



The Origin and Evolution of Plant Flavonoid Metabolism

Keiko Yonekura-Sakakibara*, Yasuhiro Higashi and Ryo Nakabayashi

RIKEN Center for Sustainable Resource Science, Yokohama, Japan

During their evolution, plants have acquired the ability to produce a huge variety of compounds. Unlike the specialized metabolites that accumulate in limited numbers of species, flavonoids are widely distributed in the plant kingdom. Therefore, a detailed analysis of flavonoid metabolism in genomics and metabolomics is an ideal way to investigate how plants have developed their unique metabolic pathways during the process of evolution. More comprehensive and precise metabolite profiling integrated with genomic information are helpful to emerge unexpected gene functions and/or pathways. The distribution of flavonoids and their biosynthetic genes in the plant kingdom suggests that flavonoid biosynthetic pathways evolved through a series of steps. The enzymes that form the flavonoid scaffold structures probably first appeared by recruitment of enzymes from primary metabolic pathways, and later, enzymes that belong to superfamilies such as 2-oxoglutarate-dependent dioxygenase, cytochrome P450, and short-chain dehydrogenase/reductase modified and varied the structures. It is widely accepted that the first two enzymes in flavonoid biosynthesis, chalcone synthase, and chalcone isomerase, were derived from common ancestors with enzymes in lipid metabolism. Later enzymes acquired their function by gene duplication and the subsequent acquisition of new functions. In this review, we describe the recent progress in metabolomics technologies for flavonoids and the evolution of flavonoid skeleton biosynthetic enzymes to understand the complicate evolutionary traits of flavonoid metabolism in plant kingdom.

Keywords: secondary metabolites, flavonoid, polyketide synthase, 2-oxoglutarate-dependent dioxygenase, cytochrome P450, short-chain dehydrogenase/reductase, plant

OPEN ACCESS

Edited by:

Kevin Davies,
The New Zealand Institute for Plant
and Food Research Ltd., New Zealand

Reviewed by:

Stefan Martens,
Fondazione Edmund Mach, Italy
John A. Morgan,
Purdue University, United States

*Correspondence:

Keiko Yonekura-Sakakibara
yskeiko@riken.jp

Specialty section:

This article was submitted to
Plant Metabolism and Chemodiversity,
a section of the journal
Frontiers in Plant Science

Received: 26 April 2019

Accepted: 08 July 2019

Published: 02 August 2019

Citation:

Yonekura-Sakakibara K, Higashi Y and
Nakabayashi R (2019) The Origin and
Evolution of Plant Flavonoid
Metabolism. *Front. Plant Sci.* 10:943.
doi: 10.3389/fpls.2019.00943

INTRODUCTION

Plants have the ability to produce a huge variety of metabolites. Over 1,000,000 metabolites are predicted to be present in the entire plant kingdom (Afendi et al., 2012). Most of these are secondary metabolites (also referred to as specialized metabolites) that play a wide range of physiological and ecological roles including defense against herbivores and pathogens, attractants for pollinators and seed carriers, and signaling. During the long process of evolution, plants have gained, expanded, and sometimes lost their capabilities to produce this huge array of metabolites, which provides the adaptive mechanisms needed for survival in changing environments.

Flavonoids form one of the major groups of specialized metabolites, and include over 9,000 compounds (Williams and Grayer, 2004; Anderson and Markham, 2006). According to Nomenclature of flavonoids (IUPAC Recommendations, 2017), the term “flavonoid” is applied to (1) compounds structurally based on derivatives of a phenyl-substituted propylbenzene having a C15 skeleton, (2) compounds with a C16 skeleton that are phenyl-substituted propylbenzene derivatives (rotenoids), (3) flavonolignans based on derivatives of phenyl-substituted propylbenzene condensed with C₆-C₃ lignan precursors (Rauter et al., 2018). In a

restricted sense, the term “flavonoid” is used only for those compounds with a C₆-C₃-C₆ carbon framework exhibiting the structure of a chromane or that of a chromene such as flavans, flavones, flavonols, and anthocyanidins. Chalcones, dihydrochalcones, and aurones are flavonoids in a broad sense, but not in a limited sense.

Flavonoids, including chalcones, flavones, flavonols, anthocyanins, and proanthocyanidins, are widely distributed in the plant kingdom, and their metabolic pathways have been extensively studied using both biochemical and molecular biological techniques. Until recently, it was believed that liverworts and mosses were the oldest flavonoid-producing plants (Rausher, 2006; Bowman et al., 2017). The genes encoding enzymes in the phenylpropanoid biosynthetic pathway, including the first two enzymes for flavonoid biosynthesis (chalcone synthase and chalcone isomerase) had not been found in the algal genera *Chlamydomonas*, *Micromonas*, *Ostreococcus*, and *Klebsormidium*, although genes encoding enzymes in the shikimate pathway were found in algae, liverworts, mosses, lycophytes, ferns and horsetails, gymnosperms, and angiosperms (Bowman et al., 2017). However, flavones, isoflavones, and flavonols were detected in microalgae from five different evolutionary lineages (*Cyanobacteria*, *Rhodophyta*, *Chlorophyta*, *Haptophyta*, and *Ochrophyta*) using ultra-high performance liquid chromatography with tandem mass spectrometry (Goiris et al., 2014). This suggests that plants may have acquired the ability to produce flavonoids earlier than we previously thought. Furthermore, in extant plants flavonoids play important roles as ultraviolet-B (UV-B) protectants, pigments that attract pollinators, phytoalexins, signaling molecules, and regulators of auxin transport and fertility (Gould and Lister, 2006). It has been proposed that defense against UV irradiation and regulation of plant hormone action were the original functions of flavonoids in the earliest flavonoid producing plants (Stafford, 1991; Shirley, 1996; Rausher, 2006). These functions have been considerably diversified during plant evolution. Thus, the study of flavonoids is a useful approach to understand how plants acquired the ability to produce specialized metabolites, and then build the metabolic pathways needed to produce such a huge variety of metabolites during the course of their evolution. This study will shed light on the relationships between genes/proteins and metabolites, and between the metabolites and their physiological functions. In this review, we describe the structural diversity of flavonoids distributed in the plant kingdom, and how plants acquired flavonoid biosynthetic genes during their evolution.

METABOLOMICS SHEDS LIGHT ON THE EVOLUTION OF FLAVONOIDS

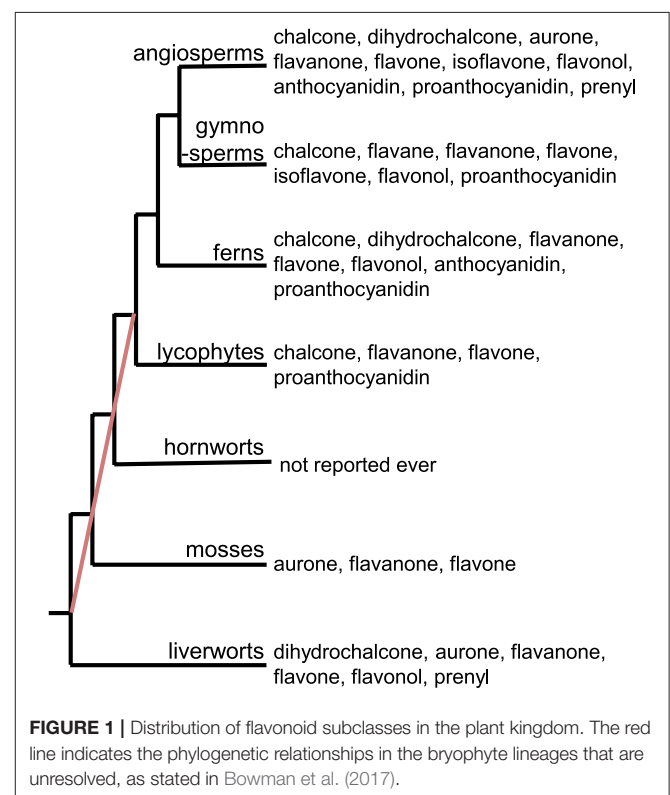
Distribution of Flavonoids in the Plant Kingdom

It is estimated that there are over 9,000 flavonoids in the plant kingdom (Williams and Grayer, 2004; Anderson and Markham, 2006). Research on flavonoids has shown that they are distributed across the plant kingdom, including angiosperms, gymnosperms, and pteridophytes (Harborne, 1988; Tohge et al., 2013). The abundance of information about flavonoids

in different species allows us to identify which flavonoid subclasses (e.g., chalcones, flavones, flavonols, anthocyanins, and proanthocyanidins) are found in each subgroup of plants (Figure 1). Flavone and flavanone are found in all plant groups except for hornworts. To our knowledge, no flavonoids have been reported in hornworts. As plant groups have evolved and diversified, so have the flavonoid subclasses produced within each group. For example, the flavonoid aglycones are most diversified in the angiosperms. In addition to flavanones and flavones, chalcones, flavonols, and proanthocyanidins are found in multiple groups. Interestingly, the prenylflavonoids are found in both liverworts and angiosperms: while more than 1,000 prenylflavonoids have been isolated from legumes (Yazaki et al., 2009), prenyldihydrochalcone has been found in the liverworts *Radula variabilis* and *Radula* spp. (Asakawa et al., 1978, 1982). These data suggest that the two groups of plants gained the ability to produce prenylflavonoids independently, or that many groups have lost the ability to produce prenylflavonoids during their evolution. Flavonoid molecules provide concrete evidence for the existence of the corresponding flavonoid biosynthetic genes in the plants. Analytical approaches for identifying flavonoids is therefore important to understand the evolution of flavonoid metabolism in the plant kingdom.

Cutting-Edge Metabolomics Technologies Contribute to Understanding the Evolution of Flavonoid Metabolism

So far, chromatographic and spectroscopic approaches have been used to analyze the structures of flavonoids and to reveal their chemical diversity. Previously, paper chromatography,



thin layer chromatography, column chromatography, and liquid chromatography (LC) were the main techniques used to study flavonoids. Crude extracts containing flavonoids are obtained through liquid-liquid partitioning. For instance, flavonoid aglycones and their mono-glycosides can be mainly extracted in ethylacetate, whereas the flavonoid di- or tri-glycosides are extracted in *n*-butanol. Chromatographic techniques are then used to purify the flavonoids from the ethylacetate or *n*-butanol fractions, in several steps. Finally, the isolated flavonoids are analyzed by nuclear magnetic resonance and mass spectrometry to elucidate their structures. This is an unequivocal and straightforward approach to identifying flavonoids, however, it is time-consuming.

With the development of metabolomics technologies, flavonoids can now be analyzed much more accurately and precisely than before. LC-tandem mass spectrometry (LC-MS/MS) has become the preferred approach for analyzing flavonoids. It had been considered that there was no flavonoids in algae. However, identification of flavonoids such as intermediates and end products by an LC-MS/MS approach proved the existence of the flavonoid biosynthetic ability in microalgae (Goiris et al., 2014). It suggests that cutting-edge metabolomics technologies in every single plant species can unveil the flavonoids that have been missed.

The integration of LC-MS/MS with cheminformatics approaches provides a powerful tool for surveying flavonoid diversity in a high-throughput way (Tsugawa et al., 2016; Akimoto et al., 2017). Glycosylated, acylated, and prenylated flavonoid molecules and their aglycones can be separated using simple combinations of solvents and LC columns. The separated molecules are then ionized for MS/MS analysis. The product ions are derived from their precursors through the cleavage of ether and ester bonds or even the prenyl moiety. These simple fragmentation steps allow us to estimate the structures of the flavonoids. Cheminformatic tools are used for high-throughput and (semi-)automatic data analysis. The results are generally classified into four classes (levels) according to the guidelines of the Metabolomics Standards Initiative: Level 1, identification using authentic standard compounds; Level 2, annotation using public databases; Level 3, characterization by deciphering MS/MS data; and Level 4, unknown (not noise) (Sumner et al., 2007). The use of these levels reduces the production of false positive data.

An integrated approach using LC-MS/MS with ^{13}C (carbon) labeling and cheminformatics can be used to assign structures to metabolites whose structures have never been recognized before. In most of the general approaches, unknown metabolites remain uncharacterized. The integrative approach is used to identify the elemental compositions of the unknown metabolites based on the numbers of C atoms, determined by comparisons between ^{13}C - and non-labeled MS spectra. The MS/MS spectra of the unknown metabolites are then subjected to fragment set enrichment analysis (FSEA) to determine candidates for each metabolite class. This integrative approach was used to characterize 1,133 metabolites, including flavonoids, in 12 angiosperm species (Tsugawa et al., 2019). The results for the characterized flavonoids are summarized in **Table 1**, along with the flavonoids found in the

liverwort *Marchantia polymorpha*, which were analyzed using the same method (Kubo et al., 2018). Our analyses using LC-MS/MS with ^{13}C labeling and cheminformatics led to the identification of more flavonoid molecules especially in the group of flavonol and flavone. Prenylated flavonoids have been analyzed by LC-MS/MS less extensively than other flavonoid groups, although a chemical assignment strategy is effective for profiling prenylated flavonoids. It is probably due to the lack of authentic standards and publicly available MS/MS spectra. The exact mass of a known structure or one that is presumed to be prenylated can also be useful for profiling prenylated flavonoids in cases where the ontologies for prenylated flavonoids does not work. This means that the observed MS/MS spectra may not be matched to the library derived from the FSEA.

Similarly, recent studies suggest that the triclin metabolism seems to specifically evolve in monocot plants. Tricin serves as a lignin co-monomer and cross couples with monolignols and γ -*p*-coumaroylated monolignols upon cell wall lignification. Recently, triclin was identified as a nucleation site for lignification in grasses (Lan et al., 2015, 2016a; Lam et al., 2017). Poales plants including *Avena sativa*, *Brachipodium distachyon*, *Oryza sativa*, *Triticum durum*, and *Zea mays* accumulate triclin, as does the Fabales plant *Medicago sativa* (Lan et al., 2016b). An integrative approach was used to detect derivatives of apigenin, luteolin, and triclin in *O. sativa*, *Z. mays*, and *M. truncatula*, and it appears that the total triclin content is much higher than the extractable triclin content in these plants. The integrative approach with an FSEA may be useful for further understanding the role of triclin and its derivatives in the cell wall.

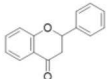
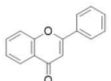
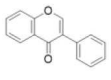
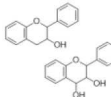
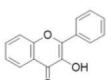
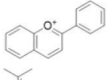
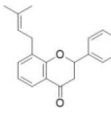
LC-MS/MS-based metabolome analyses have been used to comprehensively analyze the flavonoid molecules in individual plant species. Recent technological developments allow us to perform high-throughput LC-MS/MS analyses. By tracing the ions derived from the aglycones and modified parts of the molecules, or by finding MS/MS similarities, the chemical assignment of flavonoids is easily performed in (semi-)automatic ways with cheminformatic tools across plant species.

Biosynthetic genes for the formation of aglycone or modification can be predicted on the basis of the structure of flavonoids. Functional genomics using both cutting-edge sequencer and more comprehensive and accurate flavonoid profiling allow us to address the evolutionary questions of when and why flavonoids appeared and the biosynthetic pathway are diversified in plants. Furthermore, along with the progress of considerable plant genome projects, it will provide valuable clues to understand the evolutionary traits of flavonoid metabolism in plant kingdom as inconsistencies in the relationships between metabolites and genes that we mentioned in later sections.

THE FLAVONOID BIOSYNTHETIC PATHWAYS

The first step in flavonoid biosynthesis is catalyzed by chalcone synthase (CHS) (**Figure 2**). The substrates *p*-coumaroyl-CoA, derived from the cinnamate/monolignol (phenylpropanoid) pathway, and malonyl-CoA from the acetate-malonate

TABLE 1 | Flavonoids experimentally characterized by LC-MS/MS in 12 angiosperm species and *Marchantia polymorpha*.

Group	Representative structure	Name													
		Marchantiales		Asparagales		Poales		Solanales		Gentianales		Brassicales		Fabales	
		Marchantia polymorpha	Allium cepa	Oryza Sativa	Zea mays	Lycopersicon esculentum	Nicotiana tabacum	Solanum tuberosum	Optiorriza pumila	Arabidopsis thaliana	Glycine max	Glycyrrhiza glabra	Glycyrrhiza uralensis	Medicago truncatula	
Flavanone		Eriodictyol, hesperetin, isookanin, liquiritigenin, naringenin, sakuranetin, etc.	0 (0)	1 (0)	0 (2)	0 (0)	2 (0)	0 (0)	0 (0)	0 (0)	0 (0)	2 (0)	10 (5)	7 (17)	1 (0)
Flavone		Apigenin, baicalein, chrysoeriol, luteolin, scutellarein, tricetin, etc.	5 (3)	1 (5)	30 (19)	23 (17)	2 (0)	2 (13)	3 (0)	1 (0)	0 (0)	10 (4)	15 (2)	11 (5)	12 (0)
Isoflavone		Aformosin, biochanin, daizein, formononetin, genistein, etc.	0 (0)	0 (0)	0 (0)	1 (0)	0 (0)	0 (6)	0 (0)	8 (0)	0 (0)	18 (18)	14 (8)	17 (3)	5 (0)
Flavanol/flavandioid		Catechin, procyanidin	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	2 (0)	2 (0)	0 (0)
Flavonol		Gossypetin, isorhamnetin, kaempferide, kaempferol, laricitrin, myricetin, patuletin, quercetin, syringetin	0 (0)	10 (3)	2 (0)	6 (3)	7 (0)	3 (1)	10 (3)	0 (0)	9 (1)	6 (9)	19 (2)	19 (0)	15 (0)
Anthocyanidin		Cyanidin, delphinidin, petunidin, riccionidin, etc.	1 (2)	0 (10)	0 (0)	1 (0)	1 (0)	0 (0)	0 (6)	0 (0)	0 (3)	1 (0)	2 (0)	2 (0)	0 (0)
Prenylated flavonoid		Bolusanthol, gancaonin, glabrol, kievitone, licoisoflavone, sigmoidin, etc.	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (4)	13 (49)	10 (40)	0 (0)

The numbers of flavonoids characterized by Tsugawa et al. (2019) are shown. The numbers of flavonoids stored in the Dictionary of Natural Products are shown in parentheses. A conventional type of reverse phase LC-MS/MS was used for these analyses. On the basis of the ontology in the FSEA, the flavonoids are summarized into seven groups. Flavonoids with the prenyl moiety are grouped together as the prenylated flavonoid group. Prenylations (29 patterns) occur in chalcones/dihydrochalcones, flavanones, flavones, flavonols/dihydroflavonols, and isoflavones (Barron and Ibrahim, 1996). Flavonoids consisting of unmodified aglycones and glycosylated, acylated, etherated, and/or prenylated flavonoids were characterized in this study.

(polyketide) pathway are converted by CHS to naringenin chalcone. The stereospecific cyclization of naringenin chalcone to naringenin is catalyzed by chalcone isomerase (CHI). This step can also proceed spontaneously.

Naringenin is a general precursor for flavonols, anthocyanins, proanthocyanidins, flavones, and isoflavones. It is converted to dihydrokaempferol by flavanone 3-hydroxylase (F3H) (also referred to as flavanone 3 β -hydroxylase). Flavonoid 3'-hydroxylase (F3'H) and flavonoid F3'5'H-hydroxylase (F3'5'H) catalyze the hydroxylation of the C3' and C3'/C5' positions of dihydrokaempferol, respectively. Dihydroflavonol 4-reductase (DFR) catalyzes the reduction of the dihydroflavonols to leucoanthocyanidins, which are further converted to anthocyanidins by leucoanthocyanidin dioxygenase/anthocyanidin synthase (LDOX/ANS). The leucoanthocyanidins and anthocyanidins are reduced to flavan 3-ols (e.g., catechin and epicatechin) by leucoanthocyanidin reductase (LAR) and anthocyanidin reductase (ANR), respectively. The dihydroflavonols are also converted to flavonols by flavonol synthase (FLS).

Flavones are synthesized from naringenin by flavone synthase I (FNS I) or flavone synthase II (FNS II) (Martens and Mithofer, 2005). Flavanone 2-hydroxylase (F2H) catalyzes the hydroxylation of flavanones (including naringenin) to 2-hydroxyflavanones, which are subsequently converted to flavones, possibly by an unknown dehydratase (Akashi et al., 1998).

Isoflavone synthase (2-hydroxyisoflavanone synthase, IFS) catalyzes the first step of isoflavone biosynthesis. IFS converts flavanones (e.g., liquiritigenin and naringenin) to 2-hydroxyisoflavanones, then the 2-hydroxyisoflavanones are dehydrated to isoflavones by 2-hydroxyisoflavanone dehydratase (HID) (Akashi et al., 2005).

The flavonoid skeletons mentioned above are highly modified by enzymes such as glycosyltransferases (GTs), acyltransferases, and methyltransferases. Here, we focus on the enzymes involved in the biosynthesis of the flavonoid skeleton molecules.

EVOLUTIONARY HISTORY OF THE FLAVONOID BIOSYNTHETIC PATHWAYS

Generally, the enzymes in secondary metabolic pathways have been derived from those involved in primary metabolism (Pichersky and Gang, 2000; Moghe and Last, 2015; Carrington et al., 2018), and the enzymes in flavonoid metabolism are no exception. CHS and CHI are derived from enzymes involved in fatty acid metabolism: β -ketoacyl ACP synthase and fatty acid binding protein, respectively (Ngaki et al., 2012; Weng and Noel, 2012a). CHS and β -ketoacyl ACP synthase are members of the type III polyketide synthase (PKS) family (Schuz et al., 1983). The enzymes in later steps of the flavonoid biosynthetic pathways belong to families such as the 2OGD, CYP, and short-chain dehydrogenase/reductase (SDR) superfamilies (Table 2). The members of these superfamilies are widely involved in primary and secondary metabolisms, suggesting that plants have acquired the enzyme functions in later biosynthetic pathways

TABLE 2 | The origin or gene family of flavonoid biosynthetic genes.

Gene	Abbreviation	Origin/gene family
<i>Chalcone synthase</i>	CHS	Type III polyketide synthase
<i>Chalcone isomerase</i>	CHI	Fatty acid binding protein
<i>Chalcone isomerase-like</i>	CHIL	Fatty acid binding protein
<i>Flavanone 3-hydroxylase</i>	F3H	2-oxoglutarate-dependent dioxygenase (DOXC28)
<i>Flavone synthase I</i>	FNS I	2-oxoglutarate-dependent dioxygenase (DOXC28)
<i>Flavone synthase II</i>	FNS II	Cytochrome P450 (CYP93B)
<i>Isoflavone synthase</i>	IFS	Cytochrome P450 (CYP93C)
<i>Flavonoid 3'-hydroxylase</i>	F3'H	Cytochrome P450 (CYP75)
<i>Flavonoid 3',5'-hydroxylase</i>	F3'5'H	Cytochrome P450 (CYP75)
<i>Flavonol synthase</i>	FLS	2-oxoglutarate-dependent dioxygenase (DOXC47)
<i>Dihydroflavonol 4-reductase</i>	DFR	Short-chain dehydrogenase (SDR108E)
<i>Leucoanthocyanidin dioxygenase/anthocyanidin synthase</i>	LDOX/ANS	2-oxoglutarate-dependent dioxygenase (DOXC47)
<i>Anthocyanidin reductase</i>	ANR	Short-chain dehydrogenase (SDR108E)
<i>Leucoanthocyanidin reductase</i>	LAR	Short-chain dehydrogenase (SDR460A)
<i>Glycosyltransferase</i>	GT	Family 1 glycosyltransferase, glycoside hydrolase family 1
<i>Acyltransferase</i>	AT	BAHD acyltransferase, serine carboxypeptidase-like acyltransferase
<i>Methyltransferase</i>	MT	Methyltransferase

via gene duplication and evolution of new functions for the duplicated gene products.

Based on the distribution patterns of the flavonoid subclasses, it has been suggested that the flavonoid biosynthetic pathways may have evolved via a series of steps, and that the first flavonoid biosynthetic enzymes were CHS, CHI, and F3H (Stafford, 1991; Rausher, 2006). It has also been proposed that CHS evolved first, followed by F3H and then CHI, because CHS catalyzes the first committed step in the pathways, and the step catalyzed by CHI can also proceed spontaneously (Rausher, 2006). Flavonoids are widely distributed among mosses, liverworts, and vascular plants, but are not found in hornworts. Algae generally contain no flavonoids (Rausher, 2006), but they have been found in a few evolutionary divergent lineages of microalgae (Goiris et al., 2014). These observations suggest that the ability to produce flavonoids may have evolved multiple times, or that the ability was widely lost during evolutionary processes. An analysis of the evolutionary rates of six genes involved in anthocyanin biosynthesis indicated that the upstream genes (*CHS*, *CHI*, and *F3H*) evolved more slowly than the downstream genes (*DFR*, *LDOX/ANS*, and *UDP-glucose:flavonoid 3-O-glycosyltransferase*) (Rausher et al., 1999). The upstream genes may be evolutionarily constrained due to their profound effects on the pathways. In

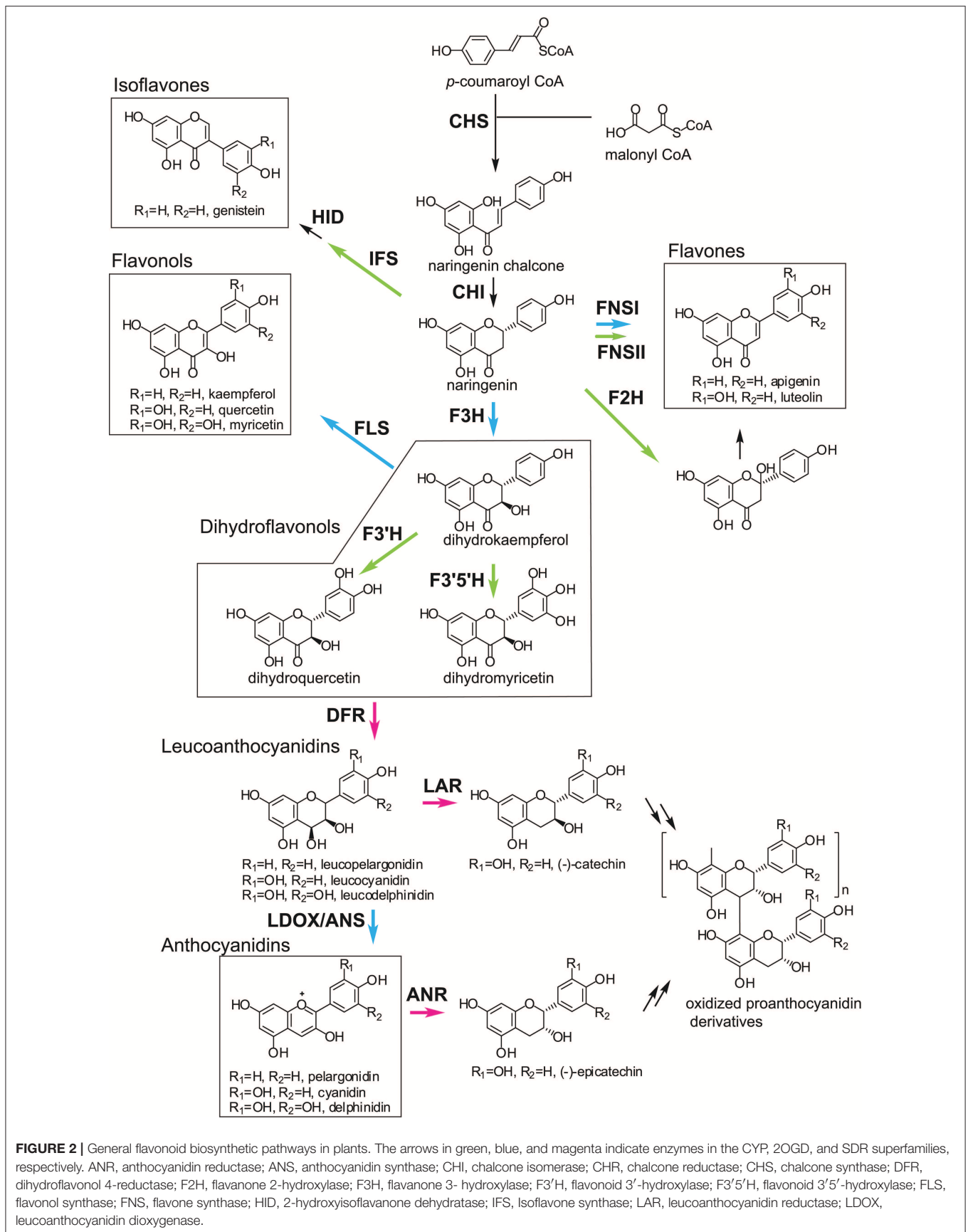


FIGURE 2 | General flavonoid biosynthetic pathways in plants. The arrows in green, blue, and magenta indicate enzymes in the CYP, 2OGD, and SDR superfamilies, respectively. ANR, anthocyanidin reductase; ANS, anthocyanidin synthase; CHI, chalcone isomerase; CHR, chalcone reductase; CHS, chalcone synthase; DFR, dihydroflavonol 4-reductase; F2H, flavanone 2-hydroxylase; F3H, flavanone 3'-hydroxylase; F3'H, flavonoid 3'-hydroxylase; F3'5'H, flavonoid 3'5'-hydroxylase; FLS, flavonol synthase; FNS, flavone synthase; HID, 2-hydroxyisoflavanone dehydratase; IFS, isoflavone synthase; LAR, leucoanthocyanidin reductase; LDOX, leucoanthocyanidin dioxygenase.

addition, the genes encoding transcription factors that regulate anthocyanin biosynthesis have evolved more rapidly than the structural genes (Rausher et al., 1999).

THE EVOLUTIONARY HISTORY OF FLAVONOID BIOSYNTHETIC GENES

CHS Is a Representative of the Type III PKS Superfamily

The *CHS* genes are widely distributed in plants, from bryophytes to angiosperms (Jiang et al., 2008; Shimizu et al., 2017; Liou et al., 2018), but they have not been found in other organisms. Every land plant species with available genomic data has at least one putative *CHS* gene (Shimizu et al., 2017). *CHS* is a member of the type III PKS superfamily, which provides diverse polyketide scaffolds of secondary metabolites (Figure 3) (Winkel-Shirley, 2001; Austin and Noel, 2003; Abe and Morita, 2010). The type III PKSs belong to the thiolase superfamily (Jiang et al., 2008). They are found in land plants, microalgae, fungi, and bacteria, but are not found in animals or archaea (Shimizu et al., 2017). The plant type III PKSs retain the overall folded protein structure and the Cys-His-Asn catalytic triad that characterize the *Escherichia coli* 3-ketoacyl-ACP synthase isoform III (KASIII) enzyme, which is involved in *de novo* fatty acid synthesis (Ferrer et al., 1999; Austin and Noel, 2003). The number of land plant type III *PKS* genes is highly variable among species in the same taxa; for instance, eudicot species have two to 42 genes, whereas the numbers in fungi and bacteria are relatively low (less than five) (Shimizu et al., 2017). The abundance of type III *PKS*s may contribute the variety of specialized plant metabolites.

The functional diversity of the type III *PKS*s is generally derived from differences in the starter molecules, the numbers of chain elongation steps, and the mechanisms of the cyclization reactions (Austin and Noel, 2003; Abe and Morita, 2010). The stilbene synthases (STSs), which produce stilbenes such as resveratrol, are also type III *PKS*s. *CHS*s and *STS*s generate the same intermediates from the same starter molecules using the same chain elongation steps, but they catalyze different intra-molecular cyclizations and produce different products (Austin et al., 2004). The *CHS*s and *STS*s from the same plant genera are usually classified as the closest neighbors in phylogenetic trees (Figure 3). Data from angiosperms (*Vitis vinifera*, *Arachis hypogaea*, and *Sorghum bicolor*), gymnosperms (*P. sylvestris*), and ferns (*Psilotum nudum*) suggest that after diverging, the *STS*s and *CHS*s evolved independently (Yu et al., 2005; Weng and Noel, 2012b).

Basal Land Plant CHSs and Non-CHSs

The eudicot *Arabidopsis thaliana* contains four type III *PKS* genes, including a single functional *CHS* (Kim et al., 2010). In contrast, the bryophyte moss *Physcomitrella patens* has around 19 *CHS* family genes and four other type III *PKS* genes (Jiang et al., 2006; Wolf et al., 2010). Five of the moss *CHS* family genes were derived from a whole genome duplication, and four are suggested to be derived from segmental duplication and transposition

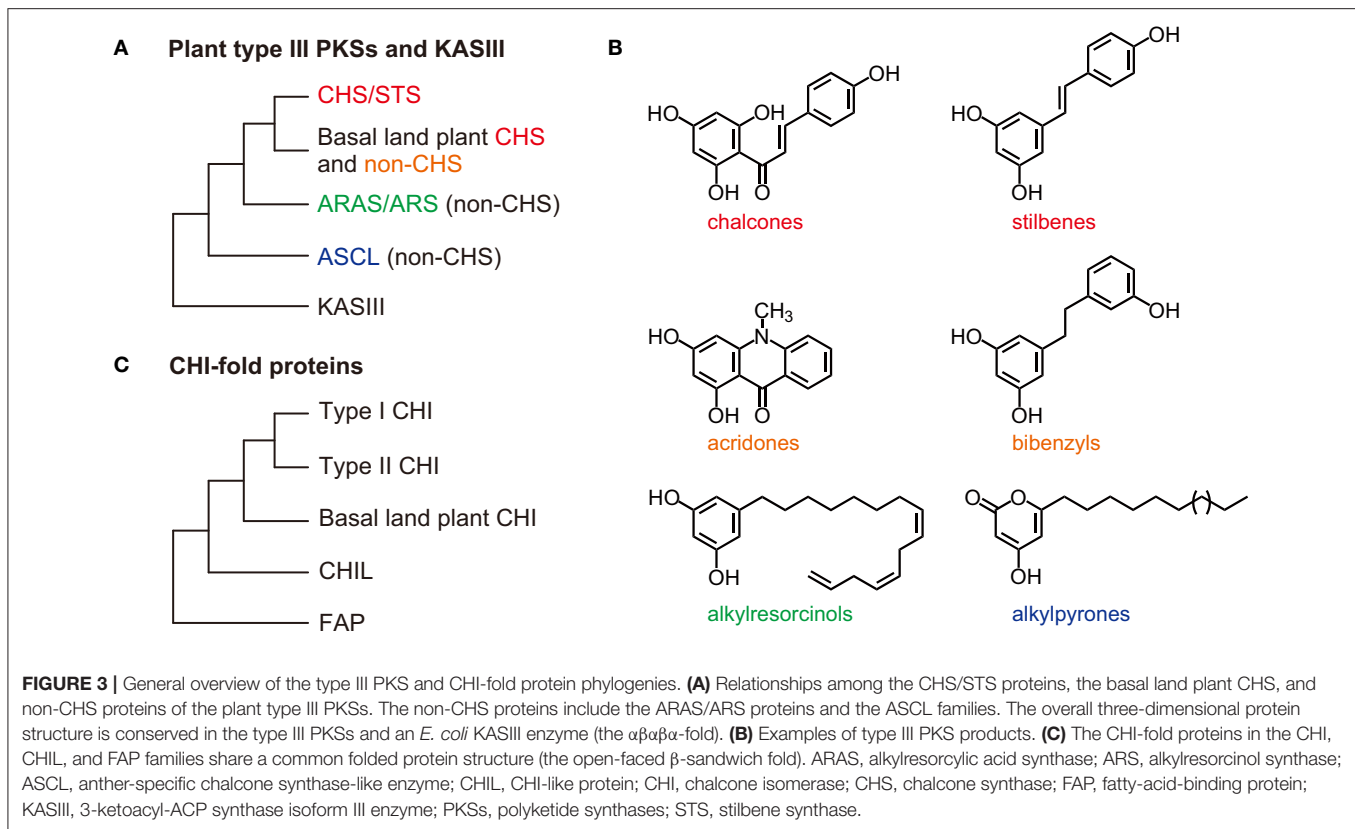
(Wolf et al., 2010). Four of the genes are upregulated by broadband UV-B irradiation, and the moss plants show increased levels of a flavonol derivative under UV-B illumination. These observations support the hypothesis that the genes and enzymes involved in the UV stress response evolved with the water-to-land transition, when early plants were exposed to increased levels of sunlight (Wolf et al., 2010).

In the bryophyte liverwort *M. polymorpha*, UV-B irradiation and nutrient deprivation significantly increase the total flavone glycoside content (Albert et al., 2018; Clayton et al., 2018). *M. polymorpha* has 24 *CHS* family genes (Bowman et al., 2017) and one of these is significantly upregulated by UV-B stress treatment. This induction was enhanced in transgenic *Marchantia* plants that overexpressed the gene encoding the UV RESISTANCE LOCUS8 (UVR8) photoreceptor (Clayton et al., 2018). Furthermore, transgenic *Marchantia* plants overexpressing a *MpMyb14* transcription factor gene showed increased expression levels of the same *CHS* gene under normal growth conditions (Albert et al., 2018). Knockout *Mpmyb14* mutants were partially impaired by an increase in *CHS* expression levels under nitrogen deficient conditions (Kubo et al., 2018). The results suggest that these liverwort species have at least one *CHS* gene that is activated by the UVR8 signal transduction pathway.

Phylogenetic analyses show close relationships between basal land plant (bryophyte and lycophyte) *CHS*s and non-*CHS* type III *PKS*s (Figure 3) (Wanibuchi et al., 2007; Yu et al., 2018). The non-*CHS* group includes many enzymes involved in the biosynthesis of secondary metabolites, such as acridone synthases, pyrone synthases, bibenzyl synthases, and *p*-coumaroyltriacetic acid synthases (Winkel-Shirley, 2001). These non-*CHS* *PKS* enzymes evolved through repeated gene duplication, mutation, and functional diversification from their ancestral plant enzymes.

Ectopic expression of a *CHS* gene from either the bryophyte *P. patens* or the lycophyte *Selaginella moellendorffii* can partially complement the phenotype of an *A. thaliana* *CHS*-null mutant, *transparent testa 4* (Liou et al., 2018). Crystal structures of *CHS*s from *P. patens*, *S. moellendorffii*, the monilophyte *Equisetum arvense*, the gymnosperm *Pinus sylvestris*, and the angiosperm *A. thaliana* revealed that the reactivity of the catalytic Cys residue (Cys164 in *M. sativa* *CHS2*) has changed during the 500 million years of evolution of land plants. The Cys residues in the three recent lineages (monilophytes, gymnosperms, and angiosperms) are present in the thiolate anion form, which gives them stronger nucleophilic power (Liou et al., 2018).

The type III *PKS*s show broad substrate promiscuity. *CHS*s do not accept bulky substrates, but the lycophyte *Huperzia serrata* HsPKS1 exhibits remarkable substrate tolerance and catalytic potential (Wanibuchi et al., 2007; Morita et al., 2011). *In vitro*, HsPKS1 produces naringenin chalcone and other polyketides, including aromatic tricyclic pyridoisoindole compounds, which are not found in natural products. A single amino acid replacement in HsPKS1 increases its active-site cavity volume and alters the product chain length and the mechanism of the cyclization reaction. This substrate promiscuity in the type III *PKS*s provides diverse polyketide scaffolds for the subsequent biosynthesis of secondary metabolites in land plants.



The A_{RAS}/A_{RS} and A_{SCL} Families

Plant non-C_{HS} type III PKSs also synthesize polyketides from fatty acyl-CoA substrates. Phenolic lipids such as alkylresorcinols are synthesized by alkylresorcinol synthases (A_{RS}s) and alkylresorcylic acid synthases (A_{RAS}s) in the monocots *S. bicolor* and *O. sativa*, respectively (Cook et al., 2010; Matsuzawa et al., 2010). Alkylresorcinols in grain crop species show anti-fungal and allelopathic activities.

Hydroxyalkyl-α-pyrone compounds, which are precursors of sporopollenin in the pollen wall exine, are synthesized by anther-specific chalcone synthase-like enzymes (A_{SCL}s) that are specifically and transiently expressed in *A. thaliana* anthers (PKSA/LAP6 and LAP5/PKSB) (Dobritsa et al., 2010; Kim et al., 2010). The *pksa pksb* double mutant plants are male sterile.

Phylogenetic analyses show that the A_{RAS}/A_{RS} and A_{SCL} families are classified into distinct groups (Figure 3) (Shimizu et al., 2017). Bacteria, fungi, and mosses also have genes encoding type III PKSs that produce these long-chain fatty acyl containing-polyketides (Colpitts et al., 2011; Shimizu et al., 2017; Li et al., 2018). Therefore, these type III PKSs are involved in lipid metabolism across three kingdoms of living organisms.

CHI-Fold Proteins: CHIs and CHILs

The CHIs and CHI-like proteins (CHILs) are members of the CHI-fold family, which also includes the fatty-acid-binding proteins (FAPs) that are involved in fatty acid biosynthesis (Figure 3). These proteins share a common folded three-dimensional structure (Jez et al., 2000; Ngaki et al., 2012; Kaltenbach et al., 2018). The FAP family is distributed in many

bacteria, fungi, and plant species. CHILs partly lack the catalytic amino acid residues conserved in CHIs, but they bind with C_{HS}s and enhance their activity (Ban et al., 2018). Phylogenetic and genomic analyses of CHI-fold proteins suggest that CHILs first appeared in mosses and evolved from FAPs, and then served as the ancestors of CHIs (Ngaki et al., 2012; Morita et al., 2014; Jiang et al., 2015). The CHILs form a group that is distinct from the CHI and FAP groups (Figure 3). The moss *P. patens* has two CHIL and four FAP genes, but does not appear to have any CHI genes (Ngaki et al., 2012; Cheng et al., 2018). The liverwort *M. paleacea* has a CHI, a CHIL, and two FAP genes (Cheng et al., 2018) and the lycophyte *S. moellendorffii* has a CHI, a CHIL, and three FAP genes. *A. thaliana* has a CHI, a CHIL, and three FAP genes whereas the legume *Glycine max* has a type I CHI, three type II CHIs, two CHILs, and six FAP genes (Dastmalchi and Dhaubhadel, 2015; Ban et al., 2018). These results suggest that the number of CHI and CHIL genes remains low in many plant lineages, but that leguminous plants have several CHI genes.

Type I and Type II CHIs in Vascular Plants

There are two types of CHI. Type I CHIs are ubiquitous in vascular plants, whereas type II CHIs are specific to legumes and are involved in flavonoid synthesis during nitrogen-fixing root nodule symbioses (Shimada et al., 2003; Subramanian et al., 2007). Phylogenetic analyses show that the type II CHIs form a distinct group from the type I CHIs (Figure 3). The type II CHIs likely evolved from ancestral CHI-fold proteins (Cheng et al., 2018).

The type II CHIs isomerize both naringenin chalcone and isoliquiritigenin (6'-deoxychalcone) to produce (2S)-naringenin and liquiritigenin (5-deoxyflavanone), respectively. The model legume *Lotus japonicus* has a type I and three type II *CHI* genes (Shimada et al., 2003). These four *CHI* genes form a tandem cluster within a 15-kb region of the genome. In soybean, a type I and two type II *CHI* genes are organized in a gene cluster on chromosome 20 (Dastmalchi and Dhaubhadel, 2015), and are probably derived from tandem gene duplications. The role of the CHIs in legume symbiosis remains unclear. Nitrogen-fixing root nodule symbioses are found in four angiosperm orders: Fabales, Fagales, Cucurbitales, and Rosales (the latter three are known as actinorhizal plants). The expression of a *CHI* gene is upregulated in nodules of the actinorhizal plant *Datisca glomerata* (Gifford et al., 2018), suggesting that these CHIs contribute to increases in flavonoid contents.

CHIs in Basal Land Plants

The *A. thaliana* *chi*-deficient *tt5* mutants produce pale yellow seeds due to significant reductions in proanthocyanidin production. This phenotype is largely complemented by ectopic expression of a *CHI* gene from either the liverwort *M. paleacea* (*MpCHI1*) or the lycophyte *S. moellendorffii* (*SmCHI1*) (Cheng et al., 2018). The CHIs from basal land plants have broad substrate specificities and are more like the type II CHIs than the type I CHIs. Phylogenetic analyses show that *MpCHI1* and *SmCHI1* form separate groups from the type I and type II CHIs (Figure 3) (Cheng et al., 2018; Kaltenbach et al., 2018).

The liverwort *M. polymorpha* also has a single *CHI* gene (Clayton et al., 2018), and *Marchantia chi* mutants do not contain detectable levels of flavone compounds. These *chi* mutants are highly sensitive to UV-B stress treatment. These results indicate that this basal plant species already has a gene encoding a *bona fide* CHI to catalyze the cyclization of naringenin chalcone in the flavonoid biosynthetic pathway.

The CHIL Family

CHILs are categorized as type IV CHI-fold proteins and are found in basal and higher plant species including mosses, liverworts, lycophytes, ferns, gymnosperms, and angiosperms. CHILs do not have *bona fide* CHI activity. However, the RNAi knockdown of *CHIL* expression in *Petunia hybrida* and *Torenia hybrida* resulted in decreased levels of total flavonoids in the flowers (Morita et al., 2014). Three independent *Ipomoea nil* (Japanese morning glory) mutants with alterations in *enhancer of flavonoid production* (a *CHIL* gene) showed pale-colored flower phenotypes (Morita et al., 2014). *CHIL* loss-of-function mutants in *A. thaliana* show reductions in the levels of proanthocyanidin and flavonols in seeds, and flavonols in leaves (Jiang et al., 2015). However, the *A. thaliana* *CHIL* gene could not rescue the phenotypes of the *tt5* mutants (Jiang et al., 2015). The liverwort *M. polymorpha* has a *CHIL* gene (Clayton et al., 2018), and the total flavone content is reduced in *Marchantia chil* mutants under normal growth conditions and under UV-B treatment. Thus, the CHILs enhance total flavonoid production but have roles that are distinct from those of the CHIs. As mentioned above, it was recently shown that CHILs from various plant lineages

(*A. thaliana*, *O. sativa*, *S. moellendorffii*, and *P. patens*) can bind CHSs and boost CHS activity (Ban et al., 2018). CHIs also bind CHSs (Jorgensen et al., 2005), and it may be that the type I and type II CHIs evolved from CHILs and gained their CHI activity during subsequent evolutionary processes (Ban et al., 2018).

Acquisition of Enzymatic CHI Activity During Evolution

The emergence of enzymatic CHIs in plants is an interesting topic in protein evolution, because the CHILs and FAPs are non-enzymatic proteins. The cleft in the CHI active site consists of three highly conserved amino acid residues (Arg36, Thr48, and Tyr106 in the *M. sativa* type II CHI sequence) and their neighboring residues are also conserved (Jez et al., 2000; Ngaki et al., 2012; Kaltenbach et al., 2018). The common ancestral proteins of the CHIs and CHILs were inferred in an extensive phylogenetic analysis, and it appears that all three key catalytic residues were conserved in the ancestral proteins but were inactive (Kaltenbach et al., 2018). The authors performed a stepwise, activity-based screening of recombinant ancestral proteins using an *E. coli* expression system. The results indicated that mutations in amino acid residues other than the catalytic residues were required to initiate CHI evolution, and to acquire CHI catalytic activity.

The 2-Oxoglutarate-Dependent Dioxygenase Family: F3H, FNS I, FLS, and LDOX/ANS

The 2OGD superfamily is one of the largest protein families in the plant kingdom. Its members are widely distributed in bacteria, fungi, plants, vertebrates, and even viruses (van den Born et al., 2008; Farrow and Facchini, 2014; Markolovic et al., 2015; Wu et al., 2016). The 2OGDs are non-heme iron containing enzymes that are localized in the cytosol. 2OGDs incorporate 2-oxoglutarate (2OG or α -ketoglutarate) and activated O₂ into a variety of substrates to form the oxidized products along with succinate and CO₂ ($\text{RH} + 2\text{OG} + \text{O}_2 \rightarrow \text{ROH} + \text{succinate} + \text{CO}_2$). 2OGDs catalyze various oxidative reactions including hydroxylation, halogenation, desaturation, and epimerization (Martinez and Hausinger, 2015) and play important roles in DNA and RNA repair, fatty acid metabolism, oxygen sensing, and biosynthesis of natural products (Farrow and Facchini, 2014; Hagel and Facchini, 2018; Herr and Hausinger, 2018; Islam et al., 2018).

In plants, 2OGDs are involved in histone demethylation, iron sensing, phytohormone metabolism, and the biosynthesis of secondary metabolites (reviewed in Farrow and Facchini, 2014). A phylogenetic analysis of plant 2OGDs from *Chlamydomonas reinhardtii*, *P. patens*, *S. moellendorffii*, *Picea abies*, *O. sativa*, and *A. thaliana* found 3 classes, which the authors named DOXA, DOXB, and DOXC (Kawai et al., 2014). Each class contains proteins from all six species. The DOXA class contains homologs of *E. coli* AlkB; these enzymes are involved in DNA repair (Lindahl et al., 1988; Meza et al., 2012; Mielecki et al., 2012). The DOXB class contains prolyl 4-hydroxylases that catalyze the hydroxylation of proline residues in plant cell

wall proteins (Hieta and Myllyharju, 2002). Proteins in the DOXC class are involved in phytohormone metabolism and the biosynthesis of secondary metabolites including flavonoids, terpenoids, alkaloids, and glucosinolates. The numbers of genes encoding DOXA and DOXB enzymes are limited in the six species, however, the DOXC genes are significantly expanded in the land plants (Kawai et al., 2014).

Four flavonoid biosynthetic enzymes, F3H, FNS I, FLS, and LDOX/ANS belong to the DOXC class. A phylogenetic analysis showed that the genes in the DOXC class can be classified into over 50 clades (Kawai et al., 2014). The *F3H* and *FNS I* genes are in the DOXC28 clade while the *FLS* and *LDOX/ANS* genes are in DOXC47. Among the flavonoid biosynthetic genes in the *2OGD* superfamily, it has been proposed that *F3H* was the first to appear (Rausher, 2006). FNS I seems to exist only in the Apiaceae, and it is likely that *FNS I* evolved from *F3H* as a paraphyletic gene (Martens et al., 2003; Gebhardt et al., 2005).

Arabidopsis thaliana contains one *F3H* gene, six *FLS* genes (*AtFLS1–AtFLS6*), one *LDOX/ANS* gene, and no FNS gene. *AtFLS1* is the major FLS (Owens et al., 2008a; Saito et al., 2013). An analysis of structural divergence between duplicated genes showed that transposed duplication (<16 million years ago) explains the relationship between *AtFLS6* (At5g43935) and *AtF3H* (At3g51240) (Wang et al., 2013). Furthermore, the relationship between *AtFLS1* (At5g08640) and *AtFLS5* (At5g63600) is likely explained by a whole genome duplication, and those between *AtFLS2* (At5g63580) and *AtFLS3* (At5g63590), *AtFLS3* and *AtFLS4* (At5g6359580), and *AtFLS4* and *AtFLS5* are most likely explained by tandem local duplications (Wang et al., 2013). The *A. thaliana* *AtLDOX/ANS* has been shown to produce flavonols *in planta* (Preuss et al., 2009). Furthermore, FLS and LDOX/ANS can partially complement *F3H* function *in vivo*, and this results in the leaky phenotype of *tt6* mutants with null mutations in *AtF3H* (Owens et al., 2008b).

Together, the results suggest that *F3H* is the ancestral *2OGD* gene for flavonoid biosynthesis, and that *FLS* and *FNS I* evolved via divergence from *F3H*. No apparent orthologs of either the DOXC28 or DOXC47 clade genes were found in *P. patens* and *S. moellendorffi*, even though these plants produce flavonols (*P. patens*) and flavones (*S. moellendorffi*) (Wolf et al., 2010; Weng and Noel, 2013). However, the liverwort *Plagiochasma appendiculatum* has an active FNS I (PaFNS I), and a phylogenetic analysis revealed that PaFNS I is related to the angiosperm FNS I and *F3H* proteins, even though it is not in the same clade as them (Han et al., 2014). These data suggest that in *Physcomitrella* and *Selaginella*, the *2OGDs* are present in distinct clade(s), or that unrelated enzymes perform the same functions as *F3H*, FNS, and/or FLS.

The PaFNS I can convert naringenin to either 2-hydroxynaringenin or apigenin (Han et al., 2014). The common horsetail *E. arvense* L also has *2OGD*-type FNS I activity (Bredbach et al., 2011). Further research in the *2OGD* genes of bryophytes, lycophytes, and ferns will help to clarify their evolutionary processes.

The Cytochrome P450 Superfamily: FNSII, F2H, IFS, F3'H, and F3'5'H

The CYPs are widely distributed in viruses, archaea, bacteria, and eukaryotes. They catalyze monooxygenase/hydroxylation reactions in various primary and secondary metabolic processes by insertion of an O atom from molecular O₂ (Mizutani and Ohta, 2010). In eukaryotes, the CYPs are heme-containing membrane proteins localized on the cytosolic surface of the endoplasmic reticulum.

In plants, the CYPs form the largest superfamily of enzymes and account for about 1% of the total number of gene products (Mizutani and Ohta, 2010; Nelson and Werck-Reichhart, 2011; Kawai et al., 2014). The CYPs are categorized into families (e.g., CYP75) that have ~40% or more amino acid sequence identity, and those with 55 % or more identity are categorized into subfamilies (e.g., CYP75A). Furthermore, the plant CYP families (CYP71 to CYP99 and CYP701–) can be classified into clans whose members are derived from single ancestors. The land plant CYPs form 11 clans and seven of these (clans 51, 74, 97, 710, 711, 727, and 746) consist of single CYP families, while the remaining four (clans 71, 72, 85, and 86) include proteins in multiple CYP families (Nelson and Werck-Reichhart, 2011). Green algae contain CYPs in five single family clans (clans 51, 97, 710, 711, and 746), and members of these clans are involved in fundamental biological processes such as biosynthesis of sterols, xanthophylls, and phytohormones (Nelson, 2006; Nelson and Werck-Reichhart, 2011). Therefore, it is likely that these clans include the ancestral CYPs. The multi-family CYP clans have become highly diversified during plant evolution (Nelson and Werck-Reichhart, 2011). Some CYP families in these clans (clans 71, 72, 85, and 86) are present in bryophytes and/or liverworts but not in green algae. Additional novel CYP families were gained in stepwise processes following the evolution of vascular plants.

Flavonoid biosynthetic enzymes (members of the CYP75 and CYP93 families) are classified into clan 71. This clan contains the largest number of CYP families, and in addition to the flavonoid biosynthetic enzymes, it includes families involved in the biosynthesis of phenylpropanoids (CYP73, CYP84, CYP98), alkaloids (CYP80, CYP82, CYP719), terpenoids (CYP76, CYP99, CYP705, CYP706, CYP726), and glucosinolates (CYP79, CYP83). The clan 71 families CYP73, CYP74, CYP78, CYP88, CYP98, CYP701, CYP703, CYP736, and CYP761 are present in bryophytes, suggesting that these may be ancestral families.

The CYP75 Family: F3'H and F3'5'H

The F3'H and F3'5'H enzymes generally belong to the CYP75B and CYP75A families, respectively. The CYP75 family enzymes from monocots and dicots form distinct clusters within each subfamily, suggesting that the F3'H and F3'5'H functions were established before the divergence of monocots and dicots. However, there are some exceptions. In the Asteraceae, F3'5'H belongs to the CYP75B subfamily and forms a distinct cluster from F3'H (Seitz et al., 2006). This suggests that the Asteraceae F3'H gained F3'5'H activity before speciation but after the separation of the monocots and dicots. Similarly, the rice

CYP75B4 catalyzes the 5'-hydroxylation of 3'-methoxyflavone chrysoeriol, and also functions as an F3'H (Lam et al., 2015). Rice contains CYP75B3 as an F3'H and CYP75A11 as a non-functional F3'5'H. Such CYP gene distributions in the CYP75A and CYP75B subfamilies are also found in other Poaceae plants. These data suggest that the CYP75A and CYP75B subfamilies separated before divergence of monocots and dicots, and that genes in the CYP75B subfamily later gained F3'5'H activity, at least in the Asteraceae and Poaceae.

Arabidopsis thaliana has a single gene for F3'H (CYP75B1, At5g07990) and no genes corresponding to F3'5'H, FNS II, or F2H. The CYP75B1 gene appears to be related to the CYP701A (At5g25900) gene, which encodes an enzyme involved in gibberellin biosynthesis, via a transposed duplication that occurred 16–107 million years ago and it was proposed that CYP75B1 is the parental locus of CYP701A (Wang et al., 2013). However, the CYP701 family is distributed among bryophytes and vascular plants while the CYP75 family is found in gymnosperms and angiosperms. In addition, a phylogenetic analysis suggested that the moss/liverwort-specific CYP761 family is closely related to the CYP75 family (Nelson and Werck-Reichhart, 2011). These data suggest that either CYP701 or CYP761 is the ancestral family of CYP75. Although 3'-hydroxylated flavone derivatives were detected in *M. polymorpha* and *S. moellendorffii* (Markham et al., 1998), CYP75 family members can be found only in gymnosperms and angiosperms (Nelson and Werck-Reichhart, 2011). Therefore, in bryophytes, lycophytes, and ferns, enzymes from other CYP families may function as F3'H and/or F3'5'H.

Such inconsistencies in the relationships between metabolites and genes are also observed in phenylpropanoid metabolism. Among the CYPs involved in phenylpropanoid metabolism from clan 71, CYP73A (cinnamate 4-hydroxylase, C4H), CYP98A (*p*-coumaroyl shikimate 3'-hydroxylase, C3'H), CYP73, and CYP98 first appeared in liverworts and mosses, while CYP84 (ferulate 5-hydroxylase, F5H) is found only in angiosperms (Nelson and Werck-Reichhart, 2011). However, the syringyl lignin units from the phenylpropanoid pathway are distributed in land plants (Bowman et al., 2017), suggesting that genes in other CYP families may have F5H activities. In fact, the *Selaginella* CYP788A1 functions as a F5H (Weng et al., 2008).

The CYP93 Family: FNS II, F2H, and IFS

A genome-wide analysis of the CYP93 family genes from 60 green plants indicated that the CYP93 family is found only in angiosperms (Du et al., 2016). Among the 10 subfamilies (CYP93A–CYP93K), CYP93A is the ancestral group distributed in both monocots and dicots; CYP93B and CYP93C are distributed only in dicots; CYP93G and CYP93J are found only in monocots; and CYP93E and CYP93F are specific to legumes and grasses, respectively. Thus, the CYP93 family shows plant lineage-specific evolution.

The CYP93 family contains enzymes involved in the biosynthesis of flavones and isoflavones (FNS II, F2H, and IFS). The monocot FNS II and F2H belong to the CYP93G subfamily, while those in the dicots are categorized in the CYP93B subfamily. Therefore, phylogenetic analyses suggest that the functions of the FNSII and F2H enzymes were established

after the divergence of monocots and dicots. The IFS enzyme in legumes is a member of the CYP93C subfamily, which may be derived from the CYP93B subfamily (Du et al., 2016).

As with the CYP75 family, there are inconsistencies in the relationships between metabolites and genes in the CYP93 family. The CYP93 family genes are found only in angiosperms (Nelson and Werck-Reichhart, 2011), but flavones are widely distributed in plants, from bryophytes to angiosperms (Figure 1). This suggests that enzymes involved in flavone biosynthesis belong to the CYP93 family in some species of angiosperms, but that different CYP families and/or other enzymes play the same roles in other plant taxa. For example, the Apiaceae family adapted the 2OGD-type FNS I rather than the CYP-type FNS II to produce flavones (Gebhardt et al., 2007).

Isoflavonoids are the typical flavonoids found in legumes, and isoflavone biosynthetic genes are found only in legumes. However, isoflavonoids are also found in non-legume plants including *Iris*, which contains a wide variety of isoflavones, some mosses (e.g., *Bryum capillare*), gymnosperms, and monocot and dicot angiosperms (Dewick, 1994; Lapcik, 2007). This suggests that plants independently acquired the ability to produce isoflavonoids during the evolution of the CYP families, except for CYP93C and/or other non-CYP enzymes that function as IFSs.

The Short-Chain Dehydrogenase/Reductase Family: DFR, ANR, and LAR

DFR, *ANR*, and *LAR* are members of the SDR superfamily, which is widely distributed in viruses, archaea, prokaryotes, and eukaryotes (Jornvall et al., 1999; Kavanagh et al., 2008). The SDRs constitute one of the largest NAD(P)(H) dependent oxidoreductase families and are involved in the primary metabolism of lipids, carbohydrates, and hormones, and the secondary metabolism of molecules such as terpenoids, alkaloids, and phenolic compounds (Jornvall et al., 1999; Kavanagh et al., 2008; Tonfack et al., 2011). In spite of the low overall sequence similarities among SDRs (15–30%), the SDRs possess a conserved 3D structure consisting of a Rossmann-fold β -sheet surrounded by α -helices for nucleotide binding. Generally, the SDRs can be classified into several types (“classical,” “extended,” “intermediate,” “divergent,” and “complex”) based on their primary structures, cofactor binding motifs, and active sites (Kavanagh et al., 2008; Moummou et al., 2012). A recent study revealed that the “intermediate” and “complex” types are not found in plants while the types “atypical” and “unknown” are found in plants. Thus, the plant SDRs can be categorized into five types: “classical,” “divergent,” “extended,” “atypical,” and “unknown” (Moummou et al., 2012). Among the plant SDRs, “classical” and “extended” are the major types, as they are in other organisms. The “classical” type is composed of about 250 amino acid residues, and the “extended” type has a domain of 100 additional amino acid residues in the C-terminal region. The “atypical” type, an uncommon type of SDRs, was included in the SDR family because of its Rossmann-fold structure, which is typical of SDRs (Moummou et al., 2012).

A genome sequence analysis using 10 species including *C. reinhardtii*, *P. patens*, *S. moellendorffii*, four dicots (*A. thaliana*,

Populus trichocarpa, *V. vinifera*, and *G. max*) and three monocots (*O. sativa*, *S. bicolor*, and *Z. mays*) showed that most plant SDRs can be classified into 49 families that are distributed among the five types mentioned above (Moummou et al., 2012). The DFR and ANR enzymes are classified into the SDR108E family in the “extended” type. The LAR enzymes belong to the SDR460A family, which is an “atypical” type (Kallberg et al., 2010; Moummou et al., 2012).

The SDR108E Family: DFR and ANR

Compared to other families that contain a few genes per species, the SDR108E family contains the largest number of SDR genes; for example, this family includes 24 genes in *A. thaliana* and 44 genes in *O. sativa*. Furthermore, the SDR108E family shows the lowest average sequence identity, indicating that the SDR108E family genes are highly diversified (Moummou et al., 2012). The distribution of SDR108E family genes in 10 species indicates that this family has expanded significantly in vascular plants.

In addition to the DFR and ANR enzymes, the SDR108E family contains other enzymes involved in secondary metabolism, including cinnamoyl-CoA reductase (CCR) for lignin biosynthesis, and phenylacetaldehyde reductase for the production of volatile 2-phenylethanol. Furthermore, enzymes involved in phytohormone metabolism, such as phaseic acid reductase for abscisic acid catabolism (Weng et al., 2016) and BEN1 for brassinosteroid homeostasis (Yuan et al., 2007) belong to this family. Each enzyme type forms a distinct cluster in the phylogenetic tree. The CCR and phenylacetaldehyde reductase clusters contain enzymes from *P. patens* and *S. moellendorffii*, whereas the clusters of DFRs, ANRs, phaseic acid reductases, and BEN1 enzymes are derived from flowering plants. These data suggest that these DFRs, ANRs, and other enzymes appeared more recently than the CCRs and phenylacetaldehyde reductases.

Other SDR families show similar gene expansion patterns to that of SDR108E. These families (SDR110C, SDR114C, SDR65C, and SDR460A) contain genes involved in the biosynthesis of alkaloids, terpenoids, phenylpropanoids, and phytohormones (Moummou et al., 2012). SDR families that are less diversified contain genes involved in primary metabolism, such as lipid and chlorophyll biosynthesis (Moummou et al., 2012).

The SDR460A Family: LAR

The SDR460A family is also referred to as the PIP family, named after the first three enzymes (pinoresinol-laricirecinol reductase, isoflavone reductase, and phenylcoumaran benzylic ether reductase) that were discovered to belong to this family (Gang et al., 1999; Min et al., 2003; Wang et al., 2006). In addition, vestitone reductase, eugenol synthase, and isoeugenol synthase also belong to this family (Koeduka et al., 2008). Pinoresinol-laricirecinol reductase and phenylcoumaran benzylic ether reductase function in the lignan biosynthetic pathway; isoflavone reductase and vestitone reductase are involved in isoflavonoid biosynthesis; and eugenol synthase and isoeugenol synthase are involved in the biosynthesis of volatile phenylpropenes. Thus, the SDR460A family members are involved in the biosynthesis of various phenolic compounds. The SDR460A family members (“atypical” type SDRs) are limited to and has greatly expanded in

vascular plants (Moummou et al., 2012), suggesting that they are needed for vascular plant prosperity.

A phylogenetic analysis of LARs from various plants revealed that the plant LARs can be classified into two clusters: proteins from dicotyledons, and proteins from monocotyledons and gymnosperms (Wang et al., 2018). Therefore, the monocotyledon LARs are more closely related to the gymnosperm LARS than the dicotyledon LARS.

FUTURE PERSPECTIVES

The progress in metabolomics technologies including chemoinformatics and the abundant genomic information of flavonoid biosynthetic genes facilitated a fully understanding of evolution of the flavonoid/phenylpropanoid metabolisms in plant kingdom. In this review, we have focused on the evolution of enzymes involved in the biosynthesis of flavonoid skeleton molecules. The basic structures of flavonoids are formed by a type III PKS, CHS. The broad substrate promiscuity and functional diversity of the type III PKSs may be a driving force for expanding the chemical variety of specialized metabolites. In other specialized metabolisms, isomerases (e.g., oxidosqualene synthase) and lyases (e.g., terpene synthase) are also involved in scaffold formation and may contribute the chemical diversity of secondary metabolites. Modification enzymes such as GTs (glycosyltransferases) and acyltransferases also contribute greatly to the huge diversity of flavonoids and other secondary metabolites. Interestingly, plants have two types of flavonoid GTs: the cytosolic family 1 GTs and the vacuolar glycoside hydrolase family 1 (GH1) (Cao et al., 2017). Acylation is also catalyzed by differentially localized enzymes: cytosolic BAHD acyltransferases and vacuolar SCPL acyltransferases, derived from serine carboxypeptidase (Milkowski and Strack, 2004; Moghe and Last, 2015). It is still unknown why plants have evolved these differentially localized enzymes for the modification of flavonoids and other specialized metabolites. The evolution of the family 1 GTs, the GH1s, and the BAHD acyltransferases in plants have been reviewed elsewhere (St-Pierre and De Luca, 2000; Yu et al., 2009; Tuominen et al., 2011; Yonekura-Sakakibara and Hanada, 2011; Caputi et al., 2012; Moghe and Last, 2015; Cao et al., 2017). Throughout their long history, plants have engineered their metabolic pathways to adapt themselves to their habitats and growth conditions in tissue and organ specific manners. A detailed understanding of the evolutionary history of metabolic enzymes involved in biosynthesis, modification, transport, secretion, transcriptional regulation, and chemodiversity will assist us in the engineering of specialized metabolic pathways to produce desirable metabolites with minimal energy expenditures.

AUTHOR CONTRIBUTIONS

KY-S proposed the concept. KY-S, YH, and RN developed it and wrote the manuscript.

FUNDING

This work was partially supported by the JSPS KAKENHI program (grant number 17K07460 to KY-S),

REFERENCES

- Abe, I., and Morita, H. (2010). Structure and function of the chalcone synthase superfamily of plant type iii polyketide synthases. *Nat. Prod. Rep.* 27, 809–838. doi: 10.1039/b909988n
- Afendi, F. M., Okada, T., Yamazaki, M., Hirai-Morita, A., Nakamura, Y., Nakamura, K., et al. (2012). Knapsack family databases: integrated metabolite-plant species databases for multifaceted plant research. *Plant Cell Physiol.* 53:e1. doi: 10.1093/pcp/pcr165
- Akashi, T., Aoki, T., and Ayabe, S. (1998). Identification of a cytochrome p450 cDNA encoding (2s)-flavanone 2-hydroxylase of licorice (*glycyrrhiza echinata* L.; *Fabaceae*) which represents licodione synthase and flavone synthase ii. *FEBS Lett.* 431, 287–290. doi: 10.1016/S0014-5793(98)00781-9
- Akashi, T., Aoki, T., and Ayabe, S. (2005). Molecular and biochemical characterization of 2-hydroxyisoflavanone dehydratase. Involvement of carboxylesterase-like proteins in leguminous isoflavone biosynthesis. *Plant Physiol.* 137, 882–891. doi: 10.1104/pp.104.056747
- Akimoto, N., Ara, T., Nakajima, D., Suda, K., Ikeda, C., Takahashi, S., et al. (2017). Flavonoidsearch: a system for comprehensive flavonoid annotation by mass spectrometry. *Sci. Rep.* 7:1243. doi: 10.1038/s41598-017-01390-3
- Albert, N. W., Thrimawithana, A. H., McGhie, T. K., Clayton, W. A., Deroles, S. C., Schwinn, K. E., et al. (2018). Genetic analysis of the liverwort *Marchantia polymorpha* reveals that r2r3myb activation of flavonoid production in response to abiotic stress is an ancient character in land plants. *New Phytol.* 218, 554–566. doi: 10.1111/nph.15002
- Anderson, Ø. M., and Markham, K. R. (2006). *Flavonoids: Chemistry, Biochemistry and Applications*. Boca Raton: CRC Press.
- Asakawa, Y., Takikawa, K., Toyota, M., and Takemoto, T. (1982). Chemosystematics of the bryophytes.12. Novel bibenzyl derivatives and ent-cuparene-type sesquiterpenoids from radula species. *Phytochemistry* 21, 2481–2490. doi: 10.1016/0031-9422(82)85245-X
- Asakawa, Y., Toyota, M., and Takemoto, T. (1978). 7 new bibenzyls and a dihydrochalcone from radula-variabilis. *Phytochemistry* 17, 2005–2010. doi: 10.1016/S0031-9422(00)88752-X
- Austin, M. B., Bowman, M. E., Ferrer, J. L., Schroder, J., and Noel, J. P. (2004). An aldol switch discovered in stilbene synthases mediates cyclization specificity of type iii polyketide synthases. *Chem. Biol.* 11, 1179–1194. doi: 10.1016/j.chembiol.2004.05.024
- Austin, M. B., and Noel, J. P. (2003). The chalcone synthase superfamily of type iii polyketide synthases. *Nat. Prod. Rep.* 20, 79–110. doi: 10.1039/b100917f
- Ban, Z. N., Qin, H., Mitchell, A. J., Liu, B. X., Zhang, F. X., Weng, J. K., et al. (2018). Noncatalytic chalcone isomerase-fold proteins in humulus lupulus are auxiliary components in prenylated flavonoid biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* 115, E5223–E5232. doi: 10.1073/pnas.1802223115
- Barron, D., and Ibrahim, R. K. (1996). Isoprenylated flavonoids - a survey. *Phytochemistry* 43, 921–982. doi: 10.1016/S0031-9422(96)00344-5
- Bowman, J. L., Kohchi, T., Yamato, K. T., Jenkins, J., Shu, S. Q., Ishizaki, K., et al. (2017). Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. *Cell.* 171, 287–304. doi: 10.1016/j.cell.2017.09.030
- Bredebach, M., Matern, U., and Martens, S. (2011). Three 2-oxoglutarate-dependent dioxygenase activities of *Equisetum arvense* L. Forming flavone and flavonol from (2s)-naringenin. *Phytochemistry* 72, 557–563. doi: 10.1016/j.phytochem.2011.01.036
- Cao, Y. Y., Yang, J. F., Liu, T. Y., Su, Z. F., Zhu, F. Y., Chen, M. X., et al. (2017). A phylogenetically informed comparison of gh1 hydrolases between *Arabidopsis* and rice response to stressors. *Front. Plant Sci.* 8:350. doi: 10.3389/fpls.2017.00350
- Caputi, L., Malnoy, M., Goremykin, V., Nikiforova, S., and Martens, S. (2012). A genome-wide phylogenetic reconstruction of family 1 UDP-glycosyltransferases revealed the expansion of the family during the adaptation of plants to life on land. *Plant J.* 69, 1030–1042. doi: 10.1111/j.1365-313X.2011.04853.x
- Carrington, Y., Guo, J., Le, C. H., Fillo, A., Kwon, J., Tran, L. T., et al. (2018). Evolution of a secondary metabolic pathway from primary metabolism: shikimate and quinate biosynthesis in plants. *Plant J.* 95, 823–833. doi: 10.1111/tpj.13990
- Cheng, A. X., Zhang, X. B., Han, X. J., Zhang, Y. Y., Gao, S., Liu, C. J., et al. (2018). Identification of chalcone isomerase in the basal land plants reveals an ancient evolution of enzymatic cyclization activity for synthesis of flavonoids. *New Phytol.* 217, 909–924. doi: 10.1111/nph.14852
- Clayton, W. A., Albert, N. W., Thrimawithana, A. H., McGhie, T. K., Deroles, S. C., Schwinn, K. E., et al. (2018). Uvr8-mediated induction of flavonoid biosynthesis for UVB tolerance is conserved between the liverwort *Marchantia polymorpha* and flowering plants. *Plant J.* 96, 503–517. doi: 10.1111/tpj.14044
- Colpitts, C. C., Kim, S. S., Posehn, S. E., Jepson, C., Kim, S. Y., Wiedemann, G., et al. (2011). Ppascl, a moss ortholog of anther-specific chalcone synthase-like enzymes, is a hydroxyalkylpyrone synthase involved in an evolutionarily conserved sporopollenin biosynthesis pathway. *New Phytol.* 192, 855–868. doi: 10.1111/j.1469-8137.2011.03858.x
- Cook, D., Rimando, A. M., Clemente, T. E., Schroder, J., Dayan, F. E., Nanayakkara, N. P. D., et al. (2010). Alkylresorcinol synthases expressed in sorghum bicolor root hairs play an essential role in the biosynthesis of the allelopathic benzoquinone sorgoleone. *Plant Cell.* 22, 867–887. doi: 10.1105/tpc.109.072397
- Dastmalchi, M., and Dhaubhadel, S. (2015). Soybean chalcone isomerase: evolution of the fold, and the differential expression and localization of the gene family. *Planta* 241, 507–523. doi: 10.1007/s00425-014-2200-5
- Dewick, P. M. (1994). "Isoflavonoids," in *The Flavonoids Advances in Research Since 1986*, ed J. B. Harborne (Chapman and Hall/CRC: United States of America), 117–238.
- Dobritsa, A. A., Lei, Z. T., Nishikawa, S., Urbanczyk-Wochniak, E., Huhman, D. V., Preuss, D., et al. (2010). Lap5 and lap6 encode anther-specific proteins with similarity to chalcone synthase essential for pollen exine development in *Arabidopsis*. *Plant Physiol.* 153, 937–955. doi: 10.1104/pp.110.157446
- Du, H., Ran, F., Dong, H. L., Wen, J., Li, J. N., and Liang, Z. (2016). Genome-wide analysis, classification, evolution, and expression analysis of the cytochrome p450 93 family in land plants. *PLoS ONE* 11:e0165020. doi: 10.1371/journal.pone.0165020
- Farrow, S. C., and Facchini, P. J. (2014). Functional diversity of 2-oxoglutarate/Fe(II)-dependent dioxygenases in plant metabolism. *Front. Plant Sci.* 5:524. doi: 10.3389/fpls.2014.00524
- Ferrer, J. L., Jez, J. M., Bowman, M. E., Dixon, R. A., and Noel, J. P. (1999). Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nat. Struct. Biol.* 6, 775–784. doi: 10.1038/11553
- Gang, D. R., Kasahara, H., Xia, Z. Q., Vander Mijnsbrugge, K., Bauw, G., Boerjan, W., et al. (1999). Evolution of plant defense mechanisms. Relationships of phenylcoumaran benzylic ether reductases to pinoresinol-laricresinol and isoflavone reductases. *J. Biol. Chem.* 274, 7516–7527. doi: 10.1074/jbc.274.11.7516
- Gebhardt, Y., Witte, S., Forkmann, G., Lukacin, R., Matern, U., and Martens, S. (2005). Molecular evolution of flavonoid dioxygenases in the family apiaceae. *Phytochemistry* 66, 1273–1284. doi: 10.1016/j.phytochem.2005.03.030

- Gebhardt, Y. H., Witte, S., Steuber, H., Matern, U., and Martens, S. (2007). Evolution of flavone synthase i from parsley flavanone 3beta-hydroxylase by site-directed mutagenesis. *Plant Physiol.* 144, 1442–1454. doi: 10.1104/pp.107.098392
- Gifford, I., Battenberg, K., Vaniya, A., Wilson, A., Tian, L., Fiehn, O., et al. (2018). Distinctive patterns of flavonoid biosynthesis in roots and nodules of *Datisca glomerata* and *medicago* spp. Revealed by metabolomic and gene expression profiles. *Front. Plant Sci.* 9:1463. doi: 10.3389/fpls.2018.01463
- Goiris, K., Muylaert, K., Voorspoels, S., Noten, B., De Paep, D., Baart, G. J. E., et al. (2014). Detection of flavonoids in microalgae from different evolutionary lineages. *J. Phycol.* 50, 483–492. doi: 10.1111/jpy.12180
- Gould, K. S., and Lister, C. (2006). “Flavonoid functions in plants,” in *Flavonoids: Chemistry, Biochemistry and Applications*, eds Ø. M. Anderson and K. R. Markham (Boca Raton: CRC Press), 397–442.
- Hagel, J. M., and Facchini, P. J. (2018). Expanding the roles for 2-oxoglutarate-dependent oxygenases in plant metabolism. *Nat. Prod. Rep.* 35, 721–734. doi: 10.1039/C7NP00060J
- Han, X. J., Wu, Y. F., Gao, S., Yu, H. N., Xu, R. X., Lou, H. X., et al. (2014). Functional characterization of a *plagiochasma appendiculatum* flavone synthase i showing flavanone 2-hydroxylase activity. *FEBS Lett.* 588, 2307–2314. doi: 10.1016/j.febslet.2014.05.023
- Harborne, J. B. (1988). *The Flavonoids. Advances in Research Since 1980*. Boston, MA: Springer US.
- Herr, C. Q., and Hausinger, R. P. (2018). Amazing diversity in biochemical roles of fe(ii)-2-oxoglutarate oxygenases. *Trends Biochem. Sci.* 43, 517–532. doi: 10.1016/j.tibs.2018.04.002
- Hieta, R., and Myllyharju, J. (2002). Cloning and characterization of a low molecular weight prolyl 4-hydroxylase from *Arabidopsis thaliana*. Effective hydroxylation of proline-rich, collagen-like, and hypoxia-inducible transcription factor alpha-like peptides. *J. Biol. Chem.* 277, 23965–23971. doi: 10.1074/jbc.M21865200
- Islam, M. S., Leissing, T. M., Chowdhury, R., Hopkinson, R. J., and Schofield, C. J. (2018). 2-oxoglutarate-dependent oxygenases. *Annu. Rev. Biochem.* 87, 585–620. doi: 10.1146/annurev-biochem-061516-044724
- Jez, J. M., Bowman, M. E., Dixon, R. A., and Noel, J. P. (2000). Structure and mechanism of the evolutionarily unique plant enzyme chalcone isomerase. *Nat. Struct. Biol.* 7, 786–791. doi: 10.1038/79025
- Jiang, C. G., Kim, S. Y., and Suh, D. Y. (2008). Divergent evolution of the thiolase superfamily and chalcone synthase family. *Mol. Phylog. Evol.* 49, 691–701. doi: 10.1016/j.ympev.2008.09.002
- Jiang, C. G., Schommer, C. K., Kim, S. Y., and Suh, D. Y. (2006). Cloning and characterization of chalcone synthase from the moss, *physcomitrella patens*. *Phytochemistry* 67, 2531–2540. doi: 10.1016/j.phytochem.2006.09.030
- Jiang, W. B., Yin, Q. G., Wu, R. R., Zheng, G. S., Liu, J. Y., Dixon, R. A., et al. (2015). Role of a chalcone isomerase-like protein in flavonoid biosynthesis in *Arabidopsis thaliana*. *J. Exp. Bot.* 66, 7165–7179. doi: 10.1093/jxb/erv413
- Jorgensen, K., Rasmussen, A. V., Morant, M., Nielsen, A. H., Bjarnholt, N., Zagrobelny, M., et al. (2005). Metabolite formation and metabolic channeling in the biosynthesis of plant natural products. *Curr. Opin. Plant Biol.* 8, 280–291. doi: 10.1016/j.pbi.2005.03.014
- Jornvall, H., Hoog, J. O., and Persson, B. (1999). Sdr and mdr: completed genome sequences show these protein families to be large, of old origin, and of complex nature. *FEBS Lett.* 445, 261–264. doi: 10.1016/S0014-5793(99)00130-1
- Kallberg, Y., Oppermann, U., and Persson, B. (2010). Classification of the short-chain dehydrogenase/reductase superfamily using hidden markov models. *FEBS J.* 277, 2375–2386. doi: 10.1111/j.1742-4658.2010.07656.x
- Kaltenbach, M., Burke, J. R., Dindo, M., Pabis, A., Munsberg, F. S., Rabin, A., et al. (2018). Evolution of chalcone isomerase from a noncatalytic ancestor. *Nat. Chem. Biol.* 14, 548–555. doi: 10.1038/s41589-018-0042-3
- Kavanagh, K. L., Jornvall, H., Persson, B., and Oppermann, U. (2008). Medium- and short-chain dehydrogenase/reductase gene and protein families: the sdr superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. *Cell. Mol. Life Sci.* 65, 3895–3906. doi: 10.1007/s00018-008-8588-y
- Kawai, Y., Ono, E., and Mizutani, M. (2014). Evolution and diversity of the 2-oxoglutarate-dependent dioxygenase superfamily in plants. *Plant J.* 78, 328–343. doi: 10.1111/tpj.12479
- Kim, S. S., Grienerberger, E., Lallemand, B., Colpitts, C. C., Kim, S. Y., Souza, C. D., et al. (2010). Lap6/polyketide synthase a and lap5/polyketide synthase b encode hydroxyalkyl alpha-pyrone synthases required for pollen development and sporopollenin biosynthesis in *Arabidopsis thaliana*. *Plant Cell* 22, 4045–4066. doi: 10.1105/tpc.110.080028
- Koeduka, T., Louie, G. V., Orlova, I., Kish, C. M., Ibdah, M., Wilkerson, C. G., et al. (2008). The multiple phenylpropene synthases in both *clarkia breweri* and *petunia hybrida* represent two distinct protein lineages. *Plant J.* 54, 362–374. doi: 10.1111/j.1365-313X.2008.03412.x
- Kubo, H., Nozawa, S., Hiwatashi, T., Kondou, Y., Nakabayashi, R., Mori, T., et al. (2018). Biosynthesis of riccionidins and marchantins is regulated by r2r3-myb transcription factors in *Marchantia polymorpha*. *J. Plant Res.* 131, 849–864. doi: 10.1007/s10265-018-1044-7
- Lam, P. Y., Liu, H., and Lo, C. (2015). Completion of tricetin biosynthesis pathway in rice: cytochrome p450 75b4 is a unique chrysoeriol 5'-hydroxylase. *Plant Physiol.* 168, 1527–1536. doi: 10.1104/pp.15.00566
- Lam, P. Y., Tobimatsu, Y., Takeda, Y., Suzuki, S., Yamamura, M., Umezawa, T., et al. (2017). Disrupting flavone synthase ii alters lignin and improves biomass digestibility. *Plant Physiol.* 174, 972–985. doi: 10.1104/pp.16.01973
- Lan, W., Lu, F. C., Regner, M., Zhu, Y. M., Rencoret, J., Ralph, S. A., et al. (2015). Tricin, a flavonoid monomer in monocot lignification. *Plant Physiol.* 167, 1284–1295. doi: 10.1104/pp.114.253757
- Lan, W., Morreel, K., Lu, F. C., Rencoret, J., del Rio, J. C., Voorend, W., et al. (2016a). Maize tricetin-oligolignol metabolites and their implications for monocot lignification. *Plant Physiol.* 171, 810–820. doi: 10.1104/pp.16.02012
- Lan, W., Rencoret, J., Lu, F. C., Karlen, S. D., Smith, B. G., Harris, P. J., et al. (2016b). Tricin-lignins: occurrence and quantification of tricetin in relation to phylogeny. *Plant J.* 88, 1046–1057. doi: 10.1111/tpj.13315
- Lapcik, O. (2007). Isoflavonoids in non-leguminous taxa: a rarity or a rule? *Phytochemistry* 68, 2909–2916. doi: 10.1016/j.phytochem.2007.08.006
- Li, L., Aslam, M., Rabbi, F., Vanderwel, M. C., Ashton, N. W., and Suh, D. Y. (2018). Ppors, an ancient type iii polyketide synthase, is required for integrity of leaf cuticle and resistance to dehydration in the moss, *physcomitrella patens*. *Planta* 247, 527–541. doi: 10.1007/s00425-017-2806-5
- Lindahl, T., Sedgwick, B., Sekiguchi, M., and Nakabeppu, Y. (1988). Regulation and expression of the adaptive response to alkylating agents. *Annu. Rev. Biochem.* 57, 133–157. doi: 10.1146/annurev.bi.57.070188.001025
- Liou, G., Chiang, Y. C., Wang, Y., and Weng, J. K. (2018). Mechanistic basis for the evolution of chalcone synthase catalytic cysteine reactivity in land plants. *J. Biol. Chem.* 293, 18601–18612. doi: 10.1074/jbc.RA118.005695
- Markham, K. R., Ryan, K. G., Bloor, S., and Mitchell, K. A. (1998). An increase in the luteolin: apigenin ratio in *Marchantia polymorpha* on UV-B enhancement. *Phytochemistry* 48, 791–794. doi: 10.1016/S0031-9422(97)00875-3
- Markolovic, S., Wilkins, S. E., and Schofield, C. J. (2015). Protein hydroxylation catalyzed by 2-oxoglutarate-dependent oxygenases. *J. Biol. Chem.* 290, 20712–20722. doi: 10.1074/jbc.R115.662627
- Martens, S., Forkmann, G., Britsch, L., Wellmann, F., Matern, U., and Lukacin, R. (2003). Divergent evolution of flavonoid 2-oxoglutarate-dependent dioxygenases in parsley. *FEBS Lett.* 544, 93–98. doi: 10.1016/S0014-5793(03)00479-4
- Martens, S., and Mithofer, A. (2005). Flavones and flavone synthases. *Phytochemistry* 66, 2399–2407. doi: 10.1016/j.phytochem.2005.07.013
- Martinez, S., and Hausinger, R. P. (2015). Catalytic mechanisms of fe(ii)- and 2-oxoglutarate-dependent oxygenases. *J. Biol. Chem.* 290, 20702–20711. doi: 10.1074/jbc.R115.648691
- Matsuzawa, M., Katsuyama, Y., Funo, N., and Horinouchi, S. (2010). Alkylresorcylic acid synthesis by type iii polyketide synthases from rice *Oryza sativa*. *Phytochemistry* 71, 1059–1067. doi: 10.1016/j.phytochem.2010.02.012
- Meza, T. J., Moen, M. N., Vagbo, C. B., Krokan, H. E., Klungland, A., Grini, P. E., et al. (2012). The DNA dioxygenase alkbh2 protects *Arabidopsis thaliana* against methylation damage. *Nucleic Acids Res.* 40, 6620–6631. doi: 10.1093/nar/gks327
- Mielecki, D., Zugaj, D. L., Muszewska, A., Piwowarski, J., Chojnacka, A., Mielecki, M., et al. (2012). Novel alkb dioxygenases—alternative models for *in silico* and *in vivo* studies. *PLoS ONE* 7:e30588. doi: 10.1371/journal.pone.0030588
- Milkowski, C., and Strack, D. (2004). Serine carboxypeptidase-like acyltransferases. *Phytochemistry* 65, 517–524. doi: 10.1016/j.phytochem.2003.12.018

- Min, T., Kasahara, H., Bedgar, D. L., Youn, B., Lawrence, P. K., Gang, D. R., et al. (2003). Crystal structures of pinoresinol-lariciresinol and phenylcoumaran benzylic ether reductases and their relationship to isoflavone reductases. *J. Biol. Chem.* 278, 50714–50723. doi: 10.1074/jbc.M308493200
- Mizutani, M., and Ohta, D. (2010). Diversification of p450 genes during land plant evolution. *Annu. Rev. Plant Biol.* 61, 291–315. doi: 10.1146/annurev-arplant-042809-112305
- Moghe, G. D., and Last, R. L. (2015). Something old, something new: conserved enzymes and the evolution of novelty in plant specialized metabolism. *Plant Physiol.* 169, 1512–1523. doi: 10.1104/pp.15.00994
- Morita, H., Yamashita, M., Shi, S. P., Wakimoto, T., Kondo, S., Kato, R., et al. (2011). Synthesis of unnatural alkaloid scaffolds by exploiting plant polyketide synthase. *Proc. Natl. Acad. Sci. U.S.A.* 108, 13504–13509. doi: 10.1073/pnas.1107782108
- Morita, Y., Takagi, K., Fukuchi-Mizutani, M., Ishiguro, K., Tanaka, Y., Nitasaka, E., et al. (2014). A chalcone isomerase-like protein enhances flavonoid production and flower pigmentation. *Plant J.* 78, 294–304. doi: 10.1111/tpj.12469
- Moummou, H., Kallberg, Y., Tonfack, L. B., Persson, B., and van der Rest, B. (2012). The plant short-chain dehydrogenase (sdr) superfamily: genome-wide inventory and diversification patterns. *BMC Plant Biol.* 12:219. doi: 10.1186/1471-2229-12-219
- Nelson, D., and Werck-Reichhart, D. (2011). A p450-centric view of plant evolution. *Plant J.* 66, 194–211. doi: 10.1111/j.1365-3113X.2011.04529.x
- Nelson, D. R. (2006). Plant cytochrome p450s from moss to poplar. *Phytochem. Rev.* 5, 193–204. doi: 10.1007/s11101-006-9015-3
- Ngaki, M. N., Louie, G. V., Philippe, R. N., Manning, G., Pojer, F., Bowman, M. E., et al. (2012). Evolution of the chalcone-isomerase fold from fatty-acid binding to stereospecific catalysis. *Nature* 485, 530–533. doi: 10.1038/nature11009
- Owens, D. K., Alerding, A. B., Crosby, K. C., Bandara, A. B., Westwood, J. H., and Winkel, B. S. (2008a). Functional analysis of a predicted flavonol synthase gene family in *Arabidopsis*. *Plant Physiol.* 147, 1046–1061. doi: 10.1104/pp.108.117457
- Owens, D. K., Crosby, K. C., Runac, J., Howard, B. A., and Winkel, B. S. (2008b). Biochemical and genetic characterization of *Arabidopsis* flavanone 3beta-hydroxylase. *Plant Physiol. Biochem.* 46, 833–843. doi: 10.1016/j.plaphy.2008.06.004
- Pichersky, E., and Gang, D. R. (2000). Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. *Trends Plant Sci.* 5, 439–445. doi: 10.1016/S1360-1385(00)01741-6
- Preuss, A., Stracke, R., Weisshaar, B., Hillebrecht, A., Matern, U., and Martens, S. (2009). *Arabidopsis thaliana* expresses a second functional flavonol synthase. *FEBS Lett.* 583, 1981–1986. doi: 10.1016/j.febslet.2009.05.006
- Rauscher, M. D. (2006). “The evolution of flavonoids and their genes,” in *The Science of Flavonoids*, ed E. Grotewold (New York, NY: Springer), 175–211.
- Rauscher, M. D., Miller, R. E., and Tiffin, P. (1999). Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. *Mol. Biol. Evol.* 16, 266–274. doi: 10.1093/oxfordjournals.molbev.a026108
- Rauter, A. P., Ennis, M., Hellwich, K. H., Herold, B. J., Horton, D., Moss, G. P., et al. (2018). Nomenclature of flavonoids (IUPAC recommendations 2017). *Pure Appl Chem.* 90, 1429–1486. doi: 10.1515/pac-2013-0919
- Saito, K., Yonekura-Sakakibara, K., Nakabayashi, R., Higashi, Y., Yamazaki, M., Tohge, T., et al. (2013). The flavonoid biosynthetic pathway in *Arabidopsis*: structural and genetic diversity. *Plant Physiol. Biochem.* 72, 21–34. doi: 10.1016/j.plaphy.2013.02.001
- Schuz, R., Heller, W., and Hahlbrock, K. (1983). Substrate specificity of chalcone synthase from petroselinum hortense. Formation of phloroglucinol derivatives from aliphatic substrates. *J. Biol. Chem.* 258, 6730–6734.
- Seitz, C., Eder, C., Deiml, B., Kellner, S., Martens, S., and Forkmann, G. (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 Mol. Biol.* 61, 365–381. doi: 10.1007/s11103-006-0012-0
- Shimada, N., Aoki, T., Sato, S., Nakamura, Y., Tabata, S., and Ayabe, S. (2003). A cluster of genes encodes the two types of chalcone isomerase involved in the biosynthesis of general flavonoids and legume-specific 5-deoxy(iso)flavonoids in *Lotus japonicus*. *Plant Physiol.* 131, 941–951. doi: 10.1104/pp.004820
- Shimizu, Y., Ogata, H., and Goto, S. (2017). Type iii polyketide synthases: functional classification and phylogenomics. *ChemBioChem.* 18, 50–65. doi: 10.1002/cbic.201600522
- Shirley, B. W. (1996). Flavonoid biosynthesis: 'new' functions for an 'old' pathway. *Trends Plant Sci.* 1, 377–382. doi: 10.1016/1360-1385(96)10040-6
- Stafford, H. A. (1991). Flavonoid evolution: an enzymic approach. *Plant Physiol.* 96, 680–685. doi: 10.1104/pp.96.3.680
- St-Pierre, B., and De Luca, V. (2000). “Evolution of acyltransferase genes: origin and diversification of the bahd superfamily of acyltransferases involved in secondary metabolism,” in *Recent Advances in Phytochemistry*, eds J. T. Romeo, R. Ibrahim, L. Varin, and V. De Luca (Amsterdam: Elsevier BV), 285–315.
- Subramanian, S., Stacey, G., and Yu, O. (2007). Distinct, crucial roles of flavonoids during legume nodulation. *Trends Plant Sci.* 12, 282–285. doi: 10.1016/j.tplants.2007.06.006
- Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., et al. (2007). Proposed minimum reporting standards for chemical analysis. *Metabolomics* 3, 211–221. doi: 10.1007/s11306-007-0082-2
- Tohge, T., Watanabe, M., Hoefgen, R., and Fernie, A. R. (2013). The evolution of phenylpropanoid metabolism in the green lineage. *Crit. Rev. Biochem. Mol. Biol.* 48, 123–152. doi: 10.3109/10409238.2012.758083
- Tonfack, L. B., Moummou, H., Latché A., Youmbi, E., Benichou, M., Pech, J.-C., et al. (2011). The plant sdr superfamily: involvement in primary and secondary metabolism. *Curr. Topics Plant Biol.* 12, 41–53. Available online at: <https://prodinra.inra.fr/record/438703>
- Tsugawa, H., Kind, T., Nakabayashi, R., Yukihira, D., Tanaka, W., Cajka, T., et al. (2016). Hydrogen rearrangement rules: computational ms/ms fragmentation and structure elucidation using ms-finder software. *Anal. Chem.* 88, 7946–7958. doi: 10.1021/acs.analchem.6b00770
- Tsugawa, H., Nakabayashi, R., Mori, T., Yamada, Y., Takahashi, M., Rai, A., et al. (2019). A cheminformatics approach to characterize metabolomes in stable-isotope-labeled organisms. *Nat. Methods.* 16, 295–298. doi: 10.1038/s41592-019-0358-2
- Tuominen, L. K., Johnson, V. E., and Tsai, C. J. (2011). Differential phylogenetic expansions in bahd acyltransferases across five angiosperm taxa and evidence of divergent expression among populus paralogues. *BMC Genomics* 12:236. doi: 10.1186/1471-2164-12-236
- van den Born, E., Omelchenko, M. V., Bekkelund, A., Leihne, V., Koonin, E. V., Dolja, V. V., et al. (2008). Viral alkB proteins repair rna damage by oxidative demethylation. *Nucleic Acids Res.* 36, 5451–5461. doi: 10.1093/nar/gkn519
- Wang, P., Zhang, L., Jiang, X., Dai, X., Xu, L., Li, T., et al. (2018). Evolutionary and functional characterization of leucoanthocyanidin reductases from *Camellia sinensis*. *Planta* 247, 139–154. doi: 10.1007/s00425-017-2771-z
- Wang, X., He, X., Lin, J., Shao, H., Chang, Z., and Dixon, R. A. (2006). Crystal structure of isoflavone reductase from alfalfa (*Medicago sativa* L.). *J. Mol. Biol.* 358, 1341–1352. doi: 10.1016/j.jmb.2006.03.022
- Wang, Y., Tan, X., and Paterson, A. H. (2013). Different patterns of gene structure divergence following gene duplication in *Arabidopsis*. *BMC Genomics* 14:652. doi: 10.1186/1471-2164-14-652
- Wanibuchi, K., Zhang, P., Abe, T., Morita, H., Kohno, T., Chen, G. S., et al. (2007). An acridone-producing novel multifunctional type iii polyketide synthase from *Hypericum serrata*. *FEBS J.* 274, 1073–1082. doi: 10.1111/j.1742-4658.2007.05656.x
- Weng, J. K., Li, X., Stout, J., and Chapple, C. (2008). Independent origins of syringyl lignin in vascular plants. *Proc. Natl. Acad. Sci. U.S.A.* 105, 7887–7892. doi: 10.1073/pnas.0801696105
- Weng, J. K., and Noel, J. P. (2012a). The remarkable pliability and promiscuity of specialized metabolism. *Cold Spring Harb. Symp. Quant. Biol.* 77, 309–320. doi: 10.1101/sqb.2012.77.014787
- Weng, J. K., and Noel, J. P. (2012b). Structure-function analyses of plant type iii polyketide synthases. *Methods Enzymol.* 515, 317–335. doi: 10.1016/B978-0-12-394290-6.00014-8
- Weng, J. K., and Noel, J. P. (2013). Chemodiversity in *Selaginella*: a reference system for parallel and convergent metabolic evolution in terrestrial plants. *Front Plant Sci.* 4:119. doi: 10.3389/fpls.2013.00119
- Weng, J. K., Ye, M., Li, B., and Noel, J. P. (2016). Co-evolution of hormone metabolism and signaling networks expands plant adaptive plasticity. *Cell* 166, 881–893. doi: 10.1016/j.cell.2016.06.027

- Williams, C. A., and Grayer, R. J. (2004). Anthocyanins and other flavonoids. *Nat. Product Rep.* 21, 539–573. doi: 10.1039/b311404j
- Winkel-Shirley, B. (2001). Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol.* 126, 485–493. doi: 10.1104/pp.126.2.485
- Wolf, L., Rizzini, L., Stracke, R., Ulm, R., and Rensing, S. A. (2010). The molecular and physiological responses of *Physcomitrella patens* to ultraviolet-b radiation. *Plant Physiol.* 153, 1123–1134. doi: 10.1104/pp.110.154658
- Wu, L. F., Meng, S., and Tang, G. L. (2016). Ferrous iron and alpha-ketoglutarate-dependent dioxygenases in the biosynthesis of microbial natural products. *Biochim. Biophys. Acta.* 1864, 453–470. doi: 10.1016/j.bbapap.2016.01.012
- Yazaki, K., Sasaki, K., and Tsurumaru, Y. (2009). Prenylation of aromatic compounds, a key diversification of plant secondary metabolites. *Phytochemistry* 70, 1739–1745. doi: 10.1016/j.phytochem.2009.08.023
- Yonekura-Sakakibara, K., and Hanada, K. (2011). An evolutionary view of functional diversity in family 1 glycosyltransferases. *Plant J.* 66, 182–193. doi: 10.1111/j.1365-313X.2011.04493.x
- Yu, C. K. Y., Springob, K., Schmidt, J. R., Nicholson, R. L., Chu, I. K., Yip, W. K., et al. (2005). A stilbene synthase gene (*sbsts1*) is involved in host and nonhost defense responses in sorghum. *Plant Physiol.* 138, 393–401. doi: 10.1104/pp.105.059337
- Yu, H. N., Liu, X. Y., Gao, S., Sun, B., Zheng, H. B., Ji, M., et al. (2018). Structural and biochemical characterization of the plant type iii polyketide synthases of the liverwort *Marchantia paleacea*. *Plant Physiol. Biochem.* 125, 95–105. doi: 10.1016/j.plaphy.2018.01.030
- Yu, X. H., Gou, J. Y., and Liu, C. J. (2009). Bahd superfamily of acyl-coa dependent acyltransferases in populus and arabidopsis: bioinformatics and gene expression. *Plant Mol. Biol.* 70, 421–442. doi: 10.1007/s11103-009-9482-1
- Yuan, T., Fujioka, S., Takatsuto, S., Matsumoto, S., Gou, X., He, K., et al. (2007). Ben1, a gene encoding a dihydroflavonol 4-reductase (*dfr*)-like protein, regulates the levels of brassinosteroids in *Arabidopsis thaliana*. *Plant J.* 51, 220–233. doi: 10.1111/j.1365-313X.2007.03129.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Yonekura-Sakakibara, Higashi and Nakabayashi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.