








COMMENTARY

# The renaissance of comparative biochemistry

Stacey D. Smith<sup>1,8</sup> , Ruthie Angelovici<sup>2</sup> , Karolina Heyduk<sup>3</sup>, Hiroshi A. Maeda<sup>4</sup> , Gaurav D. Moghe<sup>5</sup> , J. Chris Pires<sup>2</sup> ,  
Joshua R. Widhalm<sup>6</sup> , and Jennifer H. Wisecaver<sup>7</sup> 

Manuscript received 15 October 2018; revision accepted 4 December 2018.

<sup>1</sup> Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO, USA

<sup>2</sup> Division of Biological Sciences, University of Missouri, Columbia, MO, USA

<sup>3</sup> Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA

<sup>4</sup> Department of Botany, University of Wisconsin-Madison, Madison, WI, USA

<sup>5</sup> Plant Biology Section, School of Integrative Plant Sciences, Cornell University, Ithaca, NY, USA

<sup>6</sup> Department of Horticulture and Landscape Architecture and Center for Plant Biology, Purdue University, West Lafayette, IN, USA

<sup>7</sup> Department of Biochemistry and Center for Plant Biology, Purdue University, West Lafayette, IN, USA

<sup>8</sup> Author for correspondence (e-mail: stacey.d.smith@colorado.edu)

**Citation:** Smith, S. D., R. Angelovici, K. Heyduk, H. A. Maeda, G. D. Moghe, J. C. Pires, J. R. Widhalm, and J. H. Wisecaver. 2019. The renaissance of comparative biochemistry. *American Journal of Botany* 106(1): 3–13.

doi:10.1002/ajb2.1216

The study of chemical diversity has a long history in the field of plant systematics. Methods for separating and identifying chemical compounds, such as chromatography and mass spectrometry, were developed in the early 1900s (reviewed by Marston, 2007), with plant pigments such as chlorophylls and carotenoids among the first targets of analysis (Twiss, 1906). These technological advances gave way to tailored approaches designed to survey major classes of plant metabolites (e.g., phenolics, terpenoids, sugars) across diverse taxa (Harborne, 1973). The techniques developed by Harborne and others could often be implemented with relatively simple equipment and materials, opening the door for many botanists to begin exploring chemical diversity in their own systems. Indeed, by the 1950s, studies in this area had emerged as a distinct discipline, “biochemical systematics” or “chemosystematics” (Alston and Turner, 1963; Gibbs, 1963, 1974).

With a sudden wealth of new comparative data, the field of chemosystematics put forward new phylogenetic hypotheses, many of which have come to be strongly supported with other sources of information (Grayer et al., 1999). One classic example is that of the Gesneriaceae, in which early phytochemical surveys supported the division of the Old World Cyrtandroideae and the

mostly New World Gesnerioideae (including the South America/South Pacific disjunct Coronanthereae). The former subfamily produces anthocyanins and chalcones or auronones, while the latter produces novel 3-deoxyanthocyanins and flavones (Harborne, 1966, 1967). Subsequent molecular systematics studies have supported the monophyly of Gesnerioideae (Smith et al., 1997; Woo et al., 2011), consistent with a single origin of 3-deoxyanthocyanins from anthocyanin-producing ancestors.

As comparative biochemical studies expanded, many classes of plant metabolites were found to exhibit complex evolutionary histories, with multiple gains and losses across the phylogeny. For instance, mustard oils, the glucosinolate-derived compounds involved in defense against herbivores, are found in 15 plant families and appear to have evolved twice, once in the Malpighiales and once in the Brassicales, with a secondary loss in one lineage (Ettlinger and Kjaer, 1968; Rodman et al., 1998). The evolutionary pattern presented by betalain pigments is even more intriguing, wherein following a single origin near the base of the Caryophyllales, multiple lineages have lost these nitrogenous compounds and reverted to the ancestral state of producing anthocyanin pigments, seemingly after millions of years without them (Brockington et al., 2015;

Lopez-Nieves et al., 2018). Thus, even specialized metabolites assembled via intricate and novel pathways show substantial lability over broad evolutionary timescales.

The explosion of functional and comparative plant genomics, beginning with the publication of the *Arabidopsis thaliana* genome (Wortman et al., 2003), has allowed for fundamental advances in understanding the mechanisms responsible for novel metabolites and their variation across taxa. By mapping plant chemical variation onto our ever-improving phylogeny for land plants, we can localize branches of the tree along which plant chemistry has shifted. Combining this framework with targeted biochemical and genomic studies, we have begun to connect metabolic innovations with changes in pathway structure, enzyme function, and in some cases, even precise molecular changes (e.g., Huang et al., 2016; Schenck et al., 2017). The results of this highly integrative field have overturned traditional notions about biochemical evolution, while underscoring the importance of genomic processes, such as whole-genome duplication and horizontal gene transfer, in driving metabolic diversification. In this review, we trace the path of this renaissance in comparative biochemistry, considering how the integration of genomic and functional studies in a robust phylogenetic framework has brought new insights into both metabolic evolution and plant diversification. In closing, we reflect on the challenging questions that remain to be addressed in comparative biochemistry and how our teaching of botany can help to train the next generation of botanists to meet these challenges.

## CONNECTING METABOLIC PATHWAYS TO GENES

The first key step in tracing the evolution of metabolic diversity is identifying the structural (enzyme-encoding) genes that are the building blocks of biochemical pathways. The earliest genetic studies, beginning with Mendel's peas, suggested that many biochemical traits have a simple genetic architecture (Reid and Ross, 2011). The development of model systems for plant molecular biology in the 1980s and 1990s allowed scientists to begin to identify key enzymes that control biochemical phenotypes, such as starch accumulation, tannin production, and flower color (Martin et al., 1987; Lin et al., 1988; Shirley et al., 1992). Moreover, parallel work in a range of model systems, such as *Arabidopsis*, petunia, and maize, resulted in the discovery of deeply conserved biochemical pathways across plants, pointing to their origin early in plant evolution (Rausher, 2006; Weng, 2014).

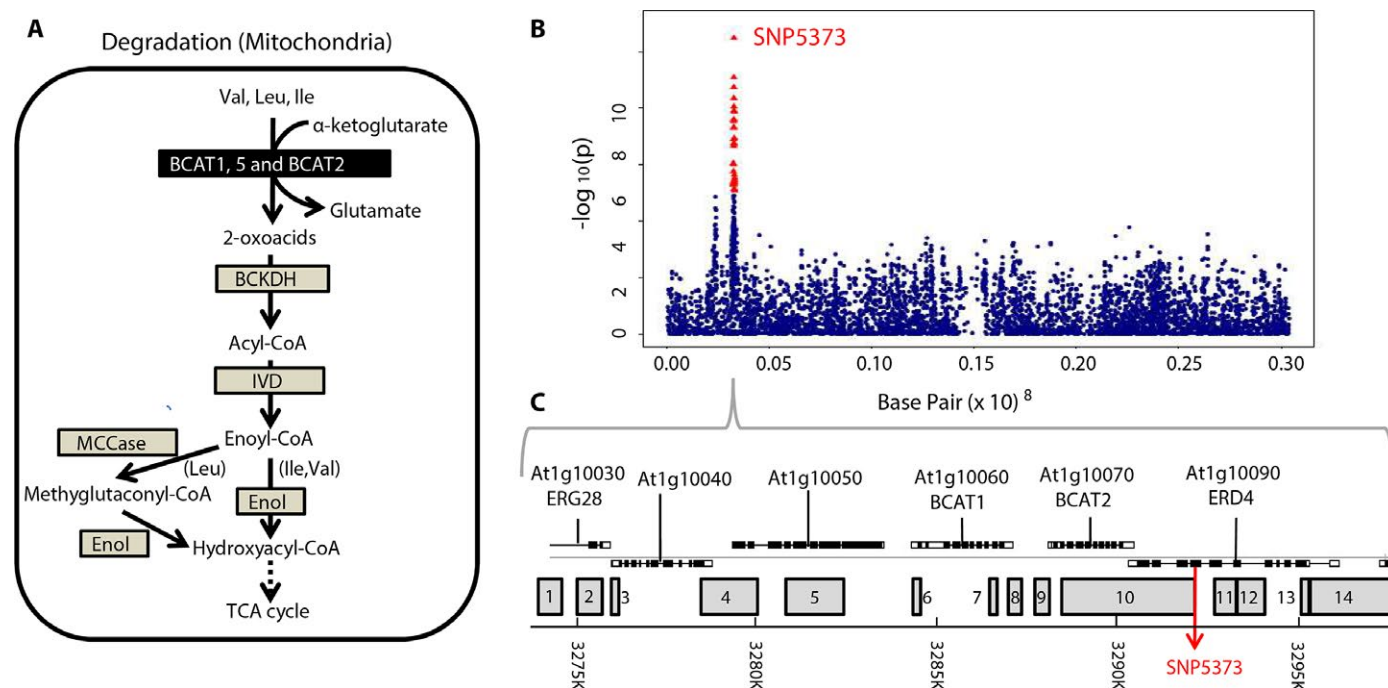
Combining molecular and genomic tools with traditional genetic approaches, plant scientists now have a vast arsenal for connecting biochemical variation to the underlying genetics. The revolutions in genomics and gene editing technologies have created an array of bottom-up approaches (reverse genetics) and top-down approaches (screening metabolic diversity across populations), which can be integrated with systems biology analyses of plant metabolism. Complementary studies that use traditional quantitative genetic approaches (crosses, recombinant inbred and introgression lines) and genome-wide association studies have identified thousands of metabolic trait loci. For example, forward genetic approaches have proved to be very successful in uncovering quantitative trait loci (QTL) controlling variation in multiple classes of metabolites, including amino acids (Fig. 1), organic acids, glucosinolates, carotenoids, and tocochromanols (Wentzell et al., 2007; Chander et al., 2008; Maloney et al., 2010; Yan et al., 2010; Kochevniko and Fernie, 2011; Wurtzel et al., 2012; Angelovici et al., 2013).

Perhaps the greatest challenge in uncovering the genetic basis for biochemical variation lies in characterizing the metabolites themselves. The plant kingdom is estimated to house over a million unique metabolites (Afendi et al., 2012), which vary qualitatively and quantitatively across species. Identification of these compounds continues to rely on relatively specialized and labor-intensive methods such as gas or liquid chromatography coupled with mass spectrometry (GC-MS, LC-MS) or nuclear magnetic resonance (NMR). Nonetheless, comparative metabolomic studies are continually building the repositories of characterized plant compounds to serve as reference libraries while new theoretical and computational approaches are increasingly aimed at creating predictive models of pathways and their products (Wang et al., 2016; Schymanski et al., 2017; Moghe and Kruse, 2018). By linking these tools for identifying plant metabolites with the currently vast repertoire of statistical and genomic approaches, plant biologists are entering a new era for investigating both the proximate and ultimate causes of biochemical diversity (Prasad et al., 2012; Sedio, 2017).

## REVISITING AND REVISING CLASSIC CONCEPTS IN PLANT BIOCHEMISTRY

One of the most fundamental concepts in plant biochemistry is the division between so-called primary and secondary metabolism. Primary metabolism (also called central or core metabolism) provides essential metabolites required for plant growth, development, and reproduction. In contrast, secondary metabolism (also called specialized metabolism) is typically described as serving an ecological function, allowing organisms to survive in the context of their abiotic and biotic environments. Given their fundamental role in essential biological processes, the genes responsible for primary metabolite production generally experience strong purifying selection and remain highly conserved across lineages (Bar-Even et al., 2011; Nam et al., 2012), while specialized metabolites are thought to experience relaxed or diversifying selection (Pichersky and Lewinsohn, 2011; Weng, 2014; Moghe and Last, 2015). Nonetheless, recent studies have uncovered surprising cases of primary metabolic diversity, which have begun to challenge and blur this traditional biochemical division.

One of the best-known examples of the diversification of primary metabolism is found in the carbon fixation pathway, where plants use  $C_3$ ,  $C_4$ , or Crassulacean acid metabolism (CAM) photosynthesis.  $C_4$  and CAM have each evolved 30 to 60 times independently (Sage et al., 2011; Edwards and Ogburn, 2012) and are often considered remarkable examples of convergent evolution. Despite the seeming complexity of  $C_4$  and CAM, all the genes involved in these pathways are present and functional in their  $C_3$  ancestors, suggesting that re-wiring of existing enzymes is responsible for the origin of  $C_4$  and CAM. The reorganization of photosynthetic primary metabolism in  $C_4$  and CAM taxa was additionally aided by traits that enabled their recurrent evolution. For example,  $C_4$  preferentially recruits genes that are inferred to have been highly expressed (Moreno-Villena et al., 2018) or induced by light (Burgess et al., 2016) in  $C_3$  ancestors. Promoter sequences necessary for cell-specific expression of  $C_4$  enzymes have been found in numerous  $C_3$  orthologs (Reyna-Llorens and Hibberd, 2017; Reyna-Llorens et al., 2018), suggesting some level of genomic exaptation is involved in the origins of  $C_4$ . Much less is known about CAM evolution, but recent genomic studies provide evidence that re-wiring of existing



**FIGURE 1.** Identification of genetic basis for natural variation in branched chain amino acid (BCAA) levels in *Arabidopsis* seeds. (A) BCAAs (isoleucine (Ile), valine (Val), and leucine (Leu)) are produced in the plastid and degraded (catabolized) in the mitochondrion to release energy through the tricarboxylic acid (TCA) cycle. The first catabolic steps are carried out by a small gene family of BCAA aminotransferases (BCATs), which therefore can mediate BCAA accumulation. Additional pathway enzymes include branched-chain keto acid dehydrogenase (BCKDH), isovaleryl-coenzyme A (CoA) dehydrogenase (IVD); enoyl-CoA hydratase (Enol), and 3-methylcrotonyl-CoA carboxylase (MCCase). (B) SNP5373 on chromosome 1 has a highly significant effect on variation in BCAA levels in seeds. Shown is the manhattan plot for one BCAA seed trait (the ratio of Ile to other BCAAs) from a genomewide association study (GWAS) (Angelovici et al., 2013). The position along chromosome 1 is plotted along the x-axis, and the negative  $\log_{10}$ -transformed  $P$ -values from the GWAS are on the y-axis. Red triangles are statistically significant at 5% FDR. (C) Genomic region surrounding SNP5373. The top section shows gene models (with exons as black boxes and untranslated regions as white boxes) and the orientation (above or below the thin gray line) indicating the strand. The gray boxes below (numbered 1–14) show haploblocks, indicating that SNP5373 is in linkage with two genes, BCAT2 and ERD4. Additional experiments demonstrated BCAT2 to be the causal gene (Angelovici et al., 2013). Figures modified from Angelovici et al. (2013).

transcriptional cascades may be responsible for transitioning from  $C_3$  to CAM (Abraham et al., 2016; Yin et al., 2018). It is therefore likely that both  $C_4$  and CAM evolve relatively frequently, despite their significant alteration to primary plant metabolism, via existing enzymatic pathways, and that certain lineages are enabled for  $C_4$  or CAM evolution.

Another emerging example of primary metabolic diversity comes from studies of amino acid biosynthesis. Amino acids are building blocks of proteins in all organisms and, in plants, also used to make various hormones and natural products (Pratelli and Pilot, 2014; Coruzzi et al., 2015). For example, L-tyrosine is an aromatic amino acid required for synthesis of plastoquinone, isoquinoline alkaloids (e.g., morphine), cyanogenic glycoside (dhurrin), rosmarinic acid, betalain pigments, and many more (Schenck and Maeda, 2018). Phylogeny-guided biochemical characterizations of the key regulatory enzyme of tyrosine biosynthesis, TyrA dehydrogenases, from diverse plant lineages recently revealed that this primary amino acid pathway operates differently in some plant lineages: legumes use two pathways to synthesize tyrosine, while the feedback regulation of tyrosine biosynthesis is relaxed in the TyrA enzymes in the order Caryophyllales (Schenck et al., 2015, 2017; Lopez-Nieves et al., 2018). The de-regulated TyrA enzymes are present in

betalain-producing Caryophyllales, but their expression appears to be downregulated or completely lost in Caryophyllales families that reverted back to anthocyanin pigmentation (e.g., Caryophyllaceae and Molluginaceae, Lopez-Nieves et al., 2018). Thus, the alteration of the primary amino acid precursor supply (e.g., tyrosine biosynthesis) likely played a role in the evolution of the novel specialized metabolic pathway, betalain biosynthesis. It will be interesting to further investigate whether there are other examples of primary metabolic diversification in coordination with their downstream specialized metabolic pathways.

## GENETIC MECHANISMS OF BIOCHEMICAL DIVERSIFICATION

The advent of new genomic, metabolomic, and computational technologies has begun to reveal the many drivers of innovation in plant biochemistry (Table 1). Mutations in the genome such as single nucleotide polymorphisms, indels, domain losses, and gene fusions may affect the coding sequence, expression pattern, or translation efficiency of a given metabolic gene, introducing variation in the downstream metabolic pathways. For example, the production of S-linalool, a monoterpene alcohol responsible for the floral scent of

**TABLE 1.** Mechanisms of evolution of novel functions in plant-specialized metabolic pathways. Listed are example cases in which novel enzyme activities have emerged in plant metabolism, giving rise to new metabolic classes or altering the diversity of an existing metabolite class. Many of these mechanisms are influenced by the process of gene duplication.

Enzyme	Mechanism of evolution	Metabolite class affected	Phylogenetic range <sup>a</sup>	Reference
Acylsugar acyltransferase 4	Expression level change	Acylsugars	<i>Solanum habrochaites</i>	(Kim et al., 2012)
Linalool synthase	Expression domain change	S-linalool (monoterpene alcohol)	<i>Clarkia</i>	(Dudareva et al., 1996)
Acylsugar acyltransferases	Variation in substrate specificity/promiscuity	Acylsugars	Solanaceae	(Moghe et al., 2017)
Indole-3-glycerol phosphate lyase	Alteration of protein-protein interactions	Benzoxazinoid alkaloids	Poaceae	(Gierl and Frey, 2001)
Class III peroxidases	Change in subcellular localization	Phenylpropanoid metabolism	<i>Populus</i>	(Ren et al., 2014)
Arogenate dehydrogenase α	Loss of allosteric regulation	Betalain alkaloids	Caryophyllales	(Lopez-Nieves et al., 2018)
Cis-prenyl transferase/terpene synthase	Gene clustering	Terpenes	Solanaceae	(Matsuba et al., 2013)
Lactobacillus-type DHNA-CoA thioesterase	New gene introduction via horizontal gene transfer	Phenylpropanoids/benzenoids	Across plants	(Widhalm et al., 2012)

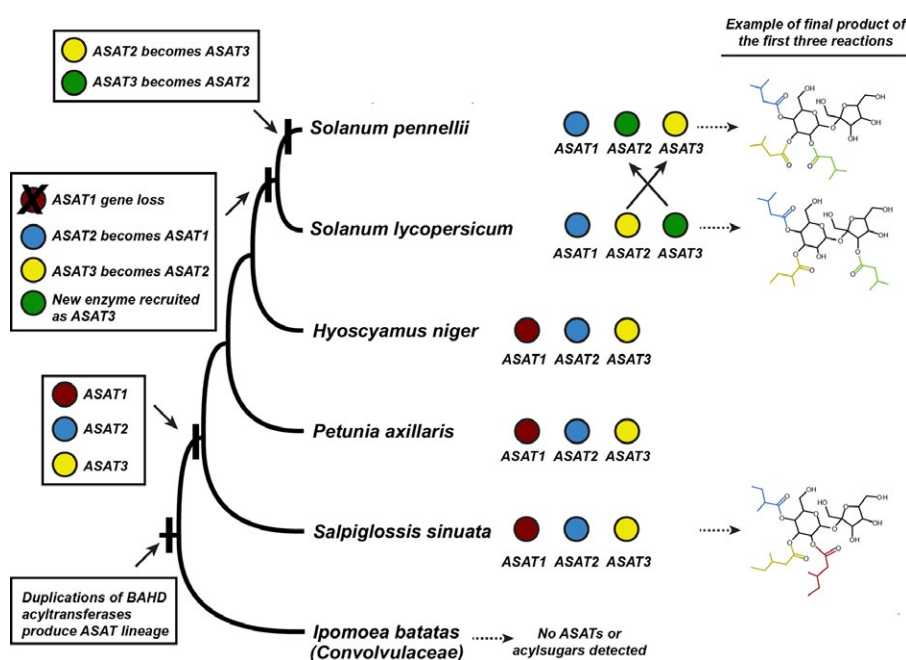
<sup>a</sup>Phylogenetic range, to the extent studied/described in the cited paper.

*Clarkia breweri* flowers, was identified to be a result of expression divergence of its biosynthetic enzyme S-linalool synthase (Dudareva et al., 1996). The mRNA of this enzyme was found to accumulate to a high level in the petals, stigma, style, and stamens of *C. breweri* flowers, but is only weakly detectable in only the stigma of the closest nonscented relative *C. concinna* (Dudareva et al., 1996), leading to very low production of S-linalool and its derivative oxides, which is not enough to attract pollinators (Raguso and Pichersky, 1995). Gene duplication underlies the diversification of many enzyme families such as cytochrome P450s oxidoreductases, hydrolases,

acyltransferases, glycosyltransferases, methyltransferases, and lipases/esterases, which play central roles in specialized metabolism. Duplicated genes may diversify in various ways to produce novel enzymatic functions (Table 1). For example, the repeated rounds of whole-genome duplications in the Brassicales expanded the gene families responsible for glucosinolate biosynthesis and spurred the expansion of these defensive compounds through subsequent functional evolution of paralogous copies of both regulatory and structural genes (Edger et al., 2015). The origin and diversification of acylsugars in the Solanaceae shows a similar pattern. The hundreds

of defensive acylsugars that can be found on a single leaf of some species emerged as a result of a complex interplay of molecular processes, including lineage-specific gene duplications, neo-functionalizations, allelic divergence, and gene losses in a single, large enzyme family (Fan et al., 2017; Moghe et al., 2017) (Fig. 2). Multi-functionality and promiscuity—hallmarks of enzyme families arising out of frequent duplications—were also found to play a role in generating acylsugar diversity (Moghe et al., 2017). These cases support the tight coupling of large- and small-scale changes in genome content with the elaboration of plant metabolic pathways.

While the role of gene duplication in metabolic evolution has long been recognized, the contribution of horizontal gene transfer (HGT), especially in plant systems, has only begun to be explored (Yue et al., 2013). Although distinguishing HGT from other sources of phylogenetic incongruence remains challenging (Kurland et al., 2003), genome-scale analyses are providing increasing support for the involvement of HGT in metabolic evolution (Keeling and Palmer, 2008). One example is that of phenylalanine ammonia lyase (PAL), the enzyme catalyzing the entry step into the phenylpropanoid pathway. This gene is homologous to the histidine ammonia-lyase (HAL) involved in histidine



**FIGURE 2.** Evolution of the first three acylsugar acyltransferase (ASAT) enzymes of the acylsugar biosynthetic pathway and its effects on the acylsugar phenotype in the Solanaceae family. Acyl chain colors in the structures correspond to colors of the ASATs (represented by circles) catalyzing their transfer. Ortholog ASATs have the same color. All ASATs are members of the BAHD acyltransferase enzyme family. Additional ASATs in each species help produce acylsugars with 4–7 acylchains as final products (not shown). The figure represents findings from Moghe et al. (2017) and Fan et al. (2017).



catabolism in prokaryotes and eukaryotes. It was thought that PAL, which contains an additional 207 N-terminal residues, evolved from HAL when fungi and plants diverged from other kingdoms (Ritter and Schulz, 2004). A more recent large-scale phylogenetic analysis, however, raised the prospect that the PAL gene was obtained via HGT from symbiotic soil bacteria or fungi early during the colonization of plants on land (Emiliani et al., 2009). Though one can only speculate on the early selective advantage conferred by PAL, its retention by modern plants for synthesis of vital cofactors (e.g., ubiquinones, Block et al., 2014), hormones (e.g., salicylic acid, Dempsey et al., 2011), lignin, anthocyanins, and thousands of specialized benzenoid/phenylpropanoids suggests that its acquisition played a key role in the evolution of the plant phenylpropanoid metabolic network.

Horizontally acquired genes often acquire new side activities (Näsvall et al., 2012); thus, HGT may result in the immediate introduction of novel metabolic capabilities to the recipient organism. One such occurrence may have come with the replacement of the 1,4-dihydroxy-2-naphthoyl-CoA (DHNA-CoA) thioesterase genes in the plant phyloquinone (vitamin K1) pathway (Fig. 3). Biosynthetic genes for phyloquinone, a one-electron transporter in photosystem I, were likely initially acquired via endosymbiosis from the cyanobacterial progenitor of plastids based on the clustering of menaquinone (vitamin K2) pathway genes in the plastids of certain red algae (van Oostende et al., 2011). Unlike their Rhodophyta counterparts, however, plant DHNA-CoA thioesterases are not of cyanobacterial ancestry. They are instead closely related to Lactobacillales orthologs encoded within clusters of menaquinone biosynthetic genes, suggesting they were obtained from an event of HGT (Widhalm et al., 2012) (Fig. 3C). Remarkably, both the cyanobacterial and plant/Lactobacillales thioesterases are selective for DHNA-CoA as a substrate, though their selectivity is achieved using divergent binding strategies (Furt et al., 2013). Moreover, while the cyanobacterial-type DHNA-CoA thioesterase strictly uses DHNA-CoA as the substrate (Widhalm et al., 2009), the Lactobacillales-type DHNA-CoA thioesterases in plants, and orthologs in *Escherichia coli*, are able to use a range of aromatic acyl-CoA substrates (Widhalm et al., 2012; Chen et al., 2013). Co-expression analysis in *Arabidopsis thaliana* further indicates that DHNA-CoA thioesterases dually function in the metabolism of benzenoid/phenylpropanoid pathway intermediates (Adebesin et al., 2018). Taken together, these studies suggest that acquisition of Lactobacillales-type DHNA-CoA thioesterases replaced cyanobacterial-type DHNA-CoA thioesterases while at the same time this HGT event may have bestowed novel activities for regulating benzenoid/phenylpropanoid metabolism. Plants gain function through duplication of genes and larger genetic units, including linkage groups, chromosomes or even genomes. Another a priori consequence of acquiring genes via HGT is that, if they fulfill the function(s) of genes in existing pathways, the replaced genes are free to evolve new functions. Indeed, vestiges of cyanobacterial-type DHNA-CoA thioesterases appear to remain in plant genomes (Widhalm et al., 2009), and though their larger functions remain enigmatic, orthologs characterized in *Solanum lycopersicum* can act as methylketone synthases for biosynthesis of 3-ketoacid volatiles (Yu et al., 2010).

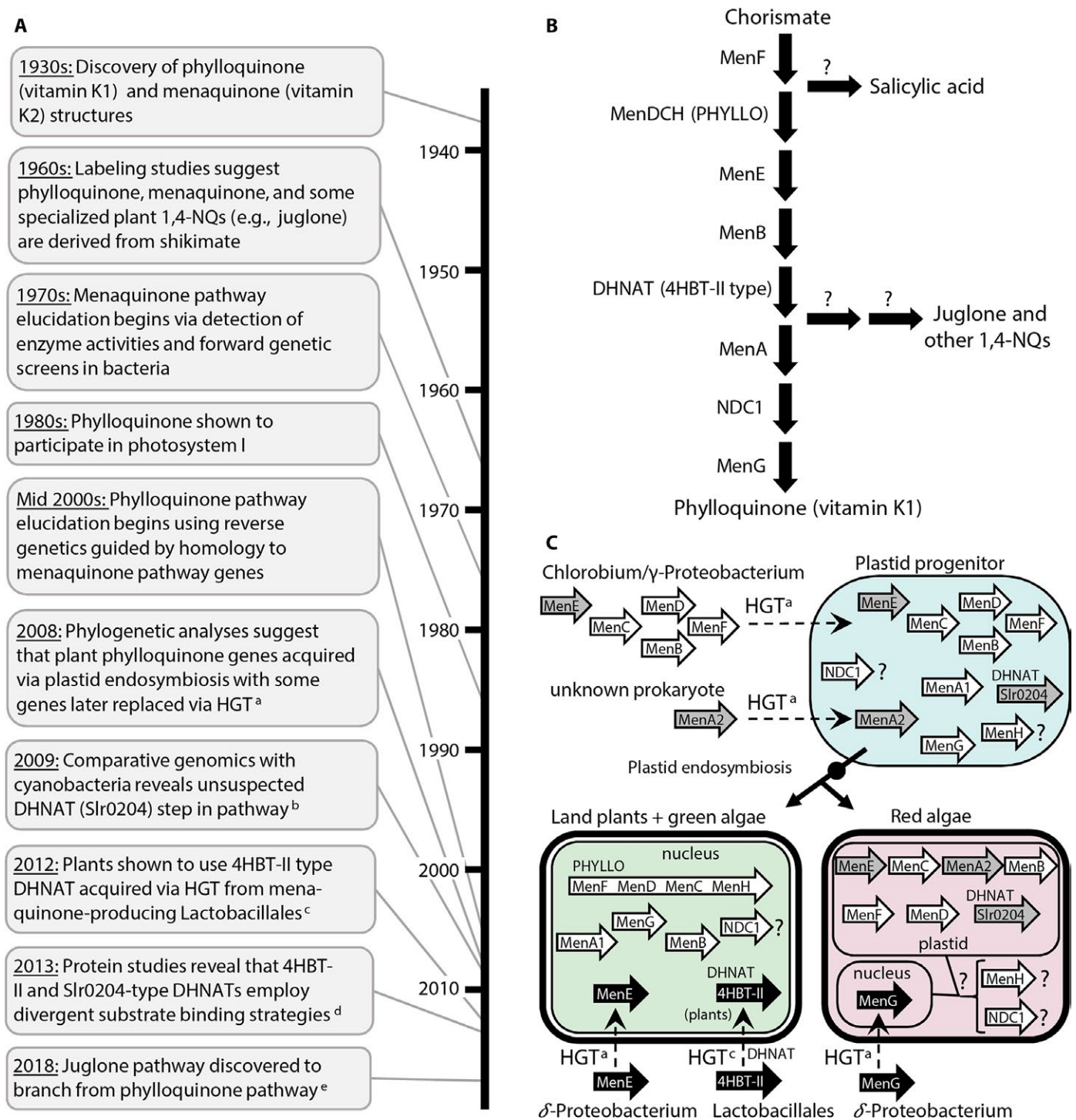
Although functional divergence is often the fate for new genes arising from gene duplication, HGT, or other mechanisms, convergent evolution is an alternate and potentially common outcome (Pichersky and Lewinsohn, 2011; Weng and Noel, 2013).

Convergent evolution at the metabolic level entails different lineages independently evolving pathways to produce identical or nearly identical compounds. Classic examples include the independent evolution of betalain biosynthesis in basidiomycete fungi and Caryophyllales (Musso, 1979; Mueller et al., 1997), caffeine biosynthesis in Malvaceae, Rubiaceae and Theaceae (Yoneyama et al., 2006; Denoeud et al., 2014), and specialized 1,4-naphthoquinones spread across several disparate plant orders (Kumara et al., 1985; Hook et al., 2014). In a few cases, the genes responsible for these convergent gains have been identified. For example, the enzyme responsible for flavanone biosynthesis (flavanone synthase, FNS) in Apiaceae was recruited from the 2-oxoglutarate-dependent dioxygenase (ODD) gene family while in other plant families, the FNS enzyme is encoded by a gene belonging to the cytochrome P450 family (Leonard et al., 2005). Similarly, enzymes with flavonoid 3',5'-hydroxylase activity have evolved at least four times independently from flavonoid 3'-hydroxylase genes within the P450 gene family (Seitz et al., 2006, 2015). Nonetheless, the mutations underlying these functional shifts and the possibility that they themselves are convergent, have only begun to be investigated (e.g., Seitz et al., 2007).

#### GENOMIC CLUSTERING OF BIOSYNTHETIC PATHWAYS

Despite the complex and dynamic histories of genome, with content frequently shuffled due to duplications, translocations, HGT, and other processes, one surprising outcome of genomic studies has been the discovery of biosynthetic gene clusters (BGCs), wherein genes belonging to the same biochemical pathway are physically co-located on a chromosome. These BGCs are common in bacteria and fungi (Osbourne, 2010; Rokas et al., 2018), but were only discovered recently in plants. The first described BGC in plants was the six-gene cluster involved in the synthesis of 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one (DIMBOA), a benzoxazinone compound that protects against insects and microbial pathogens in maize (Frey et al., 1997). Since then, over two dozen gene clusters have been reported from a range of plant taxa (Nutzmann et al., 2016). Biosynthetic gene clusters that are functionally similar appear to have evolved independently in different plant species. For example, oat and *Arabidopsis thaliana* have independently assembled BGCs for the biosynthesis of triterpenoids (avenacin in oat and thalianol in *Arabidopsis*; Field and Osbourne, 2008). In both cases, these clusters contain genes that code for oxidosqualene cyclases, cytochrome P450s, and acyltransferases (Field and Osbourne, 2008). As in nonclustered metabolic pathways, such as the flavonoid pathway (De Jong et al., 2004; Yildiz et al., 2013), gene duplication and subsequent neofunctionalization have played an important role in the innovation of BGCs (Chu et al., 2011; Boycheva et al., 2014).

While the discovery of BGCs in plants presents exciting opportunities for gene discovery and metabolic engineering, their prevalence across different taxa and different classes of metabolites remains unclear. Surveys to date suggest that the degree of clustering can vary significantly, even among closely related taxa. For example, diterpenoid pathways are clustered in some monocots (Schmelz et al., 2014) but not others (Mafu et al., 2018). Similarly, the benzoxazinone pathway appears as a cluster in maize but is spread across four chromosomes in wheat and rye (Sue et al., 2011). Beyond these detailed molecular and biochemical studies, network biology approaches suggest that BGCs may be uncommon in



**FIGURE 3.** Elucidation of plant naphthoquinone pathway. (A) Timeline of key discoveries about the metabolic pathways that give rise to phyloquinone (vitamin K1) and related specialized plant 1,4-naphthoquinones (1,4-NQs) such as juglone. (B) The angiosperm phyloquinone pathway showing branch points to salicylic acid and 1,4-NQs. Note that some 1,4-NQs are synthesized via other convergently evolved metabolic pathways. (C) Predicted evolutionary origin of *men* genes in plants via multiple horizontal gene transfer (HGT) events. Ancestral genes not retained in plants and green algae, based on clustering in modern *Cyanidiales* plastid genomes, are shaded in gray. Genes acquired via HGT into nuclear genomes are shaded in black. Genes with question marks currently lack predicted evolutionary origins. Green algae have neither the Slr0204 nor 4HBT-II type of DHNAT and presumably possess a third type. Angiosperms contain a separated *MenF* (shown in panel B but not in panel C) with the *MenF* domain in *PHYLLO* being truncated and a priori inactive. Abbreviations are as follows: DHNA-CoA, 1,4-dihydroxy-2-naphthoyl-CoA; DHNAT, DHNA-CoA thioesterase; Men, menaquinone/phyloquinone biosynthetic gene. Slr0204, cyanobacterial-type DHNAT; 4-HBT-II, 4-hydroxybenzoyl-CoA thioesterase type II, plant/Lactobacillales-type DHNAT. References indicated with subscripts are: <sup>a</sup>Gross et al., 2008; <sup>b</sup>Widhalm et al., 2009; <sup>c</sup>Widhalm et al., 2012; <sup>d</sup>Furt et al., 2013; <sup>e</sup>McCoy et al., 2018.

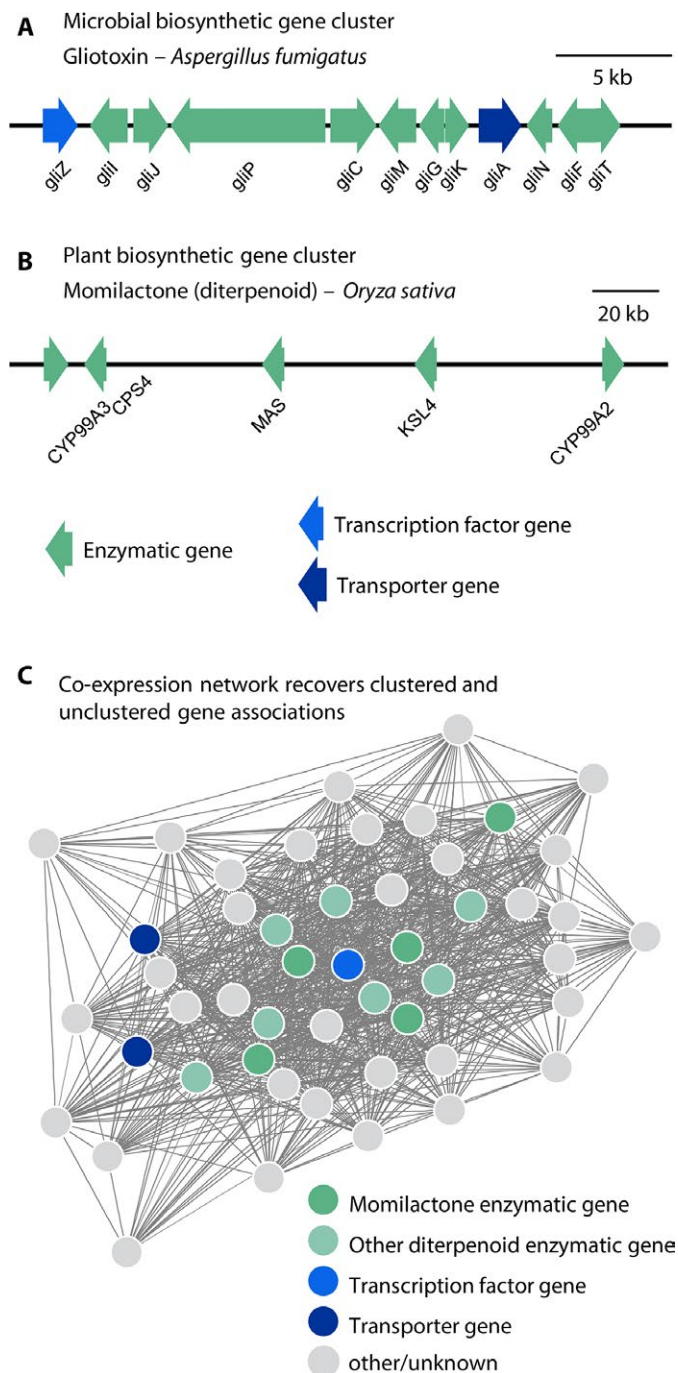


plants as the vast majority of bioinformatically predicted BGCs (i.e., those identified based only physical clustering but not associated with a known pathway or metabolite) do not exhibit co-expression (Wisecaver et al., 2017). The lack of coordinated expression of clustered genes is in stark contrast to the pattern seen in fungi and bacteria, in which both functionally characterized as well as predicted BGCs show strong co-expression (Andersen et al., 2013; Lawler et al., 2013; Lind et al., 2016). Incorporating gene co-expression into analyses of plant BGCs is helpful for eliminating false positives and has the potential to identify functionally associated genes that lie outside the BGC (Fig. 4).

In plants, as in other taxa, there is likely a selective advantage to organize metabolic pathways as BGCs in a genome. Clustering can facilitate the proper co-regulation (Yu et al., 2016) and co-inheritance of functionally associated genes (McGary et al., 2013), thereby reducing the risk of pathway dysfunction, such as the accumulation of toxic metabolic intermediates (Field and Osbourn, 2008; Mylona et al., 2008). Why then are there fewer BGCs in plants compared to fungi and bacteria, despite all having extensive metabolic diversity? Comparisons of plant genomes often show extensive macrosynteny (Shultz et al., 2007; Tang et al., 2008; Denoeud et al., 2014), indicating a relative dearth of gene rearrangement at the scale necessary for BGC formation. In contrast, the chromosomes of fungi rearrange rapidly, and a comparison of fungal genomes showed little to no conservation of gene order (macro- or macrosynteny) between species of the major fungal taxa (Hane et al., 2011). The proliferation of intronic and intergenic DNA in plant genomes due to their smaller effective population sizes compared to microbes may also play a role (Lynch and Conery, 2003); when genes do jump around in plant genomes, the likelihood that they will wind up nearby another functionally associated gene may be low.

## CHALLENGES AND FUTURE DIRECTIONS IN COMPARATIVE BIOCHEMISTRY

The advances in comparative plant biochemistry over the past few decades have provided a window in the complex dynamics of metabolic evolution and laid the foundation for tackling previously intractable questions at multiple biological scales. With new technologies for quantifying diverse specialized metabolites, it is increasingly feasible to test the ecological and environmental factors that shape biochemical variation across populations, even in non-model systems (e.g., Li et al., 2015; Berardi et al., 2016; Huber et al., 2016). These same biotic and abiotic forces can influence patterns at broad phylogenetic scales, leading to macro-evolutionary trends such as the escalation of herbivore defense compounds in milkweeds and mustards (Agrawal et al., 2009; Edger et al., 2015). Specialized metabolites have also been proposed to act as key innovations that increase diversification rates (Erich and Raven, 1964), although this hypothesis has rarely been tested (Farrell et al., 1991; Edger et al., 2015). The integration of comparative biochemistry with plant developmental biology presents an additional research area that is ripe for exploration. It has been observed that specialized structures, such as laticifers, nectaries, and trichomes are associated with the production and accumulation of specialized metabolites (e.g., Hagel et al., 2008; Schilmiller et al., 2008; Camina et al., 2018), yet little is known about the regulatory networks that allow for this tight coupling of



**FIGURE 4.** Comparison of fungal and plant gene clusters. (A) The gliotoxin gene cluster in the fungus *Aspergillus fumigatus* with 11 enzymatic genes, a pathway specific transcription factor and a transporter (Gardiner and Howlett, 2005). (B) The 5-gene cluster for momilactone, a diterpenoid, in rice, consisting solely of enzymatic genes (Shimura et al., 2007). (C) Co-expression network analyses showing strong co-expression of momilactone genes and other unlinked diterpenoid enzymes as well as currently uncharacterized transcription factor and transporter genes. Network adapted from Wisecaver et al. (2017).

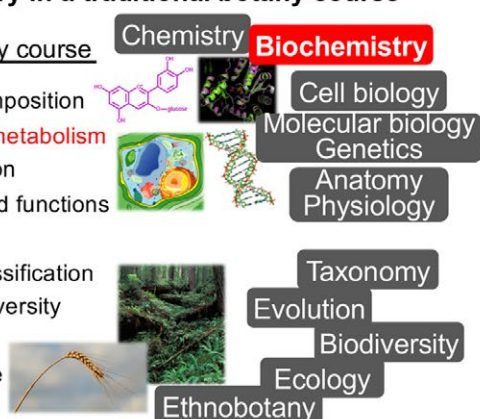
morphology and chemical composition. New genomic and computational tools, such as co-expression analyses and single cell transcriptomics, may provide the first steps toward unraveling

## A Biochemistry in a traditional botany course

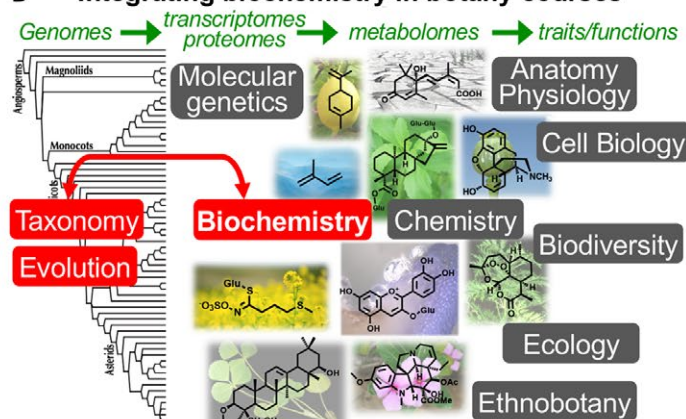
### Introductory botany course

- Plant cell and composition
- **Energy flow and metabolism**
- Genetic information
- Plant structure and functions

- Evolution and classification
- Survey of plant diversity
- Plant ecology
- Plants and people



## B Integrating biochemistry in botany courses



**FIGURE 5.** Integrating biochemistry in botany courses. (A) The subject of biochemistry (e.g. energy flow and metabolism, red letters) is often taught in an early section of introductory courses in plant biology or botany, separately from the subject of plant evolution and diversity. (B) Considering the multidisciplinary nature of modern botany, biochemistry can be integrated with other aspects of botany courses, especially plant taxonomy and evolution (red boxes and arrow). Training the next generation of botanists to make connections across disciplines is important for advancing our understanding of the plant evolution and diversity at all scales, extending across genomes to metabolomes and their physiological and ecological functions (green letters and arrows). For a list of image credits, see Appendix S1.

these developmental linkages and understanding the coordinated evolution of morphology and biochemistry.

As reviewed in this commentary, biochemistry has played a central role in botany since the early 1900s and yet the stories underlying plant biochemical diversification and its ecological significance are just beginning to be discovered. Expanding this renaissance in comparative biochemistry to new traits, new clades, and new metabolites will depend on training the next generation of botanists with a broad and integrative view of the field. In this spirit, we offer a vision of botany teaching where, instead of being relegated to background information, biochemistry is united with other subfields (e.g., cell biology, anatomy, evolution) in a comparative framework that highlights the close connections between genomic, metabolic, cellular, and functional diversity (Fig. 5). Establishing these connections is crucial for not only understanding plants' evolutionary and ecological interactions, but also to their utilization as crops and medicines. Building on case studies such as those described in this commentary, today's botany educators have the opportunity to invite students to experience the integrative nature of modern botany, where traditional disciplinary boundaries are increasingly dissolving in the pursuit of understanding plant diversity at all scales.

## ACKNOWLEDGEMENTS

We thank the Botanical Society of America, the American Society of Plant Taxonomists, and the International Society for Plant Taxonomy for providing support to organize the 2018 Botany symposium that spurred this commentary. S.D.S. thanks L. Allende, R. Deanna, A. Dunbar-Wallis, M. Larter, S. Liu, C. Pretz, and L. Wheeler for valuable feedback on the manuscript. This work was supported by start-up funds from Purdue University to J.R.W. and J.H.W. and start-up funds from Cornell University to G.D.M. H.A.M. acknowledges funding from the National Science Foundation (NSF) (IOS #1354971) and the USDA National Institute of Food and Agriculture (AFRI 2015-67013-22955); J.C.P. and S.D.S. also acknowledge NSF

support (DBI #1339156 and DEB #1553114, respectively). We also thank two anonymous reviewers for helpful comments.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

## APPENDIX S1. Attributes for photographs in Fig. 5.

## LITERATURE CITED

- Abraham, P. E., H. F. Yin, A. M. Borland, D. Weighill, S. D. Lim, H. C. De Paoli, N. Engle, et al. 2016. Transcript, protein and metabolite temporal dynamics in the CAM plant *Agave*. *Nature Plants* 2: 16178.
- Adebesin, F., J. R. Widhalm, J. H. Lynch, R. M. McCoy, and N. Dudareva. 2018. A peroxisomal thioesterase plays auxiliary roles in plant beta-oxidative benzoic acid metabolism. *Plant Journal* 93: 905–916.
- Afendi, F. M., T. Okada, M. Yamazaki, A. Hirai-Morita, Y. Nakamura, K. Nakamura, S. Ikeda, et al. 2012. KNApSACk family databases: integrated metabolite–plant species databases for multifaceted plant research. *Plant and Cell Physiology* 53: e1.
- Agrawal, A. A., J. P. Salminen, and M. Fishbein. 2009. Phylogenetic trends in phenolic metabolism of milkweeds (*Asclepias*): evidence for escalation. *Evolution* 63: 663–673.
- Alston, R. E., and B. L. Turner. 1963. *Biochemical systematics*. Prentice-Hall, Englewood Cliffs, NJ, USA.
- Andersen, M. R., J. B. Nielsen, A. Klitgaard, L. M. Petersen, M. Zachariasen, T. J. Hansen, L. H. Blicher, et al. 2013. Accurate prediction of secondary metabolite gene clusters in filamentous fungi. *Proceedings of the National Academy of Sciences, USA* 110: E99–E107.
- Angelovici, R., A. E. Lipka, N. Deason, S. Gonzalez-Jorge, H. N. Lin, J. Cepela, R. Buell, et al. 2013. Genome-wide analysis of branched-chain amino acid levels in *Arabidopsis* seeds. *Plant Cell* 25: 4827–4843.
- Bar-Even, A., E. Noor, Y. Savir, W. Liebermeister, D. Davidi, D. S. Tawfik, and R. Milo. 2011. The moderately efficient enzyme: evolutionary and



- physicochemical trends shaping enzyme parameters. *Biochemistry* 50: 4402–4410.
- Berardi, A. E., P. D. Fields, J. L. Abbate, and D. R. Taylor. 2016. Elevational divergence and clinal variation in floral color and leaf chemistry in *Silene vulgaris*. *American Journal of Botany* 103: 1508–1523.
- Block, A., J. R. Widhalm, A. Fatihi, R. E. Cahoon, Y. Wamboldt, C. Elowsky, S. A. Mackenzie, et al. 2014. The origin and biosynthesis of the benzenoid moiety of ubiquinone (coenzyme Q) in *Arabidopsis*. *Plant Cell* 26: 1938–1948.
- Boycheva, S., L. Daviet, J. L. Wolfender, and T. B. Fitzpatrick. 2014. The rise of operon-like gene clusters in plants. *Trends in Plant Science* 19: 447–459.
- Brockington, S. F., Y. Yang, F. Gandia-Herrero, S. Covshoff, J. M. Hibberd, R. F. Sage, G. K. S. Wong, et al. 2015. Lineage-specific gene radiations underlie the evolution of novel betalain pigmentation in Caryophyllales. *New Phytologist* 207: 1170–1180.
- Burgess, S. J., I. Granero-Moya, M. J. Grangé-Guermente, C. Boursnell, M. J. Terry, and J. M. Hibberd. 2016. Ancestral light and chloroplast regulation form the foundations for  $C_4$  gene expression. *Nature Plants* 2: 16161.
- Camina, J. L., J. S. Dambolena, J. A. Zygallo, and L. Ashworth. 2018. Chemical composition of essential oils of peltate glandular trichomes from leaves and flowers of *Lepechinia floribunda* (Lamiaceae). *Boletín de la Sociedad Argentina de Botánica* 53: 375–384.
- Chander, S., Y. Guo, X. Yang, J. Zhang, X. Lu, J. Yan, T. Song, et al. 2008. Using molecular markers to identify two major loci controlling carotenoid contents in maize grain. *Theoretical and Applied Genetics* 116: 223–233.
- Chen, M. J., X. Y. Ma, X. L. Chen, M. Jiang, H. G. Song, and Z. H. Guo. 2013. Identification of a hotdog fold thioesterase involved in the biosynthesis of menaquinone in *Escherichia coli*. *Journal of Bacteriology* 195: 2768–2775.
- Chu, H. Y., E. Wegel, and A. Osbourn. 2011. From hormones to secondary metabolism: the emergence of metabolic gene clusters in plants. *Plant Journal* 66: 66–79.
- Coruzzi, G., R. Last, N. Dudareva, and N. Amrhein. 2015. Amino acids. In B. B. Buchanan, W. Gruissem, and R. L. Jones [eds.], *Biochemistry and molecular biology of plants*. American Society of Plant Biologists, Rockville, MD, USA.
- De Jong, W. S., N. T. Eannetta, D. M. De Jong, and M. Bodis. 2004. Candidate gene analysis of anthocyanin pigmentation loci in the Solanaceae. *Theoretical and Applied Genetics* 108: 423–432.
- Dempsey, D. A., A. C. Vlot, M. C. Wildermuth, and D. F. Klessig. 2011. Salicylic acid biosynthesis and metabolism. *Arabidopsis Book* 9: e0156.
- Denoeud, F., L. Carretero-Paulet, A. Dereeper, G. Droc, R. Guyot, M. Pietrella, C. F. Zheng, et al. 2014. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* 345: 1181–1184.
- Dudareva, N., L. Cseke, V. M. Blanc, and E. Pichersky. 1996. Evolution of floral scent in *Clarkia*: Novel patterns of S-linalool synthase gene expression in the *C. breweri* flower. *Plant Cell* 8: 1137–1148.
- Edger, P. P., H. M. Heide-Fischer, M. Bekaert, J. Rota, G. Glockner, A. E. Platts, D. G. Heckel, et al. 2015. The butterfly plant arms-race escalated by gene and genome duplications. *Proceedings of the National Academy of Sciences, USA* 112: 8362–8366.
- Edwards, E. J., and R. M. Ogburn. 2012. Angiosperm responses to a low- $CO_2$  world: CAM and  $C_4$  photosynthesis as parallel evolutionary trajectories. *International Journal of Plant Sciences* 173: 724–733.
- Emiliani, G., M. Fondi, R. Fani, and S. Gribaldo. 2009. A horizontal gene transfer at the origin of phenylpropanoid metabolism: a key adaptation of plants to land. *Biology Direct* 4: 7.
- Erlich, P. R., and P. H. Raven. 1964. Butterflies and plants: a study in coevolution. *Evolution* 18: 586–608.
- Ettlinger, M. G., and A. Kjaer. 1968. Sulphur compounds in plants. In T. J. Mabry [ed.], *Recent advances in phytochemistry*, 59–144. North-Holland Publishing, Amsterdam, Netherlands.
- Fan, P. X., A. M. Miller, X. X. Liu, A. D. Jones, and R. L. Last. 2017. Evolution of a flipped pathway creates metabolic innovation in tomato trichomes through BAH1 enzyme promiscuity. *Nature Communications* 8: 2080.
- Farrell, B. D., D. E. Dussourd, and C. Mitter. 1991. Escalation of plant defense: Do latex and resin canals spur plant diversification? *American Naturalist* 138: 881–900.
- Field, B., and A. E. Osbourn. 2008. Metabolic diversification: independent assembly of operon-like gene clusters in different plants. *Science* 320: 543–547.
- Frey, M., P. Chomet, E. Glawischnig, C. Stettner, S. Grun, A. Winklmeier, W. Eisenreich, et al. 1997. Analysis of a chemical plant defense mechanism in grasses. *Science* 277: 696–699.
- Furt, F., W. J. Allen, J. R. Widhalm, P. Madzelan, R. C. Rizzo, G. Basset, and M. A. Wilson. 2013. Functional convergence of structurally distinct thioesterases from cyanobacteria and plants involved in phyloquinone biosynthesis. *Acta Crystallographica, D, Biological Crystallography* 69: 1876–1888.
- Gardiner, D. M., and B. J. Howlett. 2005. Bioinformatic and expression analysis of the putative gliotoxin biosynthetic gene cluster of *Aspergillus fumigatus*. *FEMS Microbiology Letters* 248: 241–248.
- Gibbs, R. D. 1963. History of chemical taxonomy. In T. Swain [ed.], *Chemical plant taxonomy*. Academic Press, NY, NY, USA.
- Gibbs, R. D. 1974. Chemotaxonomy of flowering plants. McGill-Queen's University Press, Montreal, Quebec, Canada.
- Gierl, A., and M. Frey. 2001. Evolution of benzoxazinone biosynthesis and indole production in maize. *Planta* 213: 493–498.
- Grayer, R. J., M. W. Chase, and M. S. J. Simmonds. 1999. A comparison between chemical and molecular characters for the determination of phylogenetic relationships among plant families: an appreciation of Hegnauer's "Chemotaxonomie der Pflanzen". *Biochemical Systematics and Ecology* 27: 369–393.
- Gross, J., J. Meurer, and D. Bhattacharya. 2008. Evidence of a chimeric genome in the cyanobacterial ancestor of plastids. *BMC Evolutionary Biology* 8: 117.
- Hagel, J. M., E. C. Yeung, and P. J. Facchini. 2008. Got milk? The secret life of laticifers. *Trends in Plant Science* 13: 631–639.
- Hane, J. K., T. Rouxel, B. J. Howlett, G. H. J. Kema, S. B. Goodwin, and R. P. Oliver. 2011. A novel mode of chromosomal evolution peculiar to filamentous *Ascomycete* fungi. *Genome Biology* 12: R45.
- Harborne, J. B. 1966. Comparative biochemistry of flavonoids II: 3-desoxyanthocyanins and their systematic distribution in ferns and gesnerads. *Phytochemistry* 5: 589–&.
- Harborne, J. B. 1967. Comparative biochemistry of flavonoids VI: flavonoid patterns in *Bignoniaceae* and *Gesneriaceae*. *Phytochemistry* 6: 1643–1651.
- Harborne, J. B. 1973. *Phytochemical methods*, 2nd ed. Chapman and Hall, London, UK.
- Hook, I., C. Mills, and H. Sheridan. 2014. Bioactive naphthoquinones from higher plants. In Atta-ur-Rahman [ed.], *Studies in natural products chemistry*, 119–160. Elsevier, Amsterdam, Netherlands.
- Huang, R. Q., A. J. O'Donnell, J. J. Barbolino, and T. J. Barkman. 2016. Convergent evolution of caffeine in plants by co-option of exapted ancestral enzymes. *Proceedings of the National Academy of Sciences, USA* 113: 10613–10618.
- Huber, M., Z. Bont, J. Fricke, T. Brillatz, Z. Aziz, J. Gershenzon, and M. Erb. 2016. A below-ground herbivore shapes root defensive chemistry in natural plant populations. *Proceedings of the Royal Society, B, Biological Sciences* 283: 20160285.
- Keeling, P. J., and J. D. Palmer. 2008. Horizontal gene transfer in eukaryotic evolution. *Nature Reviews Genetics* 9: 605–618.
- Kim, J., K. Kang, E. Gonzales-Vigil, F. Shi, A. D. Jones, C. S. Barry, and R. L. Last. 2012. Striking natural diversity in glandular trichome acylsugar composition is shaped by variation at the Acyltransferase2 locus in the wild tomato *Solanum habrochaites*. *Plant Physiology* 160: 1854–1870.
- Kochevenko, A., and A. R. Fernie. 2011. The genetic architecture of branched-chain amino acid accumulation in tomato fruits. *Journal of Experimental Botany* 62: 3895–3906.
- Kumara, S. J. S., E. V. Packiyasothy, and E. R. Jansz. 1985. Some studies on the effect of maturity and storage on the chlorophyll content and essential oils of the cardamom fruit (*Elettaria cardamomum*). *Journal of the Science of Food and Agriculture* 36: 491–498.
- Kurland, C. G., B. Canback, and O. G. Berg. 2003. Horizontal gene transfer: a critical view. *Proceedings of the National Academy of Sciences, USA* 100: 9658–9662.
- Lawler, K., K. Hammond-Kosack, A. Brazma, and R. M. R. Coulson. 2013. Genomic clustering and co-regulation of transcriptional networks in the pathogenic fungus *Fusarium graminearum*. *BMC Systems Biology* 7: 52.

- Leonard, E., Y. J. Yan, K. H. Lim, and M. A. G. Koffas. 2005. Investigation of two distinct flavone synthases for plant-specific flavone biosynthesis in *Saccharomyces cerevisiae*. *Applied and Environmental Microbiology* 71: 8241–8248.
- Li, D., I. T. Baldwin, and E. Gaquerel. 2015. Navigating natural variation in herbivory-induced secondary metabolism in coyote tobacco populations using MS/MS structural analysis. *Proceedings of the National Academy of Sciences, USA* 112: E4147–4155.
- Lin, T. P., T. Caspar, C. R. Somerville, and J. Preiss. 1988. A starch deficient mutant of *Arabidopsis thaliana* with low ADP glucose pyrophosphorylase activity lacks one of the two subunits of the enzyme. *Plant Physiology* 88: 1175–1181.
- Lind, A. L., T. D. Smith, T. Saterlee, A. M. Calvo, and A. Rokas. 2016. Regulation of secondary metabolism by the velvet complex is temperature-responsive in *Aspergillus*. *G3: Genes Genomes Genetics* 6: 4023–4033.
- Lopez-Nieves, S., Y. Yang, A. Timoneda, M. M. Wang, T. Feng, S. A. Smith, S. F. Brockington, and H. A. Maeda. 2018. Relaxation of tyrosine pathway regulation underlies the evolution of betalain pigmentation in *Caryophyllales*. *New Phytologist* 217: 896–908.
- Lynch, M., and J. S. Conery. 2003. The origins of genome complexity. *Science* 302: 1401–1404.
- Mafu, S., Y. Z. Ding, K. M. Murphy, O. Yaacoobi, J. B. Addison, Q. Wang, Z. X. Shen, et al. 2018. Discovery, biosynthesis and stress-related accumulation of dolabradiene-derived defenses in maize. *Plant Physiology* 176: 2677–2690.
- Maloney, G. S., A. Kochevenko, D. M. Tieman, T. Tohge, U. Krieger, D. Zamir, M. G. Taylor, et al. 2010. Characterization of the branched-chain amino acid aminotransferase enzyme family in tomato. *Plant Physiology* 153: 925–936.
- Marston, A. 2007. Role of advances in chromatographic techniques in phytochemistry. *Phytochemistry* 68: 2786–2798.
- Martin, C., E. Coen, T. Robbins, J. Bartlett, J. Alemaida, and R. Carpenter. 1987. The control of floral pigmentation in *Antirrhinum*. *Biochemical Society Transactions* 15: 14–17.
- Matsuba, Y., T. T. H. Nguyen, K. Wiegert, V. Falara, E. Gonzales-Vigil, B. Leong, P. Schafer, et al. 2013. Evolution of a complex locus for terpene biosynthesis in *Solanum*. *Plant Cell* 25: 2022–2036.
- McCoy, R. M., S. M. Utturkar, J. W. Crook, J. Thimmapuram, and J. R. Widhalm. 2018. The origin and biosynthesis of the naphthalenoid moiety of juglone in black walnut. *Horticulture Research* 5: 67.
- McGary, K. L., J. C. Slot, and A. Rokas. 2013. Physical linkage of metabolic genes in fungi is an adaptation against the accumulation of toxic intermediate compounds. *Proceedings of the National Academy of Sciences, USA* 110: 11481–11486.
- Moghe, G. D., and L. H. Kruse. 2018. The study of plant specialized metabolism: challenges and prospects in the genomics era. *American Journal of Botany* 105: 959–962.
- Moghe, G. D., and R. L. Last. 2015. Something old, something new: conserved enzymes and the evolution of novelty in plant specialized metabolism. *Plant Physiology* 169: 1512–1523.
- Moghe, G. D., B. J. Leong, S. M. Hurney, A. D. Jones, and R. L. Last. 2017. Evolutionary routes to biochemical innovation revealed by integrative analysis of a plant-defense related specialized metabolic pathway. *Elife* 6: e28468.
- Moreno-Villena, J. J., L. T. Dunning, C. P. Osborne, and P. A. Christin. 2018. Highly expressed genes are preferentially co-opted for  $C_4$  photosynthesis. *Molecular Biology and Evolution* 35: 94–106.
- Mueller, L. A., U. Hinz, M. Uze, C. Sautter, and J. P. Zryd. 1997. Biochemical complementation of the betalain biosynthetic pathway in *Portulaca grandiflora* by a fungal 3,4-dihydroxyphenylalanine dioxygenase. *Planta* 203: 260–263.
- Musso, H. 1979. Pigments of fly agaric, *Amanita muscaria*. *Tetrahedron* 35: 2843–2853.
- Mylona, P., A. Owatworakit, K. Papadopoulou, H. Jenner, B. Qin, K. Findlay, L. Hill, et al. 2008. Sad3 and Sad4 are required for saponin biosynthesis and root development in oat. *Plant Cell* 20: 201–212.
- Nam, H., N. E. Lewis, J. A. Lerman, D. H. Lee, R. L. Chang, D. Kim, and B. O. Palsson. 2012. Network context and selection in the evolution to enzyme specificity. *Science* 337: 1101–1104.
- Näsvall, J., L. Sun, J. R. Roth, and D. I. Andersson. 2012. Real-time evolution of new genes by innovation, amplification, and divergence. *Science* 338: 384–387.
- Nutzmann, H. W., A. C. Huang, and A. Osbourn. 2016. Plant metabolic clusters—from genetics to genomics. *New Phytologist* 211: 771–789.
- Osbourn, A. 2010. Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. *Trends in Genetics* 26: 449–457.
- Pichersky, E., and E. Lewinsohn. 2011. Convergent evolution in plant specialized metabolism. *Annual Review of Plant Biology* 62: 549–566.
- Prasad, K. V. S. K., B. H. Song, C. Olson-Manning, J. T. Anderson, C. R. Lee, M. E. Schranz, A. J. Windsor, et al. 2012. A gain-of-function polymorphism controlling complex traits and fitness in nature. *Science* 337: 1081–1084.
- Pratelli, R., and G. Pilot. 2014. Regulation of amino acid metabolic enzymes and transporters in plants. *Journal of Experimental Botany* 65: 5535–5556.
- Raguso, R. A., and E. Pichersky. 1995. Floral volatiles from *Clarkia breweri* and *C. concinna* (Onagraceae): recent evolution of floral scent and moth pollination. *Plant Systematics and Evolution* 194: 55–67.
- Rausher, M. D. 2006. The evolution of flavonoids and their genes. In E. Grotewold [ed.], *The science of flavonoids*. Springer, NY, NY, USA.
- Reid, J. B., and J. J. Ross. 2011. Mendel's genes: toward a full molecular characterization. *Genetics* 189: 3–10.
- Ren, L. L., Y. J. Liu, H. J. Liu, T. T. Qian, L. W. Qi, X. R. Wang, and Q. Y. Zeng. 2014. Subcellular relocalization and positive selection play key roles in the retention of duplicate genes of *Populus* class III peroxidase family. *Plant Cell* 26: 2404–2419.
- Reyna-Llorens, I., and J. M. Hibberd. 2017. Recruitment of pre-existing networks during the evolution of  $C_4$  photosynthesis. *Philosophical Transactions of the Royal Society, B, Biological Sciences* 372: 20160386.
- Reyna-Llorens, I., S. J. Burgess, G. Reeves, P. Singh, S. R. Stevenson, B. P. Williams, S. Stanley, and J. M. Hibberd. 2018. Ancient duons may underpin spatial patterning of gene expression in  $C_4$  leaves. *Proceedings of the National Academy of Sciences, USA* 115: 1931–1936.
- Ritter, H., and G. E. Schulz. 2004. Structural basis for the entrance into the phenylpropanoid metabolism catalyzed by phenylalanine ammonia-lyase. *Plant Cell* 16: 3426–3436.
- Rodman, J. E., P. S. Soltis, D. E. Soltis, K. J. Sytsma, and K. G. Karol. 1998. Parallel evolution of glucosinolate biosynthesis inferred from congruent nuclear and plastid gene phylogenies. *American Journal of Botany* 85: 997–1006.
- Rokas, A., J. H. Wisecaver, and A. L. Lind. 2018. The birth, evolution and death of metabolic gene clusters in fungi. *Nature Reviews Microbiology* 16: 731–744.
- Sage, R. F., P. A. Christin, and E. J. Edwards. 2011. The  $C_4$  plant lineages of planet Earth. *Journal of Experimental Botany* 62: 3155–3169.
- Schenck, C. A., S. Y. Chen, D. L. Siehl, and H. A. Maeda. 2015. Non-plastidic, tyrosine-insensitive prephenate dehydrogenases from legumes. *Nature Chemical Biology* 11: 52–57.
- Schenck, C. A., C. K. Holland, M. R. Schneider, Y. Men, S. G. Lee, J. M. Jez, and H. A. Maeda. 2017. Molecular basis of the evolution of alternative tyrosine biosynthetic routes in plants. *Nature Chemical Biology* 13: 1029–1035.
- Schenck, C. A., and H. A. Maeda. 2018. Tyrosine biosynthesis, metabolism, and catabolism in plants. *Phytochemistry* 149: 82–102.
- Schillmiller, A. L., R. L. Last, and E. Pichersky. 2008. Harnessing plant trichome biochemistry for the production of useful compounds. *Plant Journal* 54: 702–711.
- Schmelz, E. A., A. Huffaker, J. W. Sims, S. A. Christensen, X. Lu, K. Okada, and R. J. Peters. 2014. Biosynthesis, elicitation and roles of monocot terpenoid phytoalexins. *Plant Journal* 79: 659–678.
- Schymanski, E. L., C. Ruttkies, M. Krauss, C. Brouard, T. Kind, K. Duhrkop, F. Allen, et al. 2017. Critical assessment of small molecule identification 2016: automated methods. *Journal of Cheminformatics* 9: 22.
- Sedio, B. E. 2017. Recent breakthroughs in metabolomics promise to reveal the cryptic chemical traits that mediate plant community composition, character evolution and lineage diversification. *New Phytologist* 214: 952–958.
- Seitz, C., S. Ameres, and G. Forkmann. 2007. Identification of the molecular basis for the functional difference between flavonoid 3'-hydroxylase and flavonoid 3',5'-hydroxylase. *FEBS Letters* 581: 3429–3434.
- Seitz, C., S. Ameres, K. Schlangen, G. Forkmann, and H. Halbwirth. 2015. Multiple evolution of flavonoid 3',5'-hydroxylase. *Planta* 242: 561–573.

- Seitz, C., C. Eder, B. Deiml, S. Kellner, S. Martens, and G. Forkmann. 2006. Cloning, functional identification and sequence analysis of flavonoid 3'-hydroxylase and flavonoid 3',5'-hydroxylase cDNAs reveals independent evolution of flavonoid 3',5'-hydroxylase in the Asteraceae family. *Plant Molecular Biology* 61: 365–381.
- Shimura, K., A. Okada, K. Okada, Y. Jikumaru, K. W. Ko, T. Toyomasu, T. Sassa, et al. 2007. Identification of a biosynthetic gene cluster in rice for momilactones. *Journal of Biological Chemistry* 282: 34013–34018.
- Shirley, B. W., S. Hanley, and H. M. Goodman. 1992. Effects of ionizing radiation on a plant genome: analysis of two *Arabidopsis* transparent testa mutations. *Plant Cell* 4: 333–347.
- Shultz, J. L., J. D. Ray, and D. A. Lightfoot. 2007. A sequence based synteny map between soybean and *Arabidopsis thaliana*. *BMC Genomics* 8: 8.
- Smith, J. F., J. C. Wolfram, K. D. Brown, C. L. Carroll, and D. S. Denton. 1997. Tribal relationships in the *Gesneriaceae*: Evidence from DNA sequences of the chloroplast gene *ndhF*. *Annals of the Missouri Botanical Garden* 84: 50–66.
- Sue, M., C. Nakamura, and T. Nomura. 2011. Dispersed benzoxazinone gene cluster: molecular characterization and chromosomal localization of glucosyltransferase and glucosidase genes in wheat and rye. *Plant Physiology* 157: 985–997.
- Tang, H. B., J. E. Bowers, X. Y. Wang, R. Ming, M. Alam, and A. H. Paterson. 2008. Perspective—Synteny and collinearity in plant genomes. *Science* 320: 486–488.
- Twiss, M. 1906. Adsorptionsanalyse und chromatographische methode. *Berichte der Deutschen botanischen Gesellschaft*. 24: 384–393.
- van Oostende, C., J. R. Widhalm, F. Furt, A.-L. Ducluzeau, and G. J. Basset. 2011. Vitamin K1 (phyloquinone): Function, enzymes and genes. *Advances in Botanical Research* 59: 229–261.
- Wang, M. X., J. J. Carver, V. V. Phelan, L. M. Sanchez, N. Garg, Y. Peng, D. D. Nguyen, et al. 2016. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature Biotechnology* 34: 828–837.
- Weng, J. K. 2014. The evolutionary paths towards complexity: a metabolic perspective. *New Phytologist* 201: 1141–1149.
- Weng, J. K., and J. P. Noel. 2013. Chemodiversity in *Selaginella*: a reference system for parallel and convergent metabolic evolution in terrestrial plants. *Frontiers in Plant Science* 4: 119.
- Wentzell, A. M., H. C. Rowe, B. G. Hansen, C. Ticconi, B. A. Halkier, and D. J. Kliebenstein. 2007. Linking metabolic QTLs with network and *cis*-eQTLs controlling biosynthetic pathways. *PLoS Genetics* 3: 1687–1701.
- Widhalm, J. R., C. van Oostende, F. Furt, and G. J. C. Basset. 2009. A dedicated thioesterase of the Hotdog-fold family is required for the biosynthesis of the naphthoquinone ring of vitamin K-1. *Proceedings of the National Academy of Sciences, USA* 106: 5599–5603.
- Widhalm, J. R., A. L. Ducluzeau, N. E. Buller, C. G. Elowsky, L. J. Olsen, and G. J. C. Basset. 2012. Phyloquinone (vitamin K1) biosynthesis in plants: two peroxisomal thioesterases of lactobacillales origin hydrolyze 1,4-dihydroxy-2-naphthoyl-CoA. *Plant Journal* 71: 205–215.
- Wisecaver, J. H., A. T. Borowsky, V. Tzin, G. Jander, D. Kliebenstein, and A. Rokas. 2017. A global co-expression approach for connecting genes to specialized metabolic pathways in plants. *The Plant Cell* 29: 944–959.
- Woo, V. L., M. M. Funke, J. F. Smith, P. J. Lockhart, and P. J. Garnock-Jones. 2011. New World origins of southwest Pacific Gesneriaceae: multiple movements across and within the South Pacific. *International Journal of Plant Sciences* 172: 434–457.
- Wortman, J. R., B. J. Haas, L. I. Hannick, R. K. Smith, R. Maiti, C. M. Ronning, A. P. Chan, et al. 2003. Annotation of the *Arabidopsis* genome. *Plant Physiology* 132: 461–468.
- Wurtzel, E. T., A. Cuttriss, and R. Vallabhaneni. 2012. Maize provitamin A carotenoids, current resources, and future metabolic engineering challenges. *Frontiers in Plant Science* 3: 29.
- Yan, J. B., C. B. Kandianis, C. E. Harjes, L. Bai, E. H. Kim, X. H. Yang, D. J. Skinner, et al. 2010. Rare genetic variation at *Zea mays crtRB1* increases beta-carotene in maize grain. *Nature Genetics* 42: 322–327.
- Yildiz, M., D. K. Willis, P. F. Cavagnaro, M. Iorizzo, K. Abak, and P. W. Simon. 2013. Expression and mapping of anthocyanin biosynthesis genes in carrot. *Theoretical and Applied Genetics* 126: 1689–1702.
- Yin, H. F., H. B. Guo, D. J. Weston, A. M. Borland, P. Ranjan, P. E. Abraham, S. S. Jawdy, et al. 2018. Diel rewiring and positive selection of ancient plant proteins enabled evolution of CAM photosynthesis in *Agave*. *BMC Genomics* 19: 588.
- Yoneyama, N., H. Morimoto, C. X. Ye, H. Ashihara, K. Mizuno, and M. Kato. 2006. Substrate specificity of *N*-methyltransferase involved in purine alkaloids synthesis is dependent upon one amino acid residue of the enzyme. *Molecular Genetics and Genomics* 275: 125–135.
- Yu, G., T. T. H. Nguyen, Y. Guo, I. Schauvinhold, M. E. Auldrige, N. Bhuiyan, I. Ben-Israel, et al. 2010. Enzymatic functions of wild tomato methylketone synthases 1 and 2. *Plant Physiology* 154: 67–77.
- Yu, N., H. W. Nutzmam, J. T. MacDonald, B. Moore, B. Field, S. Berriri, M. Trick, et al. 2016. Delineation of metabolic gene clusters in plant genomes by chromatin signatures. *Nucleic Acids Research* 44: 2255–2265.
- Yue, J., X. Hu, and J. Huang. 2013. Horizontal gene transfer in the innovation and adaptation of land plants. *Plant Signaling and Behavior* 8: e24130.