

A new P450 involved in the furanocoumarin pathway underlies a recent case of convergent evolution

New Phytologist

Villard, Cloé; Munakata, Ryosuke; Kitajima, Sakihito; Velzen, Robin; Schranz, Eric M. et al

<https://doi.org/10.1111/nph.17458>

This publication is made publicly available in the institutional repository of Wageningen University and Research, under the terms of article 25fa of the Dutch Copyright Act, also known as the Amendment Taverne. This has been done with explicit consent by the author.








Article 25fa states that the author of a short scientific work funded either wholly or partially by Dutch public funds is entitled to make that work publicly available for no consideration following a reasonable period of time after the work was first published, provided that clear reference is made to the source of the first publication of the work.

This publication is distributed under The Association of Universities in the Netherlands (VSNU) 'Article 25fa implementation' project. In this project research outputs of researchers employed by Dutch Universities that comply with the legal requirements of Article 25fa of the Dutch Copyright Act are distributed online and free of cost or other barriers in institutional repositories. Research outputs are distributed six months after their first online publication in the original published version and with proper attribution to the source of the original publication.

You are permitted to download and use the publication for personal purposes. All rights remain with the author(s) and / or copyright owner(s) of this work. Any use of the publication or parts of it other than authorised under article 25fa of the Dutch Copyright act is prohibited. Wageningen University & Research and the author(s) of this publication shall not be held responsible or liable for any damages resulting from your (re)use of this publication.

For questions regarding the public availability of this publication please contact openscience.library@wur.nl

A new P450 involved in the furanocoumarin pathway underlies a recent case of convergent evolution

Cloé Villard¹ , Ryosuke Munakata² , Sakihito Kitajima^{3,4} , Robin van Velzen⁵ , Michael Eric Schranz⁵ , Romain Larbat¹  and Alain Hehn¹ 

¹LAE, Université de Lorraine-INRAE, Nancy 54000, France; ²Laboratory of Plant Gene Expression, Research Institute for Sustainable Humanosphere, Kyoto University, Uji, Kyoto 611-0011, Japan; ³Department of Applied Biology, Kyoto Institute of Technology, Matsugasaki Sakyo-ku, Kyoto 606-8585, Japan; ⁴The Center for Advanced Insect Research Promotion, Kyoto Institute of Technology, Matsugasaki Sakyo-ku, Kyoto 606-8585, Japan; ⁵Biosystematics Group, Wageningen University and Research Center, Wageningen 6708 PB, the Netherlands

Author for correspondence:

Alain Hehn

Email: alain.hehn@univ-lorraine.fr

Received: 1 February 2021

Accepted: 1 May 2021

New Phytologist (2021)

doi: 10.1111/nph.17458

Key words: convergent evolution, cytochrome P450, *Ficus carica* (fig tree), furanocoumarins, marmesin synthase.

Summary

- Furanocoumarins are phytoalexins often cited as an example to illustrate the arms race between plants and herbivorous insects. They are distributed in a limited number of phylogenetically distant plant lineages, but synthesized through a similar pathway, which raised the question of a unique or multiple emergence in higher plants.
- The furanocoumarin pathway was investigated in the fig tree (*Ficus carica*, Moraceae). Transcriptomic and metabolomic approaches led to the identification of CYP76F112, a cytochrome P450 catalyzing an original reaction. CYP76F112 emergence was inquired using phylogenetics combined with *in silico* modeling and site-directed mutagenesis.
- CYP76F112 was found to convert demethylsuberosin into marmesin with a very high affinity. This atypical cyclization reaction represents a key step within the polyphenol biosynthesis pathway. CYP76F112 evolutionary patterns suggests that the marmesin synthase activity appeared recently in the Moraceae family, through a lineage-specific expansion and diversification.
- The characterization of CYP76F112 as the first known marmesin synthase opens new prospects for the use of the furanocoumarin pathway. It also supports the multiple acquisition of furanocoumarin in angiosperms by convergent evolution, and opens new perspectives regarding the ability of cytochromes P450 to evolve new functions related to plant adaptation to their environment.

Introduction

Plants have evolved a wealth of chemical diversity which includes many lineage-specific metabolites that play important roles in their adaptation and interaction with their environment. Among such specialized metabolites, furanocoumarins (Bourgaud *et al.*, 2006, 2014) constitute an effective defense against various bioaggressors including pathogens, herbivores and other plants (Beier & Oertli, 1983; Schuler, 2011; Bourgaud *et al.*, 2014). Their distribution is restricted to phylogenetically distinct angiosperm lineages (The Angiosperm Phylogeny Group, 2016) including the Fabaceae, Moraceae, Rutaceae and Apiaceae (Pathak *et al.*, 1962; Bourgaud *et al.*, 2014; Munakata *et al.*, 2020; Fig. 1a), raising questions about the evolution of their biosynthetic pathways. The presence of similar compounds in distant lineages can be explained by two alternative hypotheses: they may have evolved only once but were subsequently massively lost, or they may have been independently acquired multiple times in distant plant families (i.e. convergent evolution).

The furanocoumarin biosynthesis pathway has been studied for decades. Historically, feeding experiments with radiolabeled precursors led to the identification of successive

metabolic intermediates, revealing a highly similar route across the investigated species (Brown & Steck, 1973; Murray *et al.*, 1982) (Fig. 1b). Molecular elucidation of the pathway started more recently with the identification of a bergaptol O-methyltransferase (BMT) that specifically converts bergaptol into bergapten (Hehmann *et al.*, 2004). Since this pioneering work, additional genes encoding dioxygenases, prenyltransferases (PTs), methyltransferases and cytochrome P450s (P450s) were characterized. Most of them were from apiaceous species (Hehmann *et al.*, 2004; Larbat *et al.*, 2007, 2009; Karamat *et al.*, 2014; Munakata *et al.*, 2016; Roselli *et al.*, 2017; Krieger *et al.*, 2018), and only three of them were isolated from Rutaceae and Moraceae (Vialart *et al.*, 2012; Limones-Mendez *et al.*, 2020; Munakata *et al.*, 2020) (Fig. 1b). Despite these steps forward, the molecular elucidation of the furanocoumarin pathway is still incomplete and its evolution history remains unclear. Thus far, the studies provided clues in favor of the multiple independent origins of the pathway. In particular, a recent study suggested that the PTs umbelliferone dimethylallyltransferases (UDTs) identified in Apiaceae and Moraceae evolved from different ancestral genes (Munakata *et al.*, 2020). Another report

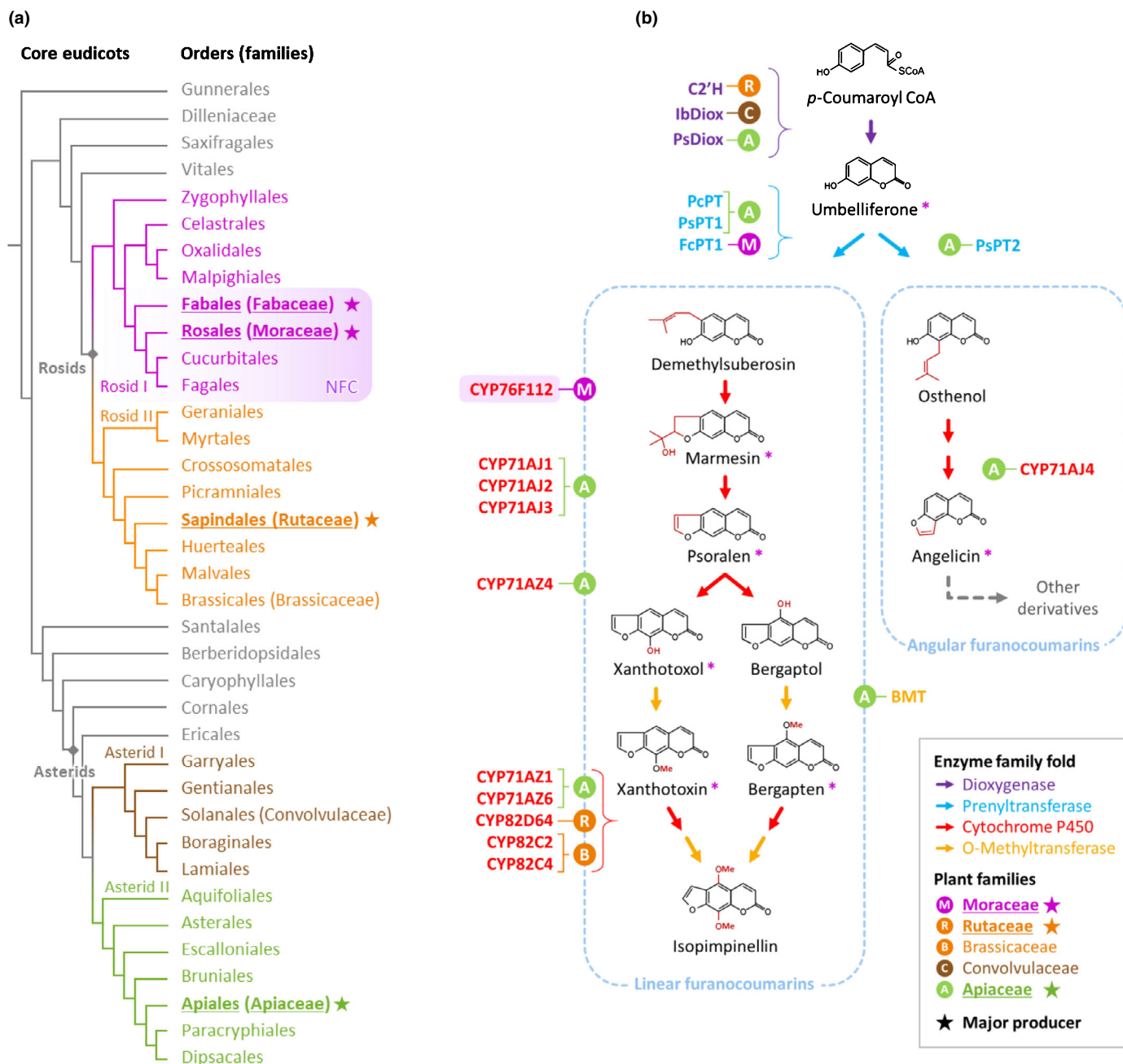


Fig. 1 Furanocoumarin production in core eudicots. (a) Phylogenetic relationship between the major furanocoumarin producers. The core eudicots phylogeny has been drawn according to the Angiosperm Phylogeny Group IV (2016). Unplaced orders are not shown. All families are not detailed. The families in which occur the major furanocoumarin producing species are highlighted with a star. The nitrogen fixing clade (NFC), containing the Fabales, Rosales, Cucurbitales and Fagales orders, is highlighted in light purple. (b) Simplified representation of the furanocoumarin biosynthesis core pathway from the plant widespread p -coumaroyl coenzyme A (CoA). The family fold of the enzymes mediating the different reaction steps are depicted by the color of the associated arrow. All enzymes described so far in the pathway are detailed, together with the associated plant families (Larbat *et al.*, 2007, 2009; Vialart *et al.*, 2012; Karamat *et al.*, 2014; Munakata *et al.*, 2016, 2020; Roselli *et al.*, 2017; Krieger *et al.*, 2018; Limones-Mendez *et al.*, 2020). The (furanocoumarins that have already been identified in *Ficus carica* (Villard *et al.*, 2019) are highlighted with a purple asterisk. It should be noted that angular furanocoumarins are trace compounds in the *Ficus* genus, and contrary to linear furanocoumarins, angelicin has only been described in *F. carica* fruits (Marrelli *et al.*, 2012). M, Moraceae; R, Rutaceae; B, Brassicaceae; C, Convolvulaceae; A, Apiaceae. The stars showed the plant families that are producing furanocoumarins.

demonstrated that the hydroxylation of xanthotoxin is catalyzed by P450s from two different families in parsnip (Apiaceae, CYP71) and grapefruit (Rutaceae, CYP82) (Limones-Mendez *et al.*, 2020).

In the late 1980s, Hamerski and Mattern found evidence that the marmesin synthase (MS) activity was catalyzed by a P450 in *Ammi majus* (Apiaceae) (Hamerski & Matern, 1988). To identify this still unknown enzyme and clarify its

emergence, we focused on P450s from *Ficus carica* (Moraceae), the common fig tree that produces large amounts of linear furanocoumarins (Villard *et al.*, 2019). Based on recent transcriptomic resources (Kitajima *et al.*, 2018), we identified and characterized CYP76F112 as a MS, which completes the furanocoumarin pathway until defensive psoralen and thus opens new prospects for future uses of this pathway. Comparative genomics and *in silico* modeling and docking analyses suggested that CYP76F112 as well as the MS activity appeared recently in the evolutionary history of Moraceae, which provides additional support for the independent acquisition hypothesis of furanocoumarins in higher plants.

Materials and Methods

Plant material and reagents

Seeds of *Ficus carica* were provided by Plant Advanced Technologies (Vandoeuvre-lès-Nancy, France). The *F. carica* plants were grown at room temperature, with natural light. Standard specimens of demethylsuberosin (DMS) and marmesin were respectively purchased from Topharman (Shanghai Co. Ltd, China) and Plant MetaChem (Gießen, Germany). Other phenolic substrates (Supporting Information Table S1) were purchased from Plant MetaChem and Extrasynthese (Lyon, France).

Identification of the P450 candidate genes

The coding sequences (CDS) were identified from a *F. carica* differential RNA-sequencing (RNA-Seq) library (Kitajima *et al.*, 2018). A tBLASTn search was performed using CYP71A/J3 (Larbat *et al.*, 2009) as query and an e-value of 1000. Resulting hits with >80 amino acids long and sharing at least 20% identity with the query were considered. Overlapping contigs were merged and sorted according to their expression level in *F. carica* petiole latex. The putative full CDS of candidates 1 and 2 were directly extracted from lengthy contigs of the *F. carica* RNA-Seq library. For candidate 3, the contigs were assembled and merged based on a homologous sequence from *Ficus religiosa* available on the OneKP (Carpenter *et al.*, 2019; One Thousand Plant Transcriptomes Initiative, 2019) database (<https://db.cngb.org/onekp/>) (Supporting Information Fig. S1).

Cloning of the CYP76F110-112 open reading frames

Total RNAs were extracted from frozen ground *F. carica* leaves according to the E.Z.N.A.[®] Plant RNA Kit protocol (Omega Bio-tek, Norcross, GA, USA). Complementary DNAs (cDNAs) were synthesized from 100 ng of extracted RNAs using the High Capacity RNA-to-cDNA[™] Kit (Applied Biosystems[™], Thermo Fisher Scientific, Waltham, MA, USA). CYP76F110-112 open reading frames (ORFs) were PCR-amplified using the PrimeSTAR[®] Max (Takara Bio Inc., Kusatsu, Japan). For subsequent Western blot analyses, we amplified the ORFs using reverse primers including a 6xHis tag extension at the 3' end

(underlined), right before the stop codon. The following pairs of primers are: 5'-ATGAAATATGTCCTTTGGAGTTGCAG-3' (forward) and 5'-CTAATGATGATGATGATGATGAGAAATTATCGCCACGGGAC-3' (reverse) for CYP76F110, 5'-ATGGATTTGATCACCTCTATATTGTGTTTTG-3' (forward) and 5'-CTAATGATGATGATGATGATGATGAGCGATTGCTATAGGGACAAC-3' (reverse) for CYP76F111, 5'-ATGGATATTTTCACCTCCTTACTGTATC-3' (forward) and 5'-CTAATGATGATGATGATGATGATGCTTTGTCGGCACGGG-3' (reverse) for CYP76F112. PCR conditions were 1 min at 98°C, 35 cycles (10 s at 98°C, 15 s at 55°C, and 20 s at 72°C), and a final 5 min at 72°C. The resulting PCR products were cloned using the pCR[™]8/GW/TOPO[®] TA Cloning Kit (Invitrogen[™], Thermo Fisher Scientific). Each plasmid was cloned and sequenced twice, from independent PCRs. The CDSs were subcloned into the yeast expression vector pYeDP60_GW[®] (Dueholm *et al.*, 2015) using the Gateway LR Clonase[™] II Enzyme Mix (Invitrogen[™], Thermo Fisher Scientific).

Synthesis of the CYP76F mutants

In order to have similar constructions, nucleotide sequences of the His-tagged CYP76F111, CYP76F112, and their associated mutants were synthesized and inserted into the yeast expression vector pYeDP60 (Urban *et al.*, 1990, 1997) by GeneCust (Boynes, France). The cloning was performed using the *Bam*HI and *Eco*RI restriction enzymes (Fig. S2).

Expression of the CYP76Fs in a yeast system

Recombinant pYeDP60_GW[®] and pYeDP60 plasmids containing the His-tagged P450s were used to transform the *Saccharomyces cerevisiae* strain WAT21 (Pompon *et al.*, 1996; Urban *et al.*, 1997). P450 heterologous expression and microsome preparation were conducted as described in Larbat *et al.* (2007) and Pompon *et al.* (1996). The expression of the P450s in the microsomal fraction was checked by performing Western-blots using 6xHistidine Epitope Tag antibodies (Acris, OriGene Technologies, Rockville, MD, USA) (Figs S3, S4). The concentration of functional P450s present in the microsomal solution were quantified using differential CO spectrum method described in Omura & Sato (1964a,b) (Fig. S5).

In vitro P450 assays and determination of the kinetic parameters

Fresh microsomal solutions were incubated into 0.1 M NaPi buffer (pH 7.0) containing 100 µM of substrate, in the presence or without 200 µM NADPH in a final volume of 100 µl. After 30 min at 27°C, under an agitation of 600 rpm, the reactions were stopped by the addition of 50 µl of acetonitrile/HCl (99:1). CYP76F112 optimal reaction temperature and pH were determined by incubating microsomes in the presence of 50 µM DMS and 200 µM NADPH for exactly

5 min, 600 rpm. To determine the optimal temperature, the NaPi buffer 0.1 M, pH 7.0 was used, and incubations were performed at temperatures ranging from 20 to 45°C. To determine the optimal pH, the incubation temperature was set to 27°C and incubations were performed using buffers with pH ranging from 4 to 10. All the incubations were performed in triplicates. CYP76F112 kinetic parameters were determined by incubating 0.54 pmol of functional P450 into 0.1 M NaPi buffer, pH 7.0 containing an excess of NADPH and 10 to 300 nM of DMS. Incubations were conducted for 1 min at 27°C, 600 rpm, and stopped by the addition of acetonitrile/HCl (99:1). Kinetic parameters of the mutants were determined following the same method, but with incubation duration, quantity of functional enzyme, and DMS range adapted to avoid the metabolization of more than 50% of the initial DMS. For every enzyme and DMS concentration, 5 to 15 incubations of 1 ml were performed, pooled together, and concentrated 50 to 150 times, which allowed the precise detection of DMS and marmesin. The marmesin present in the final solutions was quantified by converting the peak area from the MS data into a quantity of marmesin (mol). This was performed in duplicates. The specific activity of each CYP76F112 (wild-type or mutants) was calculated and expressed in mol of marmesin formed per minute per mol of functional P450 (Table S2). Finally, the SIGMAPLOT software (v.12; Systat Software Inc., San Jose, CA, USA) was used to plot a Michaelis–Menten model curve and calculate the kinetic parameters.

Identification and quantification of metabolites

Reaction products were analyzed using an ultra-high-performance liquid chromatography (UHPLC) Nexera2 (Shimadzu, Kyoto, Japan, <https://www.shimadzu.com/>), equipped with an ultraviolet (UV) detector SPD20A (Shimadzu) coupled to a simple quadrupole mass spectrometer LCMS2020 (Shimadzu). The UHPLC-mass spectrometry analyses were performed as described in Krieger *et al.* (2018), with the following differences: the mobile phase B consisted either of methanol with 0.1% formic acid, or of acetonitrile with 0.1% formic acid. Compounds were detected based on UV scans. Data were recorded and analyzed on the LABSOLUTION software (Shimadzu). Reaction products were identified by comparison of their retention time and *m/z* ratio with those of standard molecules. Additional analyses were performed by liquid chromatography (Thermo Vanquish, Thermo Scientific) coupled to a tandem mass spectrometer (Orbitrap-IX; Thermo Scientific). The column was a Phenomenex Kinetex XB-C18 (150 mm × 2.1 mm, 2.6 mm) reversed-phase column. The elution solvents consisted of ultrapure water with 0.1% formic acid (A) and methanol with 0.1% formic acid (B). The mobile gradient phase was as follows (A/B ; v/v): 10:90 at 0 min, 60:40 at 20 min, 65:35 at 32 min, 90:10 between 36 and 38.5, and 10:90 from 38.51 to 40 min. Data were recorded and analyzed on XCALIBUR™ software (v.2.1.SP1. Build1160; Thermo Fisher Scientific).

Phylogenetic analysis of the CYP76Fs in the nitrogen fixing clade

The CYP76F dataset was constituted by performing a similarity search in the genetic resources of 20 plant species (Fig. 2a) available in the public OneKP (Carpenter *et al.*, 2019; One Thousand Plant Transcriptomes Initiative, 2019) and GenBank (Benson *et al.*, 2012) databases. Genomic and transcriptomic resources were screened through BLASTN search, using *CYP76F112* as query. Resulting hits were kept if they were more than 100 nucleotides long and shared at least 55% identity with the query. Full length putative P450 genes and pseudogenes were identified and isolated from the resulting hits. The NETGENE2 Server (<http://www.cbs.dtu.dk/services/NetGene2/>) set on *Arabidopsis thaliana* helped to predict intron-exon borders (Hebsgaard, 1996). All identified putative CYP76F genes were included in the initial CYP76F dataset (Dataset S1). Putative pseudogenes containing premature stop codon(s) but no frameshifts were included in the dataset. Putative frameshift-containing pseudogenes, which could not be aligned well with other sequences to generate the subsequent partitioned files, were excluded. Additional sequences of furanocoumarin-committed P450s were added to the dataset, to constitute the outgroup of the following tree. The sequences were aligned with the GENEIOUS software (Geneious prime 2019), relying on a Translation Alignment based on MAFFT v.7.450, Auto algorithm, BLOSUM62 scoring matrix, Gap open penalty = 1.26, Offset value = 0.123. The alignment was checked and partitioned (CodonPos1 CodonPos2, CodonPos3) on MESQUITE v.3.6 (Maddison & Maddison, 2019). Bayesian inference trees were constructed using MRBAYES (Ronquist & Huelsenbeck, 2003; Ronquist *et al.*, 2012) (v.3.2.7) at CIPRES v.3.3 (Cipres Science Gateway). The Bayesian Markov Chain Monte Carlo analysis was set and tested locally, and ran on CIPRES using the following parameters: nst = mixed, rates = gamma, shape = (all), 30 million generations, two independent runs, four chains each, temperature heating 0.05, samples taken every 20 000 generations, burn-in time set at 7.5 million samples, partitioning: CodonPos1 CodonPos2, CodonPos3. The model convergence and model likelihood between runs were checked in TRACER v.1.7.1 (Rambaut *et al.*, 2018). Then, the tree was drawn and colored with FIGTREE v.1.4.4. The tree branches supported by posterior probabilities below 0.7 were considered as weak, above 0.9 as strong. This first gene tree was used to reduce the initial CYP76F dataset to a smaller subset of sequences (Dataset S1), that were realigned and used to generate a second tree, as described earlier.

Homology modeling and docking experiments

CYP76F three-dimensional (3D) homology models were generated on the basis of CYP76AH1 crystal structure (Gu *et al.*, 2019) (PDB ID: 5YLW), using the SWISS-MODEL protein homology modeling server (Guex *et al.*, 2009; Benkert *et al.*, 2011; Biasini *et al.*, 2014; Bertoni *et al.*, 2017; Bienert *et al.*, 2017;

(a) Nitrogen fixing clade

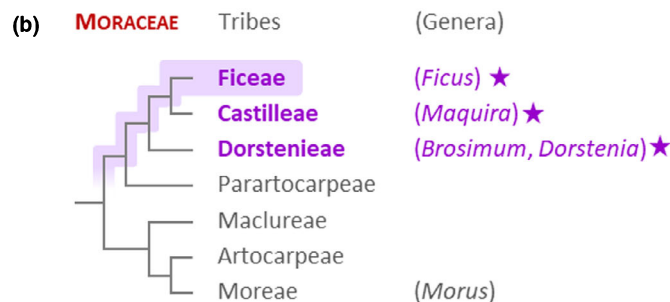
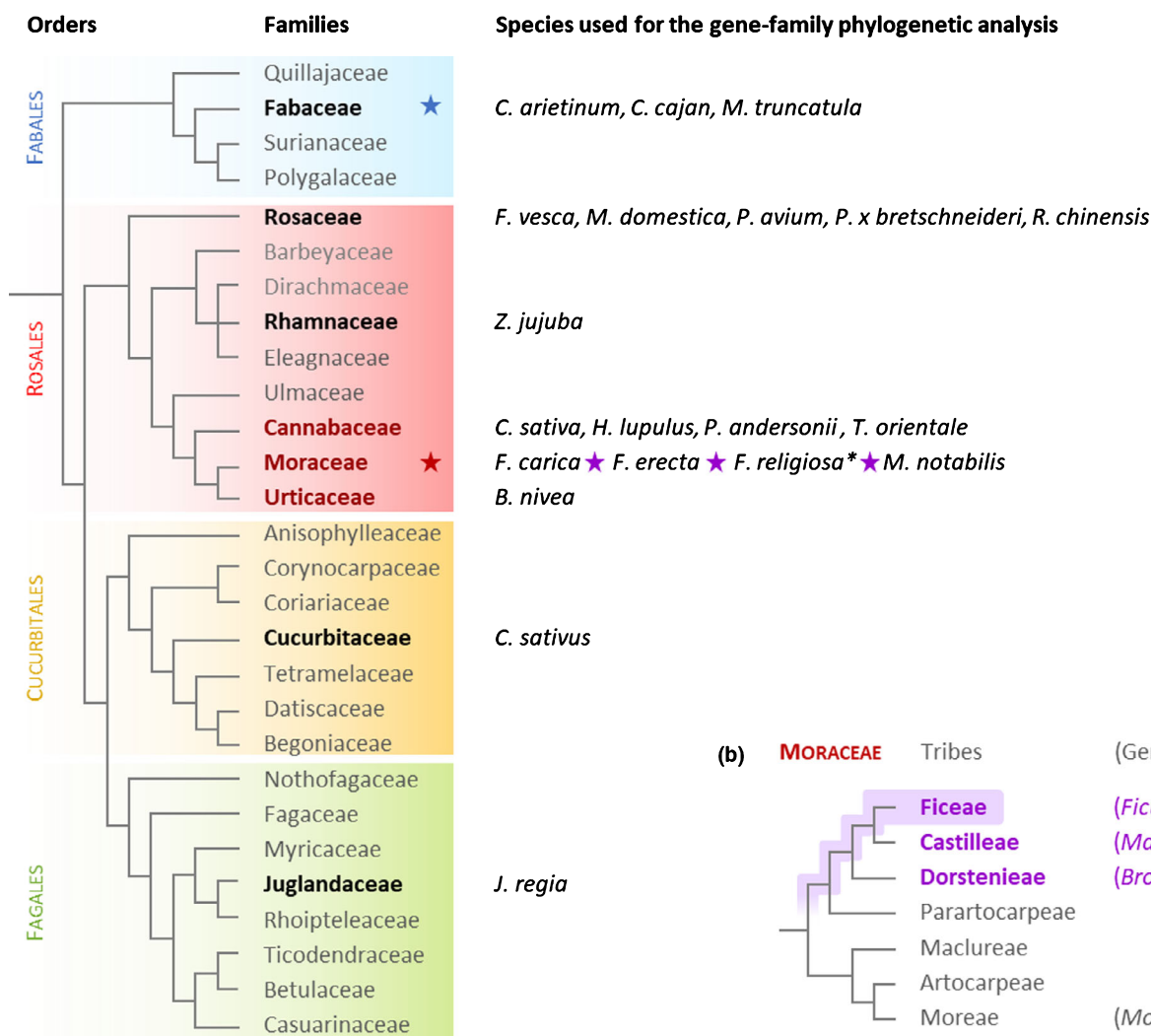


Fig. 2 Furanocoumarins repartition within the nitrogen fixing clade (NFC). (a) Species from the NFC included in the data mining. The phylogeny of the NFC has been drawn according to the Angiosperm Phylogeny Group IV (2016). The families and species in which furanocoumarins have already been described are marked with a star. In the families highlighted in bold red, a similarity search has been performed in all currently publicly available genetic resources. For the families in bold black, the search was arbitrarily restricted to a few entirely sequenced species. The families in gray were not included in the similarity search, either because no genetic resource was available, or because of their low interest (distant from the Moraceae). The species used in the similarity search are detailed. For all of them, except *Ficus religiosa* (*), (almost) complete genomes were available (scaffold to chromosome assembly level). The complete names of these species, from top to bottom, are as follows: *Cicer arietinum*, *Cajanus cajan*, *Medicago truncatula*, *Fragaria vesca* subsp. *Vesca*, *Malus domestica*, *Prunus avium*, *Pyrus x bretschnideri*, *Rosa chinensis*, *Ziziphus jujuba*, *Cannabis sativa*, *Humulus lupulus*, *Parasponia andersonii*, *Trema orientale*, *Ficus carica*, *Ficus erecta*, *Ficus religiosa*, *Morus notabilis*, *Boehmeria nivea*, *Cucumis sativus*, *Juglans regia*. (b) Putative emergence of the marmesin synthase (MS) activity in moraceous CYP76Fs. The phylogeny of the Moraceae family was adapted from Clement & Weiblen (2009) and Zerega & Gardner (2019). The tribes in which furanocoumarin producing species have already been described are in purple, the associated genera (*Ficus*, *Maquira*, *Brosimum* and *Dorstenia*) are detailed and marked with a star (Rovinski & Sneden, 1984; Vieira *et al.*, 1999; Abegaz *et al.*, 2004). Data related to the Parartocarpeae tribe are scarce; these plants may or may not contain furanocoumarins. The putative emergence of furanocoumarins and the MS activity during the evolution of the Moraceae family is highlighted with a blue path.

Waterhouse *et al.*, 2018; Studer *et al.*, 2020) (<https://swissmodel.expasy.org/>, SMTL, last update: 24 October 2019, last included PDB release: 1 October 2019). CYP76AH1 N-terminal transmembrane domain was not included in the template. Model quality was assessed using the GMQE and QMEAN scoring functions. Ramachandran plots and local quality estimate were also checked. The 3D homology models were visualized on

PyMOL (v.2.3.3, Schrödinger, LLC). Three-dimensional models of the heme (CID: 4971) and the DMS (CID: 5316525) were downloaded on PubChem (<https://pubchem.ncbi.nlm.nih.gov/>). Docking experiments were performed using AUTODOCK VINA v.1.1.2 (Trott & Olson, 2009). Unless otherwise specified, standard settings were used. To dock the heme into the CYP76Fs, the receptor was defined as rigid, except for the P450 absolutely

conserved cysteine which was treated as flexible. The receptor grid was defined as a $16 \text{ \AA} \times 16 \text{ \AA} \times 16 \text{ \AA}$ cube around the conserved cysteine. DMS was docked into the CYP76Fs in complex with its heme. The receptor was treated as rigid, except for a set of 15 flexible amino acids listed in Table S3(a). The receptor grid was defined as a $45 \text{ \AA} \times 35 \text{ \AA} \times 35 \text{ \AA}$ block around the heme. In both cases, the resulting models were ranked according to their affinity. The models described in this study correspond to the most probable binding mode (lowest affinity). Other models were also checked, they are presented in Tables S3(b)–S5. Finally, PyMOL was used to visualize the different models, compare them, measure distances, merge the receptor and the ligand, and to prepare the figures.

Results

Screening of the *F. carica* RNA-Seq library

Kitajima *et al.* recently reported the construction of a *F. carica* RNA-Seq library prepared from the latexes of three organs (petioles, trunks, fruits) producing different quantities of furanocoumarins (Kitajima *et al.*, 2018). Linear furanocoumarin concentrations were highest in petioles compared to trunk and fruit latexes (Kitajima *et al.*, 2018), thus we assumed that the corresponding biosynthetic genes were preferentially expressed in petiole latex. This was supported by the differential expression patterns of *FcPT1*, the *F. carica* UDT converting umbelliferone into DMS (Kitajima *et al.*, 2018; Munakata *et al.*, 2020). Besides, the MS was hypothesized to belong to the P450s enzyme superfamily (Hamerski & Matern, 1988). All P450s described to date in the furanocoumarin pathway belong to the CYP71AJ, CYP71AZ, CYP82C and CYP82D subfamilies (Fig. 1b) which are part of the CYP71 clan. We therefore targeted the entire CYP71 clan, that is known to be the main clan involved in the phenylpropanoid pathway (Nelson & Werck-Reichhart, 2011; Hamberger & Bak, 2013).

To identify the MS, we performed a tBLASTn search on the *F. carica* RNA-Seq library (Kitajima *et al.*, 2018) using as query *CYP71AJ3*, the psoralen synthase isolated from parsnip (Larbat *et al.*, 2009). A total of 62 contigs (Fig. S6) of various length were identified, assembled, and sorted by expression level. The sequences preferentially expressed in fruits or in trunks latexes were eliminated. This led to the identification of nine putative genes belonging to the CYP71 clan: two putative genes were part of the CYP71 family (CYP71A and CYP71B), three of the CYP76F subfamily, three of the CYP81 family (CYP81B, CYP81BN and CYP81CA) and one of the CYP82J subfamily (Table S6). Interestingly, none belonged to the subfamilies previously described in the furanocoumarin pathway (Fig. 1b). All nine putative genes were investigated in preliminary experiments: seven were successfully cloned and expressed, but only one CYP76F candidate was found to metabolize a molecule belonging to the core furanocoumarin pathway (Fig. 1b). Consequently, the CYP71, CYP81 and CYP82 putative genes were not further investigated, and the present article will exclusively focus on the CYP76 candidates.

Identification and cloning of CYP76F110–112

The three CYP76F candidate genes (candidates 1–3) were preferentially expressed (≈ 80 – 90%) in petiole latex (Fig. S1; Table S6). For candidates 1 and 2, putative full CDSs were directly isolated from contigs of the *F. carica* RNA-Seq library. For candidate 3, the reads did not allow us to construct a full length CDS. This sequence was therefore obtained by assembling partial contigs from the *F. carica* library, using a homologous contig identified in the *F. religiosa* genetic resources available in the OneKP database (Carpenter *et al.*, 2019; One Thousand Plant Transcriptomes Initiative, 2019) as a scaffold (Table S6). The predicted CDSs were experimentally confirmed through a reverse transcription polymerase chain reaction (RT-PCR) amplification and subsequent sequencing of *F. carica* messenger RNAs (mRNAs). The resulting CYP76 ORFs were classified into the CYP76F subfamily as CYP76F110–112 (P450 nomenclature committee, <https://drnelson.uthsc.edu/CytochromeP450.html>).

The three CDSs share ≈ 70 – 75% nucleotide identity and were mapped on the recently published *F. carica* genome (Usai *et al.*, 2020). This analysis revealed that *CYP76F110* is located on chromosome 2 while *CYP76F111* and *CYP76F112* are clustered on chromosome 11 (Fig. S7). The genomic analysis showed that the three genes have an intron of *c.* 250–1800 bp, framed by two exons of *c.* 900 and *c.* 600 bp (Fig. S7). The CDSs cloned from our plant samples share more than 99% identity with their corresponding genomic sequences (Usai *et al.*, 2020), but differ by a few mutations. In particular, the genomic equivalents of *CYP76F111* and *CYP76F112* both have frameshift mutations that introduce premature stop codons (Fig. S7). These frameshift mutations likely come from sequencing errors (i.e. number of T repeats for *CYP76F112*) in the genomic resource. For the rest of this work, we therefore used our CDSs, which were isolated and sequenced twice from independent RT-PCR amplifications.

CYP76F112 encodes a stable, specific and efficient marmesin synthase

To determine their activity, CYP76F110–112 were heterologously expressed in yeast (Fig. S3). The microsomal fractions were used to perform *in vitro* enzyme assays in the presence of NADPH and various putative substrates. A total of 42 putative substrates including coumarins and furanocoumarins were tested (Table S1). For CYP76F110 and CYP76F111, no product was detected. Yet, CYP76F112 could convert DMS into a unique product that was formally identified as marmesin by comparison of its retention time and tandem mass spectrometry spectrum with those of standard marmesin (Fig. 3). The synthesis of this molecule was NADPH-dependent (Fig. 3). CYP76F112 did not convert any of the other 41 tested molecules (Table S1).

Enzymatic characterization showed that CYP76F112 maximal activity was obtained at pH 7 and 27°C (Fig. S8), which is consistent with the optimal conditions described for other furanocoumarin-committed P450s (Larbat *et al.*, 2007, 2009). The apparent K_m value associated to the conversion of DMS into marmesin by CYP76F112 was determined to be $32.2 \pm 3.9 \text{ nM}$,

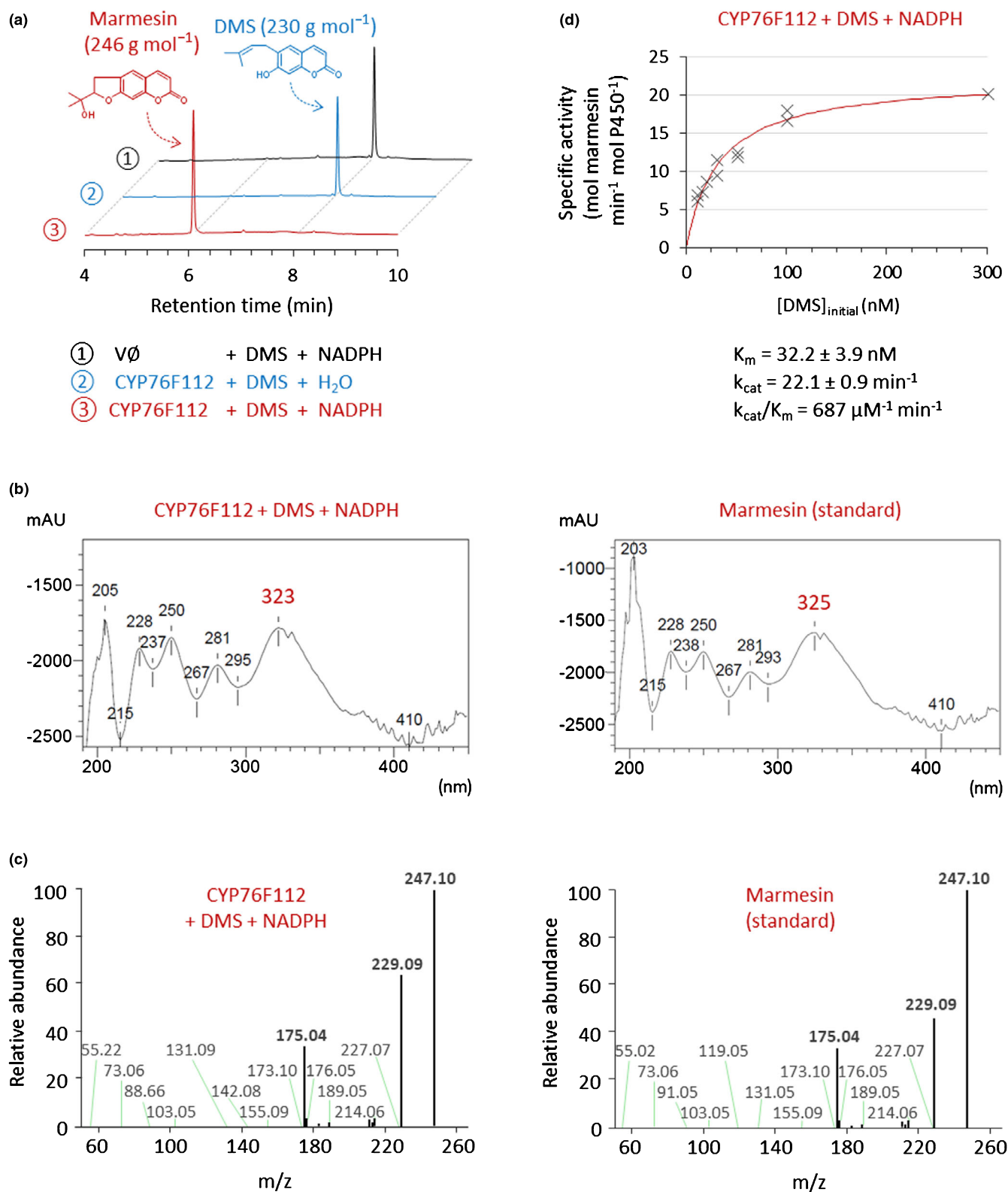


Fig. 3 Conversion of demethylsuberosin (DMS) into marmesin mediated by CYP76F112 from *Ficus carica*. (a) Ultra-high-performance liquid chromatography (UHPLC) separation profiles. Microsomes collected from yeast producing CYP76F112 or not (Vø, negative control) were incubated in the presence of DMS, with or without NADPH. The reaction products were monitored at 320 nm. The retention times of 6 and 8 min respectively correspond to marmesin and DMS. (b, c) Ultraviolet (UV) spectrum and tandem mass spectrometry fragmentation pattern of CYP76F112 reaction product, in comparison with a marmesin standard. (d) CYP76F112 specific activity, in the presence of various DMS concentrations. Experimental points are represented by black crosses. The red Michaelis–Menten model curve has been plotted on SIGMAPLOT to fit the experimental data. The kinetic parameters associated to this reaction were determined in SIGMAPLOT.

which is low for a P450. The apparent k_{cat} associated to this reaction was determined to be $22.1 \pm 0.9 \text{ min}^{-1}$, for a catalytic efficiency k_{cat}/K_m of $11.5 \mu\text{M}^{-1} \text{ s}^{-1}$ (Fig. 3; Table S2a).

CYP76F112 results from a recent Moraceae-specific expansion of CYP76Fs

To further investigate the evolution of the MS, a gene phylogenetic analysis of the CYP76Fs was performed. As *CYP76F112* was identified in *F. carica* (Moraceae, Rosales) the plants most likely to possess *CYP76F112* homologs were *F. carica* closest relatives, i.e. species from the Rosales and the neighboring Cucurbitales, Fagales and Fabales orders, which are collectively known as the nitrogen fixing clade or NFC (Soltis *et al.*, 1995; The Angiosperm Phylogeny Group, 2016) (Figs 1a, 2a). It should be noted that the Fabales order contains furanocoumarin-producing species such as *Psoralea cinerea* (Fabaceae). If the furanocoumarin pathway had emerged only once in higher plants, it may thus have been possible to trace the MS back to the root of the NFC.

CYP76F112 was used as query to perform an extensive similarity search of 20 species (including 19 genomes) from the NFC (Fig. 2a). This led to the identification of 134 putative CYP76F sequences corresponding to 123 genes and 11 pseudogenes (Dataset S1). *CYP76F110-112* and these putative CYP76Fs were used to generate a gene-family phylogenetic tree (Fig. S9), which resulting topology reflected the known NFC phylogeny with lineage-specific gene-family expansions (Fig. 2a). Using our gene-tree, we further investigated the well-supported *Ficus*-containing clade (Fig. S9), which showed that the CYP76F subfamily had dramatically expanded in the Moraceae family (Fig. 4). In particular, the analysis revealed that *CYP76F112* belongs to a lineage-specific clade exclusively composed of 21 *Ficus* sequences (the *Ficus*-clade, Fig. 4). The closest non-*Ficus* sequences belong to *Morus notabilis* (Moraceae) that does not produce furanocoumarins. The *Ficus*-clade can be subdivided into four subclades (subclades I–IV), with *CYP76F111* and *CYP76F112* being respectively included in subclades I (earliest branching) and IV (more recent branching). However, *CYP76F110* is not part of the *Ficus*-clade and had likely branched before the Moraceae-specific expansion (Fig. 4).

Genomic analysis showed that all 14 CYP76F sequences from *F. carica* (Usai *et al.*, 2020) were clustered on chromosome 11, except one sequence localized on chromosome 2. Similarly, the 16 CYP76Fs from *F. erecta* (Shirasawa *et al.*, 2019) were clustered on chromosomes 10 (14 sequences) and 9 (two sequences) (Fig. 4). The genomic organization of these sequences showed characteristic patterns (i.e. close clustering, tandem repetitions and head to tail orientation) that strongly suggests recent and repeated tandem duplications (Fig. 4). Besides, two sequences from *F. erecta* and *F. religiosa* (G01_76_F_erecta and G01_76_F_religiosa) share more than 98% identity with *CYP76F112*, which makes them putative orthologous MSs (Figs 4, S10).

The phylogenetic analysis thus revealed that the *Ficus* MS genes originated from a recent expansion and diversification of the CYP76Fs. Moreover, our data support that this taxon-

specific expansion likely occurred after the divergence of *Ficus* and *Morus* ancestors (Fig. 4).

In silico analysis identified four residues essential for the docking of demethylsuberosin in CYP76F112 active site

To obtain further insights into the evolution and acquisition of the MS activity, we modeled the active-site docking of CYP76F112 compared to CYP76F111. First, 3D homology models of CYP76F111 and CYP76F112 were generated on the basis of the CYP76AH1 crystal structure (Gu *et al.*, 2019), which permitted to establish a putative structural annotation and illustration (Fig. 5). Then, a DMS molecule was virtually docked into the heme-containing active site of both enzymes (Figs 6, S11; Table S3). In the case of CYP76F112, the DMS was docked in close vicinity to the heme, with its C2' and C3' carbons located at respectively 4.3 and 4.6 Å from the heme iron (Fig. 6). These distances are consistent with the reactive range described for other P450s (Larbat *et al.*, 2007; Fujiyama *et al.*, 2019), making the model conform to expectations for a MS. For CYP76F111, the DMS was docked in an upside-down orientation compared to CYP76F112, making the conversion into marmesin impossible (Fig. 6).

By comparing CYP76F111 and CYP76F112 active sites, we identified four residues that were essential for the positioning of the DMS within these enzymes. The respective equivalents of these four residues in CYP76F111/CYP76F112 are P105/T102, A108/S105, W120/M117 and P310/S305 (Fig. 7; Table S3a). In CYP76F112, the polar residues T102, S105, and S305 are close to the docked DMS (< 4 Å). More precisely, T102 and S105 are localized side by side and, during the docking, their side chains get reoriented toward the DMS ketone group (Figs 7, S12). S305 is localized near the other extremity of the DMS, and its side chain points toward the DMS hydroxy group (Fig. 7). These polar amino acids might establish hydrogen bonds with the ketone (T102, S105) and hydroxy group (S305) of the DMS. Contrarily, in CYP76F111, the equivalent amino acids P105, A108 and P310 are not polar. In CYP76F112, M117 is located right below the distal end of the DMS. Conversely, in CYP76F111, W120 extends slightly above the heme, reducing the substrate-binding pocket (Fig. 7).

Virtual mutants were generated by exchanging the amino acids P105/T102, A108/S105, W120/M117 and P310/S305 between CYP76F111 and CYP76F112 (Table S7). These mutants were modeled and used for new docking experiments. The results suggested that the simultaneous replacement of the four residues may impact DMS positioning within CYP76F111 and CYP76F112, possibly leading to the disappearance of the MS activity for CYP76F112-T102P, S105A, M117W, S305P and to its appearance for CYP76F111-P105T, A108S, W120 M, P310S (Fig. 7). Similarly, the docking results also suggested that the individual mutations of the residues M117 or S305 might be enough to impact the positioning of the DMS within CYP76F112-M117W and CYP76F112-S305P, while the residues T102 and S105 probably both have to be modified (CYP76F112-T102P, S105A) to impact it (Figs 7, S13; Table S4).

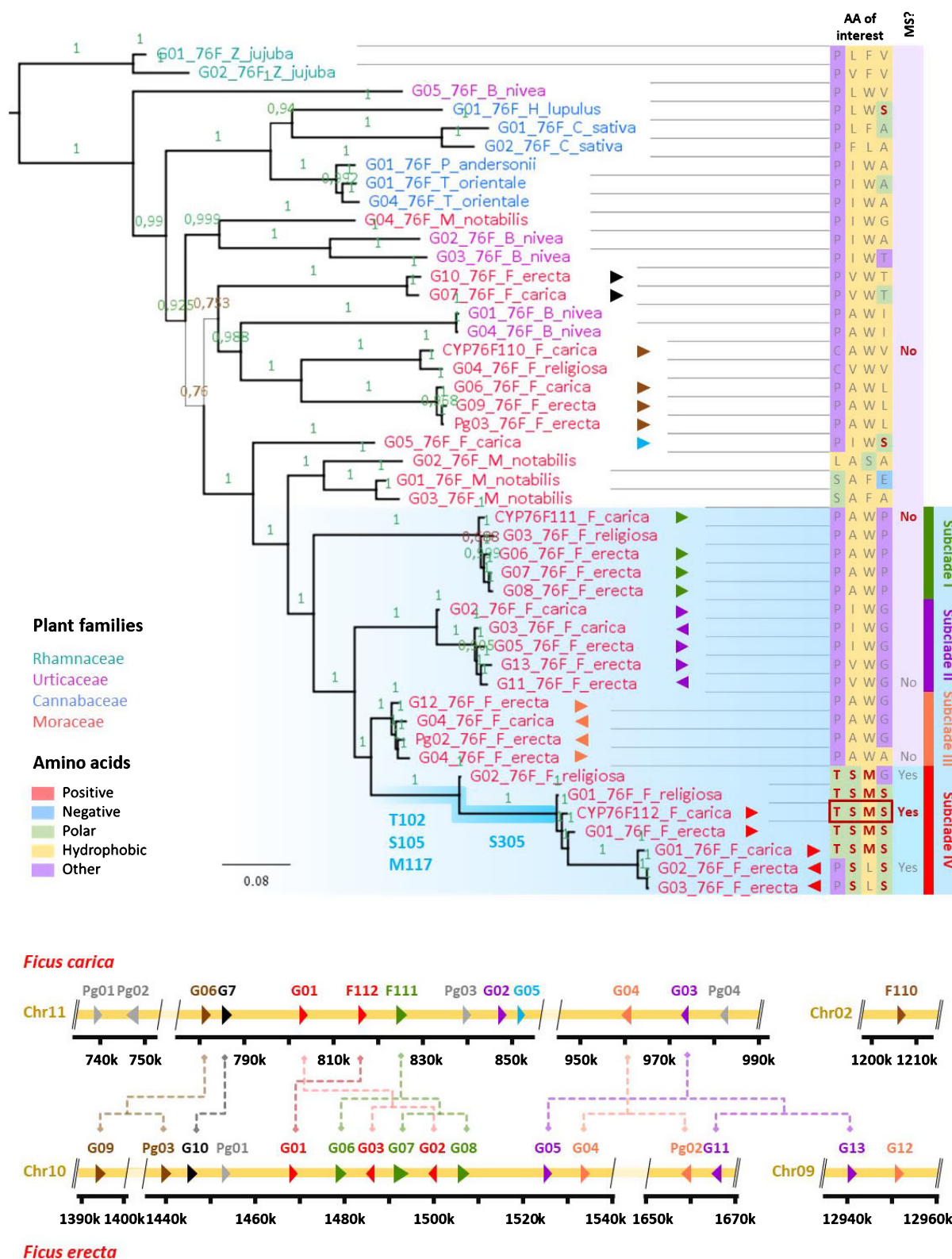


Fig. 4 Evolution of the *Ficus* CYP76Fs. (a) Gene-family phylogenetic tree of the CYP76Fs from the Cannabaceae, Moraceae and Urticaceae families. The alignment of full-length CYP76F nucleotide sequences was used to generate a Bayesian inference tree. The tree was rooted using CYP76Fs from the Rhamnaceae family. The *Ficus*-clade, which consists in a well-supported clade exclusively composed of *Ficus* sequences, is highlighted in light blue and is subdivided into four subclades (I–IV). The residues T102, S105, M117 and S305 are reported through the phylogeny. Those corresponding to CYP76F112 derived character state are highlighted in bold red, others are in gray. Putative marmesin synthases (MSs) are in light blue, others are in light purple. (b) Genomic organization of the *F. carica* (Usai *et al.*, 2020) and *F. erecta* (Shirasawa *et al.*, 2019) sequences along the chromosomes. The (pseudo)genes are depicted with triangular arrows which direction reflect the gene orientation. They are colored according to their subclades, and reported in the tree with matching arrows. For an easier understanding, *F. carica* and *F. erecta* closest relatives have been linked with dotted lines reflecting their phylogenetic relationship.

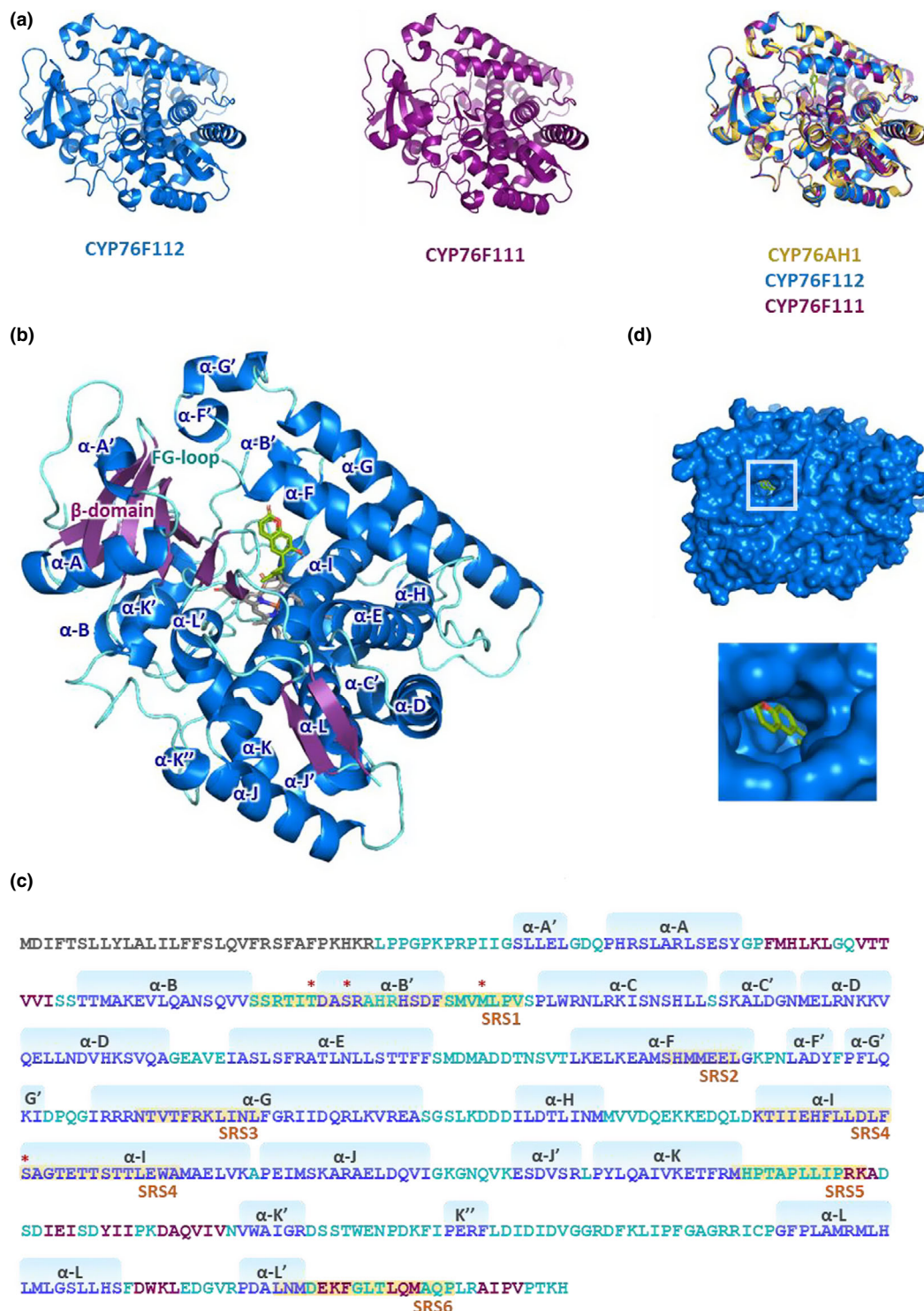


Fig. 5 Structural models of CYP76F111 and CYP76F112 from *Ficus carica*. (a) Three-dimensional (3D) homology models of CYP76F111 and CYP76F112, generated on the basis of CYP76AH1 (PDB ID: 5YLW). CYP76F111-112 3D models can both be superimposed almost perfectly with CYP76AH1. (b) Annotated 3D model of CYP76F112, complexed with an heme and a demethylsuberosin (DMS) molecule. CYP76F112 α -helices are in marine blue, the β -sheets in purple, and the loops are turquoise. The heme is in gray, the DMS is green. The α -helices and β -sheets have been annotated by analogy to CYP76AH1 annotated structure (Gu *et al.*, 2019). (c) CYP76F112 amino acid sequence, annotated like the 3D model. The beginning of the sequence, which was not modeled, is in gray. The substrates recognition sites (SRS1–SRS6) are highlighted in yellow/orange. The residues T102, S105, M117 and S305 are highlighted with red asterisks. (d) CYP76F112 3D model represented as surfaces (blue), complexed with a heme (not visible) and a DMS molecule (green). The enlargement shows the entrance of CYP76F112 wide access channel, located between the F/G loop and the β -domain, that opens onto a substrate-binding pocket in which the DMS is docked. The substrate-binding pocket is mostly made of hydrophobic residues, excepts for two polar regions that include the amino acids T102, S105, H110, S305 and T310.

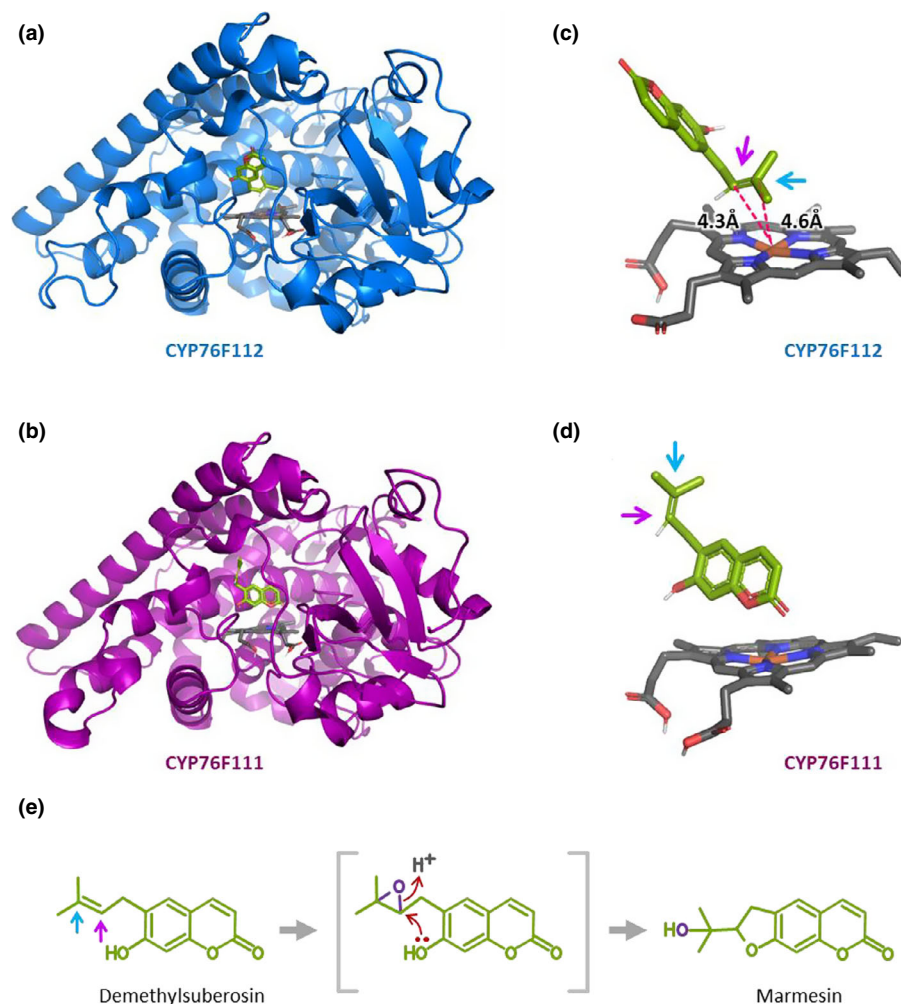


Fig. 6 Docking of the demethylsuberosin (DMS) within CYP76F111 and CYP76F112 from *Ficus carica*. (a, b) Three-dimensional (3D) models of CYP76F112 (blue) and CYP76F111 (purple) in complex with a heme (gray) and the DMS (green), reoriented to better show the heme and the DMS. (c, d) Enlargement on the heme and the DMS, as docked into CYP76F111-112 respective active sites. (e) Reactional mechanism proposed for the P450-mediated conversion of DMS into marmesin, adapted from Mizutani & Sato (2011). The carbons of the DMS that are involved in the reaction are C2' and C3', respectively highlighted with pink and blue arrows.

Site-directed mutagenesis confirmed the importance of T102, S105, M117 and S305 on CYP76F112 activity

To assess the robustness of the modeling and docking predictions, we synthesized seven single and multiple *CYP76F111* and *CYP76F112* mutant gene constructs. As for the wild-type, the corresponding proteins were expressed in yeast (Fig. S4) and used for functional characterizations (Figs 7, S5b–e; Table S2b–e). As predicted, the CYP76F112-T102P, S105A, M117W, S305P quadruple mutant but also the CYP76F112-M117W single mutant could not metabolize DMS, confirming that the M117W mutation alone is enough to disturb the positioning of the DMS. It might be deduced that replacing M117 by W creates a steric hindrance that pushes the DMS away from the heme, preventing MS activity (Fig. 7). The residue M117 therefore plays a critical role for CYP76F112 specificity, probably by shaping the substrate-binding pocket. The single mutants CYP76F112-T102P, CYP76F112-S105A and CYP76F112-S305P together with the double mutant CYP76F112-T102P, S105A retained MS activity but have modified kinetic parameters (Fig. 7). In particular, the mutations T102P, S105A and S305P significantly reduced the enzyme affinity ($\approx K_m \times 3$ for CYP76F112-T102P,

$\approx K_m \times 4$ for CYP76F112-T102P, S105A, and $\approx K_m \times 4$ for CYP76F112-S305P), although the variations in catalytic activity were less clear (Fig. 7). We hypothesize that CYP76F112 polar T102, S105 and S305 residues, by establishing hydrogen bonds with the DMS, help attract, position and stabilize the substrate in the 'correct' position above the heme iron. In this way, they increase the affinity and allow the enzyme to remain active with only a few nM of DMS, without being indispensable for the MS activity (Fig. 7). Finally, the results showed that the mutant CYP76F111-P105T, A108S, W120M, P310S did not exhibit MS activity. This implies that, even though the residues T102, S105, M117 and S305 play an essential role in CYP76F112 enzymatic activity, additional residues must also play essential roles, for instance in the folding of the transmembrane domain or the substrate access channel.

Recent evolution of T102, S105, M117 and S305 coincides with the emergence of CYP76F marmesin synthase activity

The residues in positions equivalent to that of CYP76F112 T102, S105, M117 and S305 were analyzed throughout the CYP76F phylogeny (Fig. 4). Our results showed that

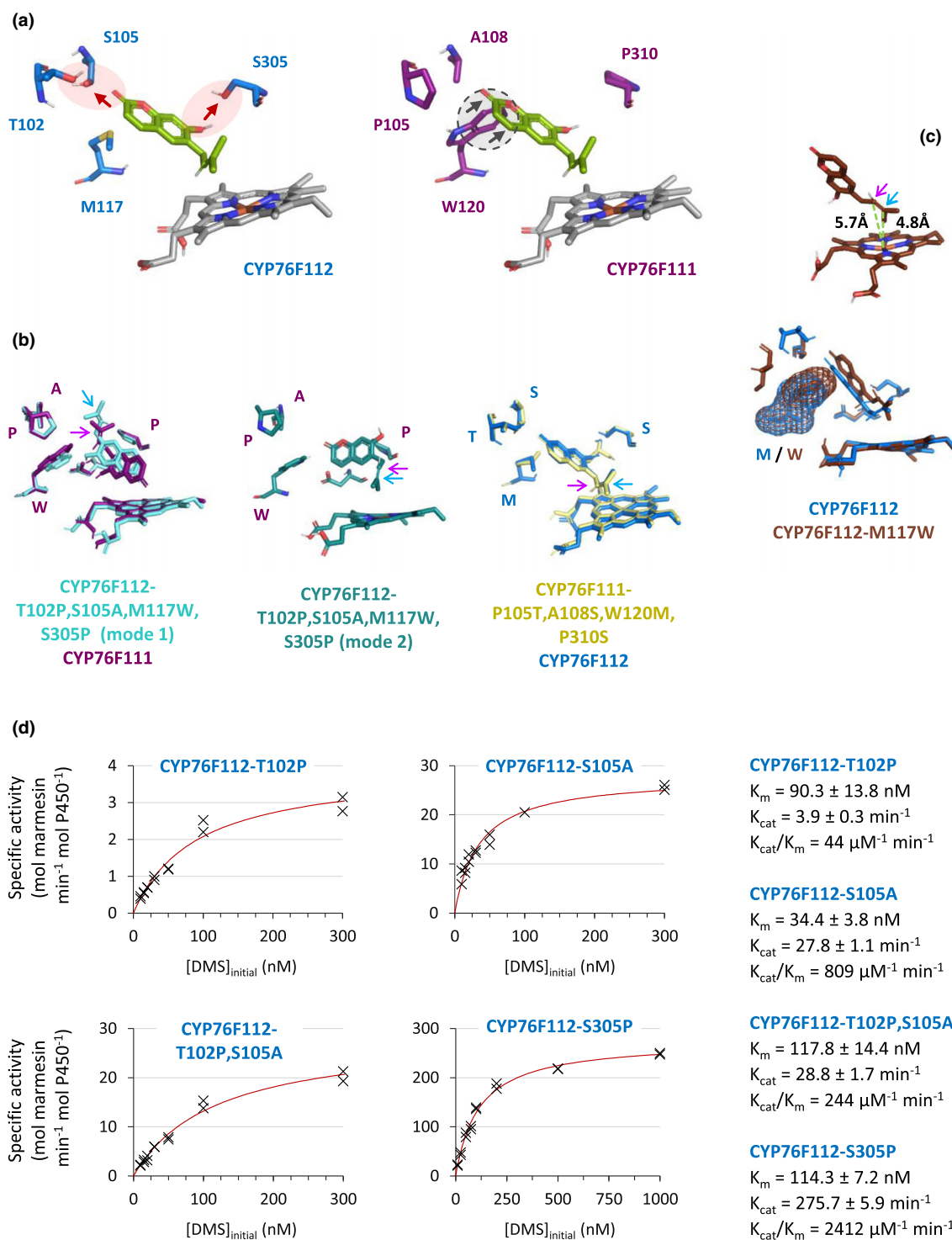


Fig. 7 Influence of the residues T102, S105, M117 and S305 on the docking of the demethylsuberosin (DMS) and the *Ficus carica* CYP76F112 marmesin synthase (MS) activity. (a) Position of the four residues of interest in CYP76F112 (blue) and CYP76F111 (purple). For an easier comparison, the heme and the DMS are shown as docked in CYP76F112. (b) Docking of the DMS within CYP76F112-T102P, S105A, M117W, S305P and CYP76F111-ABCD, in comparison to CYP76F112 and CYP76F111, respectively. The four amino acids of interest are detailed. The pink and blue arrows point toward the carbons C2' and C3' of the DMS. The position of the docked DMS into CYP76F112-T102P, S105A, M117W, S305P active site would not allow its conversion into marmesin. A, alanine; M, methionine; P, proline; S, serine; T, threonine; W, tryptophan. (c) Influence of the residue C. Docking of the DMS within CYP76F112-M117W, in comparison with CYP76F112. To better visualize their respective size, the amino acids M/W117 are shown as mesh. (d) Specific activity of the mutants CYP76F112-T102P, CYP76F112-S105A, CYP76F112-T102P, S105A and CYP76F112-S305P, in the presence of various initial DMS concentrations. Experimental points are represented by black crosses. The red Michaelis–Menten model curve has been plotted on SIGMAPLOT to fit the experimental data. The kinetic parameters associated to the reaction were determined in SIGMAPLOT, by using the Michaelis–Menten equation fitting the experimental data.

CYP76F112 amino acids do not correspond to an ancestral state, but rather recent synapomorphies that are shared only by subclade IV from the *Ficus*-specific clade (Fig. 4). Interestingly, the ancestral state of residue M117 was likely tryptophan, which was replaced by a methionine in subclade IV only. Given the importance of this residue, this observation suggests that no enzyme outside the subclade IV would be able to convert DMS into marmesin. To test this hypothesis, additional modeling and docking analyses was performed with four *F. erecta* and *F. religiosa* sequences from subclade IV and the neighboring II and III subclades (Figs 4, S14; Table S5). The results showed that the DMS docked within both enzymes from subclade IV was in the same position than within CYP76F112, suggesting that these enzymes may have MS activity. Whereas, the enzymes from subclades II and III, the DMS was docked in a position that would not allow its conversion into DMS (Figs 4, S14). Accordingly, we concluded that CYP76F MSs were restricted to subclade IV.

Finally, we can now conclude that *CYP76F112*, as well as its MS activity, evolved recently in the Moraceae family through a lineage-specific expansion and diversification of CYP76Fs. In particular, the acquisition of this activity would not have been possible without the evolution of the residue M117, and CYP76F112 remarkably high affinity is at least partially linked to the evolution of the residues T102, S105 and S305 (Figs 4, 7).

Discussion

P450s constitute an old enzyme superfamily that can be found in nearly all organisms. They typically catalyze monooxygenation/oxidation reactions, which makes them valuable tools to build complex molecules (Nelson & Werck-Reichhart, 2011; Nelson, 2018). They are involved in many complex metabolic networks, and might be considered as one of the dominant driving forces for phytochemical diversification and therefore in adaptation to fluctuating environmental constraints (Mizutani & Ohta, 2010; Nelson & Werck-Reichhart, 2011).

The identification of CYP76F112, a *F. carica* MS involved in the production of furanocoumarins, completes the set of four enzymes that allows the production of toxic psoralen from the plant wide spread *p*-coumaroyl coenzyme A (Fig. 1b). This opens many new prospects related to the study and use of the furanocoumarin biosynthesis pathway, since these bioactive metabolites can be seen as a two-face Janus provoking health issues (Pathak *et al.*, 1962; Bailey *et al.*, 2013; Villard *et al.*, 2019) because of the crop plants that produce them, and considered as promising drug candidates against diseases such as cancer and vitiligo (Chauthe *et al.*, 2015; Zabolinejad *et al.*, 2020). As furanocoumarins are natural defensive compounds, the identification of CYP76F112 may also be of particular interest for biocontrol and sustainable crop protection.

More broadly, CYP76F112 also provides new data and perspective about P450s. For instance, the reaction mediated by CYP76F112 is not a common hydroxylation but an atypical oxidation followed by a cyclization (Mizutani & Sato, 2011), which could be a subject for deeper enzymatic studies. Also, it is noteworthy that CYP76F112 was remarkably easy to produce,

resistant to freezing (> 1 yr of storage at -20°C) and highly efficient. Therefore, deeper structural analyses of CYP76F112 might help understanding P450 stability, which could then help engineering improved P450s involved in the production of molecules of interest.

But above all, the identification and evolutionary history of *CYP76F112* provides an exciting example of how a particular P450 family can expand, diversify and evolve new functions, leading to its recruitment in a new lineage-specific specialized metabolic pathway.

Thus far, all P450s described to be involved in furanocoumarin biosynthesis belong to the CYP71 and CYP82 families (Larbat *et al.*, 2007, 2009; Krieger *et al.*, 2018; Limones-Mendez *et al.*, 2020) (Fig. 1b). Interestingly, most CYP76s and all the CYP76Fs described to date have been associated with terpenoid metabolism (Diaz-Chavez *et al.*, 2013; Höfer *et al.*, 2014; Sintupachee *et al.*, 2015; Ilc *et al.*, 2017; Bathe & Tissier, 2019) (Fig. S15). CYP76F112 is therefore the first known CYP76F to metabolize a phenolic substrate. However, like other CYP76F substrates, DMS is a prenylated compound that is oxidized on its prenyl group (Diaz-Chavez *et al.*, 2013; Sintupachee *et al.*, 2015; Ilc *et al.*, 2017). DMS is also structurally similar to 7-ethoxycoumarin, a xenobiotic metabolized by CYP76B1 (Bataud *et al.*, 1998) (Fig. S15). The MS activity may thus have emerged from the CYP76F subfamily thanks to its ability to metabolize prenylated compounds, but additional experiments are required to support this hypothesis.

In previous works, CYP76s were described as an old and extensively diversified family (Nelson & Werck-Reichhart, 2011; Bathe & Tissier, 2019). Several reports highlighted that lineage-specific gene duplications and diversifications frequently occurred in the CYP76 family and subfamilies. Such expansions are commonly assumed to be associated with new lineage-specific adaptive functions (Wang *et al.*, 2012; Höfer *et al.*, 2014; Bathe & Tissier, 2019). Our results are in complete accordance with these previous studies since we showed that a Moraceae-specific expansion of the CYP76F subfamily occurred through multiple tandem-duplications, giving rise to the MS activity and therefore to the biosynthesis of defensive furanocoumarins in the *Ficus* lineage.

The genomic information we collected may also provide new leads regarding furanocoumarin biosynthesis. It is quite common for P450s belonging to the same subfamily to metabolize similar compounds or to catalyze successive reactional steps of a given pathway (Nelson & Werck-Reichhart, 2011). The physical clustering of homologous P450s, which seems to be a characteristic feature of recently tandem-duplicated P450s, often suggests active evolution dynamics in favor of the acquisition of new activities and/or that some of the clustered genes function in the same pathway (Mizutani & Ohta, 2010; Nelson & Werck-Reichhart, 2011). Such clustering has already been described for furanocoumarin biosynthesis genes, for instance with *CYP71AJ3* and *CYP71AJ4* from *Pastinaca sativa* (Roselli *et al.*, 2017). It is possible that some of the *CYP76Fs* clustered with *CYP76F112* catalyze other steps of the furanocoumarin pathway such as the formation of psoralen and/or bergaptol. However, in the *F. carica* genome

(Usai *et al.*, 2020), *CYP76F112* is localized on chromosome 11 while *FcPT1* is on chromosome 12, and no dioxygenase, prenyl-transferase or methyltransferase was found in the genomic region of the *CYP76F* cluster. Therefore, it seems that in *F. carica*, the furanocoumarin biosynthesis genes are not gathered in a single metabolic cluster. It is thus possible that other steps of the pathway are catalyzed by P450s that are not part of the *CYP76* family and are located in different genomic regions.

CYP76F112 was shown to be a selective and efficient enzyme that interacts with high specificity and affinity with DMS. Moreover, *CYP76F112* is preferentially expressed in petiole latex (Table S6), which is consistent with both the repartition of linear furanocoumarins in *F. carica* (Kitajima *et al.*, 2018) and the expression patterns of *FcPT1* (Munakata *et al.*, 2020). Taken together, these results led us to conclude that the conversion of DMS into marmesin described here might be the primary physiological function of *CYP76F112*. Additional observations are consistent with this conclusion. For instance, DMS is necessary for the biosynthesis of downstream furanocoumarins (Fig. 1b) and its biosynthesis is supported by *FcPT1* activity (Munakata *et al.*, 2020). However, it has never been described in *Ficus* phytochemical analyses. The high affinity and efficiency of *CYP76F112* for DMS (apparent $K_m = 32.2 \pm 3.9$ nM, $k_{cat}/K_m = 687 \mu\text{M}^{-1} \text{min}^{-1}$) might therefore reflect an adaptation to DMS being present *in planta* at low concentrations, allowing the conversion of DMS into marmesin as soon as it is synthesized, making it a fleeting intermediate of furanocoumarin biosynthesis. For comparison, *CYP76F112* affinity is 10 to > 100 times higher than that of the P450s previously described in the furanocoumarin pathway (Larbat *et al.*, 2007, 2009; Krieger *et al.*, 2018; Limones-Mendez *et al.*, 2020) and ≈ 1000 times higher than *FcPT1* affinity (Munakata *et al.*, 2020). But interestingly, *CYP76F112* is not the unique *CYP76F* with a very high affinity since *CYP76F45* apparent K_m value is of 66 nM for geraniol (Sintupachee *et al.*, 2015).

In this study, we identified four residues that impact *CYP76F112* specificity and affinity. These residues are included into the substrate recognition sites (SRSs) (Gotoh, 1992) of the enzyme. In particular, T102, S105 and M117 are located into SRS1 while S305 belongs to SRS4. The results we obtained thus confirmed the essential influence of SRS1,4 onto P450s catalytic properties. Interestingly, the residue in position M117, which is critical for the MS activity, seems to correspond to a hotspot position that was described for other P450s from bacteria, fungi and animals (Seifert *et al.*, 2009; Sirim *et al.*, 2010), but not yet in plant P450s. In *CYP102A1* (*Bacillus megaterium*), this hotspot position corresponds to the amino acid F87, which is located close to the heme and is assumed to shape the substrate-binding pocket, thus playing a key role for substrate specificity and selectivity (Seifert *et al.*, 2009; Sirim *et al.*, 2010) (Fig. S16).

Until now, we described the *CYP76F* expansion and the emergence of the MS activity as specific of a restricted taxon of the Moraceae family. The only moraceous species for which genetic data are publicly available so far were from the *Ficus* genus (tribe Ficeae), which produces furanocoumarins, and the *Morus* genus (tribe Moreae), which does not. Yet, other moraceous species

such as *Dorstenia* (Dorstenieae) and *Maquira* (Castilleae) produce furanocoumarins (Fig. 2b) (Rovinski & Sneden, 1984; Vieira *et al.*, 1999; Abegaz *et al.*, 2004). According to phylogenetic-bases articles, tribes Ficeae, Dorstenieae, and Castilleae comprise a clade (Clement & Weiblen, 2009; Zerega & Gardner, 2019). This suggests that the acquisition of the MS activity by the *CYP76Fs* may have occurred in an ancestor of this furanocoumarin-producing clade. Genomic data from additional tribes within the Moraceae would thus help to date the emergence of *CYP76F* MSs more precisely.

As the MS activity emerged recently within Moraceae, *CYP76F112* necessarily emerged independently from the still unknown MS genes of distant Fabaceae, Apiaceae and Rutaceae species, through convergent evolution. It was previously suggested that the UDTs from Moraceae (*FcPT1*) and Apiaceae (*PcPT1*, *PsPT1,2*) evolved from a different ancestral sequence (Munakata *et al.*, 2020). As *FcPT1* and *CYP76F112* catalyze two of the first enzymatic steps of the furanocoumarin pathway, they should have been among the first enzymes recruited in the pathway. Consequently, we can now conclude that the furanocoumarin pathway emerged recently within the Moraceae family, and this acquisition was independent from that of the Apiaceae, Rutaceae, and Fabaceae families. In other words, the furanocoumarin pathway constitutes a compelling case of convergent evolution in higher plants.

The independent emergence of furanocoumarins in the Moraceae, Fabaceae, Apiaceae and Rutaceae implies that different enzymes families and/or fold families may have been recruited in the different plant families to catalyze the same reactional steps. For instance, it is likely that the pathway of the Fabaceae does not include any *CYP71AJ*, *CYP71AZ*, *CYP76F*, *CYP82C* or *CYP82D*, but rather involves other P450 (sub)families that may be unique to this plant taxa. Consequently, to pursue the elucidation of the pathway, we would advise not to rely exclusively on similarity searches, but to favor complementary strategies such as the transcriptomic approach used to identify both *FcPT1* (Munakata *et al.*, 2020) and *CYP76F112*.

Finally, despite this independent emergence, all currently known enzymes of the pathway belong to the same fold families (i.e. dioxygenases, UbiA prenyltransferases, P450s and methyltransferases, Fig. 1b), and the apiaceous MS is assumed to be a P450 (Hamerski & Matern, 1988). The entire furanocoumarin pathway might thus be a case of parallel evolution, confirming its relevance for studying the evolutionary history of plant defense mechanisms.

Acknowledgements








The authors would like to thank Clément Charles and Jérémy Grosjean for technical assistance, Ludivine Hocq and Alexandre Olry for advices on molecular biology and enzymology, and Janani Durairaj and Freek Baker for advices on bioinformatics. CV was funded by the French research ministry and the GrandEst region. The research was supported by the Impact Biomolecules Project (ANR-15-IDEX-04-LUE Lorraine Université d'Excellence) (to AH), by JSPS Overseas Research

Fellowships (to RM), and by PRESTO, the Japan Science and Technology Agency (no. JPMJPR20D7 to RM). The authors declare that there is no competing interest.

Author contributions

CV realized the experiments. CV, RM and SK identified the candidate genes, CV, RvV and MES performed the *in silico* analyses. CV, RL and AH wrote the article. RL and AH directed the project.

ORCID

Alain Hehn  <https://orcid.org/0000-0003-4507-8031>
Sakihito Kitajima  <https://orcid.org/0000-0002-9840-6708>
Romain Lariat  <https://orcid.org/0000-0002-9060-5040>
Ryosuke Munakata  <https://orcid.org/0000-0002-7888-6281>
Michael Eric Schranz  <https://orcid.org/0000-0001-6777-6565>
Robin van Velzen  <https://orcid.org/0000-0002-6444-7608>
Cloé Villard  <https://orcid.org/0000-0001-6683-8541>

Data availability

Nucleotide sequences coding for CYP76F110-112 are openly available in the GenBank data libraries under accession numbers MW348920, MW348921, MW348922.

References

- Abegaz BM, Ngadjui BT, Folefoc GN, Fotso S, Ambassa P, Bezabih M, Dongo E, Rise F, Petersen D. 2004. Prenylated flavonoids, monoterpenoid furanocoumarins and other constituents from the twigs of *Dorstenia elliptica* (Moraceae). *Phytochemistry* 65: 221–226.
- Bailey DG, Dresser G, Arnold JMO. 2013. Grapefruit-medication interactions: forbidden fruit or avoidable consequences? *Canadian Medical Association Journal* 185: 309–316.
- Batard Y, LeRet M, Schalk M, Robineau T, Durst F, Werck-Reichhart D. 1998. Molecular cloning and functional expression in yeast of CYP76B1, a xenobiotic-inducible 7-ethoxycoumarin O-de-ethylase from *Helianthus tuberosus*. *The Plant Journal* 14: 111–120.
- Bathe U, Tissier A. 2019. Cytochrome P450 enzymes: a driving force of plant diterpene diversity. *Phytochemistry* 161: 149–162.
- Beier RC, Oertli EH. 1983. Psoralen and other linear furocoumarins as phytoalexins in celery. *Phytochemistry* 22: 2595–2597.
- Benkert P, Biasini M, Schwede T. 2011. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* 27: 343–350.
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2012. GenBank. *Nucleic Acids Research* 41: D36–D42.
- Bertoni M, Kiefer F, Biasini M, Bordoli L, Schwede T. 2017. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Scientific Reports* 7: 1–15.
- Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L *et al.* 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research* 42: W252–W258.
- Bienert S, Waterhouse A, de Beer TAP, Tauriello G, Studer G, Bordoli L, Schwede T. 2017. The SWISS-MODEL repository—new features and functionality. *Nucleic Acids Research* 45: D313–D319.
- Bourgaud F, Hehn A, Lariat R, Doerper S, Gontier E, Kellner S, Matern U. 2006. Biosynthesis of coumarins in plants: a major pathway still to be unravelled for cytochrome P450 enzymes. *Phytochemistry Reviews* 5: 293–308.
- Bourgaud F, Olry A, Hehn A. 2014. Recent advances in molecular genetics of furanocoumarin synthesis in higher plants. In: Jacob C, Kirsch G, Slusarenko A, Winyard PG, Burkholz T, eds. *Recent advances in redox active plant and microbial products: from basic chemistry to widespread applications in medicine and agriculture*. Dordrecht, the Netherlands: Springer, 363–375.
- Brown SA, Steck W. 1973. 7-Demethylsuberosin and ostenol as intermediates in furanocoumarin biosynthesis. *Phytochemistry* 12: 1315–1324.
- Carpenter EJ, Matasci N, Ayyampalayam S, Wu S, Sun J, Yu J, Jimenez Vieira FR, Bowler C, Dorrell RG, Gitzendanner MA *et al.* 2019. Access to RNA-sequencing data from 1,173 plant species: the 1000 plant transcriptomes initiative (1KP). *GigaScience* 8: 1–7.
- Chauthe SK, Mahajan S, Rachamalla M, Tikoo K, Singh IP. 2015. Synthesis and evaluation of linear furanocoumarins as potential anti-breast and anti-prostate cancer agents. *Medicinal Chemistry Research* 24: 2476–2484.
- Clement WL, Weiblen GD. 2009. Morphological evolution in the mulberry family (Moraceae). *Systematic Botany* 34: 530–552.
- Diaz-Chavez ML, Moniodis J, Madilao LL, Jancsik S, Keeling CI, Barbour EL, Ghisalberti EL, Plummer JA, Jones CG, Bohlmann J. 2013. Biosynthesis of sandalwood oil: *Santalum album* CYP76F cytochromes P450 produce santalols and bergamotol. *PLoS ONE* 8–9: 1–11 e75053.
- Dueholm B, Krieger C, Drew D, Olry A, Kamo T, Taboureau O, Weitzel C, Bourgaud F, Hehn A, Simonsen HT. 2015. Evolution of substrate recognition sites (SRs) in cytochromes P450 from Apiaceae exemplified by the CYP71AJ subfamily. *BMC Evolutionary Biology* 15: 122–136.
- Fujiyama K, Hino T, Kanadani M, Watanabe B, Jae Lee H, Mizutani M, Nagano S. 2019. Structural insights into a key step of brassinosteroid biosynthesis and its inhibition. *Nature Plants* 5: 589–594.
- Gotoh O. 1992. Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *The Journal of Biological Chemistry* 267: 83–90.
- Gu M, Wang M, Guo J, Shi C, Deng J, Huang L, Huang L, Chang Z. 2019. Crystal structure of CYP76AH1 in 4-PI-bound state from *Salvia miltiorrhiza*. *Biochemical and Biophysical Research Communications* 511: 813–819.
- Guex N, Peitsch MC, Schwede T. 2009. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis* 30: S162–S173.
- Hamberger B, Bak S. 2013. Plant P450s as versatile drivers for evolution of species-specific chemical diversity. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368: 1–16.
- Hamerski D, Matern U. 1988. Elicitor-induced biosynthesis of psoralens in *Ammi majus* L. suspension cultures – microsomal conversion of demethylsuberosin into (+)marmesin and psoralen. *European Journal of Biochemistry* 171: 369–375.
- Hebsgaard S. 1996. Splice site prediction in *Arabidopsis thaliana* pre-mRNA by combining local and global sequence information. *Nucleic Acids Research* 24: 3439–3452.
- Hehmann M, Lukaćin R, Ekiert H, Matern U. 2004. Furanocoumarin biosynthesis in *Ammi majus* L.: cloning of bergaptol O-methyltransferase. *European Journal of Biochemistry* 271: 932–940.
- Hofer R, Boachon B, Renault H, Gavira C, Miesch L, Iglesias J, Ginglinger J-f, Allouche L, Miesch M, Grec S *et al.* 2014. Dual function of the cytochrome P450 CYP76 Family from *Arabidopsis thaliana* in the metabolism of monoterpenols and phenylurea herbicides. *Plant Physiology* 166: 1149–1161.
- Ilc T, Halter D, Miesch L, Lauvoisard F, Kriegshauser L, Ilg A, Baltenweck R, Hugueney P, Werck-Reichhart D, Duchêne E *et al.* 2017. A grapevine cytochrome P450 generates the precursor of wine lactone, a key odorant in wine. *New Phytologist* 213: 264–274.
- Karamat F, Olry A, Munakata R, Koeduka T, Sugiyama A, Paris C, Hehn A, Bourgaud F, Yazaki K. 2014. A coumarin-specific prenyltransferase catalyzes the crucial biosynthetic reaction for furanocoumarin formation in parsley. *The Plant Journal* 77: 627–638.
- Kitajima S, Aoki W, Shibata D, Nakajima D, Sakurai N, Yazaki K, Munakata R, Taira T, Kobayashi M, Aburaya S *et al.* 2018. Comparative multi-omics analysis reveals diverse latex-based defense strategies against pests among latex-producing organs of the fig tree (*Ficus carica*). *Planta* 247: 1423–1438.

- Krieger C, Roselli S, Kellner-Thielmann S, Galati G, Schneider B, Grosjean J, Olry A, Ritchie D, Matern U, Bourgaud F *et al.* 2018. The CYP71AZ P450 subfamily: a driving factor for the diversification of coumarin biosynthesis in apiaceous plants. *Frontiers in Plant Science* 9: 1–15.
- Larbat R, Hehn A, Hans J, Schneider S, Jugé H, Schneider B, Matern U, Bourgaud F. 2009. Isolation and functional characterization of *CYP71AJ4* encoding for the first P450 monooxygenase of angular furanocoumarin biosynthesis. *Journal of Biological Chemistry* 284: 4776–4785.
- Larbat R, Kellner S, Specker S, Hehn A, Gontier E, Hans J, Bourgaud F, Matern U. 2007. Molecular cloning and functional characterization of psoralen synthase, the first committed monooxygenase of furanocoumarin biosynthesis. *Journal of Biological Chemistry* 282: 542–554.
- Limones-Mendez M, Dugrand-Judek A, Villard C, Coqueret V, Froelicher Y, Bourgaud F, Olry A, Hehn A. 2020. Convergent evolution leading to the appearance of furanocoumarins in citrus plants. *Plant Science* 292: 1–9.
- Maddison WP, Maddison DR. 2019. *Mesquite: a modular system for evolutionary analysis*, v.3.61. [WWW document] URL <http://www.mesquiteproject.org/>.
- Marrelli M, Menichini F, Statti GA, Bonesi M, Duez P, Menichini F, Conforti F. 2012. Changes in the phenolic and lipophilic composition, in the enzyme inhibition and antiproliferative activity of *Ficus carica* L. cultivar Dottato fruits during maturation. *Food and Chemical Toxicology* 50: 726–733.
- Mizutani M, Ohta D. 2010. Diversification of P450 genes during land plant evolution. *Annual Review of Plant Biology* 61: 291–315.
- Mizutani M, Sato F. 2011. Unusual P450 reactions in plant secondary metabolism. *Archives of Biochemistry and Biophysics* 507: 194–203.
- Munakata R, Kitajima S, Nuttens A, Tatsumi K, Takemura T, Ichino T, Galati G, Vautrin S, Bergès H, Grosjean J *et al.* 2020. Convergent evolution of the UbiA prenyltransferase family underlies the independent acquisition of furanocoumarins in plants. *New Phytologist* 225: 2166–2182.
- Munakata R, Olry A, Karamat F, Courdavault V, Sugiyama A, Date Y, Krieger C, Silie P, Foureau E, Papon N *et al.* 2016. Molecular evolution of parsnip (*Pastinaca sativa*) membrane-bound prenyltransferases for linear and/or angular furanocoumarin biosynthesis. *New Phytologist* 211: 332–344.
- Murray RDH, Mendez J, Brown RA. 1982. *The natural coumarins: occurrence, chemistry and biochemistry*. Chichester, UK: John Wiley & Sons.
- Nelson DR. 2018. Cytochrome P450 diversity in the tree of life. *Biochimica et Biophysica Acta (BBA) – Proteins and Proteomics* 1866: 141–154.
- Nelson D, Werck-Reichhart D. 2011. A P450-centric view of plant evolution: P450-centric evolution. *The Plant Journal* 66: 194–211.
- Omura T, Sato R. 1964a. The carbon monoxide-binding pigment of liver microsomes. I. Evidence for its hemoprotein nature. *The Journal of Biological Chemistry* 239: 2370–2378.
- Omura T, Sato R. 1964b. The carbon monoxide-binding pigment of liver microsomes. II. Solubilization, purification and properties. *Journal of Biological Chemistry* 239: 2379–2385.
- One Thousand Plant Transcriptomes Initiative. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574: 679–685.
- Pathak MA, Daniels F Jr, Fitzpatrick TB. 1962. The presently known distribution of furocoumarins (psoralens) in plants. *Journal of Investigative Dermatology* 39: 225–239.
- Pompon D, Louerat B, Bronine A, Urban P. 1996. Yeast expression of animal and plant P450s in optimized redox environments. In: Johnson EF, Waterman MR, eds. *Methods in enzymology*, Vol. 272. Amsterdam, the Netherlands: Academic Press, 51–64.
- Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* 67: 901–904.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61: 539–542.
- Roselli S, Olry A, Vautrin S, Coriton O, Ritchie D, Galati G, Navrot N, Krieger C, Vialart G, Bergès H *et al.* 2017. A bacterial artificial chromosome (BAC) genomic approach reveals partial clustering of the furanocoumarin pathway genes in parsnip. *The Plant Journal* 89: 1119–1132.
- Rovinski JM, Sneden AT. 1984. Furanocoumarins from *Maquira calophylla*. *Journal of Natural Products* 47: 557.
- Schuler MA. 2011. P450s in plant–insect interactions. *Biochimica et Biophysica Acta (BBA) – Proteins and Proteomics* 1814: 36–45.
- Seifert A, Vomund S, Grohmann K, Kriening S, Urlacher VB, Laschat S, Pleiss J. 2009. Rational design of a minimal and highly enriched CYP102A1 mutant library with improved regio-, stereo- and chemoselectivity. *ChemBioChem* 10: 853–861.
- Shirasawa K, Yakushiji H, Nishimura R, Morita T, Jikumaru S, Ikegami H, Toyoda A, Hirakawa H, Isobe S. 2019. The *Ficus erecta* genome to identify the Ceratocystis canker resistance gene for breeding programs in common fig (*F. carica*). *The Plant Journal* 102: 1313–1322.
- Sintupachee S, Promden W, Ngamrojanavanich N, Sitthithaworn W, De-Eknamkul W. 2015. Functional expression of a putative geraniol 8-hydroxylase by reconstitution of bacterially expressed plant CYP76F45 and NADPH-cytochrome P450 reductase CPR I from *Croton stellatopilosus* Ohba. *Phytochemistry* 118: 204–215.
- Sirim D, Widmann M, Wagner F, Pleiss J. 2010. Prediction and analysis of the modular structure of cytochrome P450 monooxygenases. *BMC Structural Biology* 10: 1–12.
- Soltis DE, Soltis PS, Morgan DR, Swensen SM, Mullin BC, Dowd JM, Martin PG. 1995. Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. *Proceedings of the National Academy of Sciences, USA* 92: 2647–2651.
- Studer G, Rempfer C, Waterhouse AM, Gumienny R, Haas J, Schwede T. 2020. QMEANDisCo—distance constraints applied on model quality estimation (A Elofsson, Ed.). *Bioinformatics* 36: 1765–1771.
- The Angiosperm Phylogeny Group. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* 181: 1–20.
- Trott O, Olson AJ. 2009. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* 31: 455–461.
- Urban P, Cullin C, Pompon D. 1990. Maximizing the expression of mammalian cytochrome P-450 monooxygenase activities in yeast cells. *Biochimie* 72: 463–472.
- Urban P, Mignotte C, Kazmaier M, Delorme F, Pompon D. 1997. Cloning, yeast expression, and characterization of the coupling of two distantly related *Arabidopsis thaliana* NADPH-cytochrome P450 reductases with P450 CYP73A5. *Journal of Biological Chemistry* 272: 19176–19186.
- Usai G, Mascagni F, Giordani T, Vangelisti A, Bosi E, Zuccolo A, Ceccarelli M, King R, Hassani-Pak K, Zambrano LS *et al.* 2020. Epigenetic patterns within the haplotype phased fig (*Ficus carica* L.) genome. *The Plant Journal* 102: 600–614.
- Vialart G, Hehn A, Olry A, Ito K, Krieger C, Larbat R, Paris C, Shimizu B-I, Sugimoto Y, Mizutani M *et al.* 2012. A 2-oxoglutarate-dependent dioxygenase from *Ruta graveolens* L. exhibits *p*-coumaroyl CoA 2'-hydroxylase activity (C2'H): a missing step in the synthesis of umbelliferone in plants: C2'H involved in umbelliferone synthesis. *The Plant Journal* 70: 460–470.
- Vieira IJC, Mathias L, Monteiro VDFF, Braz-Filho R, Rodrigues-Filho E. 1999. A new coumarin from *Brosimum gaudichaudii* Trecul. *Natural Product Letters* 13: 47–52.
- Villard C, Larbat R, Munakata R, Hehn A. 2019. Defence mechanisms of *Ficus*: pyramiding strategies to cope with pests and pathogens. *Planta* 249: 617–633.
- Wang Q, Hillwig ML, Okada K, Yamazaki K, Wu Y, Swaminathan S, Yamane H, Peters RJ. 2012. Characterization of CYP76M5-8 indicates metabolic plasticity within a plant biosynthetic gene cluster. *Journal of Biological Chemistry* 287: 6159–6168.
- Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TA P, Rempfer C, Bordoli L *et al.* 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* 46: W296–W303.
- Zabolinejad N, Maleki M, Salehi M, Ashrafi Z, Molkara S, Layegh P. 2020. Psoralen and narrowband UVB combination provides higher efficacy in treating vitiligo compared with narrowband UVB alone: a randomised clinical trial. *Australasian Journal of Dermatology* 61: e65–e69.

Zerega NJC, Gardner EM. 2019. Delimitation of the new tribe Parartocarpeae (Moraceae) is supported by a 333-gene phylogeny and resolves tribal level Moraceae taxonomy. *Phytotaxa* 388: 253–265.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Dataset S1 CYP76F datasets used to generate the gene-trees.

Fig. S1 Putative CDS of the three P450 candidates.

Fig. S2 Sequence of the mutants synthesized and inserted into the expression vector pYeDP60.

Fig. S3 Evaluation of the expression of CYP76F110-112 by immunodetection.

Fig. S4 Evaluation of the expression of the CYP76F111-112 mutants by immunodetection.

Fig. S5 Determination of the concentration of functional P450s using the differential CO spectrum method.

Fig. S6 Screening of the *Ficus carica* RNA-Seq library.

Fig. S7 CYP76F110-112 equivalents in the *Ficus carica* genome.

Fig. S8 Determination of the optimal temperature and pH for the activity of CYP76F112, incubated in presence of DMS and NADPH.

Fig. S9 Gene-family phylogenetic tree of the CYP76Fs across the nitrogen fixing clade.

Fig. S10 Alignment between CYP76F112 and the putative marmesin synthase from *Ficus erecta* and *F. religiosa*.

Fig. S11 Hemes docked within CYP76F111 and CYP76F112.

Fig. S12 Movement of CYP76F112 flexible amino acids during the docking of the DMS.

Fig. S13 Docking of the DMS within CYP76F112-T102P, CYP76F112-S105A, CYP76F112-T102P, S105A and CYP76F112-S305P.

Fig. S14 Docking of the DMS within four additional CYP76Fs of the *Ficus*-clade.

Fig. S15 Some substrates metabolized by enzymes from the CYP76 family.

Fig. S16 Identification of the CYP76F111-112 amino acids equivalent to F87 from CYP102A1.

Table S1 Putative substrates tested with CYP76F110-112.

Table S2 Determination of the specific activities.

Table S3 Docking of the DMS within CYP76F111 and CYP76F112.

Table S4 Docking of the DMS within the CYP76F111 and CYP76F112 mutants.

Table S5 Docking of the DMS within four additional CYP76Fs of the *Ficus*-clade.

Table S6 Contigs used to identify the P45 candidates and associated expression patterns.

Table S7 Description of the CYP76F111-112 mutants.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.