The evolution of chemical diversity in plants : pyrrolizidine alkaloids and cytochrome P450s in Jacobaea

Chen, Y.

Chapter 3

Diversity and evolution of cytochrome P450s of *Jacobaea vulgaris* and *Jacobaea aquatica*

Yangan Chen, Peter G.L. Klinkhamer, Johan Memelink, Klaas Vrieling

Diversity and evolution of cytochrome P450s of *Jacobaea vulgaris* and *Jacobaea aquatica*

Abstract

Collectively, plants produce a huge variety of secondary metabolites (SMs) which are involved in the adaptation of plants to biotic and abiotic stresses. The most characteristic feature of SMs is their striking inter- and intraspecific chemical diversity. Cytochrome P450 monooxygenases (CYPs) often play an important role in the biosynthesis of SMs and thus in the evolution of chemical diversity. Here we studied the diversity and evolution of CYPs of two *Jacobaea* species which contain a characteristic group of SMs namely the pyrrolizidine alkaloids (PAs). We retrieved CYPs from RNA-seq data of *J. vulgaris* and *J. aquatica*, resulting in 221 and 157 full-length CYP genes, respectively. The analyses of conserved motifs confirmed that *Jacobaea* CYP proteins share conserved motifs including the heme-binding signature, the PERF motif, the K-helix and the I-helix. KEGG annotation revealed that the CYPs assigned as being SM metabolic pathway genes were all from the CYP71 clan but no CYPs were assigned as being involved in alkaloid pathways. Phylogenetic analyses of full-length CYPs were conducted for the six largest CYP families of *Jacobaea* (CYP71, CYP76, CYP706, CYP82, CYP93 and CYP72) and were compared with CYPs of two other members of the Asteraceae, *Helianthus annuus* and *Lactuca sativa* and with the outgroups of *Arabidopsis thaliana* CYPs. The phylogenetic trees showed strong lineage specific diversification of CYPs, implying that the evolution of CYPs has been very fast even within the Asteraceae family. Only in the closely related species *J. vulgaris* and *J. aquatica*, CYPs were found often in pairs, confirming a close relationship in the evolutionary history.This study discovered 378 full-length CYPs in *Jacobaea* species, which can be used for future exploration of their functions, including possible involvement in PA biosynthesis and PA diversity.

Keywords

chemical diversity, pyrrolizidine alkaloid biosynthesis, RNA-seq, conserved motifs, phylogeny

Introduction

Plants produce a great variety of secondary metabolites (SMs) which are involved in the adaptation of plants to both biotic and abiotic stresses (Bennett and Wallsgrove, 1994; Wink, 2003; Kessler and Kalske, 2018). At present, more than 200,000 SMs have been isolated and identified, including different chemical classes such as glucosinolates, alkaloids, terpenes, and flavonoids. Typically, species within a clade share similar classes of SMs (Wink, 2003). For example, glucosinolates are major SMs near-universally in Brassicaceae, Capparidaeae and Caricaceae (Moore *et al*., 2014), and benzylisoquinoline alkaloids occur mainly in the Papaveraceae, the Ranunculaceae, the Berberidaceae and the Menispermaceae (Ziegler and Facchini, 2008), while pyrrolizidine alkaloids (PAs) distribute preferably in the Asteraceae, the Boraginaceae, the Fabaceae and the Orchidaceae families (Langel *et al*., 2011). Each class of SMs contains a number of similar molecules derived from the same skeleton mostly differing in substitution groups by addition of a number of polar and non-polar substituents. This structural diversity is well documented for PAs in *Jacobaea* species in the Asteraceae family. Thirty-seven structurally related PAs have been detected in *Jacobaea vulgaris*, *Jacobaea aquatica* and their hybrids (Cheng *et al*., 2011). As yet, it is not fully understood how secondary metabolite diversity comes about and why it is maintained in nature.

To understand the origin of SM diversity, molecular investigations of SM biosynthetic pathways are promising as it is believed that SM diversity of plants is under genetic control (van Dam and Vrieling, 1994; Hartmann and Dierich, 1998; Kliebenstein *et al*., 2001; Macel *et al*., 2004). Progress in the identification and characterization of encoding genes involved in SM pathways has provided examples of genes that derived from gene duplication and further diversification of genes which belong to large gene families, such as cytochrome P450s (CYPs) (Bak *et al*., 2006; Frey *et al*., 2009). CYP genes form a large family in any given plant species and play vital roles in many metabolic processes including secondary metabolism (Mizutani, 2012). Many CYPs are involved in biosynthesis of various SMs as they catalyze the oxidative modifications of various substrates using oxygen and NAD(P)H. Structurally, all plant CYPs found so far are membrane-bound enzymes and are mainly anchored in the endoplasmic reticulum membrane via a hydrophobic signal sequence at the N-terminus (Werck-Reichhart *et al*., 2002; Bak *et al*., 2011). CYP proteins share well-conserved motifs including the heme-binding signature, the PERF motif, the K-helix and the I-helix, which are essential for catalytic activity (Paquette *et al*., 2009). The fact that CYPs are often recruited as versatile catalysts in the biosynthesis of SMs makes these enzymes landmarks in the evolution of species-specific chemical diversity (Hamberger and Bak, 2013).

A well-curated set of CYP genes from a particular species is essential for functional identification of the encoded enzymes. In recent years, genome/transcriptome-wide identification of CYPs from plants has been performed to explore their involvement in metabolic pathways (Chen *et al*., 2014; Liao *et al*., 2017; Qi *et al*., 2017; Hori *et al*., 2018; Ilc *et al*., 2018). For example, Liao *et al*. (2017) identified 118 full-length and 175 partial CYP

genes in *Taxus chinensis* transcriptomes with the aim to discover candidate genes involved in the biosynthesis of diterpenoids including taxol. Chen *et al*. (2014) found 116 full-length and 135 partial CYP genes in *Salvia miltiorrhiza* transcriptomes with candidates for terpenoid biosynthesis.

PAs in *Jacobaea* species were selected to launch the discovery of structural genes causing SM diversity in our study. So far, the only pathway-specific enzyme of PA biosynthesis that has been identified is homospermidine synthase, which converts spermidine and putrescine into homospermidine, the first specific intermediate in the PA biosynthesis pathway (Böttcher *et al*., 1993). It is not known how homospermidine is converted to the central PA backbone structure senecionine *N*-oxide. Senecionine *N*-oxide undergoes structural transformations in a position-specific and stereoselective manner resulting in the rearrangement of the skeletal structure and oxidative modifications thereof (Hartmann and Dierich, 1998). It was shown that the diversification of PAs in *Jacobaea* species occurs in the shoots while the primary PA senecionine *N*-oxide is synthesized in the roots (Thomas Hartmann and Toppel, 1987; Thomas Hartmann *et al*., 1989). With the exception of senecivernine it was deduced that the PA diversification from senecionine *N*-oxide to other PAs is brought about via specific one- or two-step reactions including epoxidation, hydroxylation, dehydrogenation and/or *O*-acetylation (Hartmann and Dierich, 1998; Pelser *et al*., 2005). The enzymes responsible for these processes have not been identified. Candidates for the oxidative reactions are members of the CYP family. A comprehensive study and comparison of CYPs between different *Jacobaea* species can be beneficial to identify potential CYP candidates involved in PA biosynthesis.

We have established *de novo* transcriptome assemblies for *J. vulgaris* and *J. aquatica* and established comprehensive information on CYP families. These two closely related species have been well studied for their PA contrasts (Cheng *et al*., 2011; Joosten *et al*., 2011), but limited genomic or transcriptomic information is available. We first identified putative full-length CYPs classified into different CYP families and extracted the conserved motifs. Furthermore, we investigated the potential involvement of these CYPs in various metabolic pathways based on the KEGG database. We subsequently performed phylogenetic analyses of the largest CYP families in *Jacobaea* species and two other species from the Asteraceae using the CYPs from *Arabidopsis thaliana* as an outgroup to explore relatedness and evolution of CYPs across five species.

Materials and methods

Plant material

From both *J. vulgaris* and *J. aquatica* species two sets of samples were obtained (Table S1). The first *J. vulgaris* set (*Jv*1) consisted of the pooled shoots and roots of 59 individuals from nine different populations across Europe including two individuals derived from tissue culture

and one population from Canada (Table S1). Set *Jv*1 was normalized. The second *J. vulgaris* set (*Jv*2) was composed from multiple individuals, clones, of one genotype that was kept in tissue culture. For the set *Jv*2, five individuals from tissue culture derived plants of *J. vulgaris* treated with methyl jasmonate (MeJA) and five mock treated individuals were used as control. From both MeJA treated and control plants cDNA libraries were obtained that were sequenced separately. The resulting reads were pooled *in silico* in the later assembly step. Both *J. aquatica* sets (*Ja*1 and *Ja*2) were derived from the same seven individuals pooled from two populations with two individuals originating from tissue culture, of which roots were included in *Ja*1 but not in *Ja*2 (Table S1). Set *Ja*1 was normalized before sequencing while set *Ja*2 was not.

For sets *Jv*1, *Ja*1 and *Ja*2, seeds were germinated on the surface of wet potting soil covered by plastic bags and the seedlings were transferred into 9×9×10 cm pots filled with 50% sandy soil (collected from Meijendel), 50% potting soil (Slingerland Potgrond, Zoeterwoude, The Netherlands) and 1.5 g/L Osmocote slow release fertilizer (Scott, Scotts Miracle-Gro, Marysville, Ohio, USA; N: P: K = 15: 9: 11). Tissue cultured plants of *J. vulgaris* and *J. aquatica* were propagated on Murashige and Skoog (MS) medium with 0.44 mM benzylaminopurine. To induce roots plants were transferred to MS medium without hormones for two weeks. After rooting plants were transferred to pots filled with the soil mixture as indicated above. All plants were kept in a climate room for six weeks (humidity 70%, light 16 h at 20 °C, dark 8 h at 20 °C). Then the plants were separated into shoots and roots, and roots were rinsed with water. Two to three fully grown leaves and ¼ of roots from each plant were wrapped in aluminum foil and flash frozen in liquid nitrogen, respectively. Afterwards all samples were separately ground into powder with liquid nitrogen. Shoot powder was mixed with root powder in a ratio of 3:1 for each plant, and then identical amounts of powder from each individual were pooled for *Jv*1 and *Ja*1, respectively, whereas only powdered shoots were pooled for *Ja*2. All powdered materials were stored at -80 ℃ until RNA extraction.

For set *Jv*2, replicate *J. vulgaris* tissue culture plants were kept on MS medium with agar for two weeks after propagation in a climate room (50% humidity, light 16 h at 20 °C, dark 8 h at 20 °C). One hundred microliters of MeJA (Sigma-Aldrich) dissolved in 10% ethanol solution (4.5 mmol/L) was added to the surface of medium, reaching a final concentration of 90 μmol/L after diffusion in each tube, while the same volume of 10% ethanol was added to the control group under axenic condition. Shoots of five biological replicates collected at eight days after the treatment were pooled and ground into fine powder for both induced and control groups, respectively. All powder was stored at -80 ℃ until RNA extraction.

RNA isolation, normalization and transcriptome sequencing

Total RNA was extracted with the NucleoSpin® RNA Plant-Macherey-Nagel kit for five samples, namely *Jv*1, MeJA induced group of *Jv*2, control group of *Jv*2, *Ja*1 and *Ja*2. The RNA integrity Number (RIN) and RNA concentration were assessed using the Agilent 2100 Bioanalyzer. Strand specific RNAseq libraries were generated using the method described by Parkhomchuk *et al.* (2009) with minor modifications by the Leiden Genome Technology

Center. In short, polyA+ mRNA was isolated from 1 µg of total RNA using oligo-dT Dynabeads (LifeTech 61002) and fragmented to 150 - 200 nucleotides in first strand buffer for three minutes at 94 °C. Random hexamer primed first strand was generated in presence of dATP, dGTP, dCTP and dTTP. dUTP was used to tag the second strand instead of dTTP. Subsequent steps to construct the sequencing libraries were performed with the KAPA HTP Library Preparation Kit for Illumina sequencing with minor modifications. Shortly, after indexed adapter ligation to the dsDNA fragments, the libraries were treated with USER enzyme (NEB M5505L) in order to digest the second strand derived fragments. Pre-amplified library yields were quantified on an Agilent high sensitivity chip. Two of four sets were normalized with duplex-specific thermostable nuclease (DSN, Evrogen) to remove abundant library molecules. The protocol was carried out according to the Illumina guidelines for *Jv*1 and *Ja*1. After DSN treatment, a second round of PCR was performed. All samples were quantified on an Agilent high sensitivity chip prior to pooling in equimolar amounts and sequencing on a HiSeq2500 with 2x126 bp paired-end reads in the Leiden Genome Technology Center.

***De novo* assembly and evaluation**

After removal of adapter sequences, the qualities of raw reads were checked using FastQC and the bases with low quality (threshold < 30) were cut off by Trimmomatic via the Galaxy platform (Afgan *et al.*, 2016). The paired-end clean reads were used for assembly. A *de novo* assembly strategy using the Trinity program (Haas *et al*., 2013) with a k-mer size of 32 and the minimum assembled contig length to report set to 300 bp was employed to assemble the four sets (*Jv*1, *Jv*2, *Ja*1 and *Ja*2). To assess the quality of four assemblies, reads were aligned back to transcriptomes by Bowtie2 (Langmead and Salzberg, 2012). GC content and basic statistics values were calculated using the script imbedded in the Trinity suite.

***In silico* mining of CYP genes**

To identify CYP-like contigs from the four transcriptomes, the HMMER program (http://hmmer.org; Version 3.2.1b2) was used to search for homologs by the hidden Markov model against the CYP reference (PF00067) of the Pfam database, with an e-value cutoff of 1e-5. The obtained CYP-like contigs from sets *Jv*1 and *Jv*2 of *J. vulgaris* were combined and 100% identical transcripts were removed by using the CD-HIT-EST algorithm (version 4.6.8) (Li and Godzik, 2006; Fu *et al*., 2012). For *J. aquatica*, the sample approach was applied to combine CYP-like contigs from sets *Ja*1 and *Ja*2.

To obtain additional CYP-like contigs, the reads of *J. vulgaris* were mapped to all CYP-like contigs of *J. aquatica* in CLC genomics workbench and vice versa (version 8.5.1) using the following parameters: mismatch cost 2, insertion cost 3, deletion cost 3, length fraction 0.8, similarity fraction 0.97. The consensus sequences of the mapped reads were retained and assembled with the original CYP-like contigs of *J. vulgaris* in Sequencher (version 5.0), using a minimum match percentage of 97% while minimum overlap was set to 15%. Thereupon, the Sequencher assembly of CYP-like contigs were checked for redundancies using the CD-HIT-

EST algorithm with sequence identity of 97% as cutoff. Similarly, to get additional CYP-like contigs for *J. aquatica*, CYP-like contigs of *J. vulgaris* were used as references for read mapping, followed by the same steps afterwards.

The likely coding regions of the resultant CYP-like contigs of both species were predicted by TransDecoder (https://github.com/TransDecoder/TransDecoder/wiki. Version 5.5.0.). In order to recognize full-length CYP genes, all the peptide sequences were blasted against NCBI, and the information of blast hits were used to classify CYPs into different clans. Within each clan the alignment of sequences which contain at least 400 amino acids was conducted in MEGA 7 (Kumar *et al*., 2016) for manual curation of complete coding regions. The putative full-length CYP genes were identified according to the following two criteria: (1) the corresponding proteins starts with amino acid 'M' and stops before a stop codon; (2) The aligned regions within each clan cover most of the length in a blast hit to a full-length CYP at the NCBI database, where the highly conserved heme signature is about 50 amino acids from the C-terminus.

Classification and characterization of *Jacobaea* CYP genes

The final classification and nomenclature of all full-length CYP proteins were carried out by Prof. Dr. David R. Nelson through comparison with references from a well-annotated plant CYP database which includes both published and confidential sequences, following the CYP nomenclature principle (Nelson, 2009). Cutoff values for family, subfamily and allelic variants were 40%, 55% and 97% amino acid sequence identity, respectively.

The CYP assemblies were divided into A-type which only comprises the CYP71 clan, and non-A-type which includes all other plant CYP clans. The sequences of A-type and non-A-type were separately submitted to MEME to predict motifs and to Motif Alignment and Search Tool (MAST) to discover homologs (Bailey *et al*., 2009). The logos of motifs were created using WEBLOGO (Schneider and Stephens, 1990; Crooks *et al*., 2004). Furthermore, the theoretical isoelectric points (PI) and molecular weights (kDa) were predicted by the "Compute pI/Mw tool" on the ExPASy server (Gasteiger *et al*., 2003) and the subcellular locations were predicted using the TargetP1.1 server with specificity > 0.95 (Emanuelsson *et al*., 2000). KEGG Automatic Annotation Server (KAAS) (Moriya *et al*., 2007) was used for ortholog assignment and pathway mapping using the SBH (single-directional best hit) method with the BLAST program.

Phylogenetic analysis

The CYP protein sequences of *H. annuus* (Badouin *et al*., 2017) and *L. sativa* (Reyes-Chin-Wo *et al*., 2017) were retrieved from their transcriptomes using the same approach as aforementioned for *Jacobaea* species based on homologs by the HMM model. The CYP protein sequences of *A. thaliana* were downloaded from the *Arabidopsis* Cytochrome P450 database (http://www.p450.kvl.dk/p450.shtml). Multiple sequence alignments were performed respectively for putative full-length CYP genes in CYP71, CYP76, CYP706, CYP82, CYP93

and CYP72 families using the MUSCLE module imbedded in the MEGA 7 package (Kumar *et al*., 2016) using default settings followed by manual editing. Phylogenetic trees were inferred by using the maximum likelihood (ML) method. The trees were obtained with IQ-tree (Nguyen *et al*., 2015; Kalyaanamoorthy *et al*., 2017) on XSEDE through CIPRES Science Gateway (Miller *et al*., 2010). Bootstrap (BS) search was conducted using standard nonparametric bootstrap with 1000 replicates.

Results

Transcriptome sequencing and *de novo* assembly

The purpose of this study was to obtain systematic information of CYPs in *Jacobaea* species, which facilitates further exploration of possible functions in PA metabolism. Aiming for the most comprehensive CYP gene sets, multiple individuals of both *J. vulgaris* and *J. aquatica* originating from different parts of the distribution ranges (Table S1) were used for transcriptome sequencing because of the large intraspecies variation in both PA composition and concentration. It was chosen to include mainly shoots as these are the sites of PA diversification (Hartmann and Dierich, 1998). In total, two sets of samples were obtained for both *J. vulgaris* (*Jv*1 and *Jv*2) and *J. aquatica* (*Ja*1 and *Ja*2). Transcript normalization was conducted to enhance the gene discovery rate by removing abundant cDNA library molecules for one set of *J. vulgaris* (*Jv*1) and one set of *J. aquatica* (*Ja*1) prior to sequencing. After removal of adaptor sequences, ambiguous reads and low-quality reads (Q < 30), paired-end clean reads were further processed. The trimmed reads obtained in this study have been deposited in the NCBI SRA database (accession number: PRJNA561604).

For each of the four sets, more than 20 million cleaned up paired-end reads were used for the *de novo* assembly with Trinity (Table 1). The resulting assemblies of *Jv*1, *Jv*2, *Ja*1 and *Ja*2 yielded equal amounts of transcripts containing 152,286, 142,213, 118,936, 130,365 transcripts with average lengths of 936, 1,132, 1,082 and 1,062 nucleotides respectively. To evaluate the qualities of the assembled transcripts, all reads were realigned back to the assemblies using Bowtie2 (Langmead and Salzberg, 2012), and we found that between 83% to 91% of reads were mapped back as proper pairs (Table 1). This showed that these assemblies were well-qualified for further mining of CYP genes as our mapping rates were well above the required value of 70-80%.

Identification and classification of CYPs

CYP-like contigs were retrieved out of the assembly based on the homologs compared to CYP references (PF00067) from the Pfam database for each set. The CYP-like contigs of the two sets of each species were combined. Moreover *J. vulgaris* and *J. aquatica* were used as references mutually to get extra CYP-like contigs. All obtained CYP-like contigs were combined for each species, followed by redundancy check with the cutoff of 97% identity.

Table 1. Summary of Illumina sequencing and *de novo* assemblies for two *J. vulgaris* and two *J. aquatica* sets.

Sets	Total paired-end clean reads	Total assembled trinity transcripts	Transcript length range (nt [a])	GC content (%)	Contig N50 [b] (nt)	Average contig length (nt)	Reads mapped [c] (%)
*Jv*1	19,725,242	152,286	301 - 13,238	39.37	1,253	936	84.69
*Jv*2	36,359,675	142,213	301 - 13,269	39.31	1,530	1,132	83.25
*Ja*1	20,306,518	118,936	301 - 15,708	39.27	1,461	1,082	91.57
*Ja*2	27,505,944	130,365	301 - 13,309	41.23	1,441	1,062	87.41

[a] nt: nucleotide.

[b] Contig N50: length such that sequence contigs of this length or longer include half the bases of the assembly.

[c] Reads mapped: the percentage of properly paired reads mapped back to the Trinity transcriptome assembly by Bowtie2.

After removal of redundant contigs, a total of 221 full-length (Table S2) and 323 partial CYP genes were identified in *J. vulgaris*, and a total of 157 full-length (Table S3) with 247 partial CYP genes were identified in *J. aquatica*, respectively. All full-length CYPs were classified and named by Prof. Dr. David R. Nelson. The 221 full-length CYPs of *J. vulgaris* were divided into eight clans and 38 families (17 A-type families, 21 non-A-type families), while the 157 full-length CYPs of *J. aquatica* were divided into eight clans including 35 families (16 A-type families, 19 non-A-type families) (Table 2). Around half of the full-length CYP sequences of both *J. vulgaris* (53.8%) and *J. aquatica* (46.4%) were assigned to CYP71, CYP706, CYP76, CYP72, CYP82 and CYP93 families, of which only CYP72 is non-A-type. Compared with *J. vulgaris*, for *J. aquatica* less full-length CYPs were detected, which might be caused by the lower number of genotypes and the lower amount of reads in the *J. aquatica* samples. However, the proportional distributions of full-length CYPs were similar not only in each CYP clan (Chi-square = 1.6, Df = 8, NS), but also within each CYP family (Chi-square = 18.6, Df = 37, NS) of the two *Jacobaea* species (Table 2).

We compared the numbers of the detected full-length CYPs of *J. vulgaris* and *J. aquatica* with three other plant species, i.e. *Helianthus annuus*, *Lactuca sativa* and *Arabidopsis thaliana* (Table 2). All CYP genes of *H. annuus* and *L. sativa* were derived from genome sequencing projects (Badouin *et al*., 2017; Reyes-Chin-Wo *et al*., 2017) and were classified based on the best blast hits by Prof. Dr. David R. Nelson. Only CYPs longer than 400 amino acids were chosen in this study as the length of the most reliably annotated CYPs of *A. thaliana* ranges from 457 to 594 amino acids without taking pseudogenes into account. Roughly, the four species of the Asteraceae (*J. vulgaris* 544 (221 full-length and 323 partial CYPs), *J. aquatica* 404 (157 full-length and 247 partial CYP), *H. annuus* 462, *L. sativa* 374) contained more CYP genes than *A. thaliana* (244). It indicates an expansion and functional diversification of CYP genes encoding metabolic pathways in the Asteraceae during evolution and genome duplications.

Table 2. Distribution of full-length CYP450 genes over clans and families of *J. vulgaris* (*Jv*), *J. aquatica* (*Ja*), *H. annuus* (*Ha*), L. *sativa* (*Ls*) and *A. thaliana* (*At*).

clan	family	*Jv*	*Ja*	*Ha*	*Ls*	*At*
51	51	3	4	1	1	2
71	71	41	21	85	74	50
	73	2	4	3	2	1
	75	1	1	3	2	1
	76	14	12	30	25	8
	77	2	2	3	4	5
	78	5	5	8	7	6
	79	1	1	12	6	7
	80	1	0	10	5	0
	81	9	2	32	18	17
	82	11	15	26	32	5
	83	0	0	0	0	2
	84	4	3	7	2	2
	89	3	3	11	5	7
	92	5	2	2	4	0
	93	12	5	7	6	1
	98	5	6	2	2	3
	701	3	2	5	3	1
	703	0	0	1	1	1
	705	0	0	0	0	25
	706	25	12	26	27	7
	712	0	0	0	0	2
	736	0	0	2	5	0
72	72	16	13	40	25	9
	714	2	2	1	1	2
	715	0	0	1	1	1
	721	1	0	4	2	1
	734	0	0	3	2	1
	735	1	1	1	1	2
	749	7	4	6	4	0
74	74	7	4	6	7	2
85	85	1	1	2	1	2
	87	0	0	8	2	1
	88	1	0	1	2	2
	90	5	4	6	7	4
	702	0	0	0	0	6
	707	4	2	9	6	4
	708	0	0	0	0	4
	709	0	0	0	0	3
	716	2	3	24	12	2
	718	0	0	1	1	1
	720	1	1	1	1	1

85	722	1	1	1	2	1
	724	0	0	1	1	1
	728	0	0	6	4	0
	729	0	0	1	6	0
	733	0	0	1	1	0
86	86	6	5	9	7	11
	94	2	1	13	10	6
	96	3	3	15	16	13
	704	9	7	18	16	3
97	97	3	3	3	3	3
710	710	1	1	2	1	4
711	711	1	1	2	1	1
total		221	157	462	374	244

Overall, the distributions of CYPs among different CYP clans over the five species (Table 2) were comparable (Chi-square = 42.0, Df = 32, NS). However, the distributions among different CYP families were significantly different (Chi-square = 466.7, Df = 212, $P < 0.001$). Numbers of CYPs in single-family CYP clans (CYP51, CYP74, CYP97, CYP710, CYP711) were fairly consistent (Chi-square = 11.2, Df = 16, NS). The significant difference was caused by multiple-family clans (CYP71, CYP72, CYP85, CYP86) which parallel land plant evolution (Nelson and Werck-Reichhart, 2011) and which have expanded dramatically (Chi-square = 445.6, Df = 192, $P < 0.001$). In accordance with the statement of Nelson and Werck-Reichhart (Nelson and Werck-Reichhart, 2011), the youngest clan, the CYP71 clan (A-type), was dominant in all five species, of which the CYP71 family possessed the largest numbers of CYPs over all five species. Within the Asteraceae families, ten CYP families were absent in *Jacobaea* species compared with *H. annuu*s and *L. sativa*, including CYP703, CYP736, CYP715, CYP734, CYP87, CYP718, CYP724, CYP728, CYP729 and CYP733. Without further information, it is difficult to infer whether the absence/presence is an evolutionary consequence or just due to the unavailability of full-length transcripts in the transcriptomes of *Jacobaea*.

Characterization of CYP proteins

The lengths of 221 full-length proteins of *J. vulgaris* ranged from 460 to 601 amino acids, with an average length of 509 amino acids, and the lengths of 157 full-length proteins of *J. aquatica* varied from 464 to 601 amino acids with an average length of 511 amino acids. All the full-length CYP proteins (Tables S2 and S3) were subjected to Multiple Expectation Maximization for Motif Elicitation (MEME) analysis to identify motifs by A-type (the CYP71 clan) and non-A-type for each species. The sequence logos of the four typical conserved motifs including the heme-binding region, the PERF motif, the K-helix region and the I-helix region were extracted (Fig. 1). The consensus sequences of the motifs of *J. vulgaris* and *J. aquatica* were highly similar and also showed high similarities to other plant species (Chen *et al*., 2014; Liao *et al*., 2017; Qi *et al*., 2017) for both A-type and non-A-type CYP proteins. Furthermore, the

differences of signatures of typical motifs (i.e. the heme-binding region, the PERF and the I-helix) between A-type and non-A-type CYPs were also similar to those of other species. The consensus sequence of the heme-binding region of A-type CYPs was "PFGxGRRxCP", whereas "xFxxGxRxCxG" was found in non-A-type CYPs. The F, G and C residues are conserved in all plant P450s, where the C residue is universally conserved in all P450s across kingdoms and coordinates the iron in the heme. For the PERF motif, A-type CYPs displayed the consensus "FxPERF" while non-A-type CYPs showed "FxPxRx", both with one additional highly conserved "F" which exists in the majority of CYPs. The I-helix motifs of A-type and non-A-type CYPs were "AGxDT" and "AGx[D/E]TT", respectively. The consensus of the ExxR motif of A-type CYPs accorded with that of non-A-type CYPs. In line with previous studies (Chen *et al*., 2014; Liao *et al*., 2017; Qi *et al*., 2017), the results confirmed that plant CYP proteins share well-conserved motifs including the heme-binding signature, the PERF motif, the K-helix and the I-helix, which are essential for catalytic activity (Paquette *et al*., 2009).

KEGG pathway analysis of *Jacobaea* CYPs

KEGG pathway-based analysis was performed to understand the potential involvement of CYPs in various biosynthetic pathways. Hundred twenty four of the 221 (56.1%) full-length CYPs of *J. vulgaris* were designated to 37 KEGG Ortholog (KO) hierarchies (Table S2), which were distributed over 21 KEGG pathways (Fig. 2A). For *J. aquatica* 91 out of 157 (58.0%) full-length CYPs were appointed to 33 KO catalogs (Table S3) covering 20 KEGG pathways (Fig. 2B). In the class of "biosynthesis of other secondary metabolites", 21 CYPs were assigned to be involved in the biosynthesis of phenylpropanoids (K00487, K09754, K09755), stilbenoids, diarylheptanoids and gingerols (K00487, K09754), flavonoids (K00487, K05280, K09754), flavones and flavonols (K05280), isoflavonoids (K13260) and/or glucosinolates (K12153) for both *Jacobaea* species, of which some genes were assigned to more than one KEGG pathway. All these SM related CYPs belonged to the CYP71 clan. No genes were found to be involved in alkaloid biosynthesis. This does not necessarily mean that they are not involved in alkaloid biosynthesis because this may result from the fact that, although the KEGG database includes information about the alkaloid biosynthesis genes these are not specifically for PAs.

Phylogenetic analyses

Comparative sequence analysis based on an evolutionary perspective can improve functional prediction (Eisen and Wu, 2002). Therefore we performed phylogenetic analyses using the maximum likelihood method for the largest six families in *Jacobeae* species, namely, CYP71, CYP76, CYP706, CYP93, CYP82 and CYP72, based on their amino acid sequences (Fig. 3; Fig. S1-S5). Functional divergence frequently accompanies gene duplication, which was confirmed by our study. Lineage-specific expansion of CYPs was observed overall (Fig. 3; Fig. S1-S5). In all phylogenetic trees, the CYPs from the same species tended to be clustered together, resulting in many lineage-specific subfamilies and/or clades. In most CYP families,

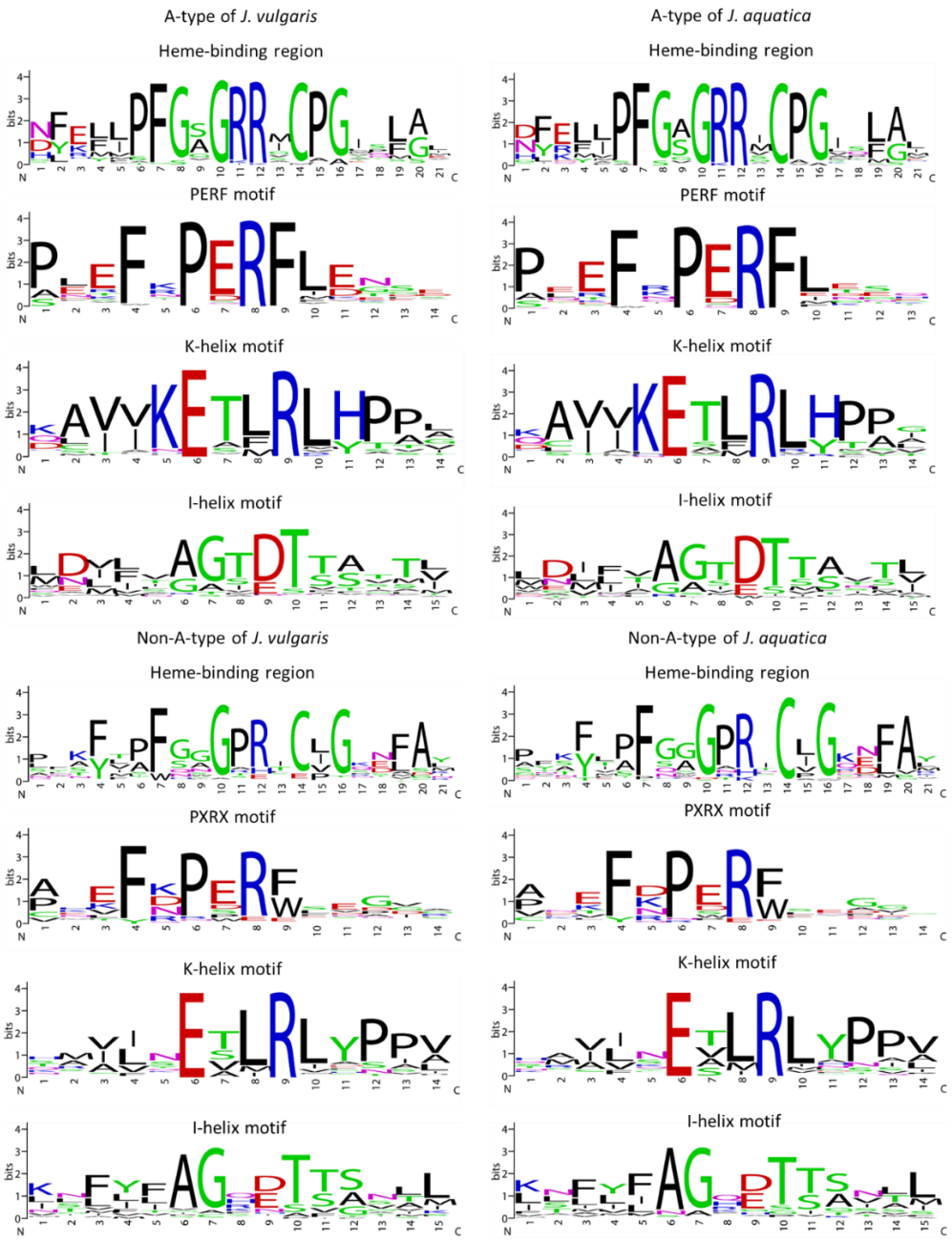

Figure 1. Weblogos of typical conserved motifs identified in the full-length CYP450s divided as A-type (the upper figure) and non-A-type (the lower figure) from *J. vulgaris* (left) and *J. aquatica* (right). The names of the motifs are shown above each logo. The bit score indicates the information content for each position in the sequence.

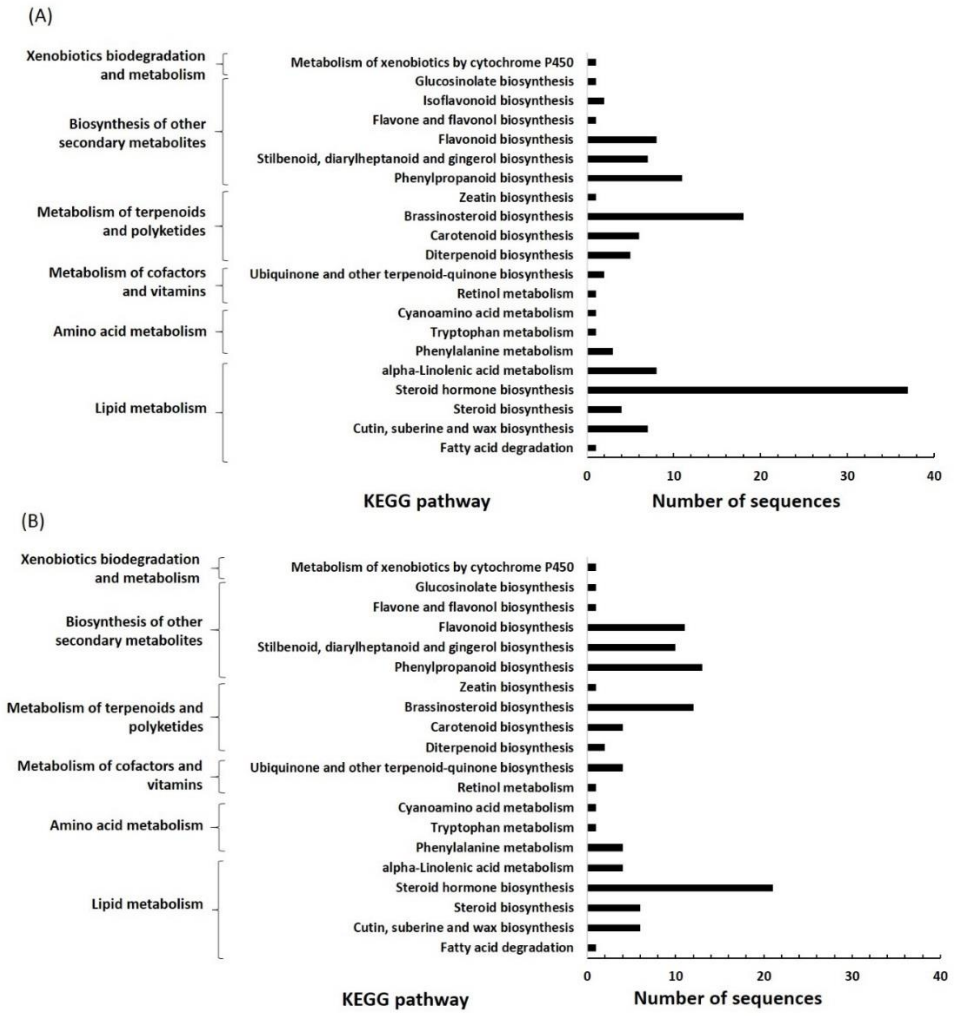

Figure 2. KEGG pathway analysis of predicted CYP450s in two *Jacobaea* species. (A) *J. vulgaris*. (B) *J. aquatica*. The numbers of CYP450 genes involved in the corresponding metabolic processes are shown.

CYPs were not equally distributed in different species, suggesting that gene duplication events happened after species divergence. Only within the *Jacobaea* species we observed that often a clade was present with a *J. vulgaris* and a *J. aquatica* CYP. Taking the CYP71 family as example, the CYPs of *A. thaliana* fell into two clades, whereas the CYPs of the Asteraceae species were divided into five distinct clades (Fig. 3). Notably, the speed of evolution of CYPs within the Asteraceae has been very fast resulting in species-specific CYPs. Particularly, the most basal clade of the Asteraceae, the CYP71AX subfamily has expanded dramatically. Even though the distributions of CYPs on the trees were more dispersed compared to those of *A.*

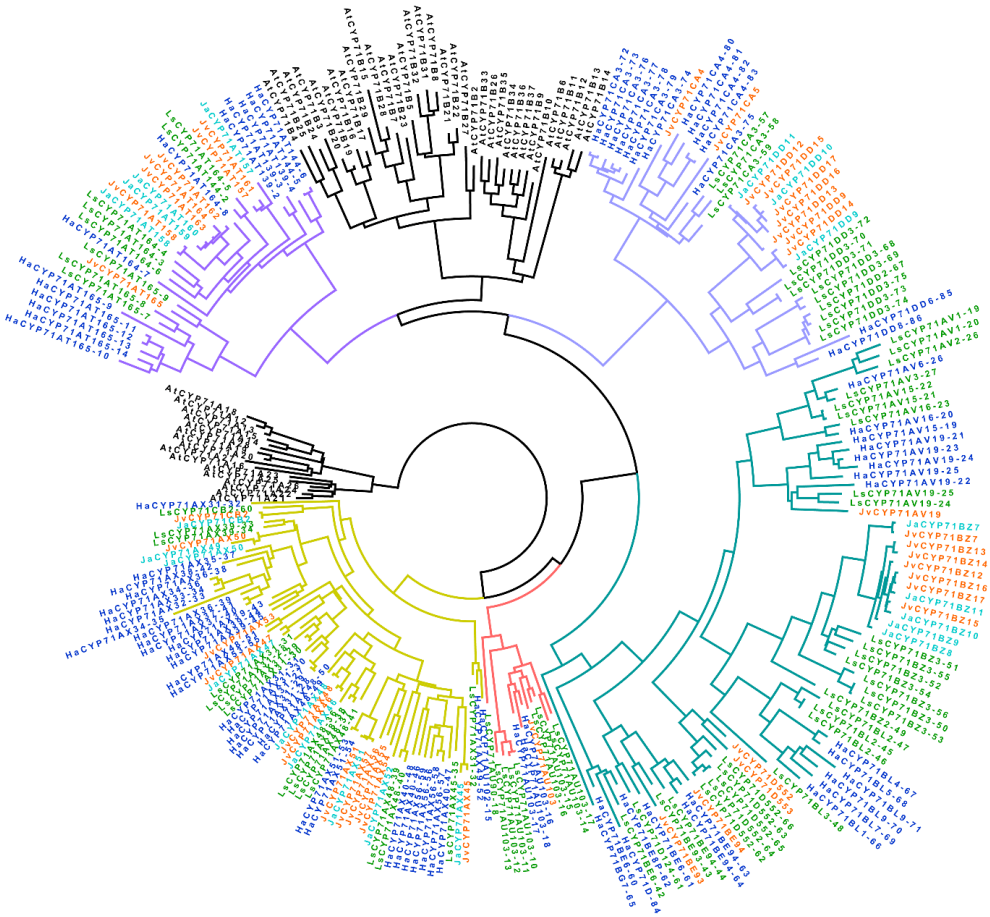

Figure 3. Phylogenetic tree of CYP71 family from five species inferred with the maximum likelihood method. CYP450s are color coded for different species: *J. vulgaris* (orange), *J. aquatica* (light blue), *H. annuus* (dark blue), *L. sativa* (green), *A. thaliana* (black). The branches of the five clades of the Asteraceae are color highlighted. The names of CYP450s of *H. annuus* and *L. sativa* were tentatively coded without nomenclature. *A. thaliana* was used as the outgroup.

thaliana, *Jacobaea* species, *H. annuus* and *L. sativa* all had their own lineage-specific subclades. Only for the closely related species *J. vulgaris* and *J. aquatica*, CYPs were found quite often in pairs, confirming a close relationship in the evolutionary history. For some CYPs of *J. vulgaris* the orthologs were missing in *J. aquatica* (Fig. 3; Figure S1), which might be caused by less available full-length CYPs of *J. aquatica* in this study or alternatively by the loss in *J. aquatica* or by the gain in *J. vulgaris* of particular CYPs during evolution.

Discussion

CYPs have an essential function in contributing to chemical diversity that is the landmark of plants (Nelson and Werck-Reichhart, 2011). However, as the largest family of enzymes engaged in primary and secondary metabolism and having a fast evolution, CYPs are notorious for their difficulty in classification and nomenclature, which hinders the study of these genes. In the current study, well-curated sets of CYPs with standard nomenclature were obtained for *J. vulgaris* and *J. aquatica*, which is vital for the functional characterization and comparison of these genes. In total, 221 and 157 full-length CYP genes were identified, classified and named from transcriptomes of *J. vulgaris* and *J. aquatica*, respectively.

KEGG pathway based annotation was performed for all full-length CYPs, and no CYPs were designated to alkaloid biosynthetic pathways. Empirically, CYPs from the same family/subfamily often catalyze similar/related reactions (Nelson and Werck-Reichhart, 2011). For example, the CYPs involved in the main reactions of benzylisoquinoline alkaloid diversity include CYP80 family (CYP80A1, CYP80B3, CYP80G2), CYP719 family (CYP719A20, CYP719A21, CYP719A25, CYP719B1) and CYP82 family (CYP82Y1, CYP82Y2, CYP82N4, CYP82X1, CYP82X2) (Dastmalchi *et al*., 2018). Nonetheless, consecutive steps in the same alkaloid pathways can be also catalyzed by CYPs from divergent families (Nelson and Werck-Reichhart, 2011). For instance, some of the functionally characterized CYPs involved in the monoterpenoid indole alkaloid pathway in *Catharanthus rosesus* are from different families: CYP71D2, CYP72A1, CYP76B6 (Schröder *et al*., 1999; Irmler S *et al*., 2000; Collu *et al*., 2001; Giddings *et al*., 2011). Alkaloids are highly species-specific SMs which are characterized by a vast structural diversity. Identifying a CYP catalyzing a particular biosynthetic step is challenging because of the homology shared by CYP proteins and the lack of correlation between primary structure and catalytic function (Mizutani and Ohta, 2010), especially since no CYPs involved in PA metabolism have been reported.

CYPs are an excellent reporter of plant evolution, especially in the evolution and role of plant metabolism. An evolutionary approach using phylogenetic trees could be beneficial to CYP function prediction (Nelson and Werck-Reichhart, 2011). The diversification of CYPs had a significant biochemical impact on the emergence of new metabolic pathways during the evolutionary process of land plants (Du *et al*., 2016). In the phylogenetic analyses of the most abundant CYP families of *Jacobaea*, a fast evolution of CYPs was observed resulting in lineage-specific expansion. Notably, CYPs do not always follow the pattern in which *H. annuus* showed a closer phylogenetic relatedness to *Jacobaea* species than *L. sativa* as indicated by Compositae metatrees (Funk *et al*., 2009), especially for CYPs in the CYP71 family. Quite often, CYPs in the CYP71 family of *H. annuus* and *L. sativa* switched phylogenetic closeness to those of *Jacobaea* species on the phylogenetic tree (Fig. 3). This suggests that species patterns in CYPs are present. Gene duplication is thought to be one of the major sources of evolutionary innovation, resulting in divergence in paralogs due to neofunctionalization or sub-functionalization (Conant and Wolfe, 2008; Nguyen *et al*., 2014). CYP members in multiple-family clans CYP71, CYP72 and CYP85 have enlarged

astonishingly, leading to the difficulty in predicting gene functions. However, those CYPs ending in the same clade/subclade in a phylogenetic tree might indicate association with metabolism of particular classes of compounds or similar reactions on different substrates (Nelson and Werck-Reichhart, 2011).

Based on our study, it is not possible to appoint CYP candidates involved in PA biosynthesis. Nonetheless, the collection of CYPs in *Jacobaea* species can speed up the exploration of function in following studies. As long as whole genome information of *Jacobaea* species is lacking, 5' Race and 3' Race techniques can be employed to obtain a more complete reservoir of full-length CYPs. The prediction of CYP candidates can be further facilitated by correlating gene expression patterns with PA abundances in plants grown under conditions that generate PA contrasts or in F_2 offspring segregating for PA profiles.

Here we detected 221 and 157 full-length CYPs for *J. vulgaris* and *J. aquatica*, respectively. Comparison of CYPs over five species showed strong lineage specific diversification of CYPs, indicating fast evolutionary speed of CYPs within the Asteraceae. Only in the closely related *J. vulgaris* and *J. aquatica*, CYPs were found quite often in pairs, confirming a close relationship in the evolutionary history. No genes were found to be involved in alkaloid biosynthesis against KEGG database. Finally, our study presents the first comprehensive overview of CYPs in *Jacobaea* species, which is beneficial for future exploration of their functions, including possible involvement in PA biosynthesis and PA diversity.

References

Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Eberhard C *et al*. 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Research* 44:W3-10.

Badouin H, Gouzy J, Grassa CJ, Murat F, Staton SE, Cottret L, Lelandais-Brière C, Owens GL, Carrère S, Mayjonade B, *et al*. 2017. The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature* 546:148-152.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research* 37:W202-208.

Bak S, Beisson F, Bishop G, Hamberger B, Höfera R, Paquette S, Werck-Reichhart. 2011. Cytochrome P450. *The Arabidopsis Book* 9:e0144.

Bennett RN, Wallsgrove RM. 1994. Secondary metabolites in plant defence mechanisms. *New Phytologist* 127:617-633.

Böttcher F, Adolph R-D, Hartmann T. 1993. Homospermidine synthase, the first pathway-specific enzyme in pyrrolizidine alkaloid biosynthesis. *Phytochemistry* 32:1373-1384.

Chen H, Wu B, Nelson DR, Wu K, Liu C. 2014. Computational identification and systematic classification of novel cytochrome P450 genes in *Salvia miltiorrhiza*. *PLOS One* 9:e115149.

Cheng D, Kirk H, Mulder PPJ, Vrieling K, Klinkhamer PGL. 2011. Pyrrolizidine alkaloid variation in shoots and roots of segregating hybrids between *Jacobaea vulgaris* and *Jacobaea aquatica*. *New Phytologist* 192:1010-1023.

Collu G, Unver N, Peltenburg-Looman AMG, van der Heijden R, Verpoorte R, Memelink J. 2001. Geraniol 10-hydroxylase, a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis. *FEBS Letters* 508:215-220.

Conant GC, Wolfe KH. 2008. Turning a hobby into a job: how duplicated genes find new functions. *Nature Reviews Genetics* 9:938-950.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Research* 14:1188-1190.

Dastmalchi M, Park MR, Morris JS, Facchini P. 2018. Family portraits: the enzymes behind benzylisoquinoline alkaloid diversity. *Phytochemistry Reviews* 17:249-277.

Du H, Ran F, Dong HL, Wen J, Li JN, Liang Z. 2016. Genome-wide analysis, classification, evolution, and expression analysis of the cytochrome P450 93 family in land plants. *PLOS One* 11:e0165020.

Eisen JA, Wu M. 2002. Phylogenetic analysis and gene functional predictions: phylogenomics in action. *Theoretical Population Biology* 61:481-487.

Emanuelsson O, Nielsen H, Brunak S, von Heijne G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology* 300:1005-1016.

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M. 2014. Pfam: the protein families database. *Nucleic Acids Research* 42:D222-230.

Frey M, Schullehner K, Dick R, Fiesselmann A, Gierl A. 2009. Benzoxazinoid biosynthesis, a model for evolution of secondary metabolic pathways in plants. *Phytochemistry* 70:1645-1651.

Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next generation sequencing data. *Bioinformatics* 28:3150-3152.

Funk VA, Anderberg AA, Baldwin BG, Bayer RJ, Bonifacino JM, Breitwieser I, Brouillet L, Carbajal R, Chan R, Coutinho AXP, *et al*. 2009. "Compositae metatrees: the next generation" in systematics, evolution, and biogeography of Compositae. In: Funk VA, Susana A, Stuessy TF, Bayer RJ, editors. Vienna, Austria: International Association for Plant Taxonomy (IAPT) p.747-777.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research* 31:3784-3788.

Giddings LA, Liscombe DK, Hamilton JP, Childs KL, DellaPenna D, Buell CR, O'Connor SE. 2011. A stereoselective hydroxylation step of alkaloid biosynthesis by a unique cytochrome P450 in *Catharanthus roseus*. *Journal of Biological Chemistry* 286:16751-16757.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, *et al*. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* 8:1494-1512.

Hamberger B, Bak S. 2013. Plant P450s as versatile drivers for evolution of species-specific chemical diversity. *Philosophical Transactions of the Royal Society B* 368:20120426.

Hartmann T, Dierich B. 1998. Chemical diversity and variation of pyrrolizidine alkaloids of the senecionine type: biological need or coincidence? *Planta* 206:443-451.

Hartmann T, Ehmke A, Eilert U, von Borstel K, Theuring C. 1989. Sites of synthesis, translocation and accumulation of pyrrolizidine alkaloid *N*-oxides in *Senecio vulgaris* L. *Planta* 177:98-107.

Hartmann T, Toppel G. 1987. Senecionine *N*-oxide, the primary product of pyrrolizidine alkaloid biosynthesis in root cultures of *Senecio vulgaris*. *Phytochemistry* 26:1639-1643.

Hori K, Yamada Y, Purwanto R, Minakuchi Y, Toyoda A, Hirakawa H, Sato F. 2018. Mining of the uncharacterized cytochrome P450 genes involved in alkaloid biosynthesis in California poppy using a draft genome sequence. *Plant Cell Physiology* 59:222-33.

Ilc T, Arista G, Tavares R, Navrot N, Duchêne E, Velt A, Choulet F, Paux E, Fischer M, Nelson DR, Hugueney P, Werch-Reichhart D, Rustenholz C. 2018. Annotation, classification, genomic organization and expression of the *Vitis vinifera* CYPome. *PLOS One* 13:e0199902.

Irmler S, Schröder G, St-Pierre B, Crouch NP, Hotze M, Schmidt J, Strack D, Matern U, Schröder J. 2000. Indole alkaloid biosynthesis in *Catharanthus roseus*: new enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase. *The Plant Journal* 24:797-804.

Joosten L, Cheng D, Mulder PPJ, Vrieling K, van Veen JA, Klinkhamer PGL. 2011. The genotype dependent presence of pyrrolizidine alkaloids as tertiary amine in *Jacobaea vulgaris*. *Phytochemistry* 72:214-222.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* 14:587-589.

Kessler A, Kalske A. 2018. Plant secondary metabolite diversity and species interactions. *Annual Review of Ecology, Evolution, and Systematics* 49:115-138.

Kliebenstein DJ, Kroymann J, Brown P, Figuth A, Pedersen D, Gershenzon J, Mitchell-Olds T. 2001. Genetic control of natural variation in *Arabidopsis* glucosinolate accumulation. *Plant Physiology* 126:811-825.

Kumar S, Stecher G, Tamura K 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* 33:1870-1874.

Langel D, Ober D, Pelser PB. 2011. The evolution of pyrrolizidine alkaloid biosynthesis and diversity in the Senecioneae. *Phytochemistry Reviews* 10:3-74.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9:357-359.

Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22:1658-1659.

Liao W, Zhao S, Zhang M, Dong K, Chen Y, Fu C, Yu L. 2017. Transcriptome assembly and systematic identification of novel cytochrome P450s in *Taxus chinensis*. *Frontiers in Plant Science* 8:1468.

Macel M, Vrieling K, Klinkhamer PGL. 2004. Variation in pyrrolizidine alkaloid patterns of *Senecio jacobaea*. *Phytochemistry* 65:865-873.

Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. New Orleans, LA: Gateway computing environments workshop (GCE) p.1-8.

Mizutani M, Ohta D. 2010. Diversification of P450 genes during land plant evolution. *Annual Review of Plant Biology* 61:291-315.

Mizutani M. 2012. Impacts of diversification of cytochrome P450 on plant metabolism. *Biological and Pharmaceutical Bulletin* 35:824-832.

Moore BD, Andrew RL, Külheim C, Foley WJ. 2014. Explaining intraspecific diversity in plant secondary metabolites in an ecological context. *New Phytologist* 201:733-750.

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research* 35:W182-185.

Nelson D, Werck-Reichhart D. 2011. A P450-centric view of plant evolution. *The Plant Journal* 66:194-211.

Nelson DR. 2009. The cytochrome P450 homepage. *Human Genomics* 4:59-60.

Nguyen Ba AN, Strome B, Hua JJ, Desmond J, Gagnon-Arsenault I, Weiss EL, Landry CR, Moses AM. 2014. Detecting functional divergence after gene duplication through evolutionary changes in posttranslational regulatory sequences. PLOS Computational Biology 10:e1003977.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Molecular Biology and Evolution* 32:268-274.

Ober D. 2010. Gene duplications and the time thereafter - examples from plant secondary metabolism. *Plant Biology* 12:570-577.

Paquette SM, Jensen K, Bak S. 2009. A web-based resource for the Arabidopsis P450, cytochromes b5, NADPH-cytochrome P450 reductases, and family 1 glycosyltransferases (http://www.P450.kvl.dk). *Phytochemistry* 70:1940-1947.

Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobitsch S, Lehrach H, Soldatov A. 2009. Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Research* 37:e123.

Pelser PB, de Vos H, Theuring C, Beuerle T, Vrieling K, Hartmann T. 2005. Frequent gain and loss of pyrrolizidine alkaloids in the evolution of *Senecio* section *Jacobaea* (Asteraceae). *Phytochemistry* 66:1285-1295.

Qi X, Yu X, Xu D, Fang H, Dong K, Li W, Liang C. 2017. Identification and analysis of CYP450 genes from transcriptome of *Lonicera japonica* and expression analysis of chlorogenic acid biosynthesis related CYP450s. *PeerJ* 5:e3781.

Reyes-Chin-Wo S, Wang Z, Yang X, Kozik A, Arikit S, Song C, Xia L, Froenicke L, Lavelle DO, Truco MJ, *et al*. 2017. Genome assembly with in vitro proximity ligation data and whole-genome triplication in lettuce. *Nature Communications* 8:14953.

Schneider TD, Stephens RM. 1990. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Research* 18:6097-6100.

Schröder G, Unterbusch E, Kaltenbach M, Schmidt J, Strack D, De Luca V, Schröder J. 1999. Light-induced cytochrome P450-dependent enzyme in indole alkaloid biosynthesis: tabersonine 16-hydroxylase. *FEBS Letters* 458:97-102.

van Dam NM, Vrieling K. 1994. Genetic variation in constitutive and inducible pyrrolizidine alkaloid levels in *Cynoglossum officinale* L. *Oecologia* 99:374-378.

Werck-Reichhart D, Bak S, Paquette S. 2002. Cytochromes P450. *The Arabidopsis Book* 1:e0028

Wink M. 2003. Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry* 64:3-19.

Ziegler J, Facchini PJ. 2008. Alkaloid biosynthesis: metabolism and trafficking. *Annual Review of Plant Biology* 59:735-769.

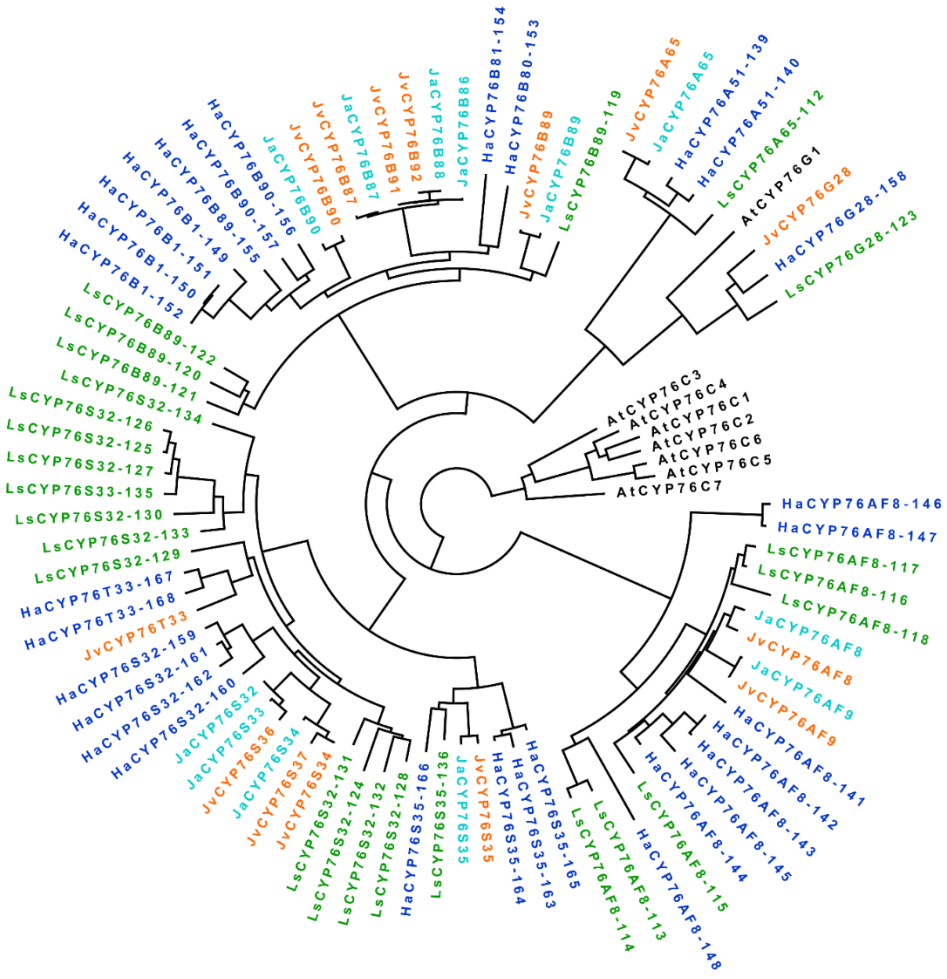

Figure S1. Phylogenetic tree of the CYP76 family from 5 species inferred with the maximum likelihood method. CYP450s are color coded for different species: *J. vulgaris* (orange), *J. aquatica* (light blue), *H. annuus* (dark blue), *L. sativa* (green), *A. thaliana* (black). The names of CYP450s of *H. annuus* and *L. sativa* were tentatively coded without nomenclature. *A. thaliana* was used as the outgroup.

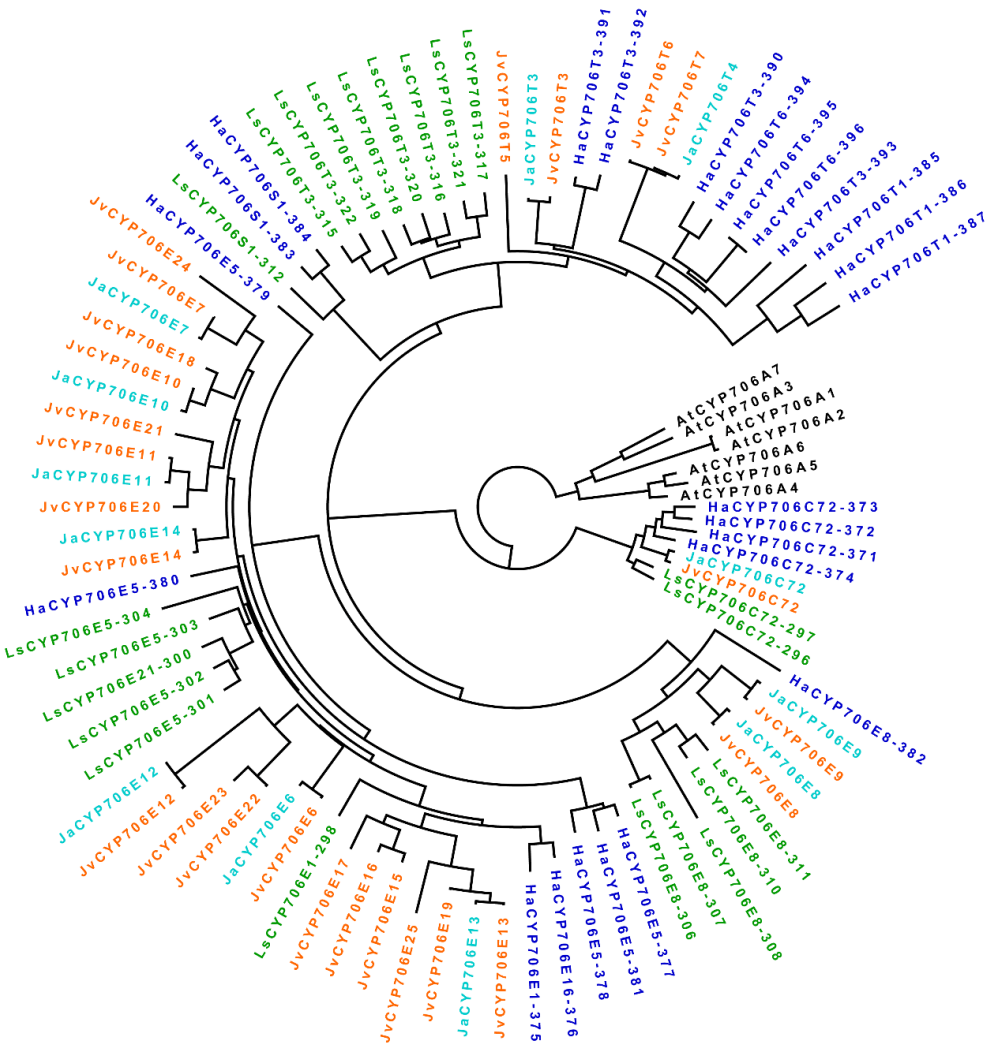

Figure S2. Phylogenetic tree of the CYP706 family from 5 species inferred with the maximum likelihood method. CYP450s are color coded for different species: *J. vulgaris* (orange), *J. aquatica* (light blue), *H. annuus* (dark blue), *L. sativa* (green), *A. thaliana* (black). The names of CYP450s of *H. annuus* and *L. sativa* were tentatively coded without nomenclature. *A. thaliana* was used as the outgroup.

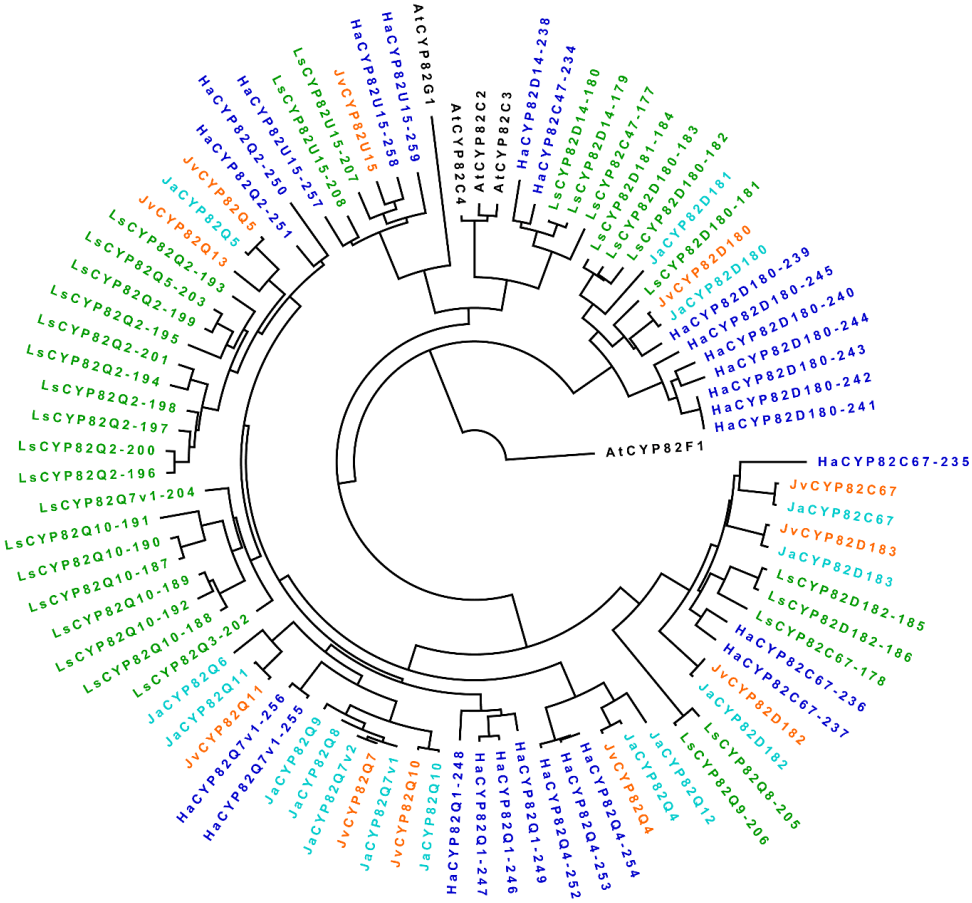

Figure S3. Phylogenetic tree of the CYP82 family from 5 species inferred with the maximum likelihood method. CYP450s are color coded for different species: *J. vulgaris* (orange), *J. aquatica* (light blue), *H. annuus* (dark blue), *L. sativa* (green), *A. thaliana* (black). The names of CYP450s of *H. annuus* and *L. sativa* were tentatively coded without nomenclature. *A. thaliana* was used as the outgroup.

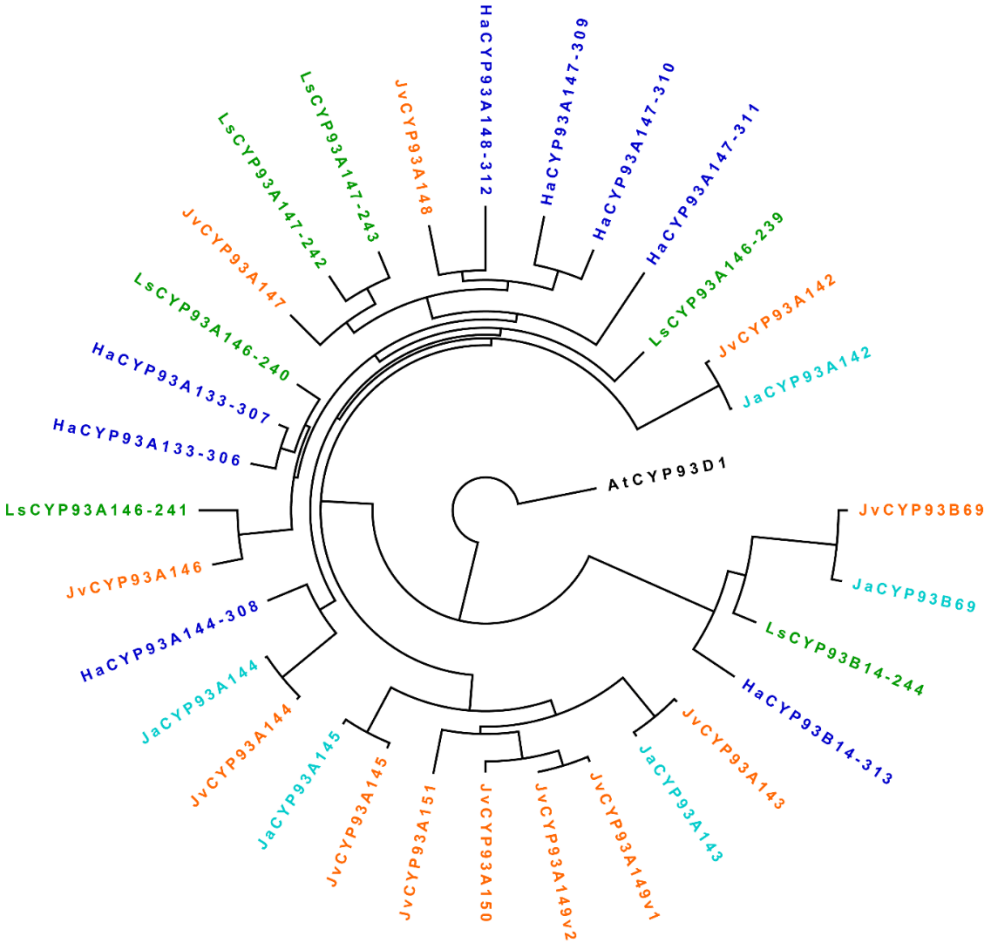

Figure S4. Phylogenetic tree of the CYP93 family from 5 species inferred with the maximum likelihood method. CYP450s are color coded for different species: *J. vulgaris* (orange), *J. aquatica* (light blue), *H. annuus* (dark blue), *L. sativa* (green), *A. thaliana* (black). The names of CYP450s of *H. annuus* and *L. sativa* were tentatively coded without nomenclature. *A. thaliana* was used as the outgroup.

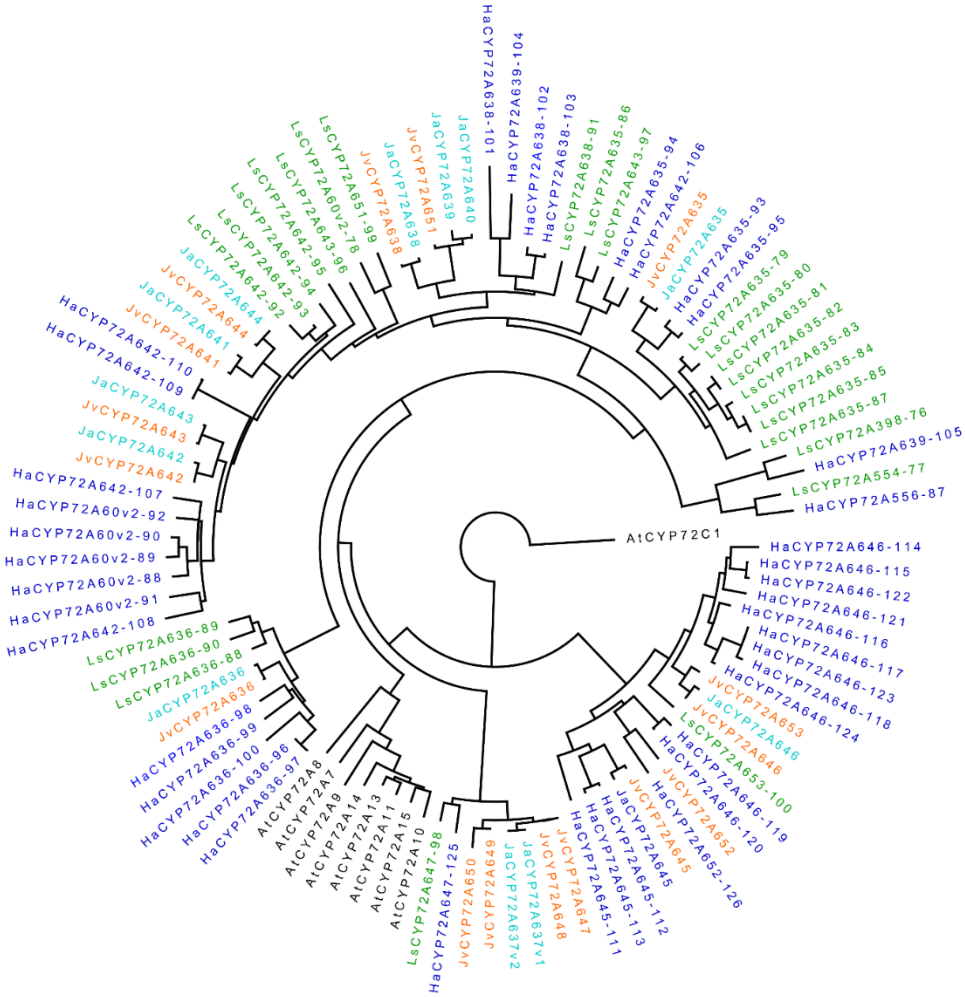

Figure S5. Phylogenetic tree of the CYP72 family from 5 species inferred with the maximum likelihood method. CYP450s are color coded for different species: *J. vulgaris* (orange), *J. aquatica* (light blue), *H. annuus* (dark blue), *L. sativa* (green), *A. thaliana* (black). The names of CYP450s of *H. annuus* and *L. sativa* were tentatively coded without nomenclature. *A. thaliana* was used as the outgroup.

Table S1. Details of sample sets for deep sequencing analysis

Species	Sets (Assemblies)	Samples	Organs	Origins	Individuals	Genotypes	cDNA library
J. vulgaris	*Jv*1	*Jv*1	Shoots, roots	Meijendel, The Netherlands (derived from tissue culture)	2	1	Normalized
				Csokvaomany, Hungary	5	5	
				Wageningen, The Netherlands	6	6	
				Bertogne, Belgium	5	5	
				Dülmen, Germany	6	6	
				Meijendel, The Netherlands	7	7	
				Gabriola island, Canada	7	7	
				Den Helder, The Netherlands	7	7	
				Bilthoven, The Netherlands	7	7	
				close to Filly, Belgium	7	7	
	*Jv*2	*Jv*-MeJA	Shoots	Meijendel, The Netherlands (tissue culture)	5	1	Non-normalized
		Jv-Control	Shoots	Meijendel, The Netherlands (tissue culture)	5	1	Non-normalized
J. aquatica	*Ja*1	*Ja*1	Shoots, roots	The Netherlands (derived from tissue culture)	2	1	Normalized
				Avon, England, UK	5	5	
	*Ja*2	*Ja*2	Shoots	The Netherlands (derived from tissue culture)	2	1	Non-normalized
				Avon, England, UK	5	5	

Table S2. List of full-length CYP450s of *J. vulgaris* identified in this study.

Name	Type	Clan	Family	Subfamily	Length in aa	Theoretical pI	Mol. Wt (kDa)	Target[a]	KO[b]
*Jv*CYP51G43v1	non-A	51	CYP51	CYP51G	484	8.65	55.0	S	K05917
*Jv*CYP51G43v2	non-A	51	CYP51	CYP51G	484	8.33	55.1	S	K05917
*Jv*CYP51G44	non-A	51	CYP51	CYP51G	485	8.83	55.2	*	K05917
*Jv*CYP71D552	A	71	CYP71	CYP71D	500	7.22	56.4	S	K00512
*Jv*CYP71D553	A	71	CYP71	CYP71D	502	5.85	56.3	S	K00512
*Jv*CYP71AT157	A	71	CYP71	CYP71AT	499	8.5	57.5	S	na[c]
*Jv*CYP71AT158	A	71	CYP71	CYP71AT	501	9.01	57.3	S	na
*Jv*CYP71AT161	A	71	CYP71	CYP71AT	501	8.54	57.7	S	na
*Jv*CYP71AT162	A	71	CYP71	CYP71AT	473	8.78	54.1	S	na
*Jv*CYP71AT163	A	71	CYP71	CYP71AT	502	6.54	57.5	S	na
*Jv*CYP71AT164	A	71	CYP71	CYP71AT	501	6.82	57.4	*	na
*Jv*CYP71AT165	A	71	CYP71	CYP71AT	496	8.56	56.6	S	na
*Jv*CYP71AU103	A	71	CYP71	CYP71AU	493	8.98	56.2	*	na
*Jv*CYP71AV19	A	71	CYP71	CYP71AV	500	8.74	56.6	*	na
*Jv*CYP71AX45	A	71	CYP71	CYP71AX	495	8.17	55.4	S	na
*Jv*CYP71AX46	A	71	CYP71	CYP71AX	499	8.44	57.4	*	na
*Jv*CYP71AX47	A	71	CYP71	CYP71AX	500	8.21	56.8	S	na
*Jv*CYP71AX48	A	71	CYP71	CYP71AX	497	8.22	56.7	*	na
*Jv*CYP71AX50	A	71	CYP71	CYP71AX	492	8.59	56.0	S	K07408
*Jv*CYP71AX51	A	71	CYP71	CYP71AX	502	8.84	56.2	S	na
*Jv*CYP71AX53	A	71	CYP71	CYP71AX	500	6.9	56.3	S	na
*Jv*CYP71AX54	A	71	CYP71	CYP71AX	504	8.46	56.7	S	na
*Jv*CYP71AX55	A	71	CYP71	CYP71AX	500	8.93	56.3	S	na
*Jv*CYP71AX56	A	71	CYP71	CYP71AX	498	5.99	56.4	S	na
*Jv*CYP71AX57	A	71	CYP71	CYP71AX	502	8.22	56.2	S	na
*Jv*CYP71BE93	A	71	CYP71	CYP71BE	502	6.91	56.0	S	K00512
*Jv*CYP71BE94	A	71	CYP71	CYP71BE	493	8.44	55.4	S	K00512
*Jv*CYP71BZ7	A	71	CYP71	CYP71BZ	500	7.2	56.8	*	K00512
*Jv*CYP71BZ12	A	71	CYP71	CYP71BZ	502	8.86	57.0	S	K00512
*Jv*CYP71BZ13	A	71	CYP71	CYP71BZ	476	8.45	54.3	S	K00512
*Jv*CYP71BZ14	A	71	CYP71	CYP71BZ	499	8.6	56.8	S	K00512
*Jv*CYP71BZ15	A	71	CYP71	CYP71BZ	499	8.54	56.5	S	K00512
*Jv*CYP71BZ16	A	71	CYP71	CYP71BZ	502	8.71	56.9	S	K00512
*Jv*CYP71BZ17	A	71	CYP71	CYP71BZ	499	8.6	56.7	S	K00512
*Jv*CYP71CA4	A	71	CYP71	CYP71CA	513	6.83	57.9	S	na
*Jv*CYP71CA5	A	71	CYP71	CYP71CA	516	7.66	58.1	S	K00512
*Jv*CYP71CB2	A	71	CYP71	CYP71CB	498	8.51	56.5	S	na
*Jv*CYP71DD9	A	71	CYP71	CYP71DD	509	8.79	58.2	*	na
*Jv*CYP71DD12	A	71	CYP71	CYP71DD	508	8.6	57.9	S	na

*Jv*CYP71DD13	A	71	CYP71	CYP71DD	508	8.49	58.0	S	na
*Jv*CYP71DD14	A	71	CYP71	CYP71DD	505	8.97	57.7	*	na
*Jv*CYP71DD15	A	71	CYP71	CYP71DD	474	8.27	53.7	S	na
*Jv*CYP71DD16	A	71	CYP71	CYP71DD	508	8.65	57.7	S	na
*Jv*CYP71DD17	A	71	CYP71	CYP71DD	512	7.59	58.1	S	na
*Jv*CYP72A635	non-A	72	CYP72	CYP72A	519	9.27	59.1	S	na
*Jv*CYP72A636	non-A	72	CYP72	CYP72A	519	9.03	59.3	S	na
*Jv*CYP72A638	non-A	72	CYP72	CYP72A	523	8.84	60.2	S	na
*Jv*CYP72A641	non-A	72	CYP72	CYP72A	516	8.9	58.9	S	na
*Jv*CYP72A642	non-A	72	CYP72	CYP72A	516	9.12	59.0	S	na
*Jv*CYP72A643	non-A	72	CYP72	CYP72A	516	8.65	59.3	S	na
*Jv*CYP72A644	non-A	72	CYP72	CYP72A	515	9.06	58.8	S	na
*Jv*CYP72A645	non-A	72	CYP72	CYP72A	515	8.37	58.4	S	na
*Jv*CYP72A646	non-A	72	CYP72	CYP72A	514	8.13	58.7	S	na
*Jv*CYP72A647	non-A	72	CYP72	CYP72A	514	9.27	59.0	S	na
*Jv*CYP72A648	non-A	72	CYP72	CYP72A	506	9.22	58.1	S	na
*Jv*CYP72A649	non-A	72	CYP72	CYP72A	514	9.29	58.9	*	na
*Jv*CYP72A650	non-A	72	CYP72	CYP72A	515	9.58	59.1	*	na
*Jv*CYP72A651	non-A	72	CYP72	CYP72A	517	9.39	59.8	S	na
*Jv*CYP72A652	non-A	72	CYP72	CYP72A	516	8.93	58.2	S	na
*Jv*CYP72A653	non-A	72	CYP72	CYP72A	516	7.29	58.8	S	na
*Jv*CYP73A213	A	71	CYP73	CYP73A	505	8.98	58.0	S	K00487
*Jv*CYP73A212	A	71	CYP73	CYP73A	505	9.08	57.9	S	K00487
*Jv*CYP74A102	non-A	74	CYP74	CYP74A	474	6.44	53.2	_	K01723
*Jv*CYP74A103	non-A	74	CYP74	CYP74A	534	8.96	60.0	*	K01723
*Jv*CYP74A104	non-A	74	CYP74	CYP74A	477	6.13	53.9	*	K01723
*Jv*CYP74A105	non-A	74	CYP74	CYP74A	482	6.4	54.4	*	K01723
*Jv*CYP74A106	non-A	74	CYP74	CYP74A	474	6.11	53.6	_	K01723
*Jv*CYP74A107	non-A	74	CYP74	CYP74A	530	9.07	59.5	C	K01723
*Jv*CYP74B35	non-A	74	CYP74	CYP74B	500	7.05	55.8	C	K10528
*Jv*CYP75B131	A	71	CYP75	CYP75B	508	8.45	55.8	*	K05280
*Jv*CYP76A65	A	71	CYP76	CYP76A	518	8.31	58.8	S	K20618
*Jv*CYP76B87	A	71	CYP76	CYP76B	494	8.72	55.2	S	K20556
*Jv*CYP76B89	A	71	CYP76	CYP76B	494	9	55.7	S	K20556
*Jv*CYP76B90	A	71	CYP76	CYP76B	497	8.83	55.5	S	K20556
*Jv*CYP76B91	A	71	CYP76	CYP76B	494	8.09	55.1	S	K20556
*Jv*CYP76B92	A	71	CYP76	CYP76B	494	7.23	55.1	S	K20556
*Jv*CYP76G28	A	71	CYP76	CYP76G	518	9.36	59.2	*	K20618
*Jv*CYP76S34	A	71	CYP76	CYP76S	497	8.47	56.5	S	K20556
*Jv*CYP76S35	A	71	CYP76	CYP76S	506	8.32	57.1	*	K20556
*Jv*CYP76S36	A	71	CYP76	CYP76S	497	8.53	56.8	*	K20556
*Jv*CYP76S37	A	71	CYP76	CYP76S	497	6.96	56.3	S	K20556
*Jv*CYP76T33	A	71	CYP76	CYP76T	506	8.49	57.9	S	K20556

*Jv*CYP76AF8	A	71	CYP76	CYP76AF	490	6.16	55.6	S	K20556
*Jv*CYP76AF9	A	71	CYP76	CYP76AF	492	6.48	55.6	*	K20556
*Jv*CYP77A54	A	71	CYP77	CYP77A	503	8.99	57.1	*	K21995
*Jv*CYP77A55	A	71	CYP77	CYP77A	506	8.85	56.8	S	K21995
*Jv*CYP78A276	A	71	CYP78	CYP78A	523	8.17	58.6	*	K20619
*Jv*CYP78A277	A	71	CYP78	CYP78A	519	6.71	57.2	*	K20619
*Jv*CYP78A278	A	71	CYP78	CYP78A	515	9.44	58.2	S	K20619
*Jv*CYP78A279	A	71	CYP78	CYP78A	500	8.96	56.5	S	na
*Jv*CYP78A280	A	71	CYP78	CYP78A	515	8.75	58.1	S	na
*Jv*CYP79D82	A	71	CYP79	CYP79D	511	8.54	58.8	*	K12153
*Jv*CYP80AA1	A	71	CYP80	CYP80AA	482	8.98	54.8	*	na
*Jv*CYP81B107	A	71	CYP81	CYP81B	503	8.32	56.6	S	na
*Jv*CYP81B108	A	71	CYP81	CYP81B	502	7.63	57.2	*	na
*Jv*CYP81B109	A	71	CYP81	CYP81B	505	7.68	57.4	S	na
*Jv*CYP81B110	A	71	CYP81	CYP81B	507	8.99	58.2	S	na
*Jv*CYP81B111	A	71	CYP81	CYP81B	499	8.5	56.7	*	K13260
*Jv*CYP81B112	A	71	CYP81	CYP81B	503	8.31	57.0	S	na
*Jv*CYP81B113	A	71	CYP81	CYP81B	505	8.01	58.0	*	na
*Jv*CYP81BG9	A	71	CYP81	CYP81BG	506	6.66	57.7	S	K13260
*Jv*CYP81BZ1	A	71	CYP81	CYP81BZ	502	8.62	56.9	S	na
*Jv*CYP82C67	A	71	CYP82	CYP82C	521	7.24	58.7	S	K00512
*Jv*CYP82D180	A	71	CYP82	CYP82D	535	7.2	60.3	S	na
*Jv*CYP82D182	A	71	CYP82	CYP82D	526	6.19	58.5	S	K00512
*Jv*CYP82D183	A	71	CYP82	CYP82D	521	8.17	58.4	*	K00512
*Jv*CYP82Q4	A	71	CYP82	CYP82Q	525	7.27	59.7	S	na
*Jv*CYP82Q5	A	71	CYP82	CYP82Q	527	8.96	59.8	S	na
*Jv*CYP82Q7	A	71	CYP82	CYP82Q	529	6.66	59.2	S	na
*Jv*CYP82Q10	A	71	CYP82	CYP82Q	529	8.41	59.1	S	na
*Jv*CYP82Q11	A	71	CYP82	CYP82Q	523	6.98	59.4	S	na
*Jv*CYP82Q13	A	71	CYP82	CYP82Q	522	8.88	59.2	S	na
*Jv*CYP82U15	A	71	CYP82	CYP82U	524	8.86	59.6	*	K17961
*Jv*CYP84A116	A	71	CYP84	CYP84A	505	6.05	56.9	S	K09755
*Jv*CYP84A117	A	71	CYP84	CYP84A	506	5.86	56.9	S	K09755
*Jv*CYP84A118	A	71	CYP84	CYP84A	506	5.95	57.1	S	K09755
*Jv*CYP84A119	A	71	CYP84	CYP84A	503	6.21	56.5	*	K09755
*Jv*CYP85A1	non-A	85	CYP85	CYP85A	467	9.13	53.8	S	K09590
*Jv*CYP86A165	non-A	86	CYP86	CYP86A	526	8.57	59.9	S	K15398
*Jv*CYP86A166	non-A	86	CYP86	CYP86A	516	9.27	58.6	S	K15401
*Jv*CYP86A167	non-A	86	CYP86	CYP86A	534	8.97	61.0	S	K15398
*Jv*CYP86B47	non-A	86	CYP86	CYP86B	547	8.9	62.5	_	K15402
*Jv*CYP86B48	non-A	86	CYP86	CYP86B	544	9.16	62.6	*	K15402
*Jv*CYP86B	non-A	86	CYP86	CYP86B	543	8.96	62.3	*	K15402
*Jv*CYP88A100	non-A	85	CYP88	CYP88A	491	9.27	56.5	S	K04123

*Jv*CYP89A176	A	71	CYP89	CYP89A	520	9.32	58.8	*	na
*Jv*CYP89A177	A	71	CYP89	CYP89A	518	8.6	58.4	*	na
*Jv*CYP89A179	A	71	CYP89	CYP89A	525	9.26	59.8	S	na
*Jv*CYP90A62	non-A	85	CYP90	CYP90A	477	9.19	54.9	S	K09588
*Jv*CYP90A61	non-A	85	CYP90	CYP90A	480	9.03	54.9	S	K09588
*Jv*CYP90B54	non-A	85	CYP90	CYP90B	491	8.63	56.3	S	K09587
*Jv*CYP90C32	non-A	85	CYP90	CYP90C	492	9.16	56.5	S	K12637
*Jv*CYP90D53	non-A	85	CYP90	CYP90D	498	9.03	57.2	S	K12638
*Jv*CYP92A161	A	71	CYP92	CYP92A	511	8.07	58.5	*	K20623
*Jv*CYP92A162	A	71	CYP92	CYP92A	500	9.14	57.1	*	K20623
*Jv*CYP92A163	A	71	CYP92	CYP92A	509	8.27	57.7	S	K20623
*Jv*CYP92A164	A	71	CYP92	CYP92A	510	8.11	58.4	*	K20623
*Jv*CYP92A165	A	71	CYP92	CYP92A	506	8.03	57.8	S	K20623
*Jv*CYP93A142	A	71	CYP93	CYP93A	505	7.67	57.3	S	na
*Jv*CYP93A143	A	71	CYP93	CYP93A	511	6.95	57.6	S	na
*Jv*CYP93A144	A	71	CYP93	CYP93A	507	8.76	57.9	S	K00512
*Jv*CYP93A145	A	71	CYP93	CYP93A	514	6.41	58.3	S	K00512
*Jv*CYP93A146	A	71	CYP93	CYP93A	508	8.96	58.0	S	na
*Jv*CYP93A147	A	71	CYP93	CYP93A	507	6.72	57.4	*	na
*Jv*CYP93A148	A	71	CYP93	CYP93A	503	9.15	56.7	*	na
*Jv*CYP93A149v1	A	71	CYP93	CYP93A	509	8.73	57.7	S	na
*Jv*CYP93A149v2	A	71	CYP93	CYP93A	510	8.58	57.5	S	na
*Jv*CYP93A150	A	71	CYP93	CYP93A	508	7.18	57.1	S	na
*Jv*CYP93A151	A	71	CYP93	CYP93A	524	7	60.0	S	na
*Jv*CYP93B69	A	71	CYP93	CYP93B	513	8.43	58.2	S	na
*Jv*CYP94A89	non-A	86	CYP94	CYP94A	494	8.84	56.9	*	K20769
*Jv*CYP94D110	non-A	86	CYP94	CYP94D	505	8.25	58.1	S	na
*Jv*CYP96A154	non-A	86	CYP96	CYP96A	509	8.56	58.2	S	na
*Jv*CYP96A153	non-A	86	CYP96	CYP96A	491	8.29	56.1	*	na
*Jv*CYP96A156	non-A	86	CYP96	CYP96A	511	9.01	59.7	*	na
*Jv*CYP97A65	non-A	97	CYP97	CYP97A	601	6.17	67.0	*	K15747
*Jv*CYP97B59	non-A	97	CYP97	CYP97B	576	6.78	64.8	C	na
*Jv*CYP97C48	non-A	97	CYP97	CYP97C	545	6.85	61.3	C	K09837
*Jv*CYP98A114	A	71	CYP98	CYP98A	508	7.25	57.5	*	K09754
*Jv*CYP98A116	A	71	CYP98	CYP98A	517	8.79	58.4	S	K09754
*Jv*CYP98A118	A	71	CYP98	CYP98A	509	8.91	57.7	S	K09754
*Jv*CYP98A119v1	A	71	CYP98	CYP98A	511	8.8	57.9	S	K09754
*Jv*CYP98A119v2	A	71	CYP98	CYP98A	511	8.66	57.9	S	K09754
*Jv*CYP701A73	A	71	CYP701	CYP701A	506	6.88	58.0	S	K04122
*Jv*CYP701A74	A	71	CYP701	CYP701A	509	6.18	57.4	*	K04122
*Jv*CYP701A75	A	71	CYP701	CYP701A	511	6.91	57.2	*	K04122
*Jv*CYP704A181	non-A	86	CYP704	CYP704A	523	8.29	60.4	S	na
*Jv*CYP704A182	non-A	86	CYP704	CYP704A	505	8.29	58.5	S	na

*Jv*CYP704A183	non-A	86	CYP704	CYP704A	498	8.65	57.4	S	na
*Jv*CYP704A185	non-A	86	CYP704	CYP704A	517	8.7	60.0	S	na
*Jv*CYP704A186	non-A	86	CYP704	CYP704A	511	8.85	59.8	S	na
*Jv*CYP704A187	non-A	86	CYP704	CYP704A	481	8.97	55.9	S	na
*Jv*CYP704A188	non-A	86	CYP704	CYP704A	482	8.9	55.9	*	na
*Jv*CYP704A189	non-A	86	CYP704	CYP704A	529	8.11	61.0	S	na
*Jv*CYP704A190	non-A	86	CYP704	CYP704A	494	7.17	56.7	S	na
*Jv*CYP706C72	A	71	CYP706	CYP706C	524	7.68	59.0	*	K00512
*Jv*CYP706E6	A	71	CYP706	CYP706E	543	5.86	62.7	*	K00512
*Jv*CYP706E7	A	71	CYP706	CYP706E	534	7.2	61.3	*	K00512
*Jv*CYP706E8	A	71	CYP706	CYP706E	523	8.66	59.7	*	na
*Jv*CYP706E9	A	71	CYP706	CYP706E	526	8.97	59.9	S	na
*Jv*CYP706E10	A	71	CYP706	CYP706E	530	7.7	60.5	_	K00512
*Jv*CYP706E11	A	71	CYP706	CYP706E	524	7.16	59.7	*	K00512
*Jv*CYP706E12	A	71	CYP706	CYP706E	516	7.19	59.3	S	K00512
*Jv*CYP706E13	A	71	CYP706	CYP706E	529	8.14	60.6	_	K00512
*Jv*CYP706E14	A	71	CYP706	CYP706E	539	8.09	61.8	*	K00512
*Jv*CYP706E15	A	71	CYP706	CYP706E	536	6.82	62.2	_	na
*Jv*CYP706E16	A	71	CYP706	CYP706E	532	6.45	61.6	_	na
*Jv*CYP706E17	A	71	CYP706	CYP706E	530	6.39	60.8	_	K00512
*Jv*CYP706E18	A	71	CYP706	CYP706E	530	5.91	60.6	_	K00512
*Jv*CYP706E19	A	71	CYP706	CYP706E	531	8.61	60.8	_	K00512
*Jv*CYP706E20	A	71	CYP706	CYP706E	551	8.3	63.0	_	K00512
*Jv*CYP706E21	A	71	CYP706	CYP706E	535	6.4	61.2	*	K00512
*Jv*CYP706E22	A	71	CYP706	CYP706E	524	6.66	60.3	_	na
*Jv*CYP706E23	A	71	CYP706	CYP706E	497	5.6	57.5	_	na
*Jv*CYP706E24	A	71	CYP706	CYP706E	526	6.59	60.5	*	K00512
*Jv*CYP706E25	A	71	CYP706	CYP706E	532	8.34	61.0	_	K00512
*Jv*CYP706T5	A	71	CYP706	CYP706T	526	8.16	60.4	*	K00512
*Jv*CYP706T3	A	71	CYP706	CYP706T	529	8.6	60.1	*	K00512
*Jv*CYP706T6	A	71	CYP706	CYP706T	533	7.1	60.2	*	K00512
*Jv*CYP706T7	A	71	CYP706	CYP706T	533	6.07	60.0	*	K00512
*Jv*CYP707A180	non-A	85	CYP707	CYP707A	464	9.34	52.9	S	K09843
*Jv*CYP707A181	non-A	85	CYP707	CYP707A	487	9.08	55.4	*	K09843
*Jv*CYP707A182	non-A	85	CYP707	CYP707A	460	9.1	52.5	S	K09843
*Jv*CYP707A183	non-A	85	CYP707	CYP707A	462	9.46	52.9	S	K09843
*Jv*CYP710A100	non-A	710	CYP710	CYP710A	508	8	57.8	S	K09832
*Jv*CYP711A77	non-A	711	CYP711	CYP711A	512	8.69	58.4	S	K20771
*Jv*CYP714A36	non-A	72	CYP714	CYP714A	523	9.07	59.4	S	K20661
*Jv*CYP714A37	non-A	72	CYP714	CYP714A	523	9.11	59.4	S	K20661
*Jv*CYP716C56	non-A	85	CYP716	CYP716C	488	9.35	55.3	S	K20667
*Jv*CYP716D65	non-A	85	CYP716	CYP716D	478	9.19	54.2	S	K20667
*Jv*CYP720A1	non-A	85	CYP720	CYP720A	476	9.08	53.9	S	na

*Jv*CYP721A64	non-A	72	CYP721	CYP721A	504	9.23	58.3	S	na
*Jv*CYP722C11	non-A	85	CYP722	CYP722C	486	9.28	55.4	S	na
*Jv*CYP735A47	non-A	72	CYP735	CYP735A	519	8.85	59.2	S	K10717
*Jv*CYP749A91	non-A	72	CYP749	CYP749A	511	9.14	58.4	S	K15639
*Jv*CYP749A92	non-A	72	CYP749	CYP749A	514	8.5	59.1	S	K15639
*Jv*CYP749A94	non-A	72	CYP749	CYP749A	515	8.88	58.7	S	K15639
*Jv*CYP749A95	non-A	72	CYP749	CYP749A	509	9.36	58.3	S	K15639
*Jv*CYP749A96	non-A	72	CYP749	CYP749A	512	8.97	58.5	S	K15639
*Jv*CYP749A97	non-A	72	CYP749	CYP749A	512	8.81	58.4	S	K15639
*Jv*CYP749A98	non-A	72	CYP749	CYP749A	503	8.95	57.4	S	K15639

[a]Cellular location of the protein predicted by TargetP. “C” chloroplast; “S” secreted ; “_” any other location; “*” unknown.
[b]KEGG Orthology.
[c]not available.

Table S3. List of full-length CYP450s of *J. aquatica* identified in this study.

Name	Type	Clan	Family	Subfamily	Length in aa	Theoretical pI	Mol. Wt (kDa)	TargetP[a]	KO[b]
*Ja*CYP51G43v1	non-A	51	CYP51	CYP51G	485	8.76	55.2	S	K05917
*Ja*CYP51G43v2	non-A	51	CYP51	CYP51G	485	8.29	55.1	S	K05917
*Ja*CYP51G44v1	non-A	51	CYP51	CYP51G	486	8.52	55.1	*	K05917
*Ja*CYP51G44v2	non-A	51	CYP51	CYP51G	486	8.86	55.2	*	K05917
*Ja*CYP71AT157	A	71	CYP71	CYP71AT	509	7.99	58.5	S	na[c]
*Ja*CYP71AT158	A	71	CYP71	CYP71AT	501	8.82	57.1	S	na
*Ja*CYP71AT159	A	71	CYP71	CYP71AT	503	6.49	57.6	S	na
*Ja*CYP71AT160	A	71	CYP71	CYP71AT	508	6.54	58.1	S	na
*Ja*CYP71AX45	A	71	CYP71	CYP71AX	501	6.72	56.2	S	na
*Ja*CYP71AX46	A	71	CYP71	CYP71AX	500	8.63	57.5	*	na
*Ja*CYP71AX47	A	71	CYP71	CYP71AX	499	9.03	56.5	S	na
*Ja*CYP71AX48	A	71	CYP71	CYP71AX	498	8.82	56.6	*	na
*Ja*CYP71AX49	A	71	CYP71	CYP71AX	495	9.2	56.2	*	na
*Ja*CYP71AX50	A	71	CYP71	CYP71AX	493	8.13	55.7	S	K07408
*Ja*CYP71AX51	A	71	CYP71	CYP71AX	503	8.7	56.0	S	na
*Ja*CYP71AX52	A	71	CYP71	CYP71AX	492	9.01	56.0	*	na
*Ja*CYP71BZ7	A	71	CYP71	CYP71BZ	500	6.52	56.8	*	K00512
*Ja*CYP71BZ8	A	71	CYP71	CYP71BZ	500	8.65	56.6	S	K00512
*Ja*CYP71BZ9	A	71	CYP71	CYP71BZ	500	8.44	56.7	S	K00512
*Ja*CYP71BZ10	A	71	CYP71	CYP71BZ	500	8.68	56.7	S	K00512
*Ja*CYP71BZ11	A	71	CYP71	CYP71BZ	498	8.41	56.5	S	K00512
*Ja*CYP71CB2	A	71	CYP71	CYP71CB	498	7.72	56.6	S	na
*Ja*CYP71DD9	A	71	CYP71	CYP71DD	519	8.9	59.4	_	na
*Ja*CYP71DD10	A	71	CYP71	CYP71DD	513	8.23	58.2	S	na
*Ja*CYP71DD11	A	71	CYP71	CYP71DD	509	8.21	58.0	S	na
*Ja*CYP72A635	non-A	72	CYP72	CYP72A	520	9.27	59.1	S	na
*Ja*CYP72A636	non-A	72	CYP72	CYP72A	519	8.93	59.1	S	na
*Ja*CYP72A637v1	non-A	72	CYP72	CYP72A	515	9.17	59.0	S	na
*Ja*CYP72A637v2	non-A	72	CYP72	CYP72A	507	9.05	58.0	S	na
*Ja*CYP72A638	non-A	72	CYP72	CYP72A	522	8.87	60.2	S	na
*Ja*CYP72A639	non-A	72	CYP72	CYP72A	518	9.23	59.5	S	na
*Ja*CYP72A640	non-A	72	CYP72	CYP72A	518	9.32	59.6	S	na
*Ja*CYP72A641	non-A	72	CYP72	CYP72A	518	8.9	58.9	S	na
*Ja*CYP72A642	non-A	72	CYP72	CYP72A	516	9.18	58.7	S	na
*Ja*CYP72A643	non-A	72	CYP72	CYP72A	516	8.51	59.2	S	na
*Ja*CYP72A644	non-A	72	CYP72	CYP72A	515	9.12	58.9	S	na
*Ja*CYP72A645	non-A	72	CYP72	CYP72A	515	7.26	58.2	S	na
*Ja*CYP72A646	non-A	72	CYP72	CYP72A	515	8.31	58.7	S	na
*Ja*CYP73A212v1	A	71	CYP73	CYP73A	506	9.08	58.1	S	K00487
*Ja*CYP73A212v2	A	71	CYP73	CYP73A	506	9.08	58.1	S	K00487

*Ja*CYP73A213v1	A	71	CYP73	CYP73A	506	9.06	58.0	S	K00487
*Ja*CYP73A213v2	A	71	CYP73	CYP73A	506	9.06	58.1	S	K00487
*Ja*CYP74A102	non-A	74	CYP74	CYP74A	475	6.44	53.5	_	K01723
*Ja*CYP74A103	non-A	74	CYP74	CYP74A	538	9.18	60.3	*	K01723
*Ja*CYP74A104	non-A	74	CYP74	CYP74A	478	6.13	53.9	*	K01723
*Ja*CYP74B35	non-A	74	CYP74	CYP74B	501	7.05	55.7	C	K10528
*Ja*CYP75B131	A	71	CYP75	CYP75B	509	8.17	55.8	*	K05280
*Ja*CYP76A65	A	71	CYP76	CYP76A	519	8.5	58.8	S	K20618
*Ja*CYP76B86	A	71	CYP76	CYP76B	495	8.55	55.3	S	K20556
*Ja*CYP76B87	A	71	CYP76	CYP76B	495	8.8	55.2	S	K20556
*Ja*CYP76B88	A	71	CYP76	CYP76B	495	8.54	55.2	S	K20556
*Ja*CYP76B89	A	71	CYP76	CYP76B	495	8.8	55.6	S	K20556
*Ja*CYP76B90	A	71	CYP76	CYP76B	502	8.73	56.4	S	K20556
*Ja*CYP76S32	A	71	CYP76	CYP76S	498	8.97	56.8	*	K20556
*Ja*CYP76S33	A	71	CYP76	CYP76S	498	8.47	56.7	*	K20556
*Ja*CYP76S34	A	71	CYP76	CYP76S	498	8.83	56.6	S	K20556
*Ja*CYP76S35	A	71	CYP76	CYP76S	508	8.77	57.2	*	K20556
*Ja*CYP76AF8	A	71	CYP76	CYP76AF	490	5.99	55.3	S	K20556
*Ja*CYP76AF9	A	71	CYP76	CYP76AF	493	6.66	55.6	*	K20556
*Ja*CYP77A54	A	71	CYP77	CYP77A	503	8.95	57.1	*	K21995
*Ja*CYP77A55	A	71	CYP77	CYP77A	507	8.85	56.5	S	K21995
*Ja*CYP78A276	A	71	CYP78	CYP78A	524	7.75	58.6	*	K20619
*Ja*CYP78A277	A	71	CYP78	CYP78A	520	6.41	57.3	*	K20619
*Ja*CYP78A278	A	71	CYP78	CYP78A	516	9.44	58.2	S	na
*Ja*CYP78A279	A	71	CYP78	CYP78A	525	9.09	59.1	S	na
*Ja*CYP78A280	A	71	CYP78	CYP78A	526	9.1	59.3	S	na
*Ja*CYP79D81	A	71	CYP79	CYP79D	539	9.1	61.5	*	K12153
*Ja*CYP81B107	A	71	CYP81	CYP81B	504	8.61	56.7	*	na
*Ja*CYP81B108	A	71	CYP81	CYP81B	502	7.17	57.2	*	na
*Ja*CYP82C67	A	71	CYP82	CYP82D	522	7.7	58.6	S	K00512
*Ja*CYP82D180	A	71	CYP82	CYP82D	534	8.74	60.1	S	na
*Ja*CYP82D181	A	71	CYP82	CYP82D	524	6.3	59.5	S	na
*Ja*CYP82D182	A	71	CYP82	CYP82D	535	6.78	59.6	S	K00512
*Ja*CYP82D183	A	71	CYP82	CYP82D	518	6.91	57.9	*	K00512
*Ja*CYP82Q4	A	71	CYP82	CYP82Q	524	7.27	59.5	S	na
*Ja*CYP82Q5	A	71	CYP82	CYP82Q	527	8.9	59.7	S	na
*Ja*CYP82Q6	A	71	CYP82	CYP82Q	526	5.97	59.0	S	na
*Ja*CYP82Q7v1	A	71	CYP82	CYP82Q	531	6.69	59.3	S	na
*Ja*CYP82Q7v2	A	71	CYP82	CYP82Q	529	7.67	59.1	S	na
*Ja*CYP82Q8	A	71	CYP82	CYP82Q	528	7.2	59.0	S	na
*Ja*CYP82Q9	A	71	CYP82	CYP82Q	523	8.13	58.7	S	na
*Ja*CYP82Q10	A	71	CYP82	CYP82Q	530	8.55	59.2	S	na
*Ja*CYP82Q11	A	71	CYP82	CYP82Q	521	7.69	59.0	S	na
*Ja*CYP82Q12	A	71	CYP82	CYP82Q	527	8.46	59.8	S	na

*Ja*CYP84A115	A	71	CYP84	CYP84A	504	6.33	56.4	*	K09755
*Ja*CYP84A116	A	71	CYP84	CYP84A	506	5.95	56.9	S	K09755
*Ja*CYP84A117	A	71	CYP84	CYP84A	507	5.95	56.9	S	K09755
*Ja*CYP85A1	non-A	85	CYP85	CYP85A	466	9.01	53.5	S	K09590
*Ja*CYP86A165	non-A	86	CYP86	CYP86A	527	8.82	59.8	S	K15398
*Ja*CYP86A166	non-A	86	CYP86	CYP86A	515	9.2	58.5	S	K15401
*Ja*CYP86A167	non-A	86	CYP86	CYP86A	537	8.92	61.2	S	K15398
*Ja*CYP86B47	non-A	86	CYP86	CYP86B	545	8.4	62.3	_	K15402
*Ja*CYP86B48	non-A	86	CYP86	CYP86B	545	9	62.5	*	K15402
*Ja*CYP89A176	A	71	CYP89	CYP89A	518	9.17	58.5	*	na
*Ja*CYP89A177	A	71	CYP89	CYP89A	519	8.13	58.5	*	na
*Ja*CYP89A178	A	71	CYP89	CYP89A	509	8.87	58.0	*	na
*Ja*CYP90A61	non-A	85	CYP90	CYP90A	481	8.83	55.0	S	K09588
*Ja*CYP90A62	non-A	85	CYP90	CYP90A	478	9.02	54.8	S	K09588
*Ja*CYP90B54	non-A	85	CYP90	CYP90B	492	8.66	56.2	S	K09587
*Ja*CYP90C32	non-A	85	CYP90	CYP90C	493	9.15	56.5	S	K12637
*Ja*CYP92A161	A	71	CYP92	CYP92A	512	6.59	58.6	*	K20623
*Ja*CYP92A162	A	71	CYP92	CYP92A	501	9.17	57.0	*	K20623
*Ja*CYP93A142	A	71	CYP93	CYP93A	506	8.33	57.3	S	na
*Ja*CYP93A143	A	71	CYP93	CYP93A	512	6.95	57.5	S	na
*Ja*CYP93A144	A	71	CYP93	CYP93A	508	na[b]	na	S	K00512
*Ja*CYP93A145	A	71	CYP93	CYP93A	513	6.72	58.3	S	K00512
*Ja*CYP93B69	A	71	CYP93	CYP93B	516	8.69	58.7	S	na
*Ja*CYP94A89	non-A	86	CYP94	CYP94A	495	8.73	56.9	*	K13407
*Ja*CYP96A153	non-A	86	CYP96	CYP96A	492	8.04	56.0	*	na
*Ja*CYP96A154	non-A	86	CYP96	CYP96A	510	8.01	58.1	S	na
*Ja*CYP96A155	non-A	86	CYP96	CYP96A	510	8.84	58.7	S	na
*Ja*CYP97A65	non-A	97	CYP97	CYP97A	602	6.04	67.2	*	K15747
*Ja*CYP97B59	non-A	97	CYP97	CYP97B	577	6.78	64.7	C	na
*Ja*CYP97C48	non-A	97	CYP97	CYP97C	551	6.49	61.9	C	K09837
*Ja*CYP98A114	A	71	CYP98	CYP98A	509	7.72	57.5	*	K09754
*Ja*CYP98A115v1	A	71	CYP98	CYP98A	512	8.91	57.9	S	K09754
*Ja*CYP98A115v2	A	71	CYP98	CYP98A	512	8.8	57.9	S	K09754
*Ja*CYP98A116	A	71	CYP98	CYP98A	513	8.91	57.9	S	K09754
*Ja*CYP98A117	A	71	CYP98	CYP98A	510	8.79	57.8	S	K09754
*Ja*CYP98A118	A	71	CYP98	CYP98A	511	8.92	57.8	S	K09754
*Ja*CYP701A73	A	71	CYP701	CYP701A	507	7.69	58.0	S	K04122
*Ja*CYP701A74	A	71	CYP701	CYP701A	510	5.92	57.3	*	K04122
*Ja*CYP704A179	non-A	86	CYP704	CYP704A	511	8.96	59.5	S	na
*Ja*CYP704A180	non-A	86	CYP704	CYP704A	512	8.44	59.7	S	na
*Ja*CYP704A181	non-A	86	CYP704	CYP704A	527	8.81	60.7	S	na
*Ja*CYP704A182	non-A	86	CYP704	CYP704A	506	8.28	58.4	S	na
*Ja*CYP704A183	non-A	86	CYP704	CYP704A	495	7.16	57.0	S	na
*Ja*CYP704A184	non-A	86	CYP704	CYP704A	526	7.21	60.6	S	na

*Ja*CYP704A185	non-A	86	CYP704	CYP704A	533	7.66	61.9	S	na
*Ja*CYP706C72	A	71	CYP706	CYP706C	525	8.41	59.0	*	K00512
*Ja*CYP706E6	A	71	CYP706	CYP706E	542	6.14	62.5	*	K00512
*Ja*CYP706E7	A	71	CYP706	CYP706E	535	6.69	61.0	*	K00512
*Ja*CYP706E8	A	71	CYP706	CYP706E	524	8.48	59.6	*	na
*Ja*CYP706E9	A	71	CYP706	CYP706E	526	8.86	59.9	S	na
*Ja*CYP706E10	A	71	CYP706	CYP706E	539	7.7	61.2	*	K00512
*Ja*CYP706E11	A	71	CYP706	CYP706E	526	8.46	59.7	*	K00512
*Ja*CYP706E12	A	71	CYP706	CYP706E	517	6.52	59.2	S	K00512
*Ja*CYP706E13	A	71	CYP706	CYP706E	518	8.43	59.1	_	K00512
*Ja*CYP706E14	A	71	CYP706	CYP706E	540	8.08	61.8	*	K00512
*Ja*CYP706T3	A	71	CYP706	CYP706T	532	8.57	60.2	S	K00512
*Ja*CYP706T4	A	71	CYP706	CYP706T	534	8.2	60.1	*	K00512
*Ja*CYP707A180	non-A	85	CYP707	CYP707A	465	9.33	52.9	S	K09843
*Ja*CYP707A181	non-A	85	CYP707	CYP707A	488	9.05	55.3	C	K09843
*Ja*CYP710A100	non-A	710	CYP710	CYP710	509	7.63	57.9	S	K09832
*Ja*CYP711A77	non-A	711	CYP711	CYP711A	511	8.73	58.1	S	K20771
*Ja*CYP714A36	non-A	72	CYP714	CYP714A	523	9.16	59.2	S	K20661
*Ja*CYP714A37	non-A	72	CYP714	CYP714A	523	9.11	59.2	S	K20661
*Ja*CYP716D63	non-A	85	CYP716	CYP716D	475	9.02	54.0	*	K20667
*Ja*CYP716D64	non-A	85	CYP716	CYP716D	485	8.86	55.4	S	K20667
*Ja*CYP716D65	non-A	85	CYP716	CYP716D	515	9.14	58.4	_	K20667
*Ja*CYP720A1	non-A	85	CYP720	CYP720A	477	9.2	53.9	S	na
*Ja*CYP722C11	non-A	85	CYP722	CYP722C	486	9.22	55.4	S	na
*Ja*CYP735A47	non-A	72	CYP735	CYP735A	519	8.77	59.2	S	K10717
*Ja*CYP749A91	non-A	72	CYP749	CYP749A	512	9.14	58.4	S	K15639
*Ja*CYP749A92	non-A	72	CYP749	CYP749A	515	8.74	59.1	S	K15639
*Ja*CYP749A93	non-A	72	CYP749	CYP749A	513	8.97	58.4	S	K15639
*Ja*CYP749A94	non-A	72	CYP749	CYP749A	515	8.97	58.7	S	K15639

[a]Cellular location of the protein predicted by TargetP. “C” chloroplast; “S” secreted ; “_” any other location; “*” unknown.
[b]KEGG Orthology.
[c]not available.