Characterization of the *PRODUCTION of ANTHOCYANIN PIGMENT 1 Arabidopsis* dominant mutant using DLEMMA dual isotope labeling approach

Yonghui Dong [a,1], Liron Feldberg [b,1], Ilana Rogachev [a], Asaph Aharoni [a,*]

[a] *Department of Plant Sciences, Weizmann Institute of Science, Rehovot, 7610001, Israel*
[b] *Department of Analytical Chemistry, Israel Institute for Biological Research, Ness Ziona, 7410001, Israel*

ARTICLE INFO

ABSTRACT

Stable isotope labeling has emerged as a valuable tool for metabolite identification and quantification. In this study, we employed DLEMMA, a dual stable isotope labeling approach to identify and track phenylpropanoid pathway in *Arabidopsis thaliana*. Three forms of phenylalanine (Phe), including unlabeled, Phe$^{13}C_6$ and Phe$^{13}C_6^2H_5$, were used as feeding precursors. The unique isotopic pattern obtained from MS spectra significantly simplified data processing and facilitated global mining of Phe-derived metabolites. Following this approach, we have identified 35 phenylalanine-derived metabolites with high confidence. We next compared phenyl-propanoids contents between leaves of wild type (WT) and the dominant *PRODUCTION OF ANTHOCYANIN PIGMENT 1 (pap1-D) Arabidopsis thaliana* mutant using a combined sample matrices and label-swap approach. This approach was designed to correct any unequal matrix effects between the two divergent samples, and any possible uneven label incorporation efficiency between the two differently labeled Phe precursors. Thirty of the 35 identified metabolites were found differential between WT and *pap1-D* leaves. Our results shown that the ectopic *PAP1* expression led to significant accumulation of cyanidin-type anthocyanins, quercetin-type flavonols and hydroxycinnamic acids and their glycosylated derivatives. While levels of kaempferol glycosides and a hydroxycinnamic acid amide were reduced in the *pap1-D* leaves.

1. Introduction

The phenylpropanoid pathway (PP) represents a rich source of metabolites critically important for plant growth, biotic and abiotic defense responses (Kurepa et al., 2018; Rasmussen and Dixon, 1999). Phenylalanine ammonia-lyase is the first committed enzyme in the pathway, which converts L-phenylalanine (Phe) from the primary metabolic pool to a myriad phenylpropanoids (Zhang and Liu, 2015). Phenylpropanoids can be classified into 5 structural groups, including flavonoids, monolignols, phenolic acids, stilbenes and coumarins (Noel et al., 2005). Although numerous studies have characterize PP genes and enzymes (Fraser and Chapple, 2011), a large portion of species- or lineage-specific phenylpropanoids still remains unidentified (Deng and Lu, 2017).

Feeding experiments using isotopically labeled precursors is commonly used to detect pathway-specific metabolites in plants and other organisms (Mahmud, 2007). In general, a stable isotope labeled precursor (tracer) is introduced into a biological system, and both labeled and unlabeled precursors are metabolized. As a consequence, downstream metabolites incorporate complete or part of the labeled elements from the precursor (Doppler et al., 2019). When analyzed by mass spectrometry (MS), characteristic and artificial isotope patterns are typically observed for the labeled metabolites. Consequently, the unique isotope patterns can be used to detect precursor-derived metabolites (Kluger et al., 2014). Apart from detecting pathway metabolites, this approach offers several additional advantages. First, it allows detection of trace level intermediate pathway metabolites because of the precursor-enrichment and second, it increases confidence in metabolite identification by efficiently eliminating the number of possible molecular formulas and chemical structures according to the labeling patterns of detected metabolites. Furthermore, it enables novel metabolic pathway discovery without prior knowledge regarding the fate of

---

precursor metabolite (Feldberg et al., 2018).

Our lab has previously extended the stable isotope tracer-based approach by simultaneously feeding whole plants or organs with the same precursor metabolite possessing two different labeling schemes (Feldberg et al., 2009). Compared to the conventional single stable isotope tracing, this approach, termed DLEMMA (Dual Labeling of Metabolites for Metabolome Analysis), largely reduces the number of plausible molecular formulas and chemical structures of the detected metabolites, thus facilitating metabolite identification. In addition, DLEMMA is also suitable for semi-quantitative comparison of different samples, as it is able to correct for possible unequal matrix effects from divergent samples and uneven label incorporation efficiency from
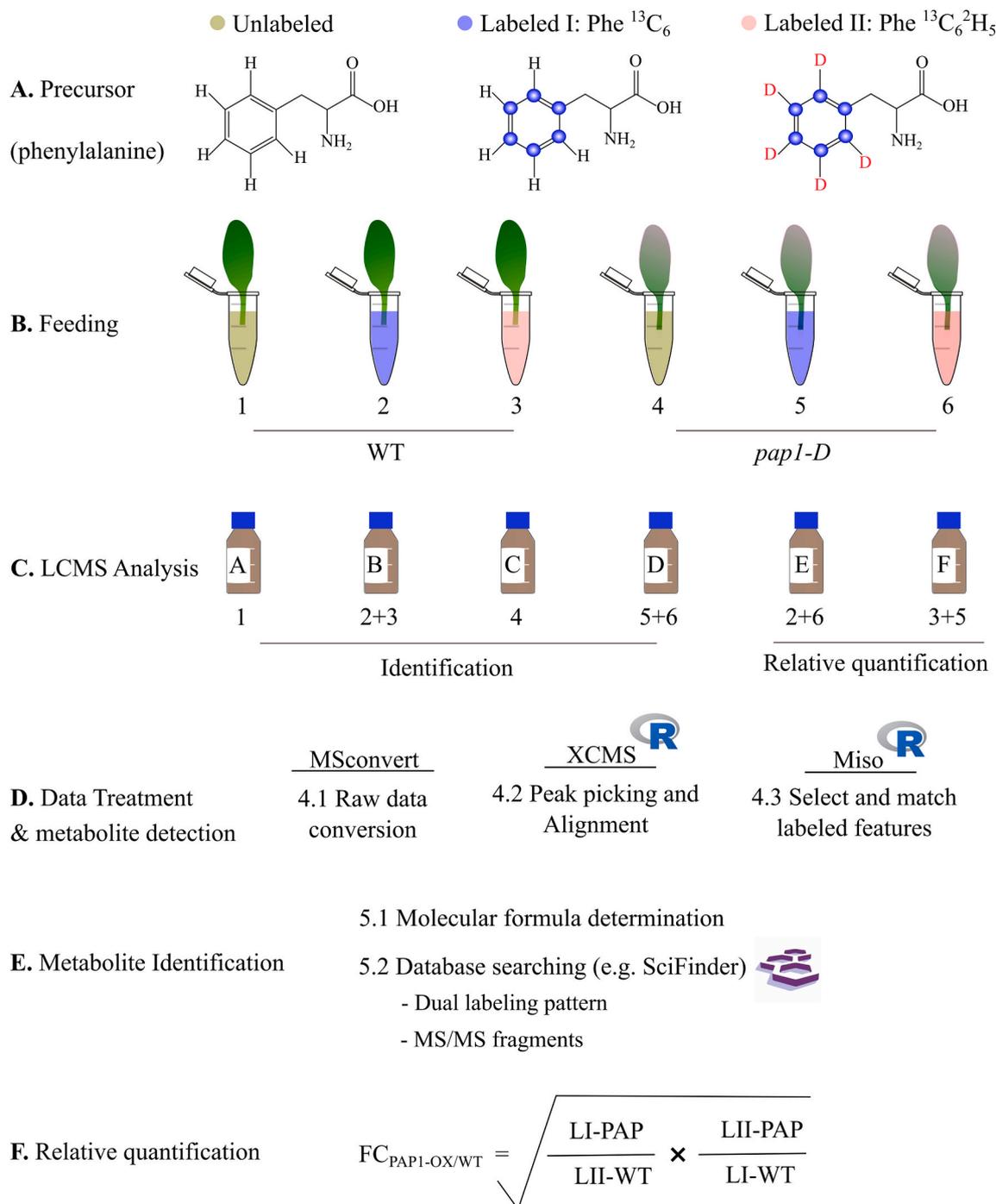
Fig. 1. Overview of the DLEMMA workflow employed in this study. (A) Three differently labeled forms of phenylalanine (Phe) were used as feeding precursors, i.e. unlabeled, Label I (Phe-$^{13}C_6$), and Label II (Phe-$^{13}C_6{}^2H_5$) Phe. (B) The two *Arabidopsis* genotype groups, WT and *pap1-D*, were fed with three forms of labeling precursors for 24h. (C) Following feeding, the six treatments were combined (i.e. combinations A to F), in which four combinations were designed for metabolite identification and another two for semi-quantitative differential metabolite analysis. The extracts were analyzed by high resolution LC-MS in both positive and negative ion modes. (D) The LC-MS raw data was preprocessed with the R package XCMS, and next the R package Miso was used to detect and extract all Phe-derived mass features. (E) Phe-derived metabolites were identified based on retention time, m/z, MS/MS spectra, and dual-labeling patterns obtained from LC-MS analysis. (F) A combined sample matrices and label-swap approach was used to semi-quantitatively compare phenylpropanoids content between the WT and *pap1-D* genotypes.

different labeled precursors. In addition to combining DLEMMA with high-resolution liquid chromatograph-mass spectrometry (HR LC-MS) for metabolite identification, semi-quantitative comparison and metabolic network construction, DLEMMA has recently been coupled with mass spectrometry imaging (DLEMMA-MS-Imaging) for accurate spatial localization analysis (Feldberg et al., 2018).

The PRODUCTION OF ANTHOCYANIN PIGMENT 1 (PAP1) is a R2-R3 MYB-type transcription factor (MYB75) regulating the biosynthesis of phenolic acids and flavonoids in *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae) (Borevitz et al., 2000). The corresponding bright-purple mutant termed *pap1-D* was originally identified through screening of an activation tagging population. Because *pap1-D Arabidopsis* plants constitutively produce a large amount of anthocyanins (Lee et al., 2016; Tohge et al., 2005), we questioned which anthocyanin derivatives are produced, and what other PP metabolites are altered due to the activation and ectopic expression of the *PAP1* transcriptional regulator in *pap1-D* mutant. To this end, we carried a detailed, high-confidence metabolites identification and semi-quantitative comparison between WT and *pap1-D Arabidopsis* plants using DLEMMA approach.

2. Results and discussion

2.1. Applying DLEMMA for studying the phenylpropanoids pathway

L-Phenylalanine (Phe), the core precursor of the phenylpropanoid pathway, and its two differently labeled forms Phe-$^{13}C_6$ (Label I) and Phe-$^{13}C_6^2H_5$ (Label II) were used for feeding experiments (Figs. 1A and 6B). Since labeled metabolites possess nearly identical physicochemical properties as their native non-labeled analogues, the Phe-derived isotopologues were detected as three co-eluting chromatographic peaks (unlabeled and two differently labeled forms), except that their *m/z* values were different. The LC-MS data were first pre-processed with XCMS for peak picking and alignment. A total of 7450 and 4822 ion peaks (retention time × *m/z*) were detected in positive and negative ion modes, respectively. The R package Miso was then used to automatically detect isotopologues (one unlabeled and two different labeled forms) and cluster them in a peak table containing retention time (RT), *m/z* values, number of labeled atoms and mean ion intensities (Fig. 1D). In total, 169 (positive ion mode) and 129 (negative mode) isotopologues were detected for wild type (WT), and167 (positive mode) and 150 (negative mode) isotopologues were detected in *pap-1D* leaves.

2.2. Improved metabolite detection and annotation confidence through feeding and dual labeling of the phenylalanine precursor

In a typical LC-MS based metabolomics study, metabolite identification is achieved through mass-based database search followed by tandem MS (MS/MS) and/or manual verification (Shahaf et al., 2016; Xiao et al., 2012). The results with Phe feeding demonstrate that labeling patterns, derived from *in vivo* dual labeling could significantly increase confidence in metabolite identification. In the first example, a Phe-derived isotopic cluster was identified in the negative ion mode with unlabeled *m/z* value being 339.0714 (Fig. 2A). The labeling patterns of this isotopic cluster were *m/z*+6 (Label I) and *m/z*+8 (Label II). Manual inspection of the LC-MS raw spectra confirmed that the isotopologues had similar RT but different *m/z* values (Fig. 2B). The

A

Unlabeled		Label I			Label II		
m/z	RT	*m/z*	RT	Label	*m/z*	RT	Label
339.0714	10.44	345.0925	10.46	6	347.1045	10.41	8
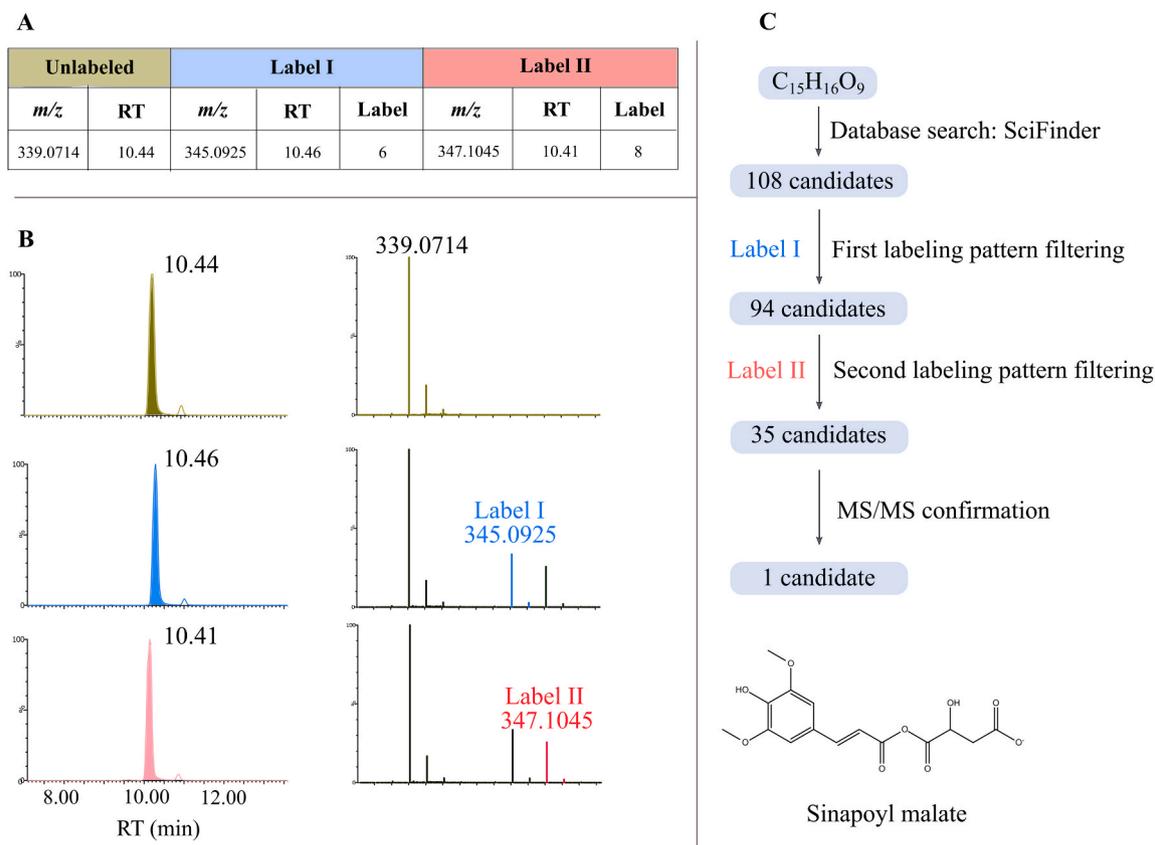
Fig. 2. A representative example of metabolite identification using DLEMMA. (A) An isotopologue cluster was identified. All information necessary for metabolite identification such as retention time (RT), *m/z* values and number of labeled atoms, was depicted in a table (B) The raw MS data was first inspected (i.e. RT and *m/z* values) to validate the results obtained by the Miso software. (C) Elemental composition was then determined using the Masslynx software based on accurate unlabeled *m/z* and natural isotope patterns. Next, elemental compositions $C_{15}H_{16}O_9$ was searched in SciFinder, and 108 possible candidates were found. The structural information obtained from the labeling patterns was then used to narrow down the number of candidates. Combined with MS/MS information, this metabolite was identified as sinapoyl malate.

elemental composition of this isotopologue cluster was determined by searching the unlabeled *m/z* value 339.0714 using the i-FIT™ algorithm (Waters Masslynx software) within a 5 ppm window. In this example, a single matched elemental composition was found, and its neutral molecular formula $C_{15}H_{16}O_9$ was searched in the SciFinder database (https://scifinder.cas.org/). In total 108 compounds were found to have this molecular formula. As this metabolite was labeled with six carbon atoms, it must contain a phenyl ring as its precursor. This sub-structure based search reduced the number from 108 to 94 candidates. Similarly, the labeling pattern *m/z*+8 indicated that this metabolite contained a phenyl ring with only 2 hydrogen atoms attached. This additional structural limitation further reduced the number of candidates to 35

compounds. Finally, based on the MS/MS information this metabolite was identified as sinapoyl malate (Fig. 2C).

In the example provided above, only dual labeling patterns were detected for the isotopic cluster. Yet, in some scenarios, we obtained complex labeling patterns that provided us with a much superior capacity and confidence to identify the metabolite structure. An isotopic cluster with an unlabeled *m/z* value being 975.2466 (positive ion mode) is one such example. Instead of dual labeling patterns being typically observed for most Phe-derived metabolites, in this case, seven labeling patterns were detected for the isotopic cluster, i.e. *m/z*+6, *m/z*+9, *m/z*+10, *m/z*+12, *m/z*+15, *m/z*+16 and *m/z*+19 (Fig. 3). Eight possible elemental compositions were obtained using the same method described
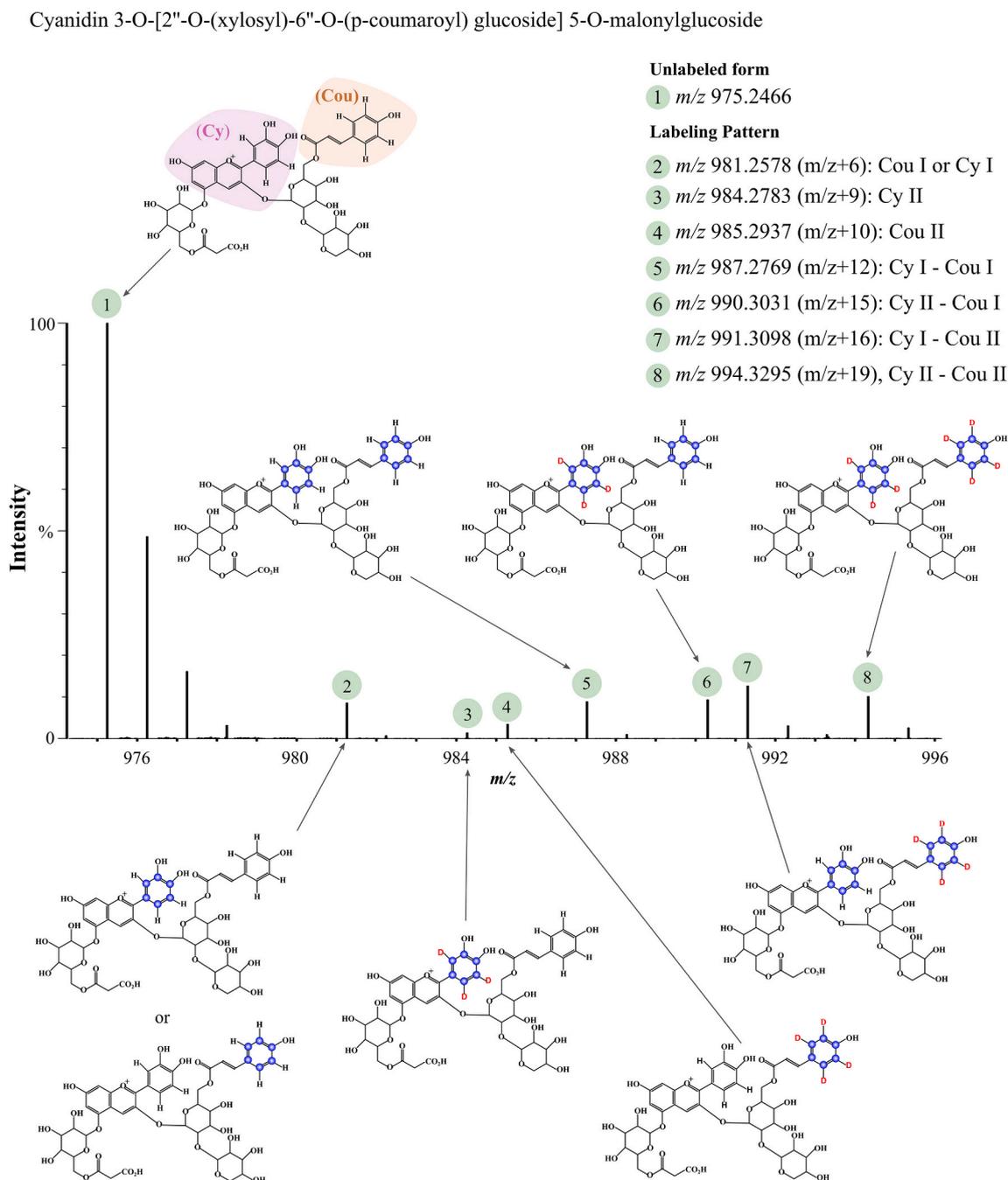
Fig. 3. A representative example demonstrating metabolite structure elucidation by DLEMMA in cases where multiple different labeling patterns are detected. An isotopologue cluster was detected with multiple different labeling patterns being observed, i.e. *m/z*+6, *m/z*+9, *m/z*+10, *m/z*+12, *m/z*+15, *m/z*+16 and *m/z*+19. Two possible elemental compositions were found for this metabolite. SciFinder database search revealed that only one structure matched all the 7 observed labeling patterns and this metabolite identified as Cyanidin 3-O-[2-O-(xylosyl)-6-O-(p-coumaroyl) glucoside] 5-O-malonylglucoside.

above, while only a single formula, $C_{44}H_{47}O_{25}$, was ranked with low i-FIT value (lower i-FIT value indicates a better match). The metabolite neutral molecular formula could be either $C_{44}H_{46}O_{25}$ (assuming it was a protonated ion $[M+H]^+$) or $C_{44}H_{47}O_{25}$ (assuming it was a radical ion $M^{+\cdot}$). When $C_{44}H_{46}O_{25}$ was searched against the SciFinder database, one possible Phe derivative was obtained; however, its theoretical labeling patterns matched only two of the observed labeling patterns. By contrast, when $C_{44}H_{47}O_{25}$ was searched in SciFinder, one possible Phe derivative was found and the theoretical labeling pattern of this compound matched the seven observed labeling patterns (Fig. 3). The MS/MS spectra confirmed its structure as a cyanidin derivative linked to a coumaric acid (Table S1). This isotopic cluster was thus putatively identified as Cyanidin 3-O-[2-O-(xylosyl)-6-O-(p-coumaroyl) glucoside] 5-O-malonylglucoside. Both coumaric acid and cyanidin are Phe-derived metabolites and there are seven possible combinations of differently labeled cyanidin and coumaric acid to form this isotopic cluster in a dual labeling experiment; all of them were detected in our study (Fig. 3). Using DLEMMA, we identified 35 Phe-derived metabolites. Table S1 provides metabolite name, accurate m/z, retention time, mass formula, and annotated fragments of these metabolites.

2.3. Phenylalanine precursor feeding enhances metabolite detection

Despite the use of highly sensitive LC-MS instrumentation, many trace level intermediate metabolites remain undetectable as they are below the instrument detection limit. Precursor feeding could promote accumulation and possibly biosynthesis of intermediate metabolites (Ouyang et al., 2005; Raghavendra et al., 2011), therefore improve their detection by LC-MS analysis. To assess the capacity of enhancing metabolite detection using precursor feeding, we compared ion intensities of the same metabolite between non-fed and unlabeled-Phe fed *Arabidopsis* leaves. In one example, two isomers of ferulic acid-hexose, eluting at 6.10 and 6.65 min, were detected in the unlabeled-Phe fed *Arabidopsis*. The detection and identification of these isomers were impossible without Phe feeding (Fig. 4A). In fact, while only trace amounts of the two isomers were found in the non-fed group, the peak areas of ferulic acid-hexose isomer I (RT = 6.10) and isomer II (RT = 6.66) was increased about 39 and 6.8 times following Phe feeding (Fig. 4).

We noticed that the degree of label incorporation was not identical among different metabolites along the metabolic pathway. Instead, it
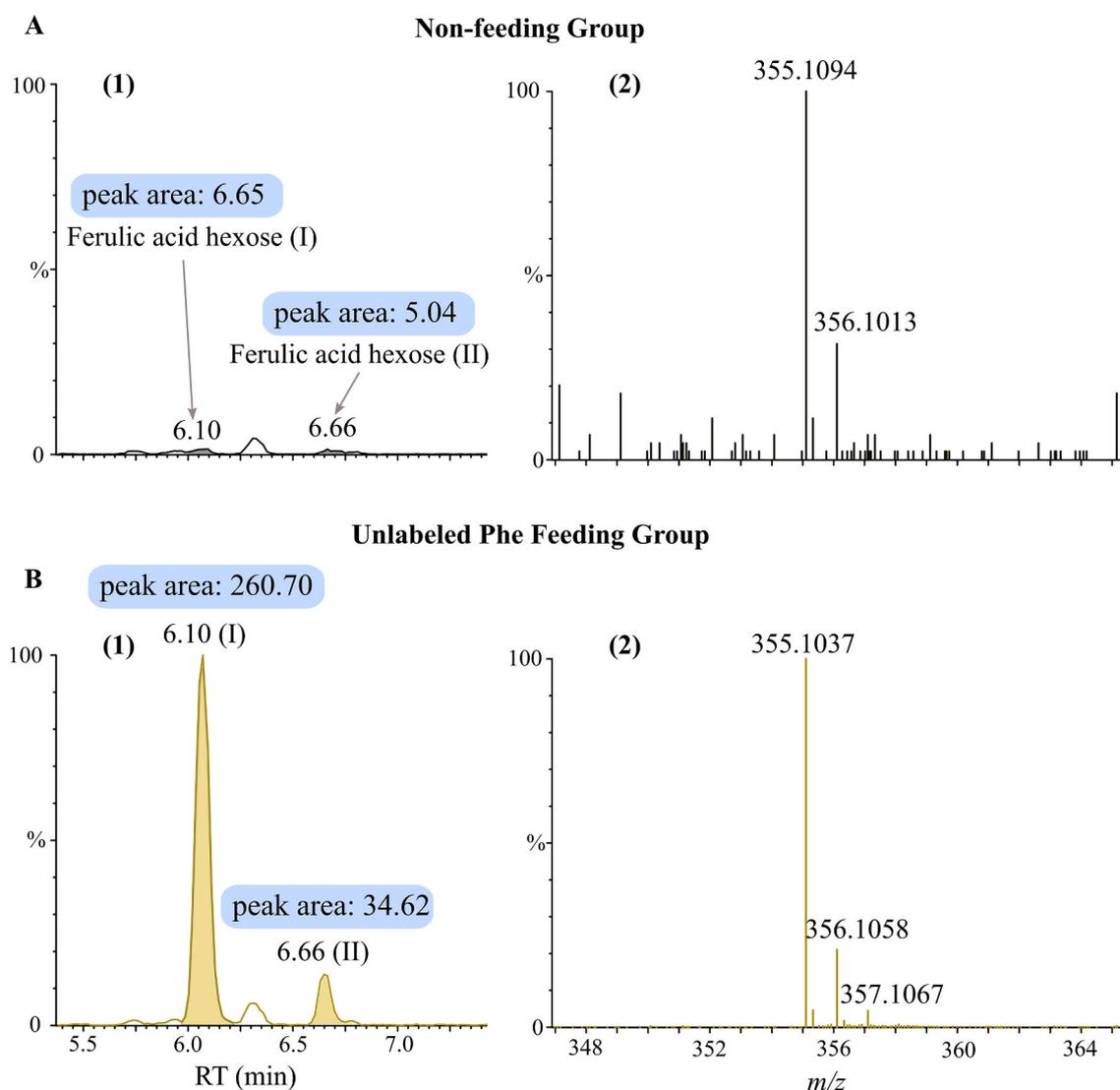
Fig. 4. Comparison of ion intensities of ferulic acid-hexose detected from phenylalanine fed and non-fed *Arabidopsis* leaves. (A) Extracted ion chromatogram (1) and mass spectrum (2) of ferulic acid-hexose from non-fed WT *Arabidopsis*. (B) Extracted ion chromatogram (1) and mass spectrum (2) of ferulic acid-hexose from unlabeled Phe fed WT *Arabidopsis*. Two isomers of ferulic acid-hexose, eluting at 6.10 and 6.66 min, were detected. The representative mass spectra are from ferulic acid-hexose isomer I (RT = 6.10 min). Peak areas of two ferulic acid-hexose isomers were used to compare their contents between non-fed and un-labeled Phe fed leaves.

was closely associated with the degree of label incorporation, defined here as labeling efficiency (LE). LE was calculated with the following equation, in which unlabeled, Label I and label II represent the ion intensities of the different forms of the metabolite from experimental group B (i.e. WT fed with Label I and Label II Phe).

$$LE = \frac{\text{Labeled I} + \text{Labeled II}}{\text{Unlabeled} + \text{Labeled I} + \text{Labeled II}} \times 100\%$$

The LE for all detected Phe-derived metabolites was calculated and ranged from 0.6% to 96.3% with an average of 33.6% and standard deviation of 34.5% (Table 1). To compare the LE among different Phe-derived metabolites, we plotted the LE as a function of metabolite location in the phenylpropanoid pathway. Metabolite location was defined by the number of enzymatic reactions (up to the metabolite intermediate) starting from Phe in the phenylpropanoid pathway. Phe-derived metabolites up to 16 enzymatic reaction steps from the Phe precursor were detected in this study (Table 1). Fig. 5 points to a strong correlation between LE and metabolite location in the metabolic pathway (R = 0.89 and p-value = 3.2×10^{-12}). In general, the lower the LE, the further the location of the metabolite in the metabolic pathway. Hence, the LE information can be used for metabolic pathway elucidation, e.g. the position of a newly identified metabolite in the pathway could be estimated according to its LE value. It is important to note that phenylalanine is the precursor of a plethora of secondary metabolites in plants. Therefore, the relationship between labeling efficiency of a metabolite and the position of its biosynthetic step in the phenylpropanoid pathway is not always linear. Nevertheless, it is arguably true that, in tracer experiments, the more biosynthetic steps are involved the lower the LE value.

2.4. Semi-quantitative differential metabolite analysis with DLEMMA

A major limitation in quantitative measurement by LC-MS is that it suffers from matrix effects, especially when using electrospray ionization (ESI) (Matuszewski et al., 2003; Zhou et al., 2017). Matrix effects refer to the alteration of ionization efficiency of a target analyte due to the presence of co-eluting compounds in the sample matrix. They often depend on the nature of samples and therefore differ from sample to sample (Silvestro et al., 2013). As a consequence, the accuracy of semi-quantitative comparison among different samples will be affected. To account for any unequal matrix effects between WT and *pap1-D*, the two samples with different labeling schemes were mixed before sample extraction and subsequent LC-MS analysis, e.g. Label I WT + Label II *pap1-D* (Group E in Fig. 1C). This approach is particularly advantageous as mixing the two different samples eliminates matrix effects variations, while the different labeling schemes allow distinguishing the same metabolites originating from the two different samples. In addition, in order to account for possible different label incorporation efficiencies between the two differently labeled-Phe precursors, we have included another 'swapped labeling' group, i.e. Label II WT + Label I *pap1-D* (Group F in Fig. 1C), for semi-quantitative comparison. Similar approach has been used to correct such experimental errors in stable isotope labeling by amino acids in cell culture (SILAC) based proteomics studies (Park et al., 2012; Chen et al., 2016). The fold change (FC) of the same metabolite from the two groups was calculated using either of the following equations. The equations are geometric mean of FC from 'swapped labeling' group:

$$FC_{PAP1}/WT = \sqrt[2]{\frac{LI - PAP1}{LII - WT} \times \frac{LII - PAP1}{LI - WT}}$$

$$FC_{WT}/PAP1 = \sqrt[2]{\frac{LI - WT}{LII - PAP1} \times \frac{LII - WT}{LI - PAP1}}$$

The regulatory gene *PAP1* has been reported to increase anthocyanin content when overexpressed in *Arabidopsis* (Borevitz et al., 2000;

Table 1

Labeling efficiency and location in the phenylpropanoid pathway of metabolites detected and identified in high confidence.

#	Pathway Location	Metabolite	Labeling Efficiency (%)	FC$_{PAP/WT}$	Class
1	L1	Phenylalanine	96.3	precursor	P1
2	L2 (1)	γ-glutamylphenylalanine	88.8	1	P2
3	L2 (2)	p-Coumaric acid	93.9	7.3	H3
4	L3 (1)	Coumaric acid hexose (I)	92.0	9.0	H2
5	L3 (2)	Coumaric acid hexose (II)	82.2	5.5	H5
6	L3 (3)	p-Coumaroylagmatine	75.0	0.025	HCAA1
7	L4 (1)	Caffeic acid hexose	70.9	3.4	H8
8	L4 (2)	Ferulic acid	73.7	3.9	H6
9	L5 (1)	Ferulic acid hexose (I)	84.6	11.5	H1
10	L5 (2)	Ferulic acid hexose (II)	47.3	6.3	H4
11	L6	Sinapinic acid	51.3	2	H
12	L7 (1)	Sinapoyl glucose (I)	47.9	3.9	H7
13	L7 (2)	Sinapoyl glucose (II)	39.1	2.1	H9
14	L8 (1)	Sinapoyl malate(I)	17.7	1.2	H
15	L8 (2)	Sinapoyl malate (II)	16.6	1.1	H
16	L9 (1)	Kaempferol 3-O-glucoside 7-O-rhamnoside	1.8	0.26	K2
17	L9 (2)	Kaempferol 3-O-rhamnoside 7-O-rhamnoside	1.4	0.036	K3
18	L10	Quercetin 3-O-rhamnoside 7-O-rhamnoside	2.2	3.7	Q3
19	L11 (1)	Quercetin 3-O-[6-O (rhamnosyl)-glucoside] 7-O-rhamnoside	2.7	7.2	Q2
20	L11 (2)	Kaempferol 3-O-[6-O-(rhamnosyl)-glucoside] 7-O-rhamnoside	2.3	0.28	K1
21	L11(3)	Quercetin 3-O-glucoside 7-O-rhamnoside	2.7	9.11	Q1
22	L12	Cyanidin 3-O-[2″-O-(xylosyl) glucoside] 5-O-glucoside	8.5	detected only in *pap1-D*	C
23	L13 (1)	Cyanidin 3-O-[2″-O-(2‴-O-(sinapoyl) xylosyl) glucoside] 5-O-glucoside	11.2	detected only in *pap1-D*	C
24	L13(2)	Cyanidin 3-O-[2″-O-(xylosyl) glucoside] 5-O-(6‴-O-malonyl) glucoside	minor amount	detected only in *pap1-D*	C
25	L13(3)	Cyanidin 3-O-[2″-O-(xylosyl) 6″-O-(p-coumaroyl) glucoside] 5-O-glucoside	10	14.1	C2
26	L14 (1)	Cyanidin 3-O-[2″-O-(xylosyl)-6″-O-(p-O-(glucosyl)-p-coumaroyl) glucoside] 5-O-glucoside	minor amount	detected only in *pap1-D*	C
27	L14 (2)	Cyanidin 3-O-[2″-O-(xylosyl)-6″-O-(p-coumaroyl) glucoside] 5-O-malonylglucoside (I)	7.4	7.7	C6
28	L14 (3)	Cyanidin 3-O-[2″-O-(xylosyl)-6″-O-(p-coumaroyl) glucoside] 5-O-malonylglucoside (II)	15.5	39.0	C1
29	L14 (4)	Cyanidin 3-O-[2″-O-(2‴-O-(sinapoyl) xylosyl) 6″-O-(p-coumaroyl) glucoside] 5-O-glucoside	12.5	6.7	C9
30	L15 (1)	Cyanidin 3-O-[2″-O-(xylosyl) 6″-O-(p-O-(glucosyl) p-coumaroyl) glucoside] 5-O-[6‴-O-(malonyl) glucoside]	5.7	7.8	C5
31	L15 (2)	Cyanidin 3-O-[2″-O-(2‴-O-(sinapoyl) xylosyl) 6″-O-(p-O-(glucosyl) p-coumaroyl) glucoside] 5-O-glucoside	3.6	6.2	C10
32	L15 (3)		9.5	8.1	C4

(continued on next page)

Table 1 (*continued*)

#	Pathway Location	Metabolite	Labeling Efficiency (%)	FC$_{PAP/WT}$	Class
33	L15 (4)	Cyanidin 3-O-[2″-O-(2‴-O-(sinapoyl) xylosyl) 6″-O-(p-O-coumaroyl) glucoside] 5-O-[6‴′-O-(malonyl) glucoside] (II) Cyanidin 3-O-[2″-O-(2‴-O-(sinapoyl) xylosyl) 6″-O-(p-O-coumaroyl) glucoside] 5-O-[6‴′-O-(malonyl) glucoside] (I)	11.6	8.6	C3
34	L16 (1)	Cyanidin 3-O-[2″-O-(6‴-O-(sinapoyl) xylosyl) 6″-O-(p-O-(glucosyl)-p-coumaroyl) glucoside] 5-O-(6‴′-O-malonyl) glucoside (I)	0.6	7.1	C8
35	L16 (2)	Cyanidin 3-O-[2″-O-(6‴-O-(sinapoyl) xylosyl) 6″-O-(p-O-(glucosyl)-p-coumaroyl) glucoside] 5-O-(6‴′-O-malonyl) glucoside (II)	3	7.6	C7

The number in the 'Pathways Location' column indicates the corresponding metabolites described in Fig. 5. 'L' is location, the number after 'L' indicates its location in the phenylpropanoid pathway. The number in the bracket is used to distinguish metabolites that are at the same pathway location; the number in the 'Class' column indicates the corresponding metabolite described in Fig. 6A. For detailed information on the identified Phe-derived metabolites, including metabolite name, accurate m/z, retention time, mass formula, and annotated fragments, see Table S1. P, phenylalanine and derivatives; H, Hydroxycinnamic acid and their glycosides; K, Kaempferol glycosides; Q, Quercetin glycosides; C, Cyanidin glycosides; HCAA, Hydroxycinnamic acid amide.

Mitsunami et al., 2014). Here 30 out of the 35 identified metabolites were found differential between *pap1-D* and WT leaves; 26 accumulated at higher level in *pap1-D* and 4 were more abundant in WT leaves. Particularly, we have identified 14 anthocyanins (Table 1 and Table S1) and all of them highly accumulated in the *pap1-D* mutant leaves (p-value < 0.05). Ten anthocyanins were also detected at trace levels in

WT leaves. FC values (FC$_{PAP1/WT}$) were then calculated for the 10 anthocyanins; they ranged from 6.2 to 39.0 (Fig. 6). Apart from anthocyanins, an additional set of 16 phenylpropanoids displayed differential accumulation between *pap1-D* and WT leaves, with FC value higher than 2 (FC$_{PAP1/WT}$ or FC$_{WT/PAP1}$). Among them, we detected six flavonoids including three quercetin and three kaempferol glycosides. Interestingly, all quercetin glycosides highly accumulated in *pap1-D* leaves with FC values ranging from 3.7 to 9.1 (FC$_{PAP1/WT}$), while kaempferol glycosides were more abundant in WT with FC values ranging from 3.6 to 27.7 (FC$_{WT/PAP-OX}$) (Fig. 6A). Dihydrokaempferol serves as a 'hub' precursor that is further metabolized by flavonol synthase (FLS) to form a variety of kaempferol glycosides. Alternatively, it can be converted to dihydroquercetin by flavonoid 3′-hydroxylase (F3′H) for quercetin and quercetin glycosides biosynthesis by FLS, or for anthocyanin biosynthesis by flavonoid 3′-hydroxylase (F3′H) (Fig. 6B). It has been reported that overexpression of *PAP1* in *Arabidopsis* resulted in elevated transcript levels of *F3′H* and dihydroflavonol reductase (DFR) (Tohge et al., 2005). As a consequence, dihydrokaempferol is directed to the quercetin and anthocyanin branch, leading to the accumulation of quercetin glycosides and anthocyanins in *pap1-D* and reduced kaempferol glycosides levels (Fig. 6B). This supports our findings here that all quercetin glycosides accumulated to higher levels in *pap1-D* as compared to WT leaves. In addition to flavonoids, nine hydroxycinnamic acids, including caffeic, *p*-coumaric, ferulic, sinapic acid, and their glycosides highly accumulated in *pap1-D* leaves with FC values ranging from 2.1 to 11.5 (FC$_{PAP1/WT}$). Accumulation of hydroxycinnamic acids is likely due to the elevated expression of *PHENYLALANINE AMMONIA-LYASE* (*PAL*) and *4-COUMARATE-COA LIGASE* (*4CL*) involved in their biosynthesis (Fig. 6B). A single hydroxycinnamic acid amide (p-coumaroylagmatine) showed significant accumulation in WT as compared to *pap1-D* mutant leaves (Fig. 6a). The *pap1-D Arabidopsis* dominant mutant has been shown to exhibit significantly improved tolerance to UV radiation, salt (Oh et al., 2011), freezing (Schulz et al., 2016) and drought stress (Nakabayashi et al., 2014), which are attributed to the accumulation of anthocyanins, quercetin and hydroxycinnamic acid derivatives (Schulz et al., 2016; Stelzner et al., 2019).

Tohge et al. (2005) have employed high-resolution LC-MS to compare phenylpropanoid levels between WT and *pap1-D Arabidopsis* leaves. They detected 17 differential metabolites, including 11 anthocyanins, 3 quercetin glycosides and 3 kaempferol glycosides (Tohge
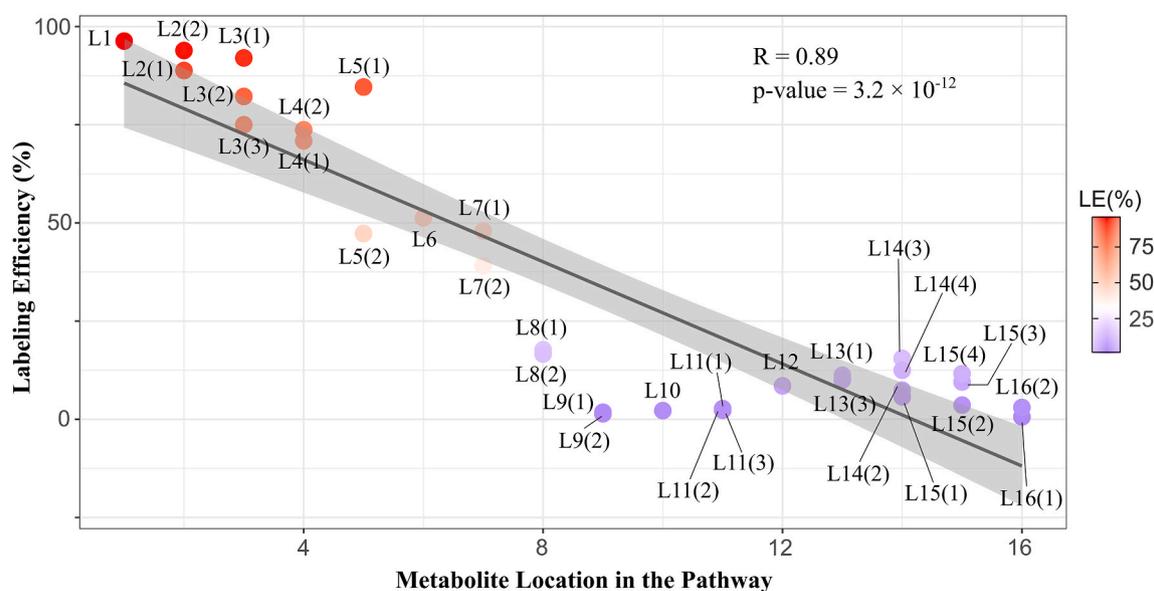
Fig. 5. Linear correlation between labeling efficiency and metabolite location in the phenylpropanoid pathway. The number next to each point corresponds to metabolite location shown in Table 1. 'L' stands for location, the number after 'L' indicates its location in the phenylpropanoid pathway. The number in the bracket is used to distinguish metabolites that are at the same pathway location; LE: labeling efficiency. The color scale represents different percentage of labeling efficiency.
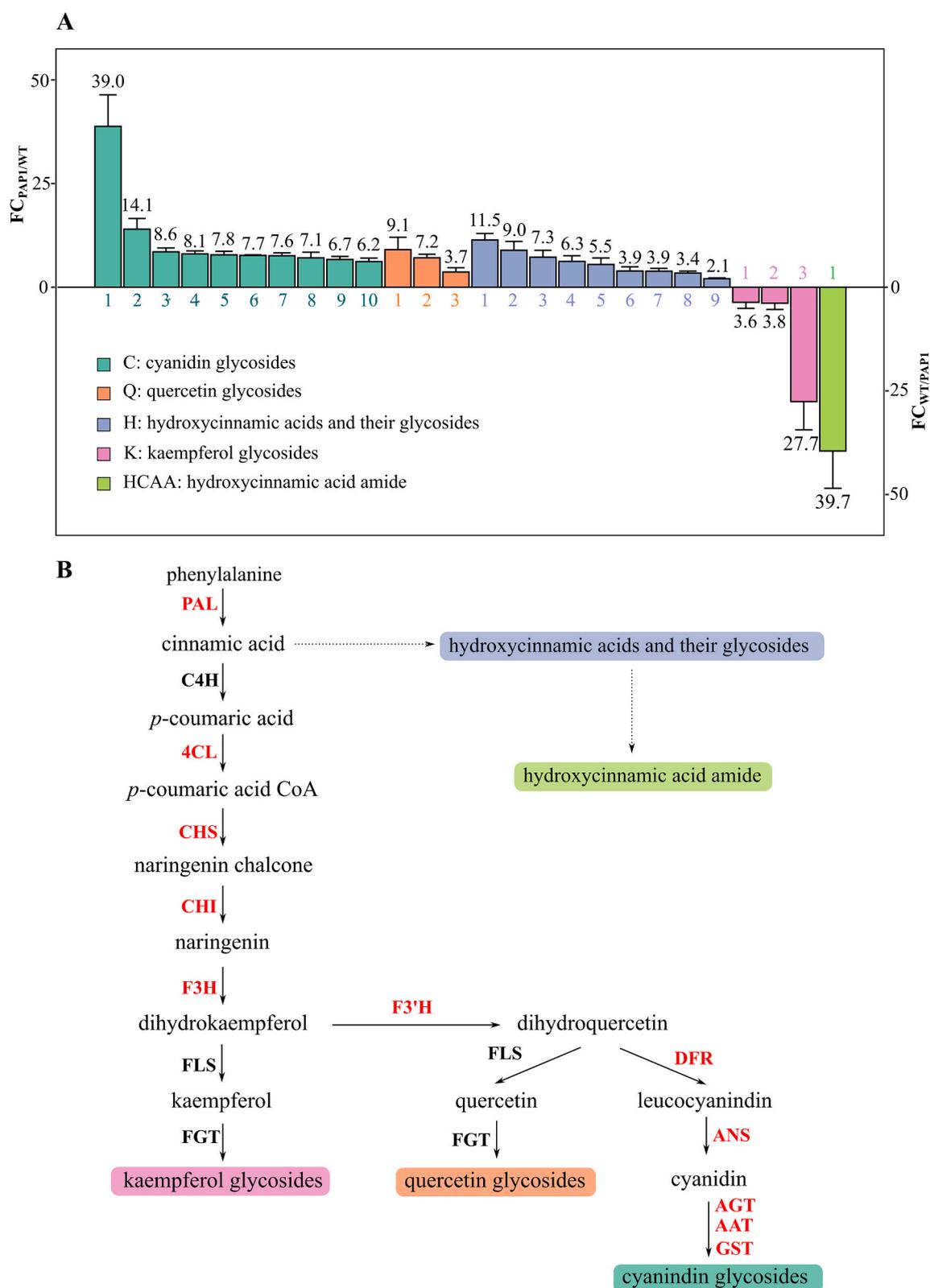
Fig. 6. Differential analysis of phenylpropanoid pathway metabolites between WT and *pap1-D* mutant *Arabidopsis* leaves. (A) Representation of differential metabolites found between WT and *pap1-D* in leaves. The data represents geometric fold change mean + standard deviation (SD) of four biological replicates. (B) The phenylpropanoid biosynthetic pathway and metabolites detected in the course of this study. The enzymes highlighted in red are those regulated by the PAP1 transcription factor according to the previous report from Tohge et al. (2005). Dash line denotes multiple enzymatic reactions in the pathway. PAL, phenylalanine ammonia-lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate-CoA ligase; CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; F3′H, flavonoid 3′-hydroxylase; FLS, flavonol synthase; FGT, flavonol glycosyltransferase; DFR, dihydroflavonol reductase; ANS, anthocyanidin synthase; AGT, anthocyani(di)n glycosyltransferase; AAT, anthocyanin acyltransferase. GST, glutathione S-transferase. The pathway was modified from Tohge et al. (2005).

et al., 2005). In this study, we have identified all 17 differential metabolites reported by Tohge et al. (2005). Moreover, we have identified 13 additional phenylpropanoids, including three anthocyanins, nine hydroxycinnamic acids and one hydroxycinnamic acid amide. Our results highlighted that DLEMMA could improve metabolite coverage and annotation.

3. Conclusion

The use of stable isotope labeling has been instrumental in circumventing major challenges in metabolomics assays, i.e. metabolite identification and quantification. In this study, we used DLEMMA to detect, identify and semi-quantitatively compare phenylpropanoids between WT and *pap1-D* mutant *Arabidopsis* leaves. By employing this approach, we have identified 35 Phe-derived metabolites in high confidence, of which 30 were differential between WT and *pap1-D* leaves. In particular, our results showed that the ectopic *PAP1* expression led to significant accumulation of cyanidin-type anthocyanins, quercetin-type flavonols and hydroxycinnamic acids and their glycosylated derivatives. While levels of kaempferol glycosides and a hydroxycinnamic acid amide were reduced in the *pap1-D* leaves. DLEMMA approach can be easily extended to other precursors and organisms for high-confidence metabolite identification and detailed metabolic pathway elucidation. Using DLEMMA requires that the precursor should be partially positionally labeled (i.e. only specific substructures of the precursor are labeled by stable isotopes). The rationale is that many precursor-derived metabolites are typically synthesized by adding up or breaking down the building blocks of the precursor. Since different substructures of the precursor can be recognized and distinguished by their unique labeling schemes, the substructures of the precursor-derived metabolites can be then inferred by comparing their labeling patterns with that of the precursor. Furthermore, in order to improve the confidence and efficiency of this approach, ideally each substructure of the precursor could be tagged with distinctive labeling schemes, thus providing more complete and definitive structural cues to the precursor-derived metabolites.

4. Experimental

4.1. Reagents

Unlabeled Phenylalanine (Phe), Label I Phe (Phe-$^{13}C_6$, all the six carbon atoms part of the phenyl ring labeled with ^{13}C) (^{13}C > 99%) and Label II Phe (Phe-$^{13}C_6$ $^2H_5$, all the six carbon atoms and five hydrogen atoms on the phenyl ring labeled with ^{13}C and ^{2}H) (^{13}C > 98%, and ^{2}H > 99%) were purchased from ISOTEC™ (Sigma-Aldrich). All other reagents were from Sigma-Aldrich Chemicals.

4.2. DLEMMA experiments set-up

Plants of *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae) ecotype Columbia were grown in a climate chamber under short-day conditions (8/16h, 22/18 °C, light/dark). One hundred milligrams leaves of 8-week-old plants were excised by cutting at the proximal side of the pedicel with scissors under water to avoid air penetration into the pedicel, which may influence the feeding efficiency. The leaves were immersed into a PCR tube with 0.5 mL aqua solution of 0.5 mg/mL feeding precursor [Phe (unlabeled), Phe $^{13}C_6$ (Label I), or Phe $^{13}C_6^2H_5$ (Label II)] (Fig. 1A). The DLEMMA feeding experimental set-up consisted of two genotype groups; three wild type (WT) and three *pap1-D* (ABRC stock number: CS3884) leaves with each group fed with unlabeled Phe (treatments #1 and #4 in WT and *pap1-D*, respectively), Label I Phe (treatments #2 and #5 in WT and *pap1-D*, respectively) and Label II Phe (treatments #3 and #6 in WT and *pap1-D*, respectively) (Fig. 1B). Feeding experiments were performed for 24 h at 25 °C under constant fluorescent illumination and humidity. After feeding, leaves were carefully rinsed with water to remove traces of external fed precursor. Next,

we combined leaves from different treatments to six experimental groups. Four of them were used for high-confidence metabolite identification: Group A (treatment #1), Group B (#2 + #3), Group C (#4) and Group D (#5 + #6). Two additional groups, i.e. Group E (treatment #2 + #6) and Group F (#3 + #5), were used for semi-quantitative differential comparison of the WT and *pap1-D*. This swapped label approach, as shown in Group E and F, was designed to account for the unequal matrix effects between WT and *pap1-D* plants, and the unequal incorporation efficiency between Label I and Label II Phenylalanine. Leaves were flash frozen with liquid N_2, ground and extracted with 0.1% formic acid in 92% MeOH:8%H_2O (v/v) (400 mL per 100 mg leaf). The suspension was sonicated for 10 min, centrifuged at 14,000 g for 5 min, and supernatant filtered through a 0.22 μm PTFE membrane filter before LC-MS analysis. Each group contain three biological replicates. LC-MS experiments were performed using a UPLC- QTOF-MS system (Waters Premier QTOF, Milford, MA, U.S.A.). Metabolites were separated on a 100 × 2.1 mm i.d., 1.7 μm UPLC BEH C18 column. The mobile phase consisted of 0.1% formic acid in Acetonitrile: Water (5:95, v/v) (phase A), and 0.1% formic acid in Acetonitrile (phase B). The linear gradient program was as follows: 100-72% A over 22 min, 72-60% A over 0.5 min, 60 - 0% A over 0.5 min, held at 100%B for a further 1.5 min, then returned to the initial conditions (100% A) in 0.5 min, and conditioning at 100% A. The total analytical run was 26 min. The flow rate was 0.3 mL/min, the column temperature was kept at 35 °C. Masses of the eluted compounds were detected by a QTOF Premier MS, equipped with ESI source. Full scan mass spectra were acquired from 50 to 1500 Da. The following settings were used for LC-MS/MS run: capillary spray at 3.0 kV; cone voltage at 30 eV; collision energies were 10–25 eV for positive mode and 15–40 eV for negative mode. Argon was used as the collision gas for collision-induced dissociation (CID) (MS/MS) experiments. The MS was calibrated using sodium formate, and leucine enkephalin used as the lock mass. A mixture of 14 standard compounds, injected after each five samples, was used for quality control.

4.3. DLEMMA data analysis and metabolic identification

The MassLynx software version 4.1 (Waters Inc.) was used to calculate accurate masses. The raw instrument data were first converted into mzML format using ProteoWizard MSConvert (Kessner et al., 2008). Peak picking was performed using XCMS (Smith et al., 2006) under R environment. The resulting XCMS set objects were then subjected to the R package Miso (Dong et al., 2019) for isotopologue cluster detection. Manual identification was performed using the SciFinder Database (https://scifinder.cas.org). Differential metabolites were selected if the fold change was greater or equal to 2 and the p-value was less than 0.05. Line plot and bar plot were produced using R package ggplot2.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.phytochem.2021.112740.

References

Borevitz, J.O., Xia, Y., Blount, J., Dixon, R.A., Lamb, C., 2000. Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. Plant Cell 12, 2383–2393. https://doi.org/10.1105/tpc.12.12.2383.

Chen, Z.A., Fischer, L., Cox, J., Rappsilber, J., 2016. Quantitative cross-linking/mass spectrometry using isotope-labeled cross-linkers and MaxQuant. Mol. Cell. Proteomics 15, 2769–2778. https://doi.org/10.1074/mcp.M115.056481.

Deng, Y., Lu, S., 2017. Biosynthesis and regulation of phenylpropanoids in plants. Crit. Rev. Plant Sci. 36, 257–290. https://doi.org/10.1080/07352689.2017.1402852.

Dong, Y., Feldberg, L., Aharoni, A., 2019. Miso: an R package for multiple isotope labeling assisted metabolomics data analysis. Bioinformatics 35, 3524–3526. https://doi.org/10.1093/bioinformatics/btz092.

Doppler, M., Bueschl, C., Kluger, B., Koutnik, A., Lemmens, M., Buerstmayr, H., Rechthaler, J., Krska, R., Adam, G., Schuhmacher, R., 2019. Stable isotope–assisted plant metabolomics: combination of global and tracer-based labeling for enhanced untargeted profiling and compound annotation. Front. Plant Sci. 10, 1366. https://doi.org/10.3389/fpls.2019.01366.

Feldberg, L., Dong, Y., Heinig, U., Rogachev, I., Aharoni, A., 2018. DLEMMA-MS-Imaging for identification of spatially localized metabolites and metabolic network map reconstruction. Anal. Chem. 90, 10231–10238. https://doi.org/10.1021/acs.analchem.8b01644.

Feldberg, L., Venger, I., Malitsky, S., Rogachev, I., Aharoni, A., 2009. Dual labeling of metabolites for metabolome analysis (DLEMMA): a new approach for the identification and relative quantification of metabolites by means of dual isotope labeling and liquid Chromatography−Mass spectrometry. Anal. Chem. 81, 9257–9266. https://doi.org/10.1021/ac901495a.

Fraser, C.M., Chapple, C., 2011. The phenylpropanoid pathway in arabidopsis. Arabidopsis Book 9, e0152. https://doi.org/10.1199/tab.0152.

Kessner, D., Chambers, M., Burke, R., Agus, D., Mallick, P., 2008. ProteoWizard: open source software for rapid proteomics tools development. Bioinformatics 24, 2534–2536. https://doi.org/10.1093/bioinformatics/btn323.

Kluger, B., Bueschl, C., Neumann, N., Stückler, R., Doppler, M., Chassy, A.W., Waterhouse, A.L., Rechthaler, J., Kampleitner, N., Thallinger, G.G., Adam, G., Krska, R., Schuhmacher, R., 2014. Untargeted profiling of tracer-derived metabolites using stable isotopic labeling and fast polarity-switching LC−ESI-HRMS. Anal. Chem. 86, 11533–11537. https://doi.org/10.1021/ac503290j.

Kurepa, J., Shull, T.E., Karunadasa, S.S., Smalle, J.A., 2018. Modulation of auxin and cytokinin responses by early steps of the phenylpropanoid pathway. BMC Plant Biol. 18, 278. https://doi.org/10.1186/s12870-018-1477-0.

Lee, W.J., Jeong, C.Y., Kwon, J., Van Kien, V., Lee, D., Hong, S.-W., Lee, H., 2016. Drastic anthocyanin increase in response to PAP1 overexpression in fls1 knockout mutant confers enhanced osmotic stress tolerance in Arabidopsis thaliana. Plant Cell Rep. 35, 2369–2379. https://doi.org/10.1007/s00299-016-2040-9.

Mahmud, T., 2007. Isotope tracer investigations of natural products biosynthesis: the discovery of novel metabolic pathways. J. Label. Compd. Radiopharm. 50, 1039–1051. https://doi.org/10.1002/jlcr.1391.

Matuszewski, B.K., Constanzer, M.L., Chavez-Eng, C.M., 2003. Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC−MS/MS. Anal. Chem. 75, 3019–3030. https://doi.org/10.1021/ac020361s.

Mitsunami, T., Nishihara, M., Galis, I., Alamgir, K.M., Hojo, Y., Fujita, K., Sasaki, N., Nemoto, K., Sawasaki, T., Arimura, G., 2014. Overexpression of the PAP1 transcription factor reveals a complex regulation of flavonoid and phenylpropanoid metabolism in nicotiana tabacum plants attacked by spodoptera litura. PloS One 9, e108849. https://doi.org/10.1371/journal.pone.0108849.

Nakabayashi, R., Yonekura-Sakakibara, K., Urano, K., Suzuki, M., Yamada, Y., Nishizawa, T., Matsuda, F., Kojima, M., Sakakibara, H., Shinozaki, K., Michael, A.J., Tohge, T., Yamazaki, M., Saito, K., 2014. Enhancement of oxidative and drought tolerance in Arabidopsis by overaccumulation of antioxidant flavonoids. Plant J. 77, 367–379. https://doi.org/10.1111/tpj.12388.

Noel, J.P., Austin, M.B., Bomati, E.K., 2005. Structure–function relationships in plant phenylpropanoid biosynthesis. Curr. Opin. Plant Biol. 8, 249–253. https://doi.org/10.1016/j.pbi.2005.03.013.

Oh, J.E., Kim, Y.H., Kim, J.H., Kwon, Y.R., Lee, H., 2011. Enhanced level of anthocyanin leads to increased salt tolerance in arabidopsis PAP1-D plants upon sucrose treatment. J. Korean Soc. Appl. Biol. Chem. 54, 79–88. https://doi.org/10.3839/jksabc.2011.011.

Ouyang, J., Wang, X., Zhao, B., Wang, Y., 2005. Enhanced production of phenylethanoid glycosides by precursor feeding to cell culture of Cistanche deserticola. Process Biochem. 40, 3480–3484. https://doi.org/10.1016/j.procbio.2005.02.025.

Park, S.-S., Wu, W.W., Zhou, Y., Shen, R.-F., Martin, B., Maudsley, S., 2012. Effective correction of experimental errors in quantitative proteomics using stable isotope labeling by amino acids in cell culture (SILAC). J. Proteomics 75, 3720–3732. https://doi.org/10.1016/j.jprot.2012.04.035.

Raghavendra, S., Ramesh, C.K., Kumar, V., Moinuddin Khan, M.H., 2011. Elicitors and precursor induced effect on L-Dopa production in suspension cultures of Mucuna pruriens L. Front. Life Sci. 5, 127–133. https://doi.org/10.1080/21553769.2011.649188.

Rasmussen, S., Dixon, R.A., 1999. Transgene-mediated and elicitor-induced perturbation of metabolic channeling at the entry point into the phenylpropanoid pathway. Plant Cell 11, 1537–1551. https://doi.org/10.1105/tpc.11.8.1537.

Schulz, E., Tohge, T., Zuther, E., Fernie, A.R., Hincha, D.K., 2016. Flavonoids are determinants of freezing tolerance and cold acclimation in Arabidopsis thaliana. Sci. Rep. 6, 34027. https://doi.org/10.1038/srep34027.

Shahaf, N., Rogachev, I., Heinig, U., Meir, S., Malitsky, S., Battat, M., Wyner, H., Zheng, S., Wehrens, R., Aharoni, A., 2016. The WEIZMASS spectral library for high-confidence metabolite identification. Nat. Commun. 7, 12423. https://doi.org/10.1038/ncomms12423.

Silvestro, L., Tarcomnicu, I., Rizea, S., 2013. Matrix effects in mass spectrometry combined with separation methods — comparison HPLC, GC and discussion on methods to control these effects. In: Coelho, A.V. (Ed.), Tandem Mass Spectrometry - Molecular Characterization. InTech. https://doi.org/10.5772/55982.

Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R., Siuzdak, G., 2006. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal. Chem. 78, 779–787. https://doi.org/10.1021/ac051437y.

Stelzner, J., Roemhild, R., Garibay-Hernández, A., Harbaum-Piayda, B., Mock, H.-P., Bilger, W., 2019. Hydroxycinnamic acids in sunflower leaves serve as UV-A screening pigments. Photochem. Photobiol. Sci. 18, 1649–1659. https://doi.org/10.1039/C8PP00440D.

Tohge, T., Nishiyama, Y., Hirai, M.Y., Yano, M., Nakajima, J., Awazuhara, M., Inoue, E., Takahashi, H., Goodenowe, D.B., Kitayama, M., Noji, M., Yamazaki, M., Saito, K., 2005. Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor: metabolomics and transcriptomics. Plant J. 42, 218–235. https://doi.org/10.1111/j.1365-313X.2005.02371.x.

Xiao, J.F., Zhou, B., Ressom, H.W., 2012. Metabolite identification and quantitation in LC-MS/MS-based metabolomics. Trac. Trends Anal. Chem. 32, 1–14. https://doi.org/10.1016/j.trac.2011.08.009.

Zhang, X., Liu, C.-J., 2015. Multifaceted regulations of gateway enzyme phenylalanine ammonia-lyase in the biosynthesis of phenylpropanoids. Mol. Plant 8, 17–27. https://doi.org/10.1016/j.molp.2014.11.001.

Zhou, W., Yang, S., Wang, P.G., 2017. Matrix effects and application of matrix effect factor. Bioanalysis 9, 1839–1844. https://doi.org/10.4155/bio-2017-0214.