



**HAL**  
open science

## The CYP71AZ P450 Subfamily: A Driving Factor for the Diversification of Coumarin Biosynthesis in Apiaceous Plants

Celia Krieger, Sandro Roselli, Sandra Kellner-Thielmann, Gianni Galati, Bernd Schneider, Jeremy Grosjean, Alexandre Olry, David Ritchie, Ulrich Matern, Frederic Bourgaud, et al.

### ► To cite this version:

Celia Krieger, Sandro Roselli, Sandra Kellner-Thielmann, Gianni Galati, Bernd Schneider, et al.. The CYP71AZ P450 Subfamily: A Driving Factor for the Diversification of Coumarin Biosynthesis in Apiaceous Plants. *Frontiers in Plant Science*, 2018, 9, pp.820. 10.3389/fpls.2018.00820 . hal-01899038

HAL Id: hal-01899038

<https://hal.science/hal-01899038v1>

Submitted on 28 Mar 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



# The CYP71AZ P450 Subfamily: A Driving Factor for the Diversification of Coumarin Biosynthesis in Apiaceous Plants

Célia Krieger<sup>1†</sup>, Sandro Roselli<sup>1†</sup>, Sandra Kellner-Thielmann<sup>2†</sup>, Gianni Galati<sup>1</sup>, Bernd Schneider<sup>3</sup>, Jérémy Grosjean<sup>1</sup>, Alexandre Olry<sup>1</sup>, David Ritchie<sup>4</sup>, Ulrich Matern<sup>2</sup>, Frédéric Bourgaud<sup>5</sup> and Alain Hehn<sup>1\*</sup>

<sup>1</sup> Laboratoire Agronomie et Environnement, Institut National de la Recherche Agronomique, Université de Lorraine, Nancy, France, <sup>2</sup> Institut für Pharmazeutische Biologie und Biotechnologie, Philipps-Universität Marburg, Marburg, Germany, <sup>3</sup> Max Planck Institute for Chemical Ecology, Jena, Germany, <sup>4</sup> INRIA Nancy, Grand-Est Research Centre, Laboratoire Lorrain De Recherche En Informatique Et Ses Applications, Nancy, France, <sup>5</sup> Plant Advanced Technologies, Nancy, France

## OPEN ACCESS

### Edited by:

Danièle Werck,  
Centre National de la Recherche  
Scientifique (CNRS), France

### Reviewed by:

Alain Tissier,  
Leibniz-Institut für Pflanzenbiochemie  
(IPB), Germany  
May Berenbaum,  
Illinois Rocstar, University of Illinois at  
Urbana–Champaign, United States

### \*Correspondence:

Alain Hehn  
alain.hehn@univ-lorraine.fr

† These authors have contributed  
equally to this work.

### Specialty section:

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

Received: 19 March 2018

Accepted: 28 May 2018

Published: 19 June 2018

### Citation:

Krieger C, Roselli S,  
Kellner-Thielmann S, Galati G,  
Schneider B, Grosjean J, Olry A,  
Ritchie D, Matern U, Bourgaud F and  
Hehn A (2018) The CYP71AZ P450  
Subfamily: A Driving Factor  
for the Diversification of Coumarin  
Biosynthesis in Apiaceous Plants.  
*Front. Plant Sci.* 9:820.  
doi: 10.3389/fpls.2018.00820

The production of coumarins and furanocoumarins (FCs) in higher plants is widely considered a model illustration of the adaptation of plants to their environment. In this report, we show that the multiplication of cytochrome P450 variants within the CYP71AZ subfamily has contributed to the diversification of these molecules. Multiple copies of genes encoding this enzyme family are found in Apiaceae, and their phylogenetic analysis suggests that they have different functions within these plants. CYP71AZ1 from *Ammi majus* and CYP71AZ3, 4, and 6 from *Pastinaca sativa* were functionally characterized. While CYP71AZ3 merely hydroxylated esculetin, the other enzymes accepted both simple coumarins and FCs. Superimposing *in silico* models of these enzymes led to the identification of different conformations of three regions in the enzyme active site. These sequences were subsequently utilized to mutate CYP71AZ4 to resemble CYP71AZ3. The swapping of these regions lead to significantly modified substrate specificity. Simultaneous mutations of all three regions shifted the specificity of CYP71AZ4 to that of CYP71AZ3, exclusively accepting esculetin. This approach may explain the evolution of this cytochrome P450 family regarding the appearance of FCs in parsnip and possibly in the Apiaceae.

**Keywords:** coumarin, cytochrome P450, furanocoumarin, *Pastinaca sativa*, psoralen, specialized metabolism

## INTRODUCTION

Plants have developed a wide range of strategies to take advantage of their environment and to adapt to many conditions by methods such as the production of defense molecules. This process has been studied for many decades (Fraenkel, 1959), but the molecular characterization of the enzymes responsible for the synthesis of the defense molecule arsenal primarily started in the 1990s (Funk and Croteau, 1993; Werck-Reichhart et al., 1997; Glawischnig et al., 1999). Beside dioxygenases (Kawai et al., 2014), the cytochrome P450s (P450s) are the most diversified enzyme family in plants (Nelson et al., 2004; Nelson and Werck-Reichhart, 2011) and are estimated to represent up to 1% of the annotated genes in plant genomes (Mizutani and Ohta, 2010). Because of

their diversity, these enzymes are considered major players in boosting metabolic pathway activity and therefore defense molecules (Glawischnig et al., 1999; Irmisch et al., 2014). Thanks to the development of next-generation sequencing methods, the number of transcriptomic databases has exploded. However, although the number of annotated P450 sequences is increasing day by day, the precise P450 functions often remain unknown. For example, more than 70% of the 245 P450s in the *Arabidopsis* genome, which is smaller than many other plant genomes, are orphans at the biochemical level (Mizutani and Sato, 2012).

Furanocoumarins (FCs) are molecules involved in defense reactions against attacks by herbivores (Berenbaum et al., 1991; Li et al., 2004). They have been detected in only a few plant families, and these families are not phylogenetically related. Linear (psoralens) and angular FCs may have arisen by convergent evolution (Bourgaud et al., 2014) and are distinguished by the position of the furan-ring on the coumarin core molecule (Figure 1). Plants first acquired the ability to produce linear isomers, and the formation of the angular analogs evolved later (Berenbaum, 2002). No plant that can synthesize only angular molecules has been described. Elicitor treatment of *A. majus* cell cultures increased the synthesis of FCs and the conversion of psoralen to 5-hydroxy-psoralen (bergaptol) (Hamerski and Matern, 1988a), and the use of radiolabeled precursors assigned the enzymatic conversion of demethyl [3-<sup>14</sup>C]suberosin into labeled (+) marmesin (Hamerski and Matern, 1988b) to P450s. However, only a few of these enzymes have been isolated and identified. The first P450 entity that was characterized at the molecular level catalyzes the synthesis of psoralen from (+)-marmesin and was assigned to the CYP71AJ P450 subfamily (Larbat et al., 2007). A variant in parsnip named CYP71AJ4 and sharing 70% amino acid identity with psoralen synthases did not accept marmesin as a substrate and was the first P450 shown to be involved in angular FC biosynthesis, converting (+)-columbianetin to angelicin (Larbat et al., 2009) (Figure 1). The assignment of two different P450s to the same subfamily but restricted to only one biosynthetic pathway each suggested that these enzymes evolved from a common ancestor (Larbat et al., 2009). A CYP71AJ-targeted search in various Apiaceae transcriptomic databases identified 36 partial or full-length paralogous genes distributed in 11 different apiaceous plants that either produce or lack FCs (Dueholm et al., 2015). A phylogenetic analysis of 19 of these CYP71AJ sequences revealed a clustering of 5 enzymes that have been described to be involved in the synthesis of FCs (Larbat et al., 2007, 2009; Dueholm et al., 2015). The functions of the 14 remaining enzymes have not been elucidated. This analysis suggests that the CYP71AJ P450 family might have evolved to synthesize these toxic defense molecules. An ancestral version of these enzymes might have been identified in plants that do not produce FCs (Dueholm et al., 2015). The spread of this gene family probably occurred through gene duplication associated with neofunctionalization. Such a mechanism has been described, for example, in *Arabidopsis* and for the glucosinolate biosynthesis pathway (Kliebenstein et al., 2001; Benderoth et al., 2006; Kliebenstein and Osbourn, 2012; Edger et al., 2015).

Here, we describe the identification and the functional characterization of several members of the CYP71AZ P450

subfamily. CYP71AZ1 was isolated by using a differential display RT-PCR approach applied to *A. majus in vitro* cultured cells that were treated with *Phytophthora sojae* extracts. CYP71AZ1 shares 33.8% identity with the psoralen synthase CYP71AJ3. BLAST searches performed on transcriptomic and genomic databases led to the identification of numerous paralogous genes that interestingly (and except one) were restricted to Apiaceae. In this report, we focused on four genes that were identified in parsnip and share 63 to 85% identity with each other. An *in vivo* investigation revealed that these four genes have different expression patterns in wounded parsnip roots. An *in vitro* functional characterization using proteins expressed in yeast showed that these enzymes are involved in the synthesis of both FCs and coumarins. A molecular shuffling of regions that change the catalytic sites of CYP71AZ4 allowed us to transform a FC/coumarin non-specific enzyme to a coumarin-specific enzyme. This study provides evidence that this multigenic P450 subfamily contributed to the diversification of coumarins and provides new insight into the evolution of higher plants.

## MATERIALS AND METHODS

### *Ammi majus* Culture and Isolation of CYP71AZ1

Total RNA and differential amplification of 3'-fragments of P450 cDNAs from elicited or water-treated (control) *A. majus* cell cultures was done as already described previously (Larbat et al., 2007).

### Parsnip Culture

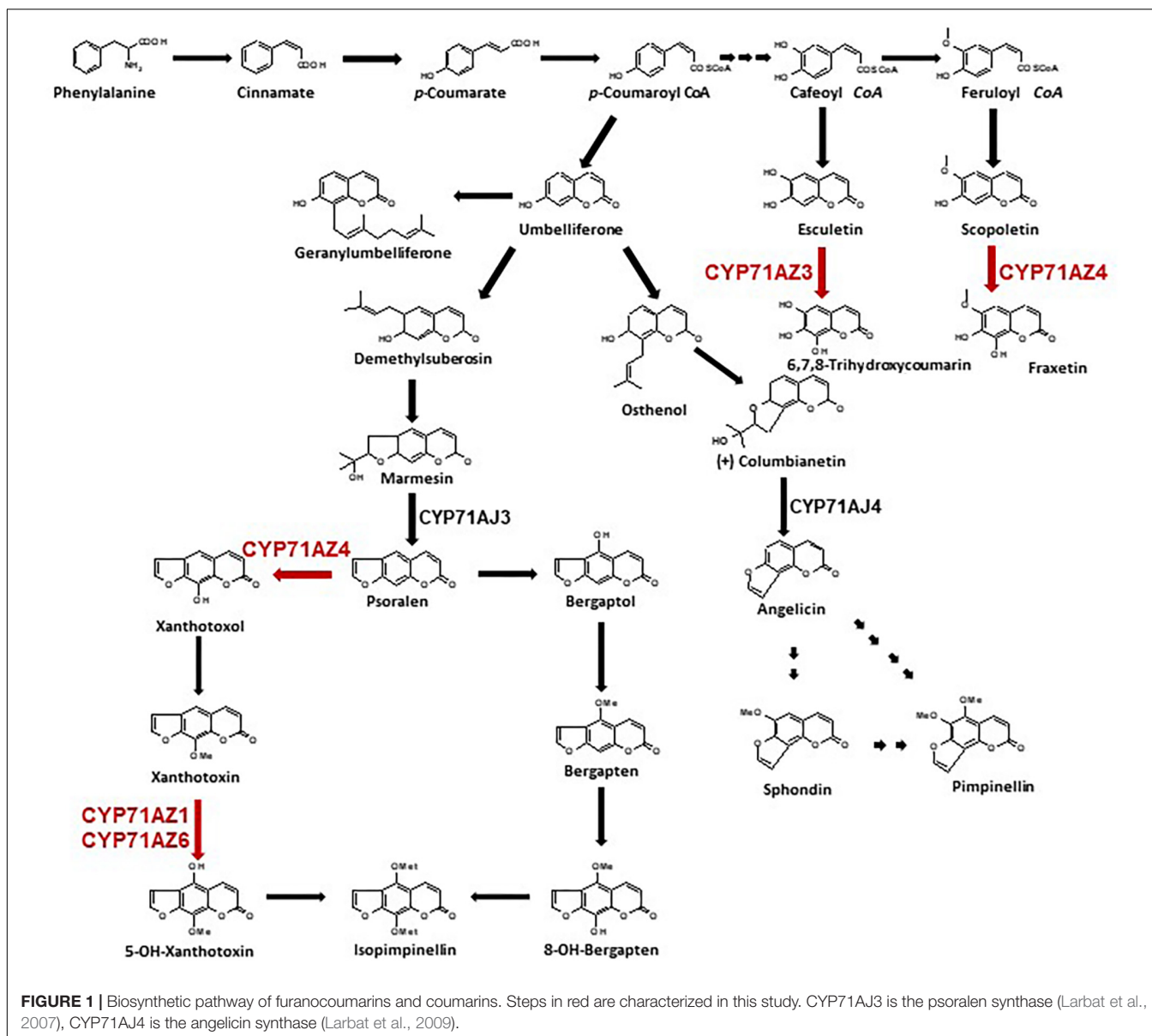
Parsnips (*Pastinaca sativa* L. subsp. *sativa* "Demi Long de Guernesey") were cultivated for 2 months as described by Munakata et al. (2016). Leaves and roots were separated from six plants used as controls and another six that endured the mechanical elicitation treatment as described by Roselli et al. (2016). Seeds were germinated in a growth chamber in soil under regulated climatic conditions, namely, temperature: 20/24°C; hygrometry: 40/70%; and light: 16 h/8 h day/night. Then, germinated seeds were transplanted into soil pots at the two leaf stage for an additional 2 months growth with the same climatic conditions.

### Wounding Induction

Plants were removed from the soil, and the roots were carefully washed with distilled water and dried. Wounding was done uniformly with a metal tip on the surfaces of the leaves, roots, and stems. Wounded tissues were frozen in liquid nitrogen and crushed with a mortar and a pestle prior to polyphenol extraction.

### Real-Time Quantitative PCR of the Target Gene Transcripts

Total RNA extraction from plants cultivated in soil was performed according the E.Z.N.A. Plant RNA kit protocol (Omega Bio-Tek, Norcross, GA, United States). Remaining DNA



was digested using an Amplification Grade DNase kit (Sigma-Aldrich, Saint-Louis, MO, United States<sup>1</sup>). The cDNA was synthesized using a High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific, Waltham, MA, United States<sup>2</sup>) and 100 ng of extracted mRNA. The cDNA was diluted fivefold before real-time quantification (StepOnePlus, Applied Biosystems). The real-time quantification was performed using a SYBR Green Master Mix: SYBR *Premix Ex Taq II* (Tli RNase H Plus) kit (Clontech, Mountain View, CA, United States<sup>3</sup>). The experiments were done on six biological replicates each.

The primers used for real-time quantification were as follows: 5.8S for 5'-AGCGGCGTCTTCCAAAAC-3' and 5.8S

rev 5'-ATATCTCGGCTCTCGCATCG-3' for the house-keeping gene 5.8S; CYP71AZ3 for 5'-TGACGACCGTAGAGCTTGCA-3' and CYP71AZ3rev 5'-ATTGGAACACCTGCTGG-3' for CYP71AZ3; CYP71AZ4 for 5'-GAAAAGAGAAATTCCAA CGAATG-3' and CYP71AZ4rev 5'-TGTTTGCTGCCTTTT TTGTTTC-3' for CYP71AZ4; CYP71AZ6 for 5'-GTTTGGG CGATGACAGCA-3' and CYP71AZ6rev 5'-TGTCTTCCTG GGCTTTCTTCA-3' for CYP71AZ6; and CYP71AJ3 for 5'-AA CAATAGCGACGGCTTTGG-3' and CYP71AJ3rev 5'-TTAAAA GTCCGGACGCCATG-3' for CYP71AJ3.

## Cloning of the CYP71AZ1 Open Reading Frame

The CYP71AZ1 open reading frame was PCR-amplified using a high-fidelity DNA polymerase (PrimeSTAR Max;

<sup>1</sup><https://www.sigmaaldrich.com/>

<sup>2</sup><https://www.thermofisher.com/>

<sup>3</sup><https://www.clontech.com/>





## Modeling

A library of 1,023 three-dimensional (3D) structures of CYP450 domains (Pfam code PF00067) was extracted from the KBDock database (Ghoorah et al., 2014), which had been built from a June 2013 snapshot of the Protein Databank (PDB). A sequence-based search of these structures using version 3.0 of the Kpax program (Ritchie et al., 2012) yielded the PDB structure 4EJI (P450 2A13) as the closest sequence homolog to CYP71AZ (39% sequence identity). Four similar structures (PDB codes 3QU8, 3T3Z, 3TK3, and 2NNJ) having no more than 90% sequence similarity to 4EJI were then extracted from the library using a structure-based search in Kpax. A multiple structure alignment of the first selected structures was built with Kpax, and the resulting multiple sequence alignment profile was used as input to the MODELLER (v9.13) homology modeling program (Sali and Blundell, 1993). Ten structural models were built using default MODELLER parameters, and the structure with the best DOPE score was retained as the final model of the CYP71AZ structure.

## In Silico Data Mining

The presence of various CYP71AZs was investigated, and their sequences were collected in transcriptomic and genomic databases [*Angelica archangelica* (onekp), *Apium graveolens* (SRR1023730 and DRR003696), *Bupleurum chinense* (SRR1002958), *Bupleurum scorzonerifolium* (SRR1002959), *Centella asiatica* (Phytometasyn), *Cimicifuga racemosa* (onekp), *Coriandrum sativum* (SRR1700630, SRR1700819, and SRR1700873), *Daucus carota* (SRR187755-1<sup>4</sup>), *Oenanthe javanica* (SRR1119291), *Petroselinum crispum* (SRR1030614), *P. sativa* (Dueholm et al., 2015; Roselli et al., 2016), *Thapsia garganica* (Henrik Toft Simonsen), and *Conium maculatum* (Heiko Rischer and (Hotti et al., 2015)]. Sequences are available in **Supplementary Figure S10**. The single reads of Illumina and 454 databases were assembled using the CLC genomic workbench (Qiagen).

The protein sequences were compared using Clustal X, and phylogenetic trees were constructed using the maximum likelihood (ML) methods included in MEGA7.

## RESULTS

### Identification of a Novel P450 Subfamily: CYP71AZ

Various studies have reported the pivotal role of P450 enzymes in the synthesis of FCs (Hamerski and Matern, 1988a,b; Larbat et al., 2007, 2009). To identify new genes involved in the synthesis of these molecules, we used a differential expression strategy that has been described previously (Larbat et al., 2007). A reverse transcription differential display approach using primers targeting a conserved P450 sequence identified a partial gene that was expressed more in *A. majus* cell

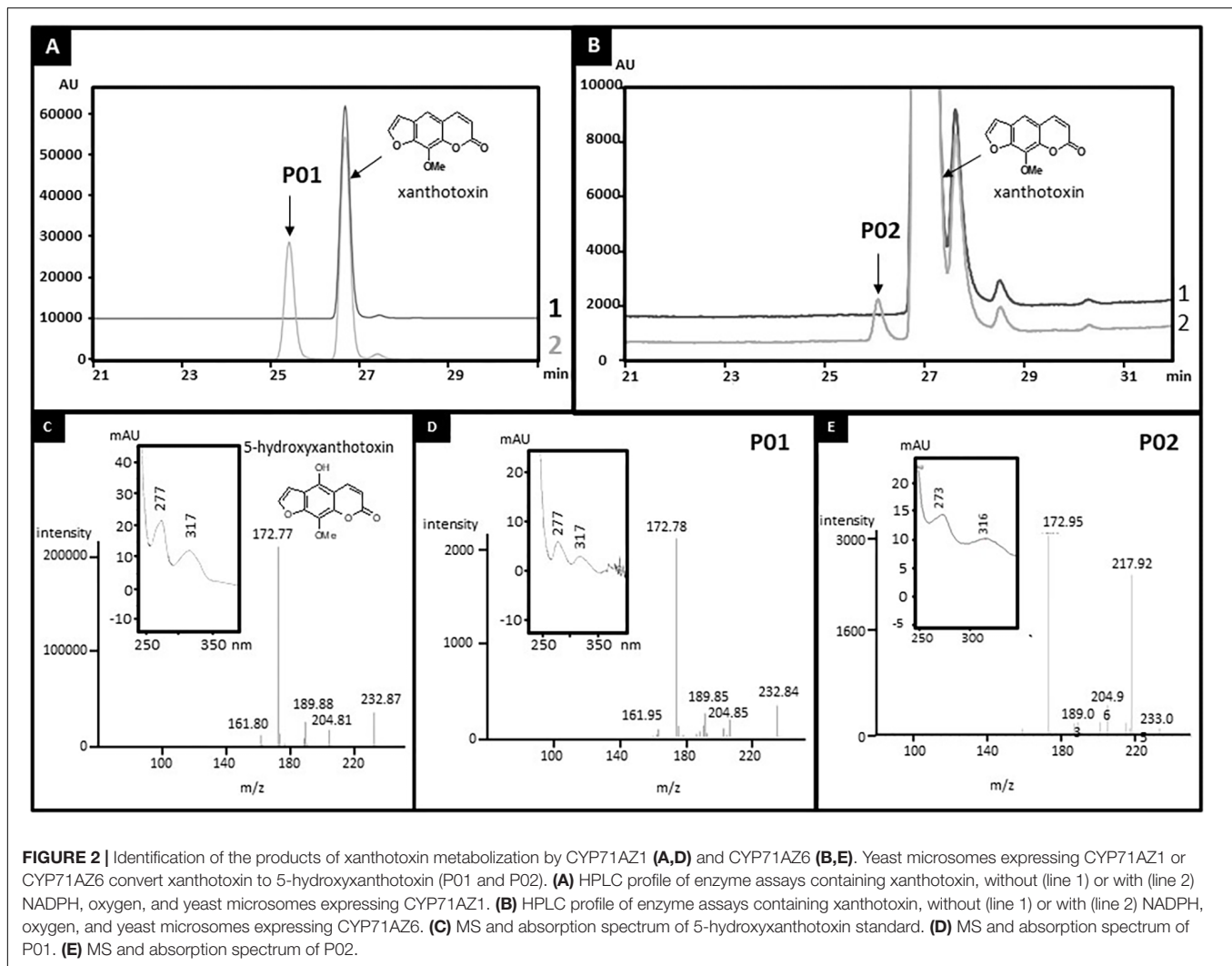
cultures elicited with crude *Phytophthora* cell wall elicitors than non-elicited cells. The full-length sequence was cloned using rapid amplification of cDNA ends (RACE) experiments, and the deduced protein sequence was submitted to the international P450 nomenclature committee and assigned as CYP71AZ1, the first member of a novel P450 subfamily (GenBank ABO32529.1)

CYP71AZ1 was expressed in a heterologous yeast expression system (Pompon et al., 1996) for functional characterization. A metabolic screening done with microsomes prepared from the yeast cells overexpressing this enzyme in the presence of FC pathway intermediates (**Figure 1**) showed the transformation of xanthotoxin into a single product, P01 (**Figure 2A**). Comparison with a set of standard compounds revealed that the elution time in HPLC and the UV spectrum of P01 matched those of 5-hydroxyxanthotoxin (**Figure 2C**), which is a precursor of isopimpinellin. The identity of the molecule was further confirmed by mass spectrometric analysis (**Figure 2D**). The apparent affinity for xanthotoxin ( $K_m = 13.1 \pm 3.2 \mu\text{M}$ ) is similar to the substrate affinities of most P450s involved in the synthesis of secondary metabolites. The isolation of CYP71AZ1 paved the way to a new class of hydroxylases relevant to the metabolism of FCs.

### CYP71AZ: A Multigenic P450 Subfamily

Databases were searched for the presence of CYP71AZ paralogs to assess the occurrence and diversity of this P450 subfamily. Coding sequences sharing high homologies with CYP71AZ1 are fairly restricted to a few spermatophytes and, except for *C. racemosa* (Ranunculaceae), were specifically reported in apiaceous plants such as *A. archangelica*, *P. sativa*, *D. carota*, *Heracleum lanatum*, *A. graveolens*, *B. chinense*, *B. scorzonerifolium*, *C. asiatica*, *C. sativum*, *O. javanica*, *P. crispum*, *C. maculatum*, and *T. garganica*. This search revealed 17 partial and full-length sequences sharing more than 55% identity and therefore belonging to the CYP71AZ subfamily (Werck-Reichhart et al., 2000). Among these 17 genes, 14 were identified in Apiaceae members known to produce FCs. However, some of them were also identified in transcriptomes of plant species that do not produce FCs, i.e., *Thapsia* spp. (Apiaceae). This result suggested that the substrate specificities of CYP71AZ-related enzymes might not be restricted to xanthotoxin (as for CYP71AZ1) and could instead provide a wider range of functions. This assumption is reinforced through a phylogenetic analysis of P450s belonging to Clan71, which shows that 10 CYP71AZ genes, including CYP71AZ1, are grouped in a cluster that were only identified in FC-producing plants (**Figure 3**). The genes identified in plants that do not produce FCs are localized out of this group. To investigate the blooming of this cluster, we further focused on *P. sativa*, which harbors four different genes. Three of them are included in the cluster and one is outside. CYP71AZ3, 4, 5, and 6 share 81.1, 71.1, 76.6, and 85.4% protein sequence identity with *A. majus* CYP71AZ1, respectively (**Supplementary Table S1**). GenBank entries for CYP71AZ3, 4, 5, and 6 are MH000218, MH000219, MH000220, and MH000221, respectively.

<sup>4</sup><http://apiaceae.njau.edu.cn>



## In Situ Expression of CYP71AZ in Parsnip Plants Suggests Their Role in the Synthesis of FCs

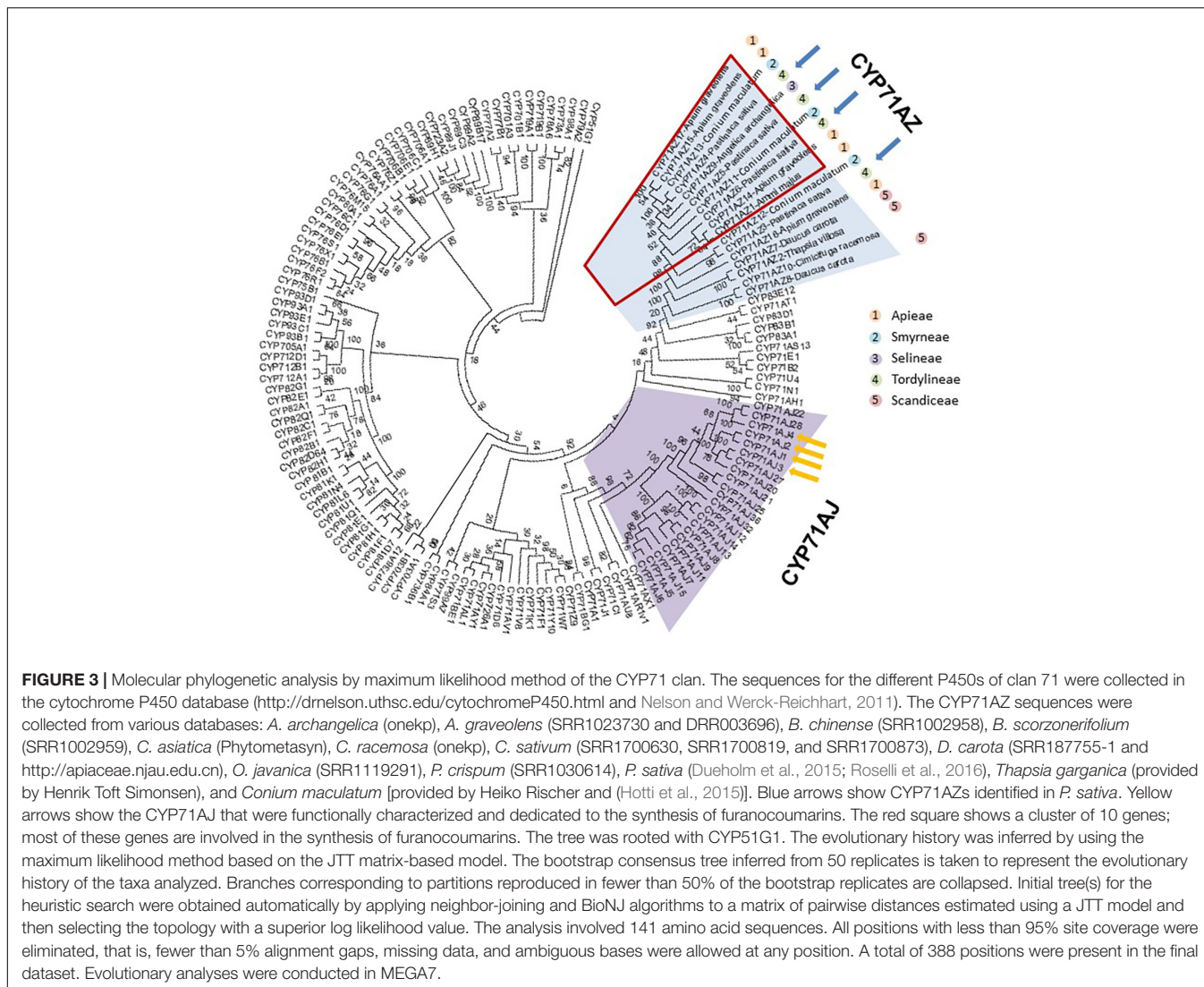
The metabolism of FCs in plants is influenced by environmental factors. Assessing the relationship between the concentration of FCs and the expression level of these different genes in *P. sativa* plants upon mechanical wounding was therefore relevant. In agreement with previously reported data (Roselli et al., 2016), the global FC concentration was higher in wounded roots than in unwounded roots (Figure 4A). The expression levels of the 4 CYP71AZ mRNAs were assessed and compared to the expression level of CYP71AJ3, which is involved in the synthesis of FCs in *P. sativa* (Larbat et al., 2007). RT-qPCR experiments (Figure 4B) highlighted different patterns of transcription for the four genes but these differences were not statistically relevant. The differential expression level of CYP71AJ3 was already demonstrated by Roselli et al. (2016). The tendency of the expression profiles of CYP71AZ4 and CYP71AZ6 matched that of CYP71J3. In contrast, the expression pattern of CYP71AZ3 was not modified in wounded plants. Although the coding

sequence of CYP71AZ5 is characterized in a genomic library (Roselli et al., 2016), we failed to amplify a full-length coding sequence corresponding to this gene. We concluded that this gene is not expressed in *P. sativa* tissues under our experimental conditions. Although not absolute proof, the diversity of the expression pattern of CYP71AZ genes suggests different roles in plant metabolism for the corresponding enzymes.

## Functional Characterization of CYP71AZs

CYP71AZ3, CYP71AZ4, and CYP71AZ6 were expressed in yeast cells, and their enzymatic activities were assayed using different intermediates of the FC pathway as substrates (Figure 1). CYP71AZ6 displayed the same enzymatic activity as CYP71AZ1, transforming xanthotoxin to 5-hydroxyxanthotoxin (product P02) with an apparent  $K_m$  of  $6.3 \pm 1.2 \mu\text{M}$  (Figures 2B,E and Table 1). This result was expected since CYP71AZ1 and CYP71AZ6 are very close from a phylogenetic point of view (Figure 3). As for CYP71AZ4, the enzymatic data provided evidence that this prefers psoralen as a substrate (Figure 5A),



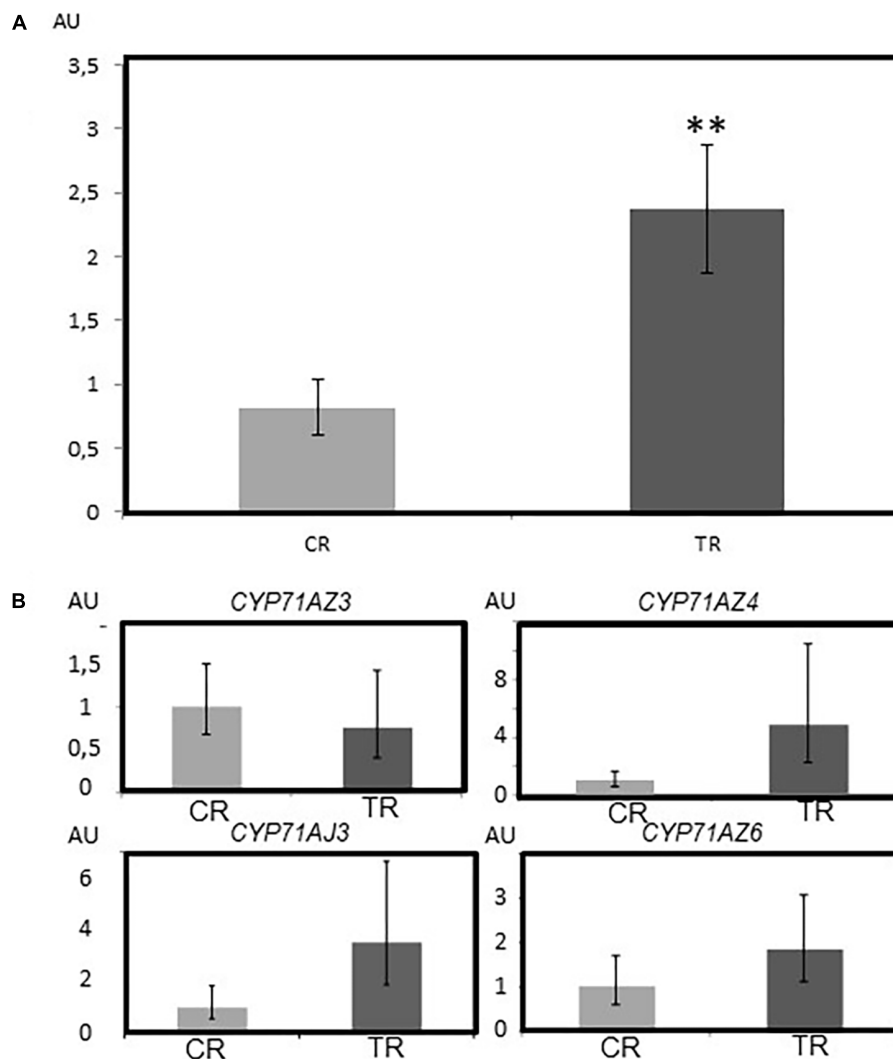


with a substrate affinity ( $K_m$ ) of  $11.3 \pm 2.2 \mu\text{M}$  (Table 1), and mass spectrometric evidence confirmed hydroxylation at the C8 position of the coumarin core structure occurred from xanthotoxol (P03; Figures 5B,C). Taken together, these first elements amend our knowledge of the psoralen metabolism: CYP71AZ4 catalyzes the step preceding *O*-methylation, and CYP71AZ1 and 6 perform the subsequent hydroxylation en route to isopimpinellin (Figure 1).

In this first set of assays with FC substrates, we could not ascribe any enzyme activity to CYP71AZ3. The metabolic screening for substrates was therefore extended to simple coumarins (Supplementary Figure S1). Some coumarins are precursors and represent the core of FCs. This additional screening was done on all the CYP71AZs. The extended screening revealed that at least five coumarins beside psoralen served as substrates for CYP71AZ4 (Supplementary Figures S2–S4), although the kinetic constants determined *in vitro* for some of these compounds were higher than commonly expected for P450s and appear to preclude an *in situ*

role as substrates. 6-Methoxycoumarin was converted to two products (Supplementary Figure S2A), one of which (P04) was identified as scopoletin (Supplementary Figures S2B,C,  $K_m = 9.5 \pm 1.5 \mu\text{M}$ , Table 1), while the other (P05) remained unassigned but showed a maximal UV absorption shift at shorter wave length than its precursor (Supplementary Figure S2D,  $K_m = 44.8 \pm 16.3 \mu\text{M}$ , Table 1). The CYP71AZ4-encoded hydroxylase slowly converted scopoletin to a product (P06) identified as fraxetin (Supplementary Figures S3A–C,  $K_m = 762.2 \pm 586.4 \mu\text{M}$ , Table 1). Moreover, 3-methyl-7-methoxycoumarin and 7-methoxycoumarin also served as substrates for CYP71AZ4, yielding product P07 (Supplementary Figure S4A) and P08 (Supplementary Figure S4B), respectively, which were recognized as 8-hydroxy-3-methyl-7-methoxycoumarin (Supplementary Figures S4C, S5) and daphnetin-7-methyl ether (Supplementary Figures S4D,E), respectively. However, the substrate affinities for both compounds are fairly high ( $K_m = 159.9 \pm 41.1$  and  $72.5 \pm 15.2 \mu\text{M}$ , Table 1) and probably



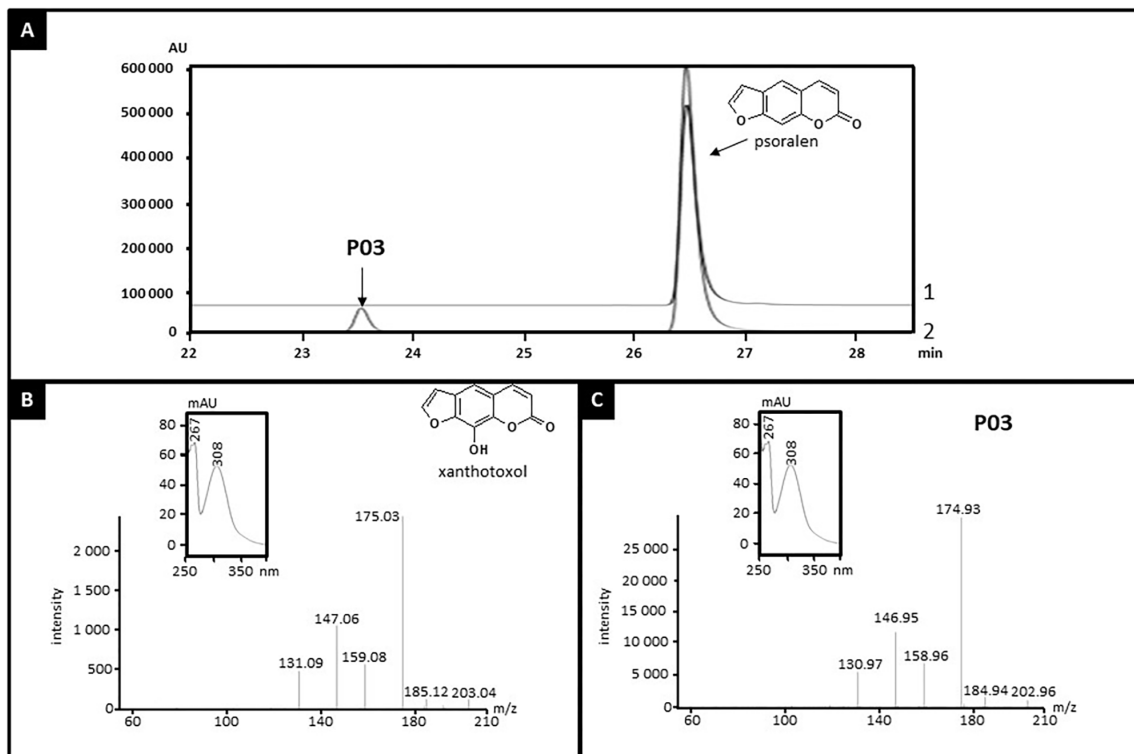


**FIGURE 4 |** Accumulation of furanocoumarins in parsnip roots and corresponding expression pattern of *CYP71AJ3*, *CYP71AZ3*, *CYP71AZ4*, and *CYP71AZ6*. **(A)** Assessment of the FC content based on UHPLC/MS as described by Munakata et al. (2016). **(B)** mRNA expression patterns of *CYP71AZ3*, *CYP71AZ4*, *CYP71AZ6*, and *CYP71AJ3*. The stars represent the significant differences in the results between the two conditions according to the Wilcoxon test (\*\**P*-value < 0.01). AU: arbitrary units. CR: control Roots (light gray). TR: treated roots (dark gray). The experiments were done on six biological replicates.

**TABLE 1 |** Metabolization pattern and kinetic parameters for CYP71AZs.

Substrate	Product	Apparent affinity ( $\mu\text{M}$ )			
		CYP71AZ1	CYP71AZ3	CYP71AZ4	CYP71AZ6
Xanthotoxin	5-hydroxyxanthotoxin	P01: $13.1 \pm 3.2$	N.D.	N.D.	P02: $6.3 \pm 1.2$
Psoralen	Xanthotoxol	N.D.	N.D.	P03: $11.3 \pm 2.2$	N.D.
6-Methoxycoumarin	Scopoletin	P11: N.Q.	N.D.	P04: $9.5 \pm 1.5$	P12: N.Q.
	$m/z = +16$	N.D.	N.D.	P05: $44.8 \pm 16.3$	N.D.
Scopoletin	Fraxetin	N.D.	N.D.	P06: $762.2 \pm 586.4$	N.D.
7-Methoxy-3-methylcoumarin	$m/z = +16$	N.D.	N.D.	P07: $159.9 \pm 41.1$	N.D.
7-Methoxycoumarin	Daphnetin-7-methylether	P09: N.Q.	N.D.	P08: $72.5 \pm 15.2$	P10: N.Q.
Esculetin	$m/z = +16$	N.D.	P13: $248.6 \pm 51.9$	N.D.	N.D.

Reactions were set up in KP<sub>i</sub> buffer (pH 7.4) at 28°C for 10 min and stopped by the addition of 10  $\mu\text{L}$  of 0.1 M HCl. Products were quantified after HPLC separation at 330 nm. N.D., not detected; N.Q., not quantifiable.



**FIGURE 5** | Identification of the product of psoralen metabolism by CYP71AZ4. Yeast microsomes expressing CYP71AZ4 convert psoralen to xanthotoxol (P03). **(A)** HPLC profile of enzyme assays containing psoralen, without (line 1) or with (line 2) NADPH, oxygen, and yeast microsomes expressing CYP71AZ4. **(B)** MS and absorption spectrum of standard xanthotoxol. **(C)** MS and absorption spectrum of P03.

also preclude any *in situ* relevance of these *in vitro* data. The mass spectrum of P07 showed a molecular ion peak at  $m/z$  207  $[M-H]^+$  (Supplementary Figure S4C), which agrees with a molecular formula of  $C_{11}H_{10}O_4$ , indicating 16 mass units more than the mass of the substrate, 3-methyl-7-methoxycoumarin. Thus, in accordance with incubation with CYP71AZ4, the product P07 was hypothesized to possess an additional oxygen atom. Subsequent NMR analysis showed that P07 was 8-hydroxy-3-methyl-7-methoxycoumarin (Supplementary Figure S5).

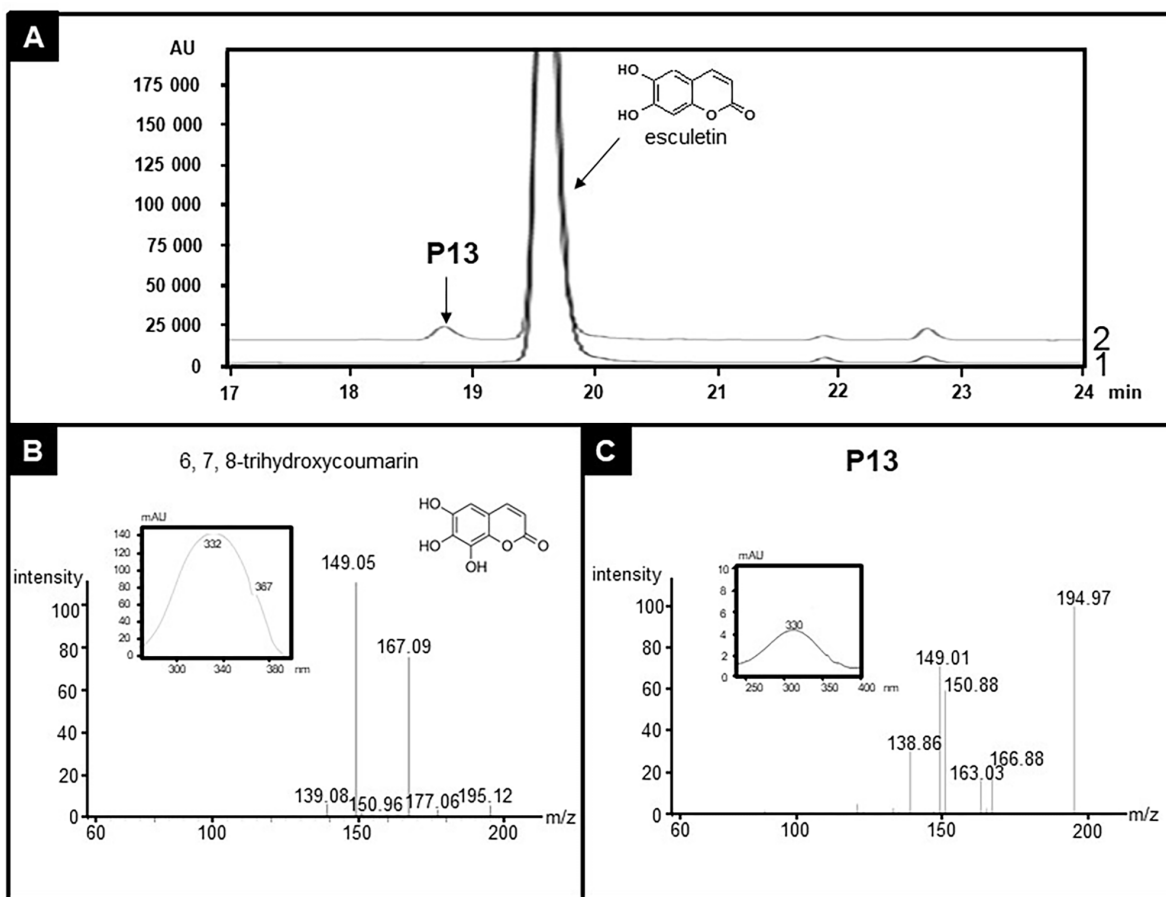
Extended enzyme assays also disclosed some side activities for CYP71AZ1 and CYP71AZ6. In addition to xanthotoxin, both enzymes hydroxylated 7-methoxycoumarin with very poor efficiencies to 5-hydroxy-7-methoxycoumarin (P09 or P10, Supplementary Figures S6A–E). Furthermore, both enzymes also slowly metabolized 6-methoxycoumarin to products P11 and P12 (Supplementary Figures S7A,C). We were not able to measure the catalytic parameters for this set of experiments with 7-methoxycoumarin and 6-methoxycoumarin as substrates. The products share their MS profiles (Supplementary Figures S7B,D) but could not be identified by NMR because of their limited availability. Finally, the screening also revealed a substrate for CYP71AZ3, the gene that did not respond to the wounding of *P. sativa* plants. This enzyme exclusively accepted esculetin among all the compounds included in the assays (Figure 6A) and formed (Table 1) 6,7,8-trihydroxycoumarin with a poor turnover

rate ( $K_m = 248.6 \pm 51.9 \mu M$ ). This product was identified by absorption spectroscopy as well as MS analyses (Figures 6B,C and Supplementary Figure S8). The mass spectrum of P13 showed a molecular ion peak at  $m/z$  195  $[M-H]^+$  (Figure 6C), which is in agreement with a molecular formula of  $C_9H_6O_5$ , indicating 16 mass units more than the mass of the substrate, esculetin. Thus, in accordance with incubation with CYP71AZ3, the product P13 was hypothesized to possess an additional oxygen atom. The structure of 6,7,8-trihydroxycoumarin was further confirmed by NMR (Supplementary Figures S8, S9).

The enzymatic and analytical data underline the significance of the CYP71AZ subfamily of enzymes for the overall synthesis of coumarins and FCs (Table 1) and adds several pieces to the FC pathway puzzle. Surprisingly, CYP71AZ4 displayed a broad range of substrate specificities *in vitro* and may be involved in multiple branches of the coumarin biosynthesis *in situ*. This result is noteworthy because the other CYP71AZs catalyze more well-defined steps in these pathways.

### Identification of Prominent Regions in the Catalytic Sites of CYP71AZ4 and CYP71AZ3

Alignments of the four complete CYP71AZ peptide sequences identified from parsnip display a moderate degree of conservation (Figure 7). To better understand the



**FIGURE 6** | Identification of the product of esculletin metabolism by CYP71AZ3. Yeast microsomes expressing CYP71AZ3 convert esculletin to 6,7,8-trihydroxycoumarin (P13). **(A)** HPLC profile of enzyme assays containing esculletin, without (line 1) or with (line 2) NADPH, oxygen, and yeast microsomes expressing CYP71AZ3. **(B)** MS and absorption spectrum of standard 6,7,8-trihydroxycoumarin. **(C)** MS and absorption spectrum of P13.

structure/activity relationship of these enzymes, we compared the highly specific for coumarin enzyme CYP71AZ3 and the broadly specific for coumarin/FC enzyme CYP71AZ4. An *in silico* model of CYP71AZ3 was constructed with esculletin fitted in the active site. This model was superimposed with sequences of the other CYP71AZs using the “Kpax” software (Ritchie et al., 2012) and led us to identify several regions that clearly differ in their active site conformations. These regions include substrate recognition sites (SRSs) (Rupasinghe et al., 2003) SRS1, SRS4 (alpha helix), and SRS5 (beta turn) (Figure 7B). Consequently, three CYP71AZ4 mutants were constructed (mut1-3, Figure 7C) to corroborate the involvement of these regions in the differential substrate specificities of CYP71AZ4 and CYP71AZ3. In these mutants, either one of these regions of wild-type CYP71AZ4 was replaced by the corresponding CYP71AZ3 peptide sequences (Nucleotide sequence 318-386 for mutant 1; 837-936 for mutant 2; and 1089-1115 for mutant 3). A fourth mutant was also constructed, hosting the three modifications in a single protein. The enzymatic characterization of the mutants expressed in yeast revealed four different substrate specificities (Figure 7C). The swapping of the mut2 sequence most likely affected the

volume of the active site pocket since psoralen was no longer metabolized. The replacement of SRS1 (mut1) narrowed the substrate specificity to solely target esculletin (in addition to psoralen), which is not even a substrate for wild-type CYP71AZ4. The exchange of SRS5 (mut3) caused an intermediate type of substrate specificity. Finally, the simultaneous swapping of the three regions (mut4) resulted in a mutant CYP71AZ4 whose specificity was identical to that of CYP71AZ3 (Figure 7C). In summary, the data provided unequivocal evidence that the SRS regions of CYP71AZs are intimately involved in the specificities of these enzymes and confirm the relevance commonly assigned to SRSs in cytochrome P450s.

## DISCUSSION

The adaptation of plants to changed environmental conditions often requires the creation of modified or novel enzymatic capabilities, and the P450 superfamily is particularly susceptible to such evolutionary processes as confirmed by the vast array of reactions catalyzed by this multigene family (Mizutani and Ohta,

**A**

```

71AZ6  --MDPVVIFL V--LAFPIAS VYLLFYHKKR VGG-LSSPPG PRGLPFIGNF FQLYKAPCIH EYLWNLSSKY GSLMTLHMGS VPILVVSPPK MAKEVLTKQD
71AZ3  --MEPVFLFL I--LAFPIAS VYLLFYHKKR VVG-LSAPPG PFGLPFIGNF HQLYKASSTH EYLWNLSSKY GSLVTLRMGS VPILVVSPPK MAKEVLTKQD
71AZ5  --MDPAVIFV VNILAFPIAS VYLLSCHKKR VDGRLSSPPG PRGLPFIGNF YQLYKAPSSH EYLWNLSSKY GSLVTLHMGS VPILVVSPPK MAQQVMKTQD
71AZ4  --MDPAAIFL I--LAIFIAS VYLLFYHKKR VNG-LSSPPG PRGLPFIGHF YQIYKSECAH EYISNLSSKY GSLMTLHLSG VPALVVSPPK MAQEVLTQD

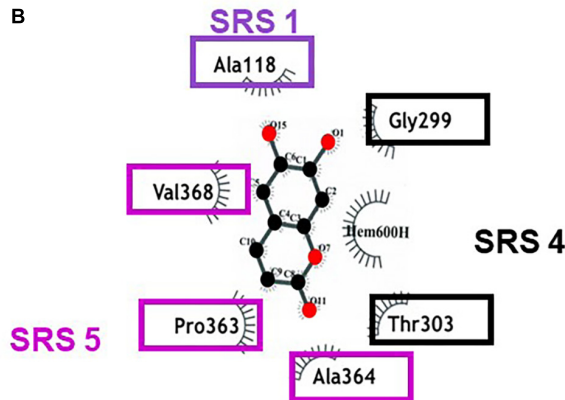
          SRS 1      Mut1
71AZ6  LIYCSRPMFT GMRKLSYDGL DVAFSTYSEH WRHVRKLCFL ELFTQKRAQL CFRLVHEQEV SRMIVRLSET AAASKDVNVA ECFSNLSISI ISRVAFGKRY
71AZ3  LIYCSRPMFT GMRKLSYDGL DVAFSTYSEH WRHVRKLCFL ELFTQKRAQL NFRFVREQEV SRMIARLSET AAASKDVNVA EWFNSLATCI ISRVAFGKRY
71AZ5  LIYCSRPMFT GMRKLSYDGL DVAFSTYSEH WRHVRKLCFL ELFTQKRAQL CFRLVHEQEV SRMIDRLSEA ASAFKDVNVA ECFSNFTTSV ISRVAFGKRY
71AZ4  LVFCSRPMFT GMRKLSYDGL DVAFSTYSEH WRHVRKLCFL ELFTQKRAQL NFRFVREDEV SRMVGRLESA AAASKDVNVA ECFTNFATSI ISRVAFGKRY

          SRS 2      SRS 3      Mut2      SRS 4
71AZ6  DEDGIGKERL QRMLSEIDAM LAGIFVSDFF PMFGWIDRLS GMRARLRTF KEMDMFYQEL IDEHLKPNRF ESPTEDLIDV MLKNKG-SSC LLTMDLTKAI
71AZ3  DEDGIGKERF QRMLSEIDAM FIGFFVSDFF PMFGWIDKLS GMRARLRTF KDLDMFYQEL IDEHLKPNRL ESPTEDLIDV MLKNKG---S SLTKDLMKAI
71AZ5  DEDMLGKQRF QRMLSEIDAV SIAFFVSDFF PMFGWIDRLS GMRARLRTF KEMDMFYQEL IDEHLKPNRF ESPTEDLIDV MLKTKDSSSF ALTMDLTKAI
71AZ4  DEDNLGKEKF QRMVADIEAM FAFFVSDFF PMFGWIDRLS GVKAVLRNF NEMDMFYQEL IDEHLKPNRF ESLNGDLIDV MLKNKG---S FLTMDLTKAI

          SRS 4      SRS 5      Mut3
71AZ6  LLNVFNGGTG TSAT LVWAM TALLRNQGVW KKAQEDIRRV IGRKGNVDED DIQNLSYLRA VVKETMRLP TGPILLPRPT MESSIIGEDK DHMYMIKPKT
71AZ3  LLNVFNGGTG TTAT LAWAM TALLRNQGVW KKAQEIIRSV IGKKGKVVDED DFPSPLYLRA VVKETMRLP PAFVLVPRPT MESSIIGEDK DHMYMVKPKT
71AZ5  LQNIFFNGGTG TSAT ALTSAM TELMRDQGVW KKAQEIIRRV IGKKGKVVDED DIQNLPLYLRA VVKETMRLP PAFVLLPRPT METSIIGEDK DHMYMIKPKT
71AZ4  LLNVFSGGIG TTGS LVFAM TALLRNQGVW KKAQEVRSV IGKKEIVDED DIQKLYLRA VVKETLRLP PGPILLPRVA MESCVLGEDE DHMYMIKPKT

          SRS 6
71AZ6  LVVYSMWAIG RDPETWKNPM EFWPERFLER PDLNYKQQQF EYIPFGARR ICAGIHGVT TVELALANLL YTFDWEPPAG TRFEDIDDET LNGLTLQKKN
71AZ3  VVYISMWAIG RDPKTIWKNPM EFWPERFLER PDINYKQQQF EYVFPFGARR ICAGINLGTV TVELALANLL YTFDWEPPAG MRFEDIDDET TNGLALQKKN
71AZ5  LVVYSMWAIG RDPETWKNPM DFVPERFMR PDIDYKQQQF EYIPFGARR ICAGIHLAVT TVELALANLL YTFDWEPPSG TRFR---DEI RN-----RN
71AZ4  IVVYNTWIG RDPKYWKNPL EFWPERFFER PDLNYTQQQF EYLPFGSRR ICAGIIIGQN NVEVGLANLL YSFDWEPPTG KTFEDIDQPP CNGLTLAKKN

71AZ6  ALYIRPKKYI CP
71AZ3  ALYIRPKKYM FP
71AZ5  KLYIRPKKYM SS
71AZ4  PLYIRPKIYV HP
    
```



**C**

Substrate	Product	71AZ4	71AZ4 mut1	71AZ4 mut2	71AZ4 mut3	71AZ4 mut4	71AZ3
Psoralen	Xanthotoxol	+	+	-	+	-	-
Scopoletin	Fraxetin	+	-	+	+	-	-
6-methoxy coumarin	Scopoletin	+	-	+	+	-	-
	N.D. m/z = +16	+	-	+	-	-	-
7-methoxy coumarin	Daphnetin-7-methylether	+	-	+	+	-	-
7-methoxy-3-methyl coumarin	N.D. m/z = +16	+	-	+	-	-	-
Esculetin	N.D. m/z = +16	-	+	+	+	+	+

**FIGURE 7 |** Sequence alignment of *P. sativa* CYP71AZ1, CYP71AZ3, CYP71A74, and CYP71AZ6 and construction of the CYP71AZ4 mutants. **(A)** Alignment of the four peptide sequences. Red amino acids are divergent among the four proteins. SRSs are substrate recognition sites (Rupasinghe et al., 2003). Yellow boxes highlight the three regions that were swapped for the mutagenesis. **(B)** Identification of the amino acids involved in the interaction with esculetin as substrate. **(C)** Summary of the metabolism of various substrates by the CYP71AZ mutants. ND: not determined.



2010; Mizutani and Sato, 2012). This adaptation is illustrated, for example, by the accumulation of FCs in umbelliferous plants infested with herbivorous insects (Berenbaum, 2001, 2002; Jogesh et al., 2015). The production and variation of FCs under these conditions were interpreted as the outcome of an “arms race” between plants and insects. The biosynthetic pathway leading to FCs has been studied for many years in Apiaceae and was shown in the 1980s to involve the sequential action of several P450s (Hamerski and Matern, 1988a,b). The first gene of this enzyme category, CYP71AJ1, was identified in 2007 (Larbat et al., 2007).

An intimate cooperation of P450s from wheat or maize has been demonstrated (Jonczyk et al., 2008), converting indole-3-glycerol phosphate (from the L-tryptophan pathway) to benzoxazinones by several variants of a single P450 subfamily (CYP71D). Such variants have likely evolved from one ancestor. The same mechanism was suspected to lead to the synthesis of FCs. This hypothesis initially gained support by the functional characterization of CYP71AJ4 (Larbat et al., 2009) catalyzing the synthesis of angelicin which is the analogous reaction in the pathway branching to angular FCs (Figure 2). Nevertheless, further studies of diversification revealed the presence of CYP71AJ paralogs in some Apiaceae species that do not produce FCs (Dueholm et al., 2015). In addition, no other CYP71AJs were related to the synthesis of FCs. With the discovery of the blooming CYP71AZ P450 subfamily, the hypothesis that only a few number of P450 families are specialized in the synthesis of FCs and coumarins had been revived. Since simple coumarins are abundantly present in spermatophytes, and FCs are present in only a few of them, the plants may have gained new CYP71AZ functionalities during evolution concomitant with or after those of the CYP71AJ family.

The study and functional characterization of five different CYP71AZs reported here provides evidence that each of them has its own evolutionary history. CYP71AZ3 is not overexpressed upon wounding of *P. sativa* plants. Our experiments show that this enzyme does not metabolize FCs but rather accepts a coumarin as substrate, although the catalytic properties determined *in vitro* suggest that esculetin is not the physiological substrate. Therefore, to identify the enzyme real substrate, the screening must be extended to other coumarins. CYP71AZ1 and 6 are very close to each other and have a restricted substrate specificity dedicated to the synthesis of an FC, which might reflect that they were selected with the appearance of the pathway and are now almost specialized for the synthesis of xanthotoxol. These reactions are catalyzed with reasonable  $K_m$  values. The evolution of this cytochrome P450 subfamily might therefore be close to that of the CYP71AJ subfamily. We can assume that the CYP71AZ and CYP71AJ P450 subfamily were originally involved in the synthesis of some common metabolites and evolved to be able to participate to the synthesis of FCs.

The emergence of the branching pathway to linear and angular FCs with CYP71AJs as the pivotal enzymes (Larbat et al., 2007, 2009; Dueholm et al., 2015) was a major breakthrough for the defense of these plants. CYP71AZ4 might thus be allocated to an intermediate stage to convert not only simple coumarins but also FCs. Regarding coumarins, CYP71AZ4 was

shown to convert 6-methoxycoumarin to scopoletin *in vitro* at a physiological acceptable  $K_m$  (9.5  $\mu$ M), and scopoletin can be further converted to fraxetin. However, this second reaction required an extraordinary high scopoletin concentration ( $K_m = 762.2 \mu$ M). CYP71AZ4 also accepts psoralen as a substrate ( $K_m = 11.3 \mu$ M, Table 1). Since the  $K_m$  values are close to each other and in an acceptable physiological range, CYP71AZ4 may be an enzyme on its way to becoming specialized in the evolutionary scheme.

Finally, CYP71AZ5 has no assigned function. The fact that this coding sequence is not transcribed and could not be cloned under our experimental conditions could reflect a pseudogenization or a gene duplication where the neofunctionalization has not yet occurred. From the evolutionary point of view, this gene might be close to CYP71AZ4, which is consistent with the phylogenetic tree. Thus, gene duplication and neofunctionalization appear to have occurred at several stages of the evolution of the CYP71AZ subfamily of genes and in close association with the evolution of CYP71AJ sequences. This evolution can be mimicked with targeted mutagenesis. Our sequence-swapping experiments showed that modifications required for modifying one enzyme into another are not limited to simple amino acid changes (mut1-3) but also involve modification of the tertiary structure of the protein (mut 4). Such a directed evolutionary strategy could be used to develop enzymes with new functionalities, as it has already been described elsewhere (Park et al., 2009).

A screening of various databases such as Onekp<sup>5</sup>, The European Nucleotide Archive<sup>6</sup>, phytometasyn<sup>7</sup> using CYP71AZ1 as target makes evidence that only two plant families harbored paralogous genes with significantly high enough homologies to belong to the same P450 subfamily. Intriguingly, no CYP71AZ could be identified in other plants producing FCs such as Rutaceae (citrus for example) or Moraceae (Ficus for example). Although we could not test all the plants belonging to these plant species because of the absence of available databases, it might be possible that the synthesis of these molecules in other plants is realized by other genes and is in favor of a convergent evolution as suggested previously (Berenbaum, 1983; Bourgaud et al., 2014; Munakata et al., 2016).

All the CYP71AZ paralogous genes, highlighted in our transcriptomic screening, were identified in Apiaceae except CYP71AZ10, which was identified in *C. racemosa*. Interestingly, Moon et al. (2011) shows that this plant is able to produce isoimperatorin. An additional search done through the Dictionary of Natural Products<sup>8</sup> also showed the presence some molecules which are structurally related to FC such as (norvisnagin: 4-hydroxy-7-methylfuro[3,2-g]chromen-5-one (chromene) and angelicin: 4-hydroxy-7-(hydroxymethyl)-2-(2-hydroxypropan-2-yl)-2,3-dihydrofuro[3,2-g]chromen-5-one). It might be possible that CYP71AZ10 could be involved

<sup>5</sup><http://onekp.com>

<sup>6</sup><https://www.ebi.ac.uk/ena>

<sup>7</sup><https://bioinformatics.tugraz.at/phytometasyn>

<sup>8</sup><http://dnp.chemnetbase.com>

in the synthesis of these molecules. A deeper analysis of new transcriptomic data will probably help, in the near future to get more tracks for identifying the function of this enzyme or other new paralogues of CYP71AZ.

To identify the function of other CYP71AZ, the emergence and publication of many metabolomics studies will help to identify potential substrates. We show that CYP71AZ4 was able to metabolize a few coumarins. Among the molecule tested, only two of them could be identified in parsnip extracts (fraxetin and osthol, **Supplementary Figure S1**). These result makes evidence that this enzyme is obviously not highly specific. An extensive screening should be realized based on phytochemicals studies. For example, very few reports showed the presence of several coumarins in the roots of *T. garganica* and are present in very low amounts. Among these molecules, various authors described scopoletin and 6-methoxy-7-geranyloxycoumarin (Larsen and Sandberg, 1970; Rubal et al., 2007). CYP71AZ2 may be involved in the synthesis of this coumarin.

Plants produce a plethora of metabolites that are required for their growth and propagation as well as for regulatory needs and adaptation to environmental conditions. Among the 370,000 predicted vascular plants, only a small part has been studied at the chemical level (Lughadha et al., 2016). According to the Royal Botanical Kew Garden, 28,000 plants are useful for medicines and so far 300,000 specialized metabolites were identified throughout the plant diversity (Hubert et al., 2017). Here, we provide a new example of the importance of cytochrome P450s, which have already been identified to be deeply involved in this diversification (Nelson and Werck-Reichhart, 2011), and suggest how they can be involved in the appearance of new molecules to meet further challenges for an ecological function (Pichersky and Gang, 2000; Lewinsohn and Gijzen, 2009).

## AUTHOR CONTRIBUTIONS

CK and SR cloned the genes (CYP71AZ3, 4, 6, and AZ4mut1-4), expressed all the proteins in yeast and realized the kinetic characterization. SK-T did the *A. majus* cell cultures and identified and isolated CYP71AZ1. GG did the induction of parsnip and the RT-qPCR experiments for CYP71AZs and CYP71AZ3. BS did the NMR analyses. JG did the LC/manuscript analyses. AO did the biochemical analyses and wrote the manuscript. DR performed the modeling and Kpax experiments and edited the manuscript. UM led the work done on the *A. majus* cell cultures and edited the manuscript. FB and AH led the work done on the CYP71AZ characterization. AH led the whole project and wrote the manuscript.

## FUNDING

SR was funded by the SAARLORBIOTEC research project (INTERREG).

## ACKNOWLEDGMENTS

The authors thank Heiko Rischert and Henrik Toft Simonsen for providing the sequences from *T. garganica* and *C. maculatum*. They acknowledge Clément Charles for his technical support. The authors Dr. D. Nelson for the assignment of CYP71AZ subfamily.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.00820/full#supplementary-material>

**FIGURE S1** | Name, molecular weight, chemical structure, and code access number of the molecules used for the *in vitro* screening of CYP71AZs.

**FIGURE S2** | Identification of the products of 6-methoxycoumarin metabolism by CYP71AZ4. Yeast microsomes expressing CYP71AZ4 convert 6-methoxycoumarin to P04 and P05. **(A)** HPLC profile of enzyme assays containing 6-methoxycoumarin, without (line 1) or with (line 2) NADPH, oxygen, and yeast microsomes expressing CYP71AZ4. **(B)** MS and absorption spectrum of standard scopoletin. **(C)** MS and absorption spectrum of P04. **(D)** MS and absorption spectrum of P05.

**FIGURE S3** | Identification of the product of scopoletin metabolism by CYP71AZ4. Yeast microsomes expressing CYP71AZ4 convert scopoletin to P06. **(A)** HPLC profile of enzyme assays containing scopoletin, without (line 1) or with (line 2) NADPH, oxygen, and yeast microsomes expressing CYP71AZ4. **(B)** MS and absorption spectrum of standard Fraxetin. **(C)** MS and absorption spectrum of P06.

**FIGURE S4** | Identification of the products of 3-methyl-7-methoxycoumarin **(A)** and 7-methoxycoumarin **(B)** metabolism by CYP71AZ4. Yeast microsomes expressing CYP71AZ4 convert 3-methyl-7-methoxycoumarin to P07 and 7-methoxycoumarin to P08. **(A)** HPLC profile of enzyme assays containing 3-methyl-7-methoxycoumarin, NADPH, oxygen, and yeast microsomes expressing CYP71AZ4. **(B)** HPLC profile of enzyme assays containing 7-methoxycoumarin, without (line 1) or with (line 2) NADPH, oxygen, and yeast microsomes expressing CYP71AZ4. **(C)** MS and absorption spectrum of P07. **(D)** MS and absorption spectrum of standard daphnetin 7-methyl ether. **(E)** MS and absorption spectrum of P08

**FIGURE S5** |  $^1\text{H}$  NMR spectra (700 MHz, acetone- $d_6$ ) of a sample containing 8-hydroxy-3-methyl-7-methoxy coumarin (P07). **(A)** Full spectrum. **(B)** Extended region showing aromatic and double bond signals of P07 (marked with green arrows). **(C)** Extended region showing the methoxy signal of P07 (marked with green arrow).  $^1\text{H}$  NMR (700 MHz, acetone- $d_6$ ):  $\delta$  7.03 (1H,  $J = 8.2$  Hz, H-5), 6.98 (1H,  $d$ ,  $J = 8.2$  Hz, H-6), 7.65 (1H,  $s$ , H-4), 3.92 (1H,  $s$ , 7-OCH $_3$ ), signal of 3-CH $_3$  not detected due to overlap with matrix signals.  $^{13}\text{C}$  NMR (175 MHz, acetone- $d_6$ , signal extracted from HSQC spectrum):  $\delta$  56.6 (7-OCH $_3$ ) HSQC cross signals of other carbon atoms were not detected due to limited amount of material. The  $^1\text{H}$  NMR spectrum of P07 (700 MHz, acetone- $d_6$ ) displays doublets ( $J = 8.2$  Hz) of an AB spin system comprising H-5 ( $\delta$  7.03) and H-6 ( $\delta$  6.98) of the aromatic ring in this figure. The proximity of H-5 and H-6 was confirmed by a strong roof effect of the two  $^1\text{H}$  NMR signals and a correlation signal in the  $^1\text{H}$ ,  $^1\text{H}$  COSY spectrum. The singlet at  $\delta$  3.92 (3H) and an HSQC correlation signal  $\delta_{\text{H}}$  3.92/ $\delta_{\text{C}}$  56.6 are assignable to the methoxy group (7-OCH $_3$ ). The signal of the methyl group protons at C-3 remained obscured due to overlap with matrix signals. Since the presence of the H-5/H-6 AB spin system ruled out hydroxylation at C-5 and C-6, positions 4 and 8 remained possibilities for the new hydroxyl group. Thus, the singlet of the isolated aromatic proton resonating at  $\delta$  7.65 (1H) had to be assigned to either to H-4 or H-8. A proton at C-8, corresponding to hydroxylation at C-4, could be disregarded because an  $m$ -coupling with H-6 was missing and the chemical shift of H-8 would be expected near  $\delta$  6.5. The latter value is largely different from the observed chemical shift of  $\delta$  7.65, which is consistent with the

value expected for H-4. Finally, hydroxylation at the 2-methyl group could also be excluded because the  $^1\text{H}$  NMR signal of a prochiral hydroxymethylene group with an expected chemical shift of approximately  $\delta$  4.5 was absent from the spectrum. From these data, P07 was identified as 8-hydroxy-3-methyl-7-methoxycoumarin

**FIGURE S6** | Identification of the products of 7-methoxycoumarin metabolism by CYP71AZ1 (**A and D**) and CYP71AZ6 (**B and E**). Yeast microsomes expressing CYP71AZ1 and CYP71AZ6 convert 7-methoxycoumarin to P09 and P10, respectively. (**A**) HPLC profile of enzyme assays containing 7-methoxycoumarin, without (line 1) or with (line 2) NADPH, oxygen, and yeast microsomes expressing CYP71AZ1. (**B**) HPLC profile of enzyme assays containing 7-methoxycoumarin, without (line 1) or with (line 2) NADPH, oxygen, and yeast microsomes expressing CYP71AZ1. (**C**) MS and absorption spectrum of 5-hydroxy-7-methoxycoumarin. (**D**) MS and absorption spectrum of P09. (**E**) MS and absorption spectrum of P10.

**FIGURE S7** | Identification of the products of 6-methoxycoumarin metabolism by CYP71AZ1 (**A and B**) and CYP71AZ6 (**C and D**). Yeast microsomes expressing CYP71AZ1 and CYP71AZ6 convert 6-methoxycoumarin to P11 and P12, respectively. (**A**) HPLC profile of enzyme assays containing 6-methoxycoumarin, without (line 1) or with (line 2) NADPH, oxygen, and yeast microsomes expressing CYP71AZ1. (**B**) MS and absorption spectrum of P11. (**C**) HPLC profile of enzyme assays containing 6-methoxycoumarin, without (line 1) or with (line 2) NADPH, oxygen, and yeast microsomes expressing CYP71AZ6. (**D**) MS and absorption spectrum of P12.

**FIGURE S8** |  $^1\text{H}$  NMR spectra (700 MHz, MeOH- $d_4$ ) of a sample containing 6,7,8-trihydroxycoumarin (P13). (**A**) Full spectrum. (**B**) Extended region showing

signals of P13 (marked with green arrows).  $^1\text{H}$  NMR (700 MHz, MeOH- $d_4$ ):  $\delta$  7.86 (1H,  $J$  = 8.2 Hz, H-4), 6.23 (1H, d,  $J$  = 8.2 Hz, H-3), 6.82 (1H, s, H-5).  $^{13}\text{C}$  NMR (175 MHz, MeOH- $d_4$ , signals extracted from HSQC spectrum):  $\delta$  112.6 (C-4), 146.0 (C-5), 103.6 (C-6). The  $^1\text{H}$  NMR spectrum of P13 (700 MHz, MeOH- $d_4$ ) showed two doublets (at  $\delta$  7.86 and 6.23), which according to their chemical shifts and coupling constant of  $J$  = 8.2 Hz were assigned to the *Z*-configured double bond protons H-4 and H-5. A  $^1\text{H}$ ,  $^1\text{H}$ -COSY spectrum confirmed the AX spin system of these two protons. Hence, the new hydroxyl group cannot be located in the pyrone ring but must be attached to either position 5 or 8 of the aromatic ring. The signal of the remaining proton was detected at  $\delta$  7.68, and assignment to H-5 was achieved by comparing the chemical shift with reported data of 6,7,8-trihydroxycoumarin (Kayser and Kolodziej, 1995; Yasuda et al., 2006). Furthermore, the  $^{13}\text{C}$  chemical shifts of the protonated carbon atoms C-3, C-4, and C-5 were extracted from the HSQC spectrum of P13 (**Supplementary Figure S9**) and agreed with reported data (Kayser and Kolodziej, 1995; Yasuda et al., 2006) of 6,7,8-trihydroxycoumarin.

**FIGURE S9** | Partial HSQC NMR spectrum (700 MHz, MeOH- $d_4$ ) of a sample containing 6,7,8-trihydroxycoumarin (P13). Top:  $^1\text{H}$  NMR spectrum.  $^1\text{H}$  NMR signals and  $^1\text{H}/^{13}\text{C}$  correlation signals of P13 are marked with green arrows.

**FIGURE S10** | Sequences of the different CYP71AZs collected in databases.

**TABLE S1** | Sequence homologies between different CYP71AZs from *A. majus* (CYP71AZ1) and *P. sativa* (AZ3, AZ4, AZ5, and AZ6). Comparison of the full length sequences are in white squares. Comparison restricted to the six substrate recognition sites (SRS) are in gray squares.

## REFERENCES

- Benderoth, M., Textor, S., Windsor, A. J., Mitchell-Olds, T., Gershenzon, J., and Kroymann, J. (2006). Positive selection driving diversification in plant secondary metabolism. *Proc. Natl. Acad. Sci. U.S.A.* 103, 9118–9123. doi: 10.1073/pnas.0601738103
- Berenbaum, M. (1983). Coumarins and caterpillars: a case for evolution. *Evol. Int. J. Org. Evol.* 37, 163–179. doi: 10.1111/j.1558-5646.1983.tb05524.x
- Berenbaum, M. (2001). Chemical mediation of coevolution: phylogenetic evidence for Apiaceae and associates. *Ann. Mo. Bot. Gard.* 88, 45–59. doi: 10.2307/2666131
- Berenbaum, M. (2002). Postgenomic chemical ecology: from genetic code to ecological interactions. *J. Chem. Ecol.* 28, 873–896. doi: 10.1023/A:1015260931034
- Berenbaum, M., Nitao, J., and Zangerl, A. (1991). Adaptive significance of furanocoumarin diversity in *Pastinaca sativa* (Apiaceae). *J. Chem. Ecol.* 17, 207–215. doi: 10.1007/BF00994434
- Bourgaud, F., Olry, A., and Hehn, A. (2014). “Recent advances in molecular genetics of furanocoumarin synthesis in higher plants,” in *Recent Advances in Redox Active Plant and Microbial Products*, eds C. Jacob, G. Kirsch, A. Slusarenko, P. Winyard, and T. Burkholz (Dordrecht: Springer), 363–375. doi: 10.1007/978-94-017-8953-0\_14
- Dueholm, B., Krieger, C., Drew, D., Olry, A., Kamo, T., Taboureau, O., et al. (2015). Evolution of substrate recognition sites (SRSs) in cytochromes P450 from Apiaceae exemplified by the CYP71AJ subfamily. *BMC Evol. Biol.* 15:122. doi: 10.1186/s12862-015-0396-z
- Edger, P. P., Heidel-Fischer, H. M., Bekaert, M., Rota, J., Glöckner, G., Platts, A. E., et al. (2015). The butterfly plant arms-race escalated by gene and genome duplications. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8362–8366. doi: 10.1073/pnas.1503926112
- Fraenkel, G. S. (1959). Raison D'être of secondary plant substances. *Science* 129, 1466–1470. doi: 10.1126/science.129.3361.1466
- Funk, C., and Croteau, R. (1993). Induction and characterization of a Cytochrome-P-450-dependent camphor hydroxylase in tissue-cultures of common sage (*Salvia officinalis*). *Plant Physiol.* 101, 1231–1237. doi: 10.1104/pp.101.4.1231
- Ghoorah, A. W., Smail-Tabbone, M., Devignes, M.-D., and Ritchie, D. W. (2014). KBDock 2013: a spatial classification of 3D protein domain family interactions. *Nucleic Acids Res.* 42, 389–395. doi: 10.1093/nar/gkt1199
- Glawischig, E., Grun, S., Frey, M., and Gierl, A. (1999). Cytochrome P450 monooxygenases of DIBOA biosynthesis: specificity and conservation among grasses. *Phytochemistry* 50, 925–930. doi: 10.1016/S0031-9422(98)00318-5
- Hamerski, D., and Matern, U. (1988a). Biosynthesis of psoralens. Psoralen 5-monooxygenase activity from elicitor-treated *Ammi majus* cells. *FEBS Lett.* 239, 263–265.
- Hamerski, D., and Matern, U. (1988b). Elicitor-induced biosynthesis of psoralens in *Ammi majus* L. suspension cultures. Microsomal conversion of demethylsuberosin into (+) marmesin and psoralen. *Eur. J. Biochem.* 171, 369–375.
- Hotti, H., Seppanen-Laakso, T., Arvas, M., Teeri, T., and Rischer, H. (2015). Polyketide synthases from poison hemlock (*Conium maculatum* L.). *FEBS J.* 282, 4141–4156. doi: 10.1111/febs.13410
- Hubert, J., Nuzillard, J.-M., and Renault, J.-H. (2017). Dereplication strategies in natural product research: how many tools and methodologies behind the same concept? *Phytochem. Rev.* 16, 55–95. doi: 10.1007/s11101-015-9448-7
- Irmisch, S., McCormick, A. C., Guenther, J., Schmidt, A., Boeckler, G. A., Gershenzon, J., et al. (2014). Herbivore-induced poplar cytochrome P450 enzymes of the CYP71 family convert aldoximes to nitriles which repel a generalist caterpillar. *Plant J.* 80, 1095–1107. doi: 10.1111/tj.12711
- Jogesh, T., Stanley, M. C., and Berenbaum, M. R. (2015). Evolution of tolerance in an invasive weed after reassociation with its specialist herbivore. *J. Evol. Biol.* 27, 2334–2346. doi: 10.1111/jeb.12469
- Jonczyk, R., Schmidt, H., Osterrieder, A., Fiesselmann, A., Schullehner, K., Haslbeck, M., et al. (2008). Elucidation of the final reactions of DIMBOA-glucoside biosynthesis in maize: characterization of Bx6 and Bx7. *Plant Physiol.* 146, 1053–1063. doi: 10.1104/pp.107.11.1237
- Kawai, Y., Ono, E., and Mizutani, M. (2014). Expansion of specialized metabolism-related superfamily genes via whole genome duplications during angiosperm evolution. *Plant Biotechnol.* 31, 579–584. doi: 10.5511/plantbiotechnology.14.0901a
- Kayser, O., and Kolodziej, H. (1995). Highly oxygenated coumarins from *Pelargonium sidoides*. *Phytochemistry* 39, 1181–1185. doi: 10.1016/0031-9422(95)00166-5
- Kliebenstein, D. J., Lambrich, V. M., Reichelt, M., Gershenzon, J., and Mitchell-Olds, T. (2001). Gene duplication in the diversification of secondary metabolism: tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in arabidopsis. *Plant Cell* 13, 681–693. doi: 10.1105/tpc.13.3.681

- Kliebenstein, D. J., and Osbourn, A. (2012). Making new molecules - evolution of pathways for novel metabolites in plants. *Curr. Opin. Plant Biol.* 15, 415–423. doi: 10.1016/j.pbi.2012.05.005
- Larbat, R., Hehn, A., Hans, J., Schneider, S., Jugdé, H., Schneider, B., et al. (2009). Isolation and functional characterization of CYP71AJ4 encoding for the first P450 monooxygenase of angular furanocoumarin biosynthesis. *J. Biol. Chem.* 284, 4776–4785. doi: 10.1074/jbc.M807351200
- Larbat, R., Kellner, S., Specker, S., Hehn, A., Gontier, E., Hans, J., et al. (2007). Molecular cloning and functional characterization of psoralen synthase, the first committed monooxygenase of furanocoumarin biosynthesis. *J. Biol. Chem.* 282, 542–554. doi: 10.1074/jbc.M604762200
- Larsen, P. K., and Sandberg, F. (1970). Constituents of umbelliferous plants. XV. Coumarins from *Thapsia garganica* L. The structure of new coumarin. *Acta Chem. Scand.* 24:1113. doi: 10.3891/acta.chem.scand.24-1113
- Lewinsohn, E., and Gijzen, M. (2009). Phytochemical diversity: the sounds of silent metabolism. *Plant Sci.* 176, 161–169. doi: 10.1016/j.plantsci.2008.09.018
- Li, W., Zangerl, A. R., Schuler, M. A., and Berenbaum, M. R. (2004). Characterization and evolution of furanocoumarin-inducible cytochrome P450s in the parsnip webworm, *Depressaria pastinacella*. *Insect Mol. Biol.* 13, 603–613. doi: 10.1111/j.0962-1075.2004.00518.x
- Lughadha, E. N., Govaerts, R., Belyaeva, I., Black, N., Lindon, H., Allkin, R., et al. (2016). Counting counts: revised estimates of numbers of accepted species of flowering plants, seed plants, vascular plants and land plants with a review of other recent estimates. *Phytotaxa* 272, 82–88. doi: 10.11646/phytotaxa.272.1.5
- 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
- Mizutani, M., and Sato, F. (2012). Unusual P450 reactions in plant secondary metabolism. *Arch. Biochem. Biophys.* 507, 194–203. doi: 10.1016/j.abb.2010.09.026
- Moon, L., Ha, Y. M., Jang, H. J., Kim, H. S., Jun, M. S., Kim, Y. M., et al. (2011). Isoimperatorin, cimicidine E and 23-O-acetylshengmanol-3-xyloside from *Cimicifugae* rhizome inhibit TNF-alpha-induced VCAM-1 expression in human endothelial cells: involvement of PPAR-gamma upregulation and PI3K, ERK1/2, and PKC signal pathways. *J. Ethnopharmacol.* 133, 336–344. doi: 10.1016/j.jep.2010.10.004
- Munakata, R., Olry, A., Karamat, F., Courdavault, V., Sugiyama, A., Date, Y., et al. (2016). Molecular evolution of parsnip membrane-bound prenyltransferases for linear and/or angular furanocoumarin biosynthesis. *New Phytol.* 211, 332–344. doi: 10.1111/nph.13899
- Nelson, D., and Werck-Reichhart, D. (2011). A P450-centric view of plant evolution. *Plant J.* 66, 194–211. doi: 10.1111/j.1365-313X.2011.04529.x
- Nelson, D. R., Schuler, M. A., Paquette, S. M., Werck-Reichhart, D., and Bak, S. (2004). Comparative genomics of rice and *Arabidopsis*. Analysis of 727 cytochrome P450 genes and pseudogenes from a monocot and a dicot. *Plant Physiol.* 135, 756–772. doi: 10.1104/pp.104.039826
- Park, S.-H., Park, H.-Y., Sohng, J. K., Lee, H. C., Liou, K., Yoon, Y. J., et al. (2009). Expanding substrate specificity of GT-B fold glycosyltransferase via domain swapping and high-throughput screening. *Biotechnol. Bioeng.* 102, 988–994. doi: 10.1002/bit.22150
- 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
- Pompon, D., Louerat, B., Bronnie, A., and Urban, P. (1996). Yeast expression of Animal and Plant P450s in optimized redox environments. *Methods Enzymol.* 272, 51–64. doi: 10.1016/S0076-6879(96)72008-6
- Ritchie, D., Ghoorah, A., Mavridis, L., and Venkatraman, V. (2012). Fast protein structure alignment using gaussian overlap scoring of backbone peptide fragment similarity. *Bioinformatics* 28, 3278–3281. doi: 10.1093/bioinformatics/bts618
- Roselli, S., Olry, A., Vautrin, S., Coriton, O., Ritchie, D., Navrot, N., et al. (2016). Genomic approach to decipher specialized plant pathways: a case study with furanocoumarins in parsnip. *Plant J.* 89, 1119–1132. doi: 10.1111/tpj.13450
- Rubal, J. J., Moreno-Dorado, F. J., Guerra, F. M., Jorge, Z. D., Saouf, A., Akssira, M., et al. (2007). A pyran-2-one and four meroterpenoids from *Thapsia transtagana* and their implication in the biosynthesis of transtaganolides. *Phytochemistry* 68, 2480–2486. doi: 10.1016/j.phytochem.2007.06.023
- Rupasinghe, S., Baudry, J., and Schuler, M. A. (2003). Common active site architecture and binding strategy of four phenylpropanoid P450s from *Arabidopsis thaliana* as revealed by molecular modeling. *Protein Eng.* 16, 721–731. doi: 10.1093/protein/gzg094
- Sali, A., and Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779–815. doi: 10.1006/jmbi.1993.1626
- Werck-Reichhart, D., Hehn, A., and Didierjean, L. (2000). Cytochromes P450 for engineering herbicide tolerance. *Trends Plant Sci.* 5, 116–123. doi: 10.1016/S1360-1385(00)01567-3
- Werck-Reichhart, D., Schalk, M., Nedelkina, S., Batard, Y., and Durst, F. (1997). Modification of critical residues in the active site of a plant P450: CYP73A1. *FASEB J.* 11, A793–A793.
- Yasuda, T., Fukui, M., Nakazawa, T., Hoshikawa, A., and Ohsawa, K. (2006). Metabolic fate of fraxin administered orally to rats. *J. Nat. Prod.* 69, 755–757. doi: 10.1021/np0580412

**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 © 2018 Krieger, Roselli, Kellner-Thielmann, Galati, Schneider, Grosjean, Olry, Ritchie, Matern, Bourgaud and Hehn. 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 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.