). Their work addressed many aspects of the genetics of metabolism, including a comparative analysis of population types, and the evaluation of heterosis, heritability and the environmental plasticity of the plant metabolome (Joseph *et al.*, 2015b). Work in the Kliebenstein laboratory additionally quantified the influence of genetic information from the organellar genomes on the metabolome (Joseph *et al.*, 2015a,b), studied epistatic interactions and defined novel potential biochemical networks (Rowe *et al.*, 2008).

The study on tomato was based on a very well characterized *Solanum pennelli* introgression line population, for which a massive volume of yield-associated and developmental data is available (Lippman *et al.*, 2007). Thus the study was able not only to define almost 900 QTL for primary metabolites (including several that were subsequently cloned; Kochevenko and Fernie, 2011; Quadrana *et al.*, 2014), but also to link fruit amino acid content to the harvest index (Schauer *et al.*, 2006; Do *et al.*, 2010). This work, alongside follow-up papers on specialized metabolites, volatile organic compounds and acyl sugars (Tieman *et al.*, 2006; Schilmiller *et al.*, 2010; Alseekh *et al.*, 2015), led to the identification of more than 2000 metabolic QTL. Although a recently created backcrossed inbred line generated from the same parental lines affords far greater genetic resolution, and has already been applied to the study of cuticle composition, acyl sugar and primary metabolism (Ning *et al.*, 2015; Fan *et al.*, 2016; Ofner *et al.*, 2016; Alseekh *et al.*, 2017). QTL analysis has subsequently been carried out for a range of other important crops, including *Oryza sativa* (rice; Matsuda *et al.*, 2012), *Triticum aestivum* (wheat; Hill *et al.*, 2013), *Zea mays* (maize; de Abreu E Lima *et al.*, 2018; Wen *et al.*, 2018), *Hordeum vulgare* (barley; Templer *et al.*, 2017) and potato (Carreno-Quintero *et al.*, 2012). These studies collectively identified a large number of structural and regulatory genes involved in the control of metabolite abundance in crops, and massively improved our understanding of the structure of the metabolic pathways as well as defined important leads for metabolic engineering (Wen *et al.*, 2016).

The third commonly taken approach is that of assessing the natural variance of the metabolome. Whilst the earliest studies in this vein (Bentsink *et al.*, 2000; Schauer *et al.*, 2005b), proceeded the utilization of metabolomics in the advanced breeding populations described above, the true power of this approach has only more recently been realized with the widespread adoption of GWAS approaches. The fundamental premise of GWAS, which was developed for use in medical genetics, is that the incidence of nucleotide polymorphisms is associated with the variance of a given trait (Fernie and Tohge, 2017). Early applications in plants included the study of the efficiency of nitrogen use (Loudet *et al.*, 2003) and glucosinolate metabolism (Kliebenstein *et al.*, 2001). Broader metabolite profiling studies followed these. First, performing GC-MS on a set of 94 ecotypes provided important insight into metabolic components of the genetic control of plant growth, suggesting that starch and protein are major determinants of growth

(Sulpice *et al.*, 2009, 2010). The evaluation of specialized metabolites in a subset of these revealed that approximately half of them contained a set of 18 previously unidentified compounds, which were revealed, by a battery of chemical analyses, to be phenylacylated flavonols, and facilitated the identification and functional analysis of the gene responsible for their production (Tohge *et al.*, 2016). Further studies in Arabidopsis have focused on the genetic architecture of glucosinolates (Chan *et al.*, 2011) and branched chain amino acids (Angelovici *et al.*, 2017), as well as looking at primary and specialized metabolism (Wu *et al.*, 2016, 2018), with these studies improving their resolution by combining metabolomics with transcriptomics and/or network analyses.

Genome-wide association studies (GWASs) have also recently been used in combination with metabolomics in tomato (Sauvage *et al.*, 2014; Tieman *et al.*, 2017; Ye *et al.*, 2017), maize (Wen *et al.*, 2014) and rice (Matsuda *et al.*, 2015). These studies all identified candidate genes underlying important nutritional and or taste traits. Alongside the identification of genes underlying quantitative traits, the study of broad natural variation has recently additionally been used to better understand the evolution of metabolism, and specifically to address the question of how metabolism has altered during crop domestication (Chan *et al.*, 2010; Kliebenstein, 2013a,b). Two studies best illustrate this. First, a deep evaluation of changes in primary metabolism in cultivated wheat, and its progenitor species, demonstrated that a reduction in unsaturated fatty acids was associated with selection during the primary domestication of emmer wheat, but that selection-driven changes in amino acid content mark the domestication of durum wheat (Beleggia *et al.*, 2016). Secondly, a combined analysis of the genomes, transcriptomes and metabolomes of several hundred tomato fruit has illustrated how global breeding has altered the metabolite content of the fruit. Several features of domestication were associated with changes in metabolite content, namely the selection of alleles of genes associated with the increase in fruit size, those for the reduction of anti-nutritional steroidal glycoalkaloids, the breeding of pink tomatoes for the Asian market and the introgression of resistance genes from wild cultivars (Zhu *et al.*, 2018). As such, this study was able to provide a very comprehensive view of the metabolic breeding history of the tomato. The suggestion that metabolic genes had hitchhiked during selection had been made prior to this study (Bellucci *et al.*, 2014), and seems likely to have occurred during the domestication of all of our crops (Giovannoni, 2018).

IMPROVING OUR COVERAGE OF THE PLANT METABOLOME

In parallel with the considerable advances in the application of metabolomics to address important biology questions, the last 20 years have been characterized by impressive advances in coverage. This is best defined as the number of metabolites that we are able to measure (which is best assessed by looking at the number of known or anticipated metabolites that a given organism contains and the number that can be accurately quantified). As stated above, this is in part a result of the improved machine performance afforded by the development of UPLC coupled with high-resolution MS, but has also relied on increased efforts in the collection of standard compounds and the sharing of reference extracts (for example, see Shahaf *et al.*, 2016) for use in peak annotation authentication, and the increased sophistication of computation approaches for compound annotation (compare the web resources listed in Tohge and Fernie, 2009 with those listed in de Souza *et al.*, 2017). As mentioned above, and detailed in Figure 1 and Tables S1 and S2, metabolomics techniques can now afford relatively good coverage of prokaryotes such as *Escherichia coli*, and even unicellular eukaryotes such as *Saccharomyces cerevisiae*. The coverage of multicellular eukaryotes is relatively poor, however. For example, the coverage of the human metabolome, which can be considered to be representative of the animal kingdom, is in the region of 60–70%. Similarly, we are able to measure only a few thousand plant metabolites, whereas between 100 000 and 1 million are estimated to be extant within the plant kingdom. It is important to note that these numbers are in no way exact and are instead best regarded as rough estimates, as much of the information found in the literature is rather nebulous. Despite this progress, even using the most optimistic estimates of our coverage of the plant metabolome it is clear that the majority of metabolites are not covered in current profiling methods. Experimentally this is being tackled by the collection of large-scale libraries of metabolites, such as the WEIZMASS library (Shahaf *et al.*, 2016); however, although these libraries certainly increase the numbers of metabolites that we can detect, a huge number of unknown analytes remain unaccounted for, and many known metabolites cannot be accessed by current metabolomics technologies. It seems likely that the diversification of extraction protocols to mirror the diversity of the chemistry of metabolites will be necessary as we try to bridge the gap between what is there and what we can measure.

In addition, effectively using the vast computational resources will be paramount to meeting this challenge. A huge number of databases have arisen that aid in metabolite annotation (for a review, de Souza *et al.*, 2017). Two of particular note are the Golm Metabolome Database (Kopka *et al.*, 2005) and MassBank (Horai *et al.*, 2010), covering primary and specialized and lipid metabolism, respectively. The first of these demonstrated the robustness of GC-MS output by illustrating that, if the same chromatography column is used, metabolites can be annotated in highly

diverse biological samples irrespective of the machine used (Schauer *et al.*, 2005a). For LC-MS, the situation is more complex; however, the MassBank database has proven a highly useful tool for sharing mass spectral information from LC-MS instrumentation, and several guides for using high-resolution MS are also highly helpful (Kind and Fiehn, 2007). In the case of lipids, the coverage is considerably better, possibly because of the high degree of conservation and the fact that within a compound class the structures are largely predictable (Kind *et al.*, 2013). The database BinBase contains 1561 studies and 114 795 samples and a total of 9563 unique metabolites, 1020 of which have been identified. Furthermore, it was recently demonstrated that many unknowns could be effectively identified by reference to 14 metabolome databases, on the basis of their elemental formulae and *in silico* fragmentation (Lai *et al.*, 2018), highlighting the power of this approach.

TOWARDS METABOLITE FUNCTION

Although the research described above details how metabolomics has been used to define the way in which the general metabolic landscape changes in response to genetic or environmental perturbation. The first of these was the very early GC-MS-based example that enhanced glucose levels was the underlying mechanism producing the metabolic phenotypes observed in transgenic potatoes exhibiting enhanced sucrose cleavage (Roessner *et al.*, 2001b). The second was related to the LC-MS-based characterization of the 18 novel phenylacylated flavonoids in Arabidopsis, which were named saiginols, and were demonstrated to confer enhanced UV resistance (Tohge *et al.*, 2016). A similar LC-MS-based study in rice was able to identify a different modified phenylpropanoid, namely a glycosylated flavone, that exhibits a similar function (Peng *et al.*, 2017). Many metabolite-mediated defense responses have also been identified via metabolomics approaches, including those mediated by coniferin and scopolin (Sonderby *et al.*, 2010; Ward *et al.*, 2011; Pichersky and Raguso, 2016). Similarly, the targeted profiling of glucosinolates recently uncovered a newly evolved regulation of the anciently conserved target of rapamycin (TOR) pathway in energy regulation (Malinovsky *et al.*, 2017), whereas a role for trehalose 6-phosphate has been suggested in a wide range of developmental processes (Figueroa and Lunn, 2016). A novel approach to understand metabolite function that has recently emerged is that of probing metabolite–protein interactions (Veyel *et al.*, 2017; Piazza *et al.*, 2018), which is commonly performed by the co-elution of metabolites and proteins following the separation of the proteins. It would seem reasonable to anticipate that such studies, which admittedly remain in their infancy, are likely to prove an important source of information regarding how metabolites function in cellular regulation.

WHAT HURDLES REMAIN?

Having documented the development of plant metabolomics from its inception to the present day, we would like to use the rest of this review to address the second part of the question raised in the title. There are two major challenges facing plant metabolomics: the technical challenge of improving coverage and the biological challenge of improving our understanding of metabolite function (see Box 2).

There are multiple facets to the technical problem. Given the wide diversity in the dynamic range of abundance and in the chemistry of metabolomics, it is currently difficult to envisage a (nearly) catch-all approach analogous to that afforded by RNA sequencing. On the one hand, machine improvements could allow for higher sensitivity and long-heralded approaches such as hyphenated-NMR technologies could result in massive gains in coverage. That said, coverage is likely to be dramatically increased with the technology that we currently have to hand; however, this would require a considerable coordinated effort, whereby multiple tissues from multiple species are evaluated by carrying out many different types of extraction and analyzing them (all) on all currently available platforms. Such open exchange of materials would surely also enhance the cross-comparability of results, which is an important prerequisite if we want to tap into the fantastic computational resources that have been developed for this discipline. They would also potentially allow the construction of a searchable database that warehouses changes in metabolite levels on the scale of Genevesigator (https://genevestigator.com), for example.

A couple of emergent strategies warrant a mention here, namely the use of isotope labeling and the development of metabolite imaging techniques. The first approach has already been demonstrated to be a powerful way of improving the annotation of plant metabolites (Feldberg *et al.*, 2009; Giavalisco *et al.*, 2011; Wang and Jones, 2014), as well as providing powerful information concerning both the elucidation of pathway structures (e.g. Arrivault *et al.*, 2017) and gene function annotations (e.g. Dal Cin *et al.*, 2011). For these methods the coverage is essentially as high as that for the analytical technique in use. Metabolite imaging techniques meanwhile provide high or even ultra-high spatial resolution of metabolite abundances (Dong *et al.*, 2016); however, to date their coverage is currently relatively limited.

Moving to the second challenge, a vital question is why has the specific function of most plant metabolites not yet been characterized? One problem is that their biological properties are largely described for a class of compounds rather than for individual metabolites. Furthermore, the promiscuity of many enzymes of specialized metabolism (which produce the majority of plant metabolites), alongside the general intricacy of most

metabolic networks, renders it difficult to modify the content of a single metabolite without affecting the others. These features of metabolism hence make it very difficult to directly assess metabolite function via conventional reverse-genetics strategies. For this purpose, more sophisticated methods, in which the kinetic properties of the respective enzymes are altered, are likely to be required. Although approaches to tackle this are likely to involve relatively complex strategies, we believe that they will be instrumental in helping us to ascertain the precise function of specific metabolites and the biological circumstances under which they play the most important role. First steps towards the latter have been made in an elegant study, in which the roles of flavonoids in the core pathway were tested in a range of mutants under conditions of oxidative and drought stress (Nakabayashi *et al.*, 2014; Cao *et al.*, 2017); however, these authors were only able to draw conclusions concerning the role of flavonoids as a compound class, and further studies are required to dissect the quantitative contributions of individual flavonoid species to the conferment of stress tolerance. We would argue that the identification of metabolite function on a metabolite-by-metabolite basis arguably represents the real grand challenge as metabolomics enters its third decade.

Box 2 Open questions

- How best do we further advance the coverage of the metabolome?
- Would the open exchange of metabolite extracts between researchers enhance the cross-comparability of results?
- Although an equivalent of the botany array resource gene expression viewer exists for metabolites, would it not be beneficial to have an analog of genevestigator, whereby changes in metabolite abundance can be tracked under a range of cellular and genetic circumstances?
- Where are all the plant metabolite receptors? Despite a growing appreciation of the importance of interactions between metabolites and other molecular entities, by contrast to other systems, very few metabolite receptors have, as yet, been reported in plants
- What is the best strategy to dissect the function of individual metabolites? This question applies equally to *in planta* functions and bioactive pharmacological functions

ACKNOWLEDGEMENTS

Work on metabolomics in our laboratories is currently supported by funding from the Max-Planck-Society and the European Union project PlantaSyst (SGA-CSA No 664621 and No 739582 under FPA No. 664620).

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

SUPPORTING INFORMATION

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

Table S1. Number of metabolites per compound class that are reported in the literature, and that we can measure.

Table S2. Estimated number of metabolites measured by GC-MS and LC-MS in bacteria, humans and higher plants.

REFERENCES

de Abreu E Lima, F., Li, K., Wen, W., Yan, J., Nikoloski, Z., Willmitzer, L. and Brotman, Y. (2018) Unraveling the lipid metabolism in maize with time-resolved multi-omics data. *Plant J.* **93**, 1102–1115.

Achnine, L., Huhman, D.V., Farag, M.A., Sumner, L.W., Blount, J.W. and Dixon, R.A. (2005) Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume Medicago truncatula. *Plant J.* **41**, 875–887.

Afendi, F.M., Okada, T., Yamazaki, M. *et al.* (2012) KNApSAcK family databases: Integrated metabolite-plant species databases for multifaceted plant research. *Plant Cell Physiol.* **53**, e1, 1–12.

Aharoni, A., De Vos, C.H.R., Verhoeven, H.A., Maliepaard, C.A., Kruppa, G., Bino, R. and Goodenowe, D.B. (2002) Nontargeted metabolome analysis by use of fourier transform ion cyclotron mass spectrometry. *OMICS*, **6**, 217–234.

Alejandro, S., Lee, Y., Tohge, T. *et al.* (2012) AtABCG29 is a monolignol transporter involved in lignin biosynthesis. *Current Biol.* **22**, 1207–1212.

Allwood, J.W., Ellis, D.I., Heald, J.K., Goodacre, R. and Mur, L.A. (2006) Metabolomic approaches reveal that phosphatidic and phosphatidyl glycerol phospholipids are major discriminatory non-polar metabolites in responses by Brachypodium distachyon to challenge by Magnaporthe grisea. *Plant J.* **46**, 351–368.

Alseekh, S., Tohge, T., Wendenberg, R. *et al.* (2015) Identification and mode of inheritance of quantitative trait loci for secondary metabolite abundance in tomato. *Plant Cell*, **27**, 485–512.

Alseekh, S., Tong, H., Scossa, F., Brotman, Y., Vigroux, F., Tohge, T., Ofner, I., Zamir, D., Nikoloski, Z. and Fernie, A.R. (2017) Canalization of tomato fruit metabolism. *Plant Cell*, **29**, 2753–2765.

Angelovici, R., Batushansky, A., Deason, N., Gonzalez-Jorge, S., Gore, M.A., Fait, A. and DellaPenna, D. (2017) Network-guided GWAS improves identification of genes affecting free amino acids. *Plant Physiol.* **173**, 872–886.

Araujo, W.L., Tohge, T., Ishizaki, K., Leaver, C.J. and Fernie, A.R. (2011) Protein degradation - an alternative respiratory substrate for stressed plants. *Trends Plant Sci.* **16**, 489–498.

Arrivault, S., Obata, T., Szecowka, M., Mengin, V., Guenther, M., Hoehne, M., Fernie, A.R. and Stitt, M. (2017) Metabolite pools and carbon flow during C4 photosynthesis in maize: 13CO2 labeling kinetics and cell type fractionation. *J. Exp. Bot.* **68**, 283–298.

Beleggia, R., Rau, D., Laido, G. *et al.* (2016) Evolutionary metabolomics reveals domestication-associated changes in tetraploid wheat kernels. *Mol. Biol. Evol.* **33**, 1740–1753.

Bellucci, E., Bitocchi, E., Ferrarini, A. *et al.* (2014) Decreased nucleotide and expression diversity and modified coexpression patterns characterize domestication in the common bean. *Plant Cell*, **26**, 1901–1912.

Bentsink, L., Alonso-Blanco, C., Vreugdenhil, D., Tesnier, K., Groot, S.P.C. and Koornneef, M. (2000) Genetic analysis of seed-soluble oligosaccharides in relation to seed storability of Arabidopsis. *Plant Physiol.* **124**, 1595–1604.

Bovy, A., de Vos, R., Kemper, M. *et al.* (2002) High-flavonol tomatoes resulting from heterologous expression of the maize transcription factor gene LC and C1. *Plant Cell*, **14**, 2509–2526.

Cajka, T. and Fiehn, O. (2014) Comprehensive analysis of lipids in biological systems by liquid chromatography-mass spectrometry. *Trac-Trends in Analytical Chemistry*, **61**, 192–206.

Cao, D., Lutz, A., Hill, C.B., Callahan, D.L. and Roessner, U. (2017) A quantitative profiling method of phytohormones and other metabolites applied to barley roots subjected to salinity stress. *Front Plant Sci.* **7**, 2070.

Carreno-Quintero, N., Acharjee, A., Maliepaard, C., Bachem, C.W.B., Mumm, R., Bouwmeester, H., Visser, R.G.F. and Keurentjes, J.J.B. (2012) Untargeted metabolic quantitative trait loci analyses reveal a relationship between primary metabolism and potato tuber quality. *Plant Physiol.* **158**, 1306–1318.

Catchpole, G.S., Beckmann, M., Enot, D.P. *et al.* (2005) Hierarchical metabolomics demonstrates substantial compositional similarity between genetically modified and conventional potato crops. *Proc. Natl Acad. Sci. USA*, **102**, 14458–14462.

Chan, E.K.F., Rowe, H.C. and Kliebenstein, D.J. (2010) Understanding the evolution of defense metabolites in Arabidopsis thaliana using genome-wide association mapping. *Genetics*, **185**, 991–1007.

Chan, E.K.F., Rowe, H.C., Corwin, J.A., Joseph, B. and Kliebenstein, D.J. (2011) Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in Arabidopsis thaliana. *PLoS Biol.* **9**, e1001125.

Chen, F., Duran, A.L., Blount, J.W., Sumner, L.W. and Dixon, R.A. (2003) Profiling phenolic metabolites in transgenic alfalfa modified in lignin biosynthesis. *Phytochemistry*, **64**, 1013–1021.

Dal Cin, V., Tieman, D.M., Tohge, T. *et al.* (2011) Identification of genes in the phenylalanine metabolic pathway by ectopic expression of a MYB transcription factor in tomato fruit. *Plant Cell*, **23**, 2738–2753.

D'Auria, J.C. and Gershenzon, J. (2005) The secondary metabolism of Arabidopsis thaliana: Growing like a weed. *Curr. Opin. Plant Biol.* **8**, 308–316.

De Vos, R.C.H., Moco, S., Lommen, A., Keurentjes, J.J.B., Bino, R.J. and Hall, R.D. (2007) Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* **2**, 778–791.

Dixon, R.A. and Strack, D. (2003) Phytochemistry meets genome analysis, and beyond. *Phytochemistry*, **62**, 815–816.

Do, P.T., Prudent, M., Sulpice, R., Causse, M. and Fernie, A.R. (2010) The influence of fruit load on the tomato pericarp metabolome in a Solanum chmielewskii introgression line population. *Plant Physiol.* **154**, 1128–1142.

Dong, Y., Li, B. and Aharoni, A. (2016) More than pictures: When MS imaging meets histology. *Trends Plant Sci.* **21**, 686–698.

Dong, Y., Li, B., Malitsky, S., Rogachev, L., Aharoni, A., Kaftan, F., Svatos, A. and Franceschi, P. (2016) Sample preparation for mass spectrometry imaging of plant tissues: A review. *Front Plant Sci.* **7**, 60.

Duenas, M.E., Klein, A.T., Alexander, L.E., Yandeau-Nelson, M.D., Nikolau, B.J. and Lee, Y.J. (2017) High spatial resolution mass spectrometry imaging reveals the genetically programmed, developmental modification of the distribution of thylakoid membrane lipids among individual cells of maize leaf. *Plant J.* **89**, 825–838.

Fan, P.X., Miller, A.M., Schilmiller, A.L., Liu, X.X., Ofner, I., Jones, A.D., Zamir, D. and Last, R.L. (2016) *In vitro* reconstruction and analysis of evolutionary variation of the tomato acylsucrose metabolic network. *Proc. Natl Acad. Sci. USA*, **113**, E239–E248.

Feldberg, L., Venger, I., Malitsky, S., Rogachev, I. and Aharoni, A. (2009) Dual labeling of metabolites for metabolome analysis (DLEMMA): a new approach for the identification and relative quantification of metabolites by means of dual isotope labeling and liquid chromatography-mass spectrometry. *Anal. Chem.* **81**, 9257–9266.

Fernie, A.R. and Tohge, T. (2017) The genetics of plant metabolism. *Annu. Rev. Genet.* **51**, 287–310.

Fernie, A.R., Trethewey, R.N., Krotzky, A.J. and Willmitzer, L. (2004) Innovation - Metabolite profiling: From diagnostics to systems biology. *Nat. Rev. Mol. Cell Biol.* **5**, 763–769.

Feussner, I. and Polle, A. (2015) What the transcriptome does not tell - proteomics and metabolomics are closer to the plants' patho-phenotype. *Curr. Opin. Plant Biol.* **26**, 26–31.

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

Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R.N. and Willmitzer, L. (2000) Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* **18**, 1157–1161.

Fiehn, O., Putri, S.P., Saito, K., Salek, R.M. and Creek, D.J. (2015) Metabolomics continues to expand: highlights from the 2015 metabolomics conference. *Metabolomics*, **11**, 1036–1040.

Figueroa, C.M. and Lunn, J.E. (2016) A Tale of Two Sugars: Trehalose 6-Phosphate and Sucrose(1 OPEN). *Plant Physiol.* **172**, 7–27.

Fraser, P.D., Pinto, M.E.S., Holloway, D.E. and Bramley, P.M. (2000) Application of high-performance liquid chromatography with photodiode array detection to the metabolic profiling of plant isoprenoids. *Plant J.* **24**, 551–558.

Fraser, P.D., Enfissi, E.M., Goodfellow, M., Eguchi, T. and Bramley, P.M. (2007) Metabolite profiling of plant carotenoids using the matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Plant J.* **49**, 552–564.

Fridman, E. and Pichersky, E. (2005) Metabolomics, genomics, proteomics, and the identification of enzymes and their substrates and products. *Curr. Opin. Plant Biol.* **8**, 242–248.

Gaquerel, E., Gulati, J. and Baldwin, I.T. (2014) Revealing insect herbivory-induced phenolamide metabolism: from single genes to metabolic network plasticity analysis. *Plant J.* **79**, 679–692.

Geu-Flores, F., Sherden, N.H., Courdavault, V., Burlat, V., Glenn, W.S., Wu, C., Nims, E., Cui, Y. and O'Connor, S.E. (2012) An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. *Nature*, **492**, 138–142.

Giavalisco, P., Li, Y., Matthes, A., Eckhardt, A., Hubberten, H.M., Hesse, H., Segu, S., Hummel, J., Kohl, K. and Willmitzer, L. (2011) Elemental formula annotation of polar and lipophilic metabolites using (13) C, (15) N and (34) S isotope labelling, in combination with high-resolution mass spectrometry. *Plant J.* **68**, 364–376.

Giovannoni, J. (2018) Tomato multiomics reveals consequences of crop domestication and improvement. *Cell*, **172**, 6–8.

Guo, A.C., Jewison, T., Wilson, M., Liu, Y.F., Knox, C., Djoumbou, Y., Lo, P., Mandal, R., Krishnamurthy, R. and Wishart, D.S. (2013) ECMDB: the E-coli metabolome database. *Nucleic Acids Res.* **41**, D625–D630.

Hill, C.B., Taylor, J.D., Edwards, J., Mather, D., Bacic, A., Langridge, P. and Roessner, U. (2013) Whole-genome mapping of agronomic and metabolic traits to identify novel quantitative trait loci in bread wheat grown in a water-limited environment. *Plant Physiol.* **162**, 1266–1281.

Hirai, M.Y., Klein, M., Fujikawa, Y. *et al.* (2005) Elucidation of gene-to-gene and metabolite-to-gene networks in Arabidopsis by integration of metabolomics and transcriptomics. *J. Biol. Chem.* **280**, 25590–25595.

Horai, H., Arita, M., Kanaya, S. *et al.* (2010) MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* **45**, 703–714.

Joseph, B., Corwin, J.A. and Kliebenstein, D.J. (2015a) Genetic variation in the nuclear and organellar genomes modulates stochastic variation in the metabolome, growth, and defense. *PLoS Genet.* **11**, e1004779.

Joseph, B., Lau, L. and Kliebenstein, D.J. (2015b) Quantitative variation in responses to root spatial constraint within Arabidopsis thaliana. *Plant Cell*, **27**, 2227–2243.

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

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

Kind, T. and Fiehn, O. (2007) Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics*, **8**, 105.

Kind, T., Liu, K.H., Lee, D.Y., DeFelice, B., Meissen, J.K. and Fiehn, O. (2013) LipidBlast *in silico* tandem mass spectrometry database for lipid identification. *Nat. Methods*, **10**, 755–758.

Kliebenstein, D.J. (2013a) Making new molecules - evolution of structures for novel metabolites in plants. *Curr. Opin. Plant Biol.* **16**, 112–117.

Kliebenstein, D.J. (2013b) New synthesis-regulatory evolution, the veiled world of chemical diversification. *J. Chem. Ecol.* **39**, 349–349.

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

Kochevenko, A. and Fernie, A.R. (2011) The genetic architecture of branched-chain amino acid accumulation in tomato fruits. *J. Exp. Bot.* **62**, 3895–3906.

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

Kopka, J., Schauer, N., Krueger, S. *et al.* (2005) GMDCSB.DB: the Golm metabolome database. *Bioinformatics*, **21**, 1635–1638.

Kruger, N.J., Troncoso-Ponce, M.A. and Ratcliffe, R.G. (2008) H-1 NMR metabolite fingerprinting and metabolomic analysis of perchloric acid extracts from plant tissues. *Nat. Protoc.* **3**, 1001–1012.

Lai, Z.J., Tsugawa, H., Wohlgemuth, G. *et al.* (2018) Identifying metabolites by integrating metabolome databases with mass spectrometry cheminformatics. *Nat. Methods*, **15**, 53–56.

Lippman, Z.B., Semel, Y. and Zamir, D. (2007) An integrated view of quantitative trait variation using tomato interspecific introgression lines. *Curr. Opin. Genet. Dev.* **17**, 545–552.

Lisec, J., Schauer, N., Kopka, J., Willmitzer, L. and Fernie, A.R. (2006) Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nat. Protoc.* **1**, 387–396.

Lisec, J., Meyer, R.C., Steinfath, M. *et al.* (2008) Identification of metabolic and biomass QTL in Arabidopsis thaliana in a parallel analysis of RIL and IL populations. *Plant Journal*, **53**, 960–972.

Loudet, O., Chaillou, S., Merigout, P., Talbotec, J. and Daniel-Vedele, F. (2003) Quantitative trait loci analysis of nitrogen use efficiency in Arabidopsis. *Plant Physiol.* **131**, 345–358.

Lu, W., Su, X., Klein, M.S., Lewis, I.A., Fiehn, O. and Rabinowitz, J.D. (2017) Metabolite measurement: pitfalls to avoid and practices to follow. *Annu. Rev. Biochem.* **86**, 277–304.

Lugan, R., Niogret, M.F., Leport, L., Guegan, J.P., Larher, F.R., Savoure, A., Kopka, J. and Bouchereau, A. (2010) Metabolome and water homeostasis analysis of Thellungiella salsuginea suggests that dehydration tolerance is a key response to osmotic stress in this halophyte. *Plant J.* **64**, 215–229.

Malinovsky, F.G., Thomsen, M.L.F., Nintemann, S.J., Jagd, L.M., Bourgine, B., Burow, M. and Kliebenstein, D.J. (2017) An evolutionarily young defense metabolite influences the root growth of plants via the ancient TOR signaling pathway. *Elife* **6**, pii: e29353.

Matsuda, F., Okazaki, Y., Oikawa, A., Kusano, M., Nakabayashi, R., Kikuchi, J., Yonemaru, J.I., Ebana, K., Yano, M. and Saito, K. (2012) Dissection of genotype-phenotype associations in rice grains using metabolome quantitative trait loci analysis. *Plant J.* **70**, 624–636.

Matsuda, F., Nakabayashi, R., Yang, Z.G., Okazaki, Y., Yonemaru, J., Ebana, K., Yano, M. and Saito, K. (2015) Metabolome-genome-wide association study dissects genetic architecture for generating natural variation in rice secondary metabolism. *Plant J.* **81**, 13–23.

Meyer, R.C., Steinfath, M., Lisec, J. *et al.* (2007) The metabolic signature related to high plant growth rate in Arabidopsis thaliana. *Proc. Natl Acad. Sci. USA*, **104**, 4759–4764.

Nakabayashi, R., Yonekura-Sakakibara, K., Urano, K. *et al.* (2014) Enhancement of oxidative and drought tolerance in Arabidopsis by overaccumulation of antioxidant flavonoids. *Plant J.*, **77**, 367–379.

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

Ning, J., Moghe, G.D., Leong, B., Kim, J., Ofner, I., Wang, Z.Z., Adams, C., Jones, A.D., Zamir, D. and Last, R.L. (2015) A feedback-insensitive isopropylmalate synthase affects acylsugar composition in cultivated and wild tomato. *Plant Physiol.* **169**, 1821–1835.

Ofner, I., Lashbrooke, J., Pleban, T., Aharoni, A. and Zamir, D. (2016) Solanum pennellii backcross inbred lines (BILs) link small genomic bins with tomato traits. *Plant J.* **87**, 151–160.

Oliver, S.G., Winson, M.K., Kell, D.B. and Baganz, F. (1998) Systematic functional analysis of the yeast genome. *Trends Biotechnol.* **16**, 373–378.

Peng, M., Shahzad, R., Gul, A. *et al.* (2017) Differentially evolved glucosyltransferases determine natural variation of rice flavone accumulation and UV-tolerance. *Nat. Commun.* **8**, 1975.

Persson, S., Wei, H.R., Milne, J., Page, G.P. and Somerville, C.R. (2005) Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. *Proc. Natl Acad. Sci. USA*, **102**, 8633–8638.

Piazza, I., Kochanowski, K., Cappelletti, V., Fuhrer, T., Noor, E., Sauer, U. and Picotti, P. (2018) A map of protein-metabolite interactions reveals principles of chemical communication. *Cell*, **172**, 358–372.e23.

Pichersky, E. and Raguso, R.A. (2016) Why do plants produce so many terpenoid compounds? *New Phytol.* 1–11. https://doi.org/10.1111/nph.14178

Pick, T.R., Brautigam, A., Schulz, M.A., Obata, T., Fernie, A.R. and Weber, A.P.M. (2013) PLGG1, a plastidic glycolate glycerate transporter, is required for photorespiration and defines a unique class of metabolite transporters. *Proc. Natl Acad. Sci. USA*, **110**, 3185–3190.

Quadrana, L., Almeida, J., Asis, R. *et al.* (2014) Natural occurring epialleles determine vitamin E accumulation in tomato fruits. *Nat. Commun.* **5**, 3027.

Rai, A., Saito, K. and Yamazaki, M. (2017) Integrated omics analysis of specialized metabolism in medicinal plants. *Plant J.* **90**, 764–787.

Roberts, J.K.M. (2000) NMR adventures in the metabolic labyrinth within plants. *Trends Plant Sci.* **5**, 30–34.

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

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

Roessner, U., Willmitzer, L. and Fernie, A.R. (2001b) High-resolution metabolic phenotyping of genetically and environmentally diverse potato tuber systems. Identification of phenocopies. *Plant Physiol.* **127**, 749–764.

Rowe, H.C., Hansen, B.G., Halkier, B.A. and Kliebenstein, D.J. (2008) Biochemical networks and epistasis shape the Arabidopsis thaliana metabolome. *Plant Cell*, **20**, 1199–1216.

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

Sajed, T., Marcu, A., Ramirez, M., Pon, A., Guo, A.C., Knox, C., Wilson, M., Grant, J.R., Djoumbou, Y. and Wishart, D.S. (2016) ECMDB 2.0: A richer resource for understanding the biochemistry of E. coli. *Nucleic Acids Res.* **44**, D495–D501.

Sarry, J.E., Kuhn, L., Ducruix, C. *et al.* (2006) The early responses of Arabidopsis thaliana cells to cadmium exposure explored by protein and metabolite profiling analyses. *Proteomics*, **6**, 2180–2198.

Sauvage, C., Segura, V., Bauchet, G., Stevens, R., Do, P.T., Nikoloski, Z., Fernie, A.R. and Causse, M. (2014) Genome-wide association in tomato reveals 44 candidate loci for fruit metabolic traits. *Plant Physiol.* **165**, 1120–1132.

Schauer, N., Steinhauser, D., Strelkov, S. *et al.* (2005a) GC-MS libraries for the rapid identification of metabolites in complex biological samples. *FEBS Lett.***579**, 1332–1337.

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

Schauer, N., Semel, Y., Roessner, U. *et al.* (2006) Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nat. Biotechnol.* **24**, 447–454.

Schilmiller, A., Shi, F., Kim, J., Charbonneau, A.L., Holmes, D., Jones, A.D. and Last, R.L. (2010) Mass spectrometry screening reveals widespread diversity in trichome specialized metabolites of tomato chromosomal substitution lines. *Plant J.* **62**, 391–403.

Shahaf, N., Rogachev, I., Heinig, U., Meir, S., Malitsky, S., Battat, M., Wyner, H., Zheng, S., Wehrens, R. and Aharoni, A. (2016) The WEIZMASS spectral library for high-confidence metabolite identification. *Nat. Commun.* **7**, 12423.

Showalter, M.R., Cajka, T. and Fiehn, O. (2017) Epimetabolites: discovering metabolism beyond building and burning. *Curr. Opin. Chem. Biol.* **36**, 70–76.

Sonderby, I.E., Geu-Flores, F. and Halkier, B.A. (2010) Biosynthesis of glucosinolates - gene discovery and beyond. *Trends Plant Sci.* **15**, 283–290.

de Souza, L.P., Naake, T., Tohge, T. and Fernie, A.R. (2017) From chromatogram to analyte to metabolite. How to pick horses for courses from the massive web resources for mass spectral plant metabolomics. *Gigascience*, **6**, 1–20.

Sulpice, R., Pyl, E.T., Ishihara, H. *et al.* (2009) Starch as a major integrator in the regulation of plant growth. *Proc. Natl Acad. Sci. USA*, **106**, 10348–10353.

Sulpice, R., Trenkamp, S., Steinfath, M. *et al.* (2010) Network analysis of enzyme activities and metabolite levels and their relationship to biomass in a large panel of Arabidopsis accessions. *Plant Cell*, **22**, 2872–2893.

Sumner, L.W., Mendes, P. and Dixon, R.A. (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry*, **62**, 817–836.

Sumner, L.W., Lei, Z., Nikolau, B.J. and Saito, K. (2015) Modern plant metabolomics: advanced natural product gene discoveries, improved technologies, and future prospects. *Natural Product Reports*, **32**, 212–229.

Suzuki, H., Reddy, M.S.S., Naoumkina, M., Aziz, N., May, G.D., Huhman, D.V., Sumner, L.W., Blount, J.W., Mendes, P. and Dixon, R.A. (2005) Methyl jasmonate and yeast elicitor induce differential transcriptional and metabolic re-programming in cell suspension cultures of the model legume Medicago truncatula. *Planta*, **220**, 696–707.

Sweetlove, L.J. and Fernie, A.R. (2013) The spatial organization of metabolism within the plant cell. *Annu. Rev. Plant Biol.* **64**, 723–746.

Sweetlove, L.J., Nielsen, J. and Fernie, A.R. (2017) Engineering central metabolism - a grand challenge for plant biologists. *Plant J.* **90**, 749–763.

Templer, S.E., Ammon, A., Pscheidt, D. *et al.* (2017) Metabolite profiling of barley flag leaves under drought and combined heat and drought stress reveals metabolic QTLs for metabolites associated with antioxidant defense. *J. Exp. Bot.* **68**, 1697–1713.

Tenenboim, H. and Brotman, Y. (2016) Omic relief for the biotically stressed: metabolomics of plant biotic interactions. *Trends Plant Sci.* **21**, 781–791.

Thimmappa, R., Geisler, K., Louveau, T., O'Maille, P. and Osbourn, A. (2014) Triterpene biosynthesis in plants. In *Annual Review of Plant Biology, Vol 65* (Merchant, S.S. ed). Palo Alto: Annual Reviews, pp. 225–257.

Tholl, D., Boland, W., Hansel, A., Loreto, F., Rose, U.S. and Schnitzler, J.P. (2006) Practical approaches to plant volatile analysis. *Plant J.* **45**, 540–560.

Tieman, D.M., Zeigler, M., Schmelz, E.A., Taylor, M.G., Bliss, P., Kirst, M. and Klee, H.J. (2006) Identification of loci affecting flavour volatile emissions in tomato fruits. *J. Exp. Bot.* **57**, 887–896.

Tieman, D., Zhu, G.T., Resende, M.F.R. *et al.* (2017) PLANT SCIENCE A chemical genetic roadmap to improved tomato flavor. *Science*, **355**, 391–394.

Tohge, T. and Fernie, A.R. (2009) Web-based resources for mass-spectrometry-based metabolomics: a user's guide. *Phytochemistry*, **70**, 450–456.

Tohge, T. and Fernie, A.R. (2010) Combining genetic diversity, informatics and metabolomics to facilitate annotation of plant gene function. *Nat. Protoc.* **5**, 1210–1227.

Tohge, T., Nishiyama, Y., Hirai, M.Y. *et al.* (2005) Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. *Plant J.* **42**, 218–235.

Tohge, T., Alseekh, S. and Fernie, A.R. (2014) On the regulation and function of secondary metabolism during fruit development and ripening. *J. Exp. Bot.* **65**, 4599–4611.

Tohge, T., Wendenburg, R., Ishihara, H. *et al.* (2016) Characterization of a recently evolved flavonol-phenylacyltransferase gene provides signatures of natural light selection in Brassicaceae. *Nat. Commun.* **7**, 12399.

Trenkamp, S., Eckes, P., Busch, M. and Fernie, A.R. (2009) Temporally resolved GC-MS-based metabolic profiling of herbicide treated plants treated reveals that changes in polar primary metabolites alone can distinguish herbicides of differing mode of action. *Metabolomics*, **5**, 277–291.

Urbanczyk-Wochniak, E. and Fernie, A.R. (2005) Metabolic profiling reveals altered nitrogen nutrient regimes have diverse effects on the metabolism of hydroponically-grown tomato (Solanum lycopersicum) plants. *J. Exp. Bot.* **56**, 309–321.

Verpoorte, R. and Alfermann, A.W. (2000) *Metabolic Engineering of the Plant Secondary Metabolism*. New York: Springer, pp. 1–29.

Veyel, D., Kierszniowska, S., Kosmacz, M., Sokolowska, E.M., Michaelis, A., Luzarowski, M., Szlachetko, J., Willmitzer, L. and Skirycz, A. (2017) System-wide detection of protein-small molecule complexes suggests extensive metabolite regulation in plants. *Sci. Rep.* **7**, 42387.

Wang, Z. and Jones, A.D. (2014) Profiling of stable isotope enrichment in specialized metabolites using liquid chromatography and multiplexed nonselective collision-induced dissociation. *Anal. Chem.* **86**, 10600–10607.

Ward, J.L., Baker, J.M., Llewellyn, A.M., Hawkins, N.D. and Beale, M.H. (2011) Metabolomic analysis of Arabidopsis reveals hemiterpenoid glycosides as products of a nitrate ion-regulated, carbon flux overflow. *Proc. Natl Acad. Sci. USA*, **108**, 10762–10767.

Weckwerth, W. and Fiehn, O. (2002) Can we discover novel pathways using metabolomic analysis? *Curr. Opin. Biotechnol.* **13**, 156–160.

Wen, W.W., Li, D., Li, X. *et al.* (2014) Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nat. Commun.* **5**, 3438.

Wen, W.W., Brotman, Y., Willmitzer, L., Yan, J.B. and Fernie, A.R. (2016) Broadening our portfolio in the genetic improvement of maize chemical composition. *Trends Genet.* **32**, 459–469.

Wen, W., Jin, M., Li, K. *et al.* (2018) An integrated multi-layered analysis of the metabolic networks of different tissues uncovers key genetic components of primary metabolism in maize. *Plant J.* **93**, 1116–1128.

Wink, M. (2004). Phytochemical diversity of secondary metabolites. In *Encyclopedia of Plant and Crop Science* (Goodman, R.M., ed.). New York: Marcel Dekker, pp. 915–919.

Wink, M. (2015) Modes of action of herbal medicines and plant secondary metabolites. *Medicines (Basel, Switzerland)*, **2**, 251–286.

Wishart, D.S., Feunang, Y.D., Marcu, A. *et al.* (2018) HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.* **46**, D608–D617.

Wu, S., Alseekh, S., Cuadros-Inostroza, A., Fusari, C.M., Mutwil, M., Kooke, R., Keurentjes, J.B., Fernie, A.R., Willmitzer, L. and Brotman, Y. (2016) Combined use of genome-wide association data and correlation networks unravels key regulators of primary metabolism in Arabidopsis thaliana. *PLoS Genet.* **12**, e1006363.

Wu, S., Tohge, T., Cuadros-Inostroza, A. *et al.* (2018) Mapping the Arabidopsis metabolic landscape by untargeted metabolomics at different environmental conditions. *Molecular Plant*, **11**, 118–134.

Ye, J., Wang, X., Hu, T.X. *et al.* (2017) An InDel in the promoter of Al-ACTIVATED MALATE TRANSPORTER9 selected during tomato domestication determines fruit malate contents and aluminum tolerance. *Plant Cell*, **29**, 2249–2268.

Zhu, G.T., Wang, S.C., Huang, Z.J. *et al.* (2018) Rewiring of the fruit metabolome in tomato breeding. *Cell*, **172**, 249–261.e12.

Ziegler, J. and Facchini, P.J. (2008) Alkaloid biosynthesis: metabolism and trafficking. *Annu. Rev. Plant Biol.* **59**, 735–769.