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

Chen, Y.

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

by Yangan Chen

陈燕柑

Chen, Yangan

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

PhD thesis Leiden University, The Netherlands

Front cover: adapted from Figure 3 of Chapter 2

Thesis lay-out by Yangan Chen

Printing and binding: Boekendeal.nl

ISBN: 978-94-61550-64-4

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

Proefschrift

ter verkrijging van

de graad van Doctor aan de Universiteit Leiden,

op gezag van de Rector Magnificus prof.mr. C.J.J.M. Stolker,

volgens besluit van het College voor Promoties

te verdedigen op woensdag 29 januari 2020

klokke 13.45 uur

door

Yangan Chen

geboren te Zhangping, Fujian, China

in 1987

Promotores

prof.dr. J. Memelink, prof.dr. P.G.L. Klinkhamer

Co-promotor

dr. K. Vrieling

Promotiecommissie

prof.dr. G.P. van Wezel (voorzitter)

prof.dr. R. Offringa (secretaris)

prof.dr. N.M. van Dam (iDIV, Leipzig, Germany)

dr. D. Cheng (China University of Geosciences, Wuhan, China)

Contents

Chapter 1

General introduction

General introduction

Secondary metabolites (SMs) are ubiquitous in the plant kingdom. So far, an estimated 200,000 SMs have been isolated and identified from plants (Kessler and Kalske, 2018). The total number of SMs in plants is estimated to be more than 500,000, so many more SMs await to be discovered (Hadacek, 2002). Often a distinction is made between primary metabolites which are directly involved in plants' growth, development and reproduction and SMs. The latter, also referred to as small molecules are known to play a role in a plant's defense against abiotic and biotic stresses although for many of SMs the function is still unknown (Fig. 1; Hartmann, 1996; 1999). Plant SMs function as defense compounds deterring herbivores and pathogens, and as signaling compounds attracting pollinators, predators and parasitoids against herbivores and mediating interactions with mycorrhizal fungi and beneficial bacterial and with neighboring plants (Kessler and Halitschke, 2007). In addition, SMs also protect plants against abiotic stressors such as UV light, drought and frost (Isah, 2019). Due to the large number, high structural diversity and multifunctionality of SMs, it is still an ongoing challenge to understand how this SM diversity comes about, and why such a large diversity is maintained in nature (Moore *et al.*, 2014; Kessler and Kalske, 2018).

In this thesis this question was studied using the pyrrolizidine alkaloids (PAs) of *Jacobaea* species as the study system from an evolutionary and biosynthetic perspective.

1. Secondary metabolite (SMs)

1.1 SM diversity

1.1.1 SM diversity among chemical classes

The isolated and identified SMs have been assigned into different chemical classes (Fig. 1; Hartmann, 1996). The SMs containing nitrogen are classified as glucosinolates, cyanogenic glycosides or alkaloids. Chemical classes of SMs without nitrogen contain amongst others terpenes, polyketides, flavonoids, quinones, coumarins and tannins. These chemical classes are further divided into subgroups. For example, alkaloids are one of the most diverse compound classes with approximately over 25,000 structures found in higher plants. They are usually classified according to the nature of basic chemical structures. Following this system, heterocyclic alkaloids are further subdivided into groups such as isoquinoline, indole, quinolone, quinazoline, pyrrolizidine, and tropane alkaloids. (Funayama and Cordell, 2015).

1.1.2 SM diversity within chemical classes

SM diversity within compound classes is very common. Each class of SMs mostly possesses a number of similar molecules showing the same skeleton while differing in substitution groups by adding a number of polar and non-polar substituents. This structural diversity is well documented by abundant examples. For instance, 34 glucosinolates were reported in *Arabidopsis thaliana* (Kliebenstein *et al.*, 2001), which are molecules consisting of a *β*-

thioglucose moiety, a sulfonated oxime, and a variable side chain derived from various amino acids. Thirty-seven structurally related PAs have been detected in *Jacobaea vulgaris*, *Jacobaea aquatica* and their hybrids (Cheng *et al.*, 2011a), where most PAs are assumed to be derived from senecionine *N*-oxide with minor chemical modifications (Hartmann and Dierich, 1998; Pelser *et al.*, 2005). Apart from the structural diversity, SMs often show remarkable quantitative variation. This is well exemplified by the abovementioned glucosinolates which showed about 20-fold difference in total concentrations among leaves of different ecotypes (Kliebenstein *et al.*, 2001).

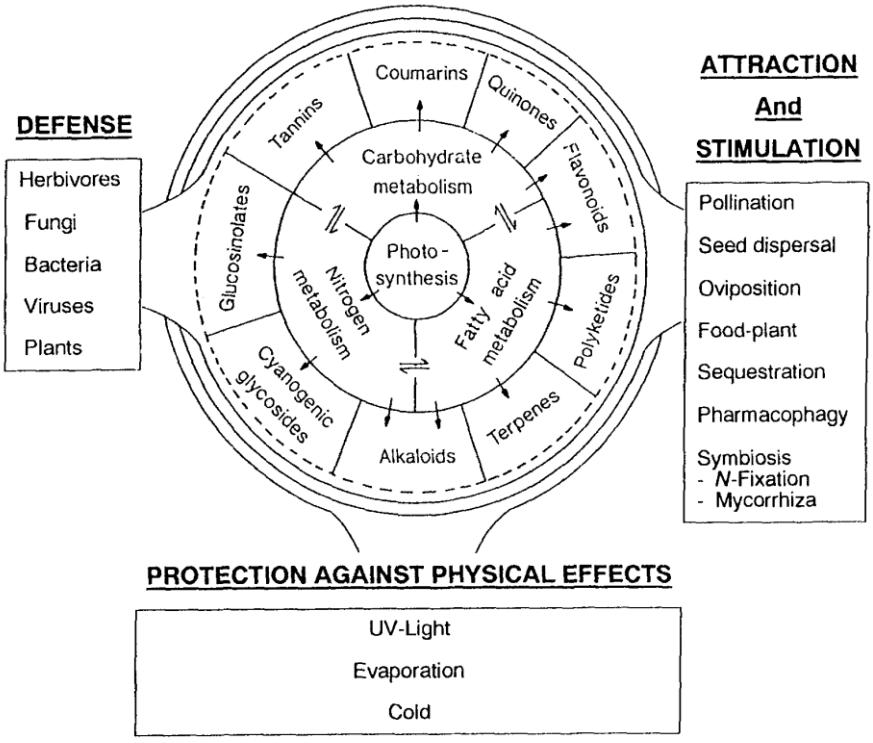

Figure 1. Classification and ecological functions of plant secondary metabolites (from Hartmann, 1996).

1.2 Distribution patterns of SMs

1.2.1 Distribution patterns of particular chemical classes

Some classes of SMs have a wide distribution among members of different plant phyla. For example, terpenes (isoprenes) responsible for plants' fragrance comprise one of the most diverse compound classes (Kessler and Kalske, 2018). Flavonoids comprise over 6000 different chemical moieties found in virtually all plants and fruits (Havsteen, 2002; Cvorovi *et al.*, 2018). Nonetheless, most classes of SMs are restricted to specific plants or plant lineages

(Ober, 2010), implying a strong phylogenetic signal although some exceptions can be observed. For instance, glucosinolates are major SMs near-universally in Brassicaceae, 1Capparidaeae 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 PAs distribute preferably in the Asteraceae, the Boraginaceae, the Fabaceae and the Orchidaceae families (Hartmann and Witte, 1995; Langel *et al.*, 2011).

1.2.2 Distribution patterns of SMs within a particular class

Commonly within families or clades, the sort of SM class used by plants tends to be conserved so that most genera and species produce chemicals that fall into a particular conserved class. However, the presence or absence of a particular SM within a particular class in phylogenetically related taxa is commonplace (Pelser *et al.*, 2005; Mint Evolutionary Genomics Consortium, 2018). In another words, the distribution of different chemical modifications of the basic structures on phylogenetic trees are often random. For example, the presence of monoterpenes exhibited a random distribution across the phylogeny of Lamiaceae, where 14 out of 57 showed significant phylogenetic signals (Mint Evolutionary Genomics Consortium, 2018). Whether this patchy distribution of SMs is the result of convergent evolution or of differential gene regulation is still open to debate.

1.3 Biochemical roots of plant SM diversity

With the development of biochemistry and molecular biology, questions about how such high SM diversity come about is partially answered. Scientists have summarized five hypotheses regarding to biochemical processes explaining the mechanisms underlying biosynthetic diversity and unique SM bouquets among and within plant individuals and species (reviewed by Kessler and Kalske, 2018): (i) simple chemical precursors originating from primary metabolism enable many possible combinations based on those subunits, (ii) a great diversity of genes are encoding a similar type of functional enzymes, catalyzing a diverse bouquet of SMs from common precursors, (iii) some multifunctional biosynthetic enzymes can generate a number of products out of the same precursor as a plastic system, (iv) the low substrate specificity and rapid functional divergence of modifying enzymes that are from large gene families, (v) spatially and temporally differential gene expression of biosynthesis genes can influence SM diversity of a plant quantitatively and qualitatively. These five mechanisms may contribute to SM diversity individually or additively.

1.3 Hypotheses explaining plant SM diversity

SMs are essential for plants' survival and reproductive fitness. SMs are adaptive traits that are subject to natural selection during evolution (Wink, 2003). Although how and why chemical diversity is maintained over evolutionary time is still not well understood, several hypotheses have been put forward to explain SM diversity. The screening hypothesis (Jones and Firn, 1991; Firn and Jones, 2003; 2009) states that the likelihood of producing new bioactive

compounds is enhanced with a great diversity of mostly inactive and inexpensive compounds. This hypothesis states that SMs arise via mutation with an inherently low probability of possessing any biological activity. However, genes encoding certain SM biosynthetic pathways are found to be grouped in gene clusters within plant genomes, which suggests strong selection in driving functional modularity of SM biosynthesis that is not derived from a simple breakdown process (Nützmann *et al.*, 2016). The "evolutionary arms race" between plants and herbivores is also a widely known theory (Ehrlich and Raven, 1964). Plants can escape from herbivory by evolving novel defense compounds, entering a new adaptive zone. Conversely, herbivores adapted to the novel SMs gain greater fitness and reciprocal evolutionary changes occur subsequently in the plants. Here we can find a striking analogy with mankind being in an evolutionary arms race with specialist herbivores to protect our crops which requires the continuous development of new and more diverse insecticides because the herbivores rapidly develop resistance. The idea of coevolution between phytochemicals and herbivores is broadly accepted (Futuyma and Agrawal, 2009; Maron *et al.*, 2019). Berenbaum and Zangerl (1996) proposed the "interaction diversity" hypothesis to explain SM diversity. It is assumed that different SMs have different biological functions. Plant SM diversity is beneficial in a diverse community and thus is the ecological consequence of the plant's interaction with diverse biotic and abiotic environments. This hypothesis is in line with the findings that a more diverse and more specialized community of herbivores is associated with an elevated chemical diversity of *Piper* species (Richards *et al.*, 2015). SM diversity is explained by synergism, where the effectiveness of a defense compound increases in the presence of another (Berenbaum *et al.*, 1991; Rasmann and Agrawal, 2009). Synergistic effects of plant SMs have been demonstrated in a number of cases (Dyer *et al.* 2003; Leckie *et al.*, 2016; Liu *et al.*, 2017).

2. Pyrrolizidine alkaloids (PAs)

In this thesis, we focus on SMs involved in plant protection against herbivores. PAs comprise a typical class of SMs that are constitutively formed in plants containing them and are thought to mediate the interactions between plants and herbivores (Hartmann, 1999). PAs are ester alkaloids containing a necine base which is esterified with one or more necic acids, forming monoesters, open-chain diesters or even triesters and macrocyclic diesters. Chemical modifications of both the necine base and the necic acid moiety result into a large number of PA structures found in plants (Fig. 2; Hartmann and Witte, 1995). PAs occur as both tertiary amines (free bases) and *N*-oxides (Fig. 2; Joosten *et al.*, 2011; Patrick *et al.*, 2018), where the polar salt-like *N*-oxide form is more common. Thus far, more than 400 PAs have been found in over 6,000 plant species (Chou and Fu, 2006). Of these ca. 6,000 angiosperm species in distantly related families more than 95% belong to four families, i.e. the Asteraceae (the tribes Senecioneae and Eupatorieae), the Boraginaceae (most genera), the Fabaceae (mainly the genus *Crotalaria*) and the Orchidaceae (ca. 40 species) (Hartmann and Witte, 1995; Langel *et al.*, 2011). Within Senecioneae, approximately 190 PAs have been detected in circa 300 species, of which more than 100 PAs are of the macrocyclic senecionine-type containing four

structural groups, i.e. the senecionine group, the senecivernine group, the nemorensine group and the rosmarinine group (Fig. 3; Hartmann and Witte, 1995; Langel *et al.*, 2011). The senecionine group were further subdivided into different structural subgroups based on their structural characteristics, including senecionine-like PAs, jacobine-like PAs, erucifoline-like PAs and otosenine-like PAs (Pelser *et al.*, 2005; Cheng *et al.*, 2011).

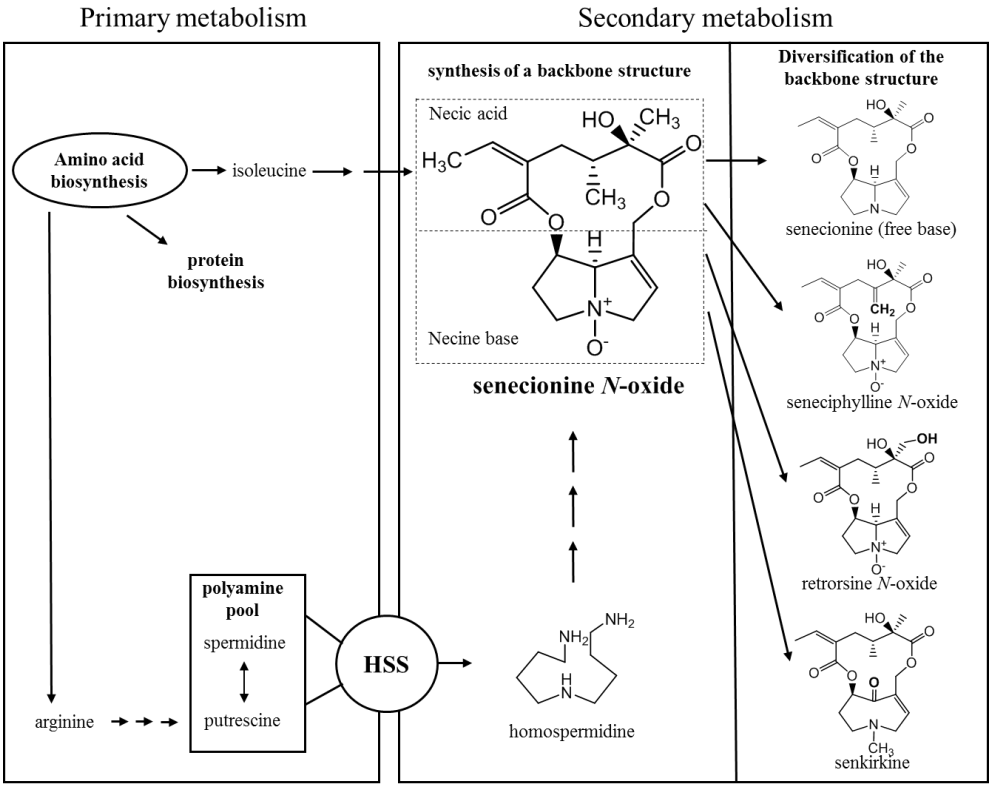

Figure 2. Simplified scheme of PA biosynthesis in *Senecio* and *Jacobaea* (from Langel *et al.*, 2011 with minor modification).

2.1 PA biosynthesis, translocation and accumulation in *Senecio* and *Jacobaea* species

Senecio and *Jacobaea* species of the tribe Senecioneae have been often used to study PA biosynthesis (Hartmann and Toppel, 1987; Toppel *et al.*, 1987; Hartmann *et al.*, 1989; Sander and Hartmann, 1989; Hartmann and Dierich, 1998). Earlier ^{14}C-labelled tracer studies revealed that PA biosynthesis is closely related to polyamine metabolism (Hartmann and Toppel, 1987; Hartmann *et al.*, 1988), where arginine is converted successively to putrescine and spermidine which are the two substrates for the formation of homospermidine (Fig. 2; Böttcher *et al.*, 1994), the first specific intermediate in the PA biosynthesis pathway (Böttcher *et al.*, 1993). The enzyme responsible for this conversion, namely homospermidine synthase (HSS), is the

only PA pathway-specific enzyme that has been identified so far (Böttcher *et al.*, 1993; Böttcher *et al.*, 1994; Langel *et al.*, 2011). It has been suggested that the HSS encoding gene originated from the gene encoding deoxyhypusine synthase by means of gene duplication and diversification based on similarities in their sequences and biochemical reactions (Ober and Hartmann, 1999). Homospermidine is exclusively incorporated into the necine base moiety of PAs without any turnover (Fig. 2; Böttcher *et al.*, 1993). Senecionine *N*-oxide (Fig. 2) is found as the first alkaloid of PA biosynthesis which is synthesized only in roots of *Senecio* and *Jacobaea* species (Hartmann and Toppel, 1987; Toppel *et al.*, 1987; Hartmann *et al.*, 1989; Hartmann and Dierich, 1998). It is not known how homospermidine is converted to senecionine *N*-oxide.

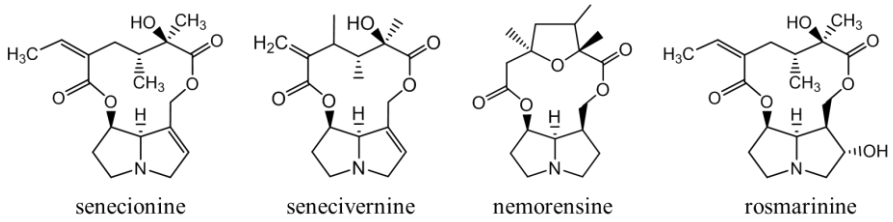

Figure 3. Pyrrolizidine alkaloids of the macrocyclic senecionine-type encompassing the senecionine group (12-membered), the senecivernine group (12-membered), the nemorensine group (13-membered) and the rosmarinine group (12-membered).

In *Senecio* and *Jacobaea* plants that were studied, senecionine *N*-oxide synthesized in roots is via the phloem allocated all over the plant and is accumulated at the preferential storage sites, i.e. inflorescences and peripheral tissues of leaves and stems (Hartmann *et al.*, 1989; Witte *et al.*, 1990). Phloem loading and unloading of polar salt-like PA *N*-oxides is predicted to be carrier-mediated (Ehmke *et al.*, 1987; 1988) since species which do not produce PAs were shown to be unable to translocate PAs via the phloem (Hartmann *et al.*, 1989). In shoot organs, the backbone structure senecionine *N*-oxide is transformed into structurally related PAs (Fig. 2; Fig. 4) without any turnover or degradation (Sander and Hartmann, 1989; Hartmann and Dierich, 1998). PAs are assumed to be selectively taken up and stored in cell vacuoles by a specific *N*-oxide carrier (Ehmke *et al.*, 1987, 1988). Consequently, *Senecio* and *Jacobaea* species have qualitative and quantitative PA profiles resulting from several interacting processes, i.e. (i) *de novo* synthesis of senecionine *N*-oxide in roots, (ii) dynamic long-distance translocation of senecionine *N*-oxide into shoots, (iii) structural diversification of senecionine *N*-oxide with varying efficiency between different organs, (iv) continuous allocation of PAs in the plant, (v) selective uptake and storage of PAs in vacuoles (Hartmann and Dierich, 1998).

2.2 PA diversification

Most of our current knowledge of PA biosynthetic diversification has come from the studies of the 12-membered macrocyclic senecionine-type PAs (Fig. 3) in the Senecioneae (Langel *et*

al., 2011). Senecionine *N*-oxide synthesized in roots has been identified as the backbone structure of PA conversion by label tracer studies (Hartmann and Toppel, 1987; Hartmann *et al.*, 1988; Hartmann *et al.*, 1989; Hartmann and Dierich, 1998). To a limited extent in roots senecionine *N*-oxide may be converted to structurally related derivatives such as seneciphylline *N*-oxide (Toppel et a., 1987; Sander and Hartmann, 1989). However, shoots have been identified as the major sites for PA transformations in *Senecio* and *Jacobaea* species (Hartmann and Dierich, 1998). Tracer experiments revealed that senecionine *N*-oxide undergoes structural transformations in a position-specific and stereoselective manner resulting in species-specific PA bouquets (Fig. 4; Hartmann and Dierich, 1998). With the exception of senecivernine transformations between PAs can be explained by two main reactions (i.e. conversion of retronecine to otonesine and site-specific epoxidation) and simple one- or two-step reactions (20-*Z*/*E*-isomerization, 13,19-dehydroxylation, site-specific hydroxylation, hydrolysis of 15,20-epoxide, chlorolysis of 15,20-epoxide, site-specific *O*-acetylation and *N*-deoxidation; Pelser *et al.*, 2005). This suggests that PA diversification is a highly plastic process (Hartmann and Dierich, 1998). The enzymes underlying these elaborate steps of chemical modifications have not been identified.

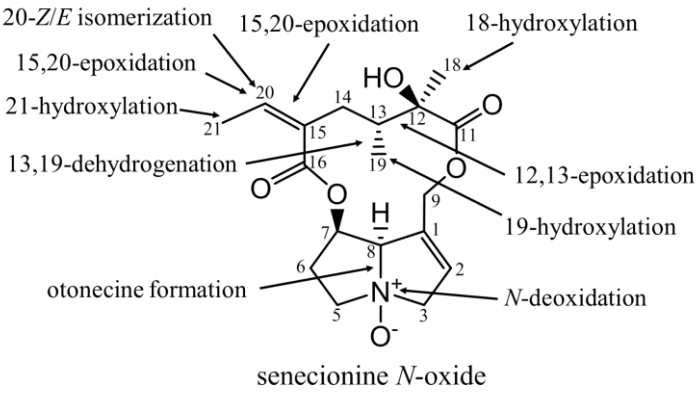

Figure 4. Structural modifications of senecionine *N*-oxide in position-specific and stereoselective manners.

2.3 PA diversity in *Senecio* and *Jacobaea* species

2.3.1 Interspecific diversity

A great diversity of PA patterns have been found in *Senecio* and *Jacobaea* species (Langel *et al.*, 2011). PA profiles are often species-specific (Hartmann and Dierich, 1998), but phylogenetic relationships are not correlated with PA composition at the level of individual PAs (Langel *et al.*, 2011). The 26 species of the genus *Jacobaea* (formerly a section of the genus *Senecio*) all produce PAs (Pelser *et al.*, 2005; Langel *et al.*, 2011) but their PA profiles

are believed to be species-specific (Soldaat *et al.*, 1996; Hartmann and Dierich, 1998; Langel *et al.*, 2011). PAs appear to be incidentally distributed over *Jacobeae* species, demonstrating little phylogenetic signal. Even phylogenetically close species can differ extensively in their PA compositions (Pelser *et al.*, 2005). For example, PA profiles of *J. vulgaris* and *J. aquatica* are generally different qualitatively and quantitatively (Pelser *et al.*, 2005; Langel *et al.*, 2011; Cheng *et al.*, 2011a), although these two species are closely related (Pelser *et al.*, 2004).

2.3.2 Intraspecific diversity

Besides interspecific variation, a large variety of PA profiles within species were abundantly found. Population-level studies revealed that different populations within species often show considerable differences in their PA compositions and concentrations (Witte *et al.*, 1992; Macel *et al.*, 2004; Joshi and Vrieling, 2005). This intraspecific diversity is well documented by different chemotypes of *J. vulgaris* and *J. erucifolia*. By evaluating populations originating from different geographic locations, 'jacobine chemotype' and 'erucifoline chemotype' were found for *J. vulgaris*, while 'erucifoline chemotype' and 'eruciflorine chemotype' were found for *J. erucifolia* (Witte *et al.*, 1992). More chemotypes were found for *J. vulgaris* in a later study conducted by Macel *et al.* (2004), in which they distinguished senecionine, jacobine, erucifoline and mixed chemotypes. In addition to the difference between populations, variation in PA composition between individual plants within populations is also considerable. The relative abundances of jacobine in 412 plants from half-sib families of the population collected from Meijendel (The Netherlands) varied from 41 to 100% of total PA, whereas that of erucifoline ranged from 0 to 19% (Macel *et al.*, 2004).

2.3 Driving factors of PA variation

2.3.1 Genetic bases

Both PA concentration and composition are under genetic control. As aforementioned, HSS is so far the only enzyme identified that is involved in PA biosynthesis. It is believed that the total PA concentration of a plant is closely related to HSS as it catalyzes the formation of the first PA intermediate, homospermidine, and PAs do not undergo turnover except for chemical diversification (Ober and Hartmann, 1999). It was indicated that under controlled conditions 50-100% of the total variation in PA concentration of *J. vulgaris* is due to genetic differences in a diallel cross and a half-sib analysis (Vrieling *et al.*, 1993). The comparison of PA profiles between 10 clonal families of *J. vulgaris* from different populations showed that the variation in PA composition within clonal families was smaller than that among these families, implying that PA composition is genetically determined (Macel *et al.*, 2004). The finding that senecionine *N*-oxide is transformed into unique PAs in different *Senecio* and *Jacobeae* species clearly indicates that species-specific PA profiles are brought about by genetically controlled specific processes (Hartmann and Dierich, 1998). It is not clear whether quantitative and qualitative variations of PAs are controlled by the same genetic mechanism or distinct mechanisms.

2.3.2 Environmental factors

Vrieling *et al.* (1993) suggested that PA profiles in populations of *J. vulgaris* are under natural selection. The potential roles of herbivores in PA variation as selective forces have been extensively investigated (Vrieling and de Boer, 1999; Macel *et al.*, 2002; Joshi and Vrieling, 2005; Cheng *et al.*, 2011b; Kostenko *et al.*, 2012; Cheng *et al.*, 2013). It was hypothesized that a complex PA profile with high structural diversity among and within species forms a powerful defense barrier which cannot be easily overcome by generalist herbivores since they need to cope with numerous chemical structures (Hartmann and Dierich, 1998; Pelser *et al.*, 2005). In contrast specialist herbivores are attracted by PAs. The cinnabar moth, a specialist herbivore causing the majority of leaf loss in *J. vulgaris* preferred to oviposit on plants with more tertiary amines of the jacobine-like and otosenine-like PAs (Cheng *et al.*, 2013). In addition, both the cinnabar moth and the garden tiger moth sequester PAs in their bodies (Aplin and Rothschild, 1972). These findings indicate that specialist herbivores too may potentially act as a selective force on variation in PA composition and concentration. Besides herbivores, soil-borne microorganisms also affect PA composition of *J. vulgaris* (Joosten *et al.*, 2009).

PA patterns in plants are also affected by abiotic factors such as nutrients, water and light. PA concentrations in the shoots and roots of *J. vulgaris*, *J. aquatica* and *S. vulgaris* were significantly reduced with increasing nutrient supplies. This decreased PA concentration was assumed to be due to the dilution effect from increased biomass following high nutrient supplies (Hol *et al.*, 2003; Hol, 2011). Conversely, *J. vulgaris* plants grown under nutrient stress tend to have higher PA concentrations compared with those grown under normal conditions (Vrieling and van Wijk, 1994). Water availability and light intensity are two other factors which affect PA concentrations. PA concentrations of *J. vulgaris* differed significantly under drought-stressed (Vrieling *et al.*, 1993) and light-limited conditions (Vrieling and van Wijk, 1994) compared to controls.

3. Cytochrome P450s (CYPs)

Genes involved in SM biosynthesis have often evolved from genes involved in primary metabolism by gene duplication with successive diversification (Ober, 2010; Moore *et al.*, 2014). Many of these genes involved in SM pathways belong to large gene families (Kessler and Kalske, 2018), such as cytochrome P450s (Bak *et al.*, 2006; Frey *et al.*, 2009). CYP genes form a large family in any given plant species (Mizutani, 2012). CYPs catalyze a wide range of regiospecific, stereospecific and irreversible steps in plant SM biosynthesis (Renault *et al.*, 2014), thus playing an important role in the evolution of chemical diversity. CYP enzymes are a major class of oxidative enzymes in eukaryotes. Given the oxidative transformations from the primary PA senecionine *N*-oxide to derived PAs, CYPs are possible candidates for performing these oxidations. Although amino acid sequences of CYP proteins are diverse, their overall topology and structural fold have remained conserved during evolution (Werck-Reichhart and Feyereisen, 2000; Bak *et al.*, 2011). CYP proteins have a number of conserved

motifs (Fig. 5; Bak *et al.*, 2011; personal communication of Prof. Dr. David R. Nelson) which make the recognition of CYPs in novel non-annotated transcriptomes possible.

3.1 Motifs of CYP proteins

The motifs of CYPs are few in the N-terminal region and more abundant in the C-terminus (Fig. 5; Nelson, 2004). CYPs all share a highly conserved catalytic center, where heme with iron is coordinated to the thiolate of a cysteine that is conserved in all CYP sequences. The heme-binding signature is located about 50 aa from the C-terminus. Upstream from the heme-binding domain is the PERF/W motif followed by a small PKG motif in a distance of ca. 20 aa. N-terminal of the PKG motif is the most conserved region in CYP proteins, the ExxR motif (K-helix). The E and R amino residues of the ExxR motif form a salt bridge with the R amino acid residue of the PERF/W motif, which is generally considered to be essential for the stability of the core structure (Hasemann *et al.*, 1995). This ERR triad is conserved in all plant CYP sequences. At about 200 aa from the C-terminus, I-helix (AGx[E,D]T) is regarded as an oxygen-binding motif containing the conserved glycine pointing at the heme center and the conserved threonine pointing to oxygen binding site (Li *et al.*, 2008). While at the N-terminus there is a proline rich membrane hinge at about 30 aa, and often but not always a C-helix (Wxxx[R,K]) at about 130 aa. All plants described so far are bound to a membrane, usually the ER membrane, through a short hydrophobic segment at their N-terminus (Fig. 5; Bak *et al.*, 2011).

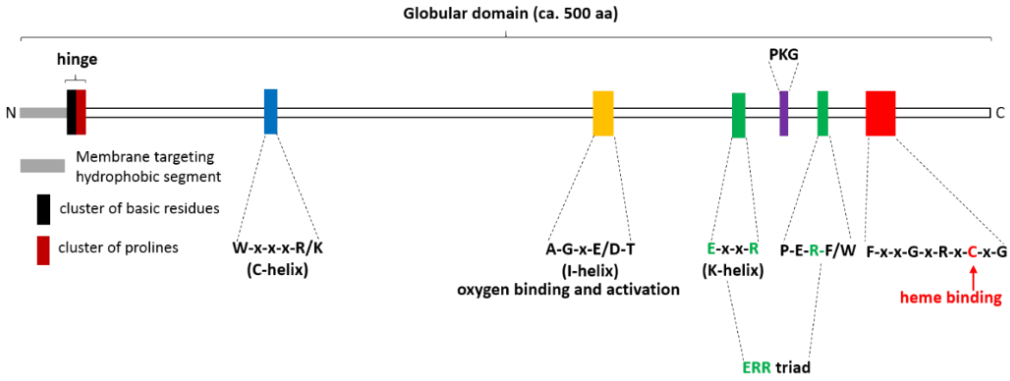

Figure 5. Conserved structures and sequences in classic CYP proteins..

3.2 Organization and evolution of CYPs in plants

CYPs are found in virtually all organisms, but the numbers have exploded in plants (Werck-Reichhart and Feyereisen, 2000). CYPs are classified and named following recommendations of a nomenclature committee (Nelson, 2009). So far, there are 127 CYP families in the plant kingdom (Nelson and Werck-Reichhart, 2011), of which the numbers of CYP families have plateaued at 59 in angiosperms (Hamberger and Bak, 2013). CYPs of land plants are grouped in 11 distinct clans according to sequence similarities. Furthermore, plant CYPs are divided

into A-type and non-A-type based on phylogenetic relationships. The A-type CYPs (CYP71 clan) form a monophyletic clade, whereas the non-A-type CYPs (the remaining 10 clans) are more diverse and do not form a coherent group in a phylogenetic sense (Hamberger and Bak, 2013). The A-type CYP71 clan is the youngest evolved clan as indicated by phylogenetic analyses of plant CYPs (Nelson and Werck-Reichhart, 2011; Bak *et al.*, 2011). This clan appears to have evolved via successive gene duplication events (Sezutsu *et al.*, 2013), leading to often species-specific expanded families (Hamberger and Bak, 2013). The CYP71 clan comprises more than half of all plant CYPs and consequently, harbors a great diversity of functions (Nelson and Werck-Reichhart, 2011). CYP72 and CYP85 are two other clans that have expanded dramatically in plants. It is more challenging to predict the functions or substrate preferences of members in the CYP71, CYP72 and CYP85 clans due to their dramatic expansion, even though phylogeny within families or subfamilies can serve as a guide to function prediction (Nelson and Werck-Reichhart, 2011).

3.3 Heterologous expression of plant CYPs

All plant CYPs are membrane-anchored by hydrophobic N-terminal regions, usually anchored on the cytoplasmic surface of the ER. They use molecular oxygen as the oxygen donor, which is coordinated by an iron atom in the heme prosthetic group. To be active, CYPs need to be coupled with electron-donating proteins, CYP reductases or cytochrome b_5, which are also anchored to the surface of the ER. Most commonly, via the NADPH-dependent CYP reductase (CPR) heme-bound O_2 is activated by the successive transfer of two electrons from NADPH (Bak *et al.*, 2011), leading to regiospecific and stereospecific oxidative attack of a plethora of substrates.

Heterologous expression of CYP proteins is an important step for their functional characterization. The heterologous expression of plant CYPs has mostly been performed in yeast, bacterial or insect cells for heterologous expression, with yeast being the most frequently used organism (Schuler and Werch-Reichhart, 2003; Duan and Schuler, 2006). There are several advantages for using the yeast *Saccharomyces cerevisiae* as heterologous expression host for plant CYPs, i.e. (i) the presence of an ER membrane environment and post-transcriptional modification systems, (ii) the availability of modified yeast strains expressing plant CPR genes (e.g. WAT11; Pompon *et al.*, 1996), (iii) the availability of vectors (e.g. pYeDP60) with high yielding galactose-inducible GAL10-CYC1 hybrid promoters (Urban *et al.*, 1990; Pompon *et al.*, 1996), (iv) relatively low costs, high efficiency and rapid growth (Pompon *et al.*, 1996; Hamann and Møller, 2007).

4. Research questions

The research described in this thesis aimed at understanding how PA diversity comes about from the perspectives of evolution and biosynthesis of PAs. In particular, the distribution patterns of PAs were studied in a phylogenetic context. CYPs were studied for involvement in PA biosynthesis and PA diversity. By using *Jacobaea* species as the model system, the following questions were addressed:

(i) Do *Jacobaea* species indeed have species-specific PA profiles at a larger scale including among species, among populations and among individuals with respect to both concentration and composition? Do distributions of individual PAs among *Jacobaea* species show phylogenetic signals?

(ii) How diverse are CYPs of *Jacobaea* species? What are evolutionary patterns of CYPs between two *Jacobaea* species, and two other members of the Asteraceae, *Helianthus annuus* and *Lactuca sativa*, and the outgroup *Arabidopsis thaliana*?

(iii) Are abundance patterns of PAs with site-specific oxidative modifications related to expression patterns of particular CYPs? Is it in this way possible to identify CYP candidates for specific oxidation steps?

(iv) Are identified candidate CYPs indeed involved in PA biosynthesis when functionally tested via heterologous expression in yeast and *in vitro* enzyme assays?

5. Outline of this thesis

In **Chapter 2**, both qualitative and quantitative PA variations for leaves of 17 *Jacobaea* species including different individuals and populations grown under controlled conditions were studied. Within the phylogenetic context of *Jacobaea* species, the evolutionary histories and phylogenetic signals of individual PAs were investigated in order to understand how PA diversity is related to species phylogeny.

CYPs often perform oxidative reactions in the biosynthesis of SMs and thus are crucial players in the evolution of chemical diversity. In **Chapter 3** the diversity and evolution of CYPs of *J. vulgaris* and *J. aquatica* were evaluated. The resulting database of CYPs was used for future exploration of their functions, including possible involvement in PA biosynthesis and PA diversity in later chapters.

In **Chapter 4**, PA profiles and expression profiles of CYPs of *J. vulgaris*, *J. aquatica*, and their hybrids with contrasting PA profiles were associated to discover candidate CYPs potentially involved in PA biosynthesis. Subsequently, in **Chapter 5** eight CYP candidates were tested using heterologous expression in yeast and *in vitro* enzyme assays.

Finally in **Chapter 6** the discussion and conclusions of the findings described in this thesis are presented.

References

Aplin RT, Rothschild M. 1972. Poisonous alkaloids in the body tissue of the garden tiger moth (*Arctia caja* L.) and the cinnabar moth (*Tyria jacobaea* L.) (Lepidoptera). In: De Vries A, Kochva E (eds) Toxins of animal and plant origin. New York: Gordon & Breach Sci Pub Vol 2, pp.579-595.

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

Berenbaum MR, Nitao JK, Zangerl AR. 1991. Adaptive significance of furanocoumarin diversity in *Pastinaca sativa* (Apicaceae). *Journal of Chemical Ecology* 17:207-215.

Berenbaum MR, Zangerl AR. 1996. Phytochemical diversity: adaptation or random variation. In: Remeo JT, Saunders JA, Barbosa P (eds) Phytochemical Diversity and Redundancy in Ecological Interactions. New York: Plenum pp.1-24.

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

Böttcher F, Ober D, Hartmann T. 1994. Biosynthesis of pyrrolizidine alkaloids: putrescine and spermidine are essential substrates of enzymatic homospermidine formation. *Canadian Journal of Chemistry* 72:80-85.

Chang A, Hartmann T. 1998. Solubilization and characterization of a senecionine *N*-oxygenase from *Crotalaria scassellatii* seedlings. *Phytochemistry* 49:1859-1866.

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

Cheng D, Kirk H, Vrieling K, Mulder PPJ, Klinkhamer PGL. 2011b. The relationship between structurally different pyrrolizidine alkaloids and western flower thrips resistance in F_2 hybrid of *Jacobaea vulgaris* and *Jacobaea aquatica*. *Journal of Chemical Ecology* 37:1071-1080.

Cheng D, van der Meijden E, Mulder PPJ, Vrieling K, Klinkhamer PGL. 2013. Pyrrolizidine alkaloid composition influences cinnabar moth oviposition preferences in *Jacobaea* hybrids. *Journal of Chemical Ecology* 39:430-437.

Chou MW, Fu PP. 2006. Formation of DHP-derived DNA adducts in vivo from dietary supplements and Chinese herbal plant extracts containing carcinogenic pyrrolizidine alkaloids. *National Center for Toxicological Research* 22:321-327.

Cvorovi J, Ziberna L, Fornasaro S, Tramer F, Passamonti S. 2018. Bioavailability of flavonoids: the role of cell membrane transporters. In: Polyphenols: Mechanisms of Action in Human Health and Disease. 2nd Edition. Academic Press Chapter 22, pp.295-320.

Duan H, Schuler MA. 2006. Heterologous expression and strategies for encapsulation of membrane-localized plant P450s. *Phytochemistry Reviews* 5:507-523.

Dyer LA, Dodson CD, Stireman JO, Tobler MA, Smilanich AM, Fincher RM, Letourneau DK. 2003. Synergistic effects of three *Piper* amides on generalists and specialist herbivores. *Journal of Chemical Ecology* 29:2499-2514.

Ehmke A, von Borstel K, Hartmann T. 1987. Specific uptake of the *N*-oxides of pyrrolizidine alkaloids by cells, protoplasts, and vacuoles from *Senecio* cell cultures. In: Marin BP (ed) Plant vacuoles. New York: Plenum Press pp. 301-304.

Ehmke A, von Borstel K, Hartmann T. 1988. Alkaloid *N*-oxides as transport and vacuolar storage compounds of pyrrolizidine alkaloids in *Senecio vulgaris* L. *Planta* 176:83-90.

Ehrlich PR, Raven PH. 1964. Butterflies and plants: a study of coevolution. *Evolution* 18:586-608.

Firn RD, Jones CG. 2003. Natural products - a simple model to explain chemical diversity. *Natural Product Reports* 20:382-391.

Firn RD, Jones CG. 2009. A Darwinian view of metabolism: molecular properties determine fitness. *Journal of Experimental Botany* 60:719-726.

Funayama S, Cordell GA. 2015. Alkaloids: a treasury of poisons and medicines. Academic Press pp.1-20.

Futuyma DJ, Agrawal AA. 2009. Macroevolution and the biological diversity of plants and herbivores. *Proceedings of the National Academy of Sciences of the United States of America* 106:18054-18061.

Hadacek F. 2002. Secondary metabolites as plant traits: current assessment and future perspectives. *Critical Reviews in Plant Sciences* 21:273-322.

Hamann T, Møller BL. 2007. Improved cloning and expression of cytochrome P450s and cytochrome P450 reductase in yeast. *Protein Expression And Purification* 56:121-127.

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. Site of synthesis, translocation and accumulation of pyrrolizidine alkaloid *N*-oxides in *Senecio vulgaris* L. *Planta* 177:98-107.

Hartmann T, Sander H, Adolph R, and Toppel G. 1988. Metabolic links between the biosynthesis of pyrrolizidine alkaloids and polyamines in root cultures of *Senecio vulgaris*. *Planta* 175:82-90.

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.

Hartmann T, Witte L. 1995. Chemistry, biology and chemoecology of the pyrrolizidine alkaloids. In Pelletier SW (ed) Alkaloids: chemical and biological perspective. Oxford: Pergamon Vol. 9, pp.155-233.

Hartmann T. 1996. Diversity and variability of plant secondary metabolism: a mechanistic view. *Entomologia Experimentalis et Applicata* 80:177-188.

Hartmann T. 1999. Chemical ecology of pyrrolizidine alkaloids. *Planta* 207:483-495.

Hasemann CA, Kurumbail RG, Boddupalli SS, Peterson JA, Deisenhofer J. 1995. Structure and function of cytochromes P450: a comparative analysis of thee crystal structures. *Structure* 3:41-62.

Havsteen BH. 2002. The biochemistry and medical significance of the flavonoids. *Pharmacology & Therapeutics* 96:67-202.

Hol WHG, Vrieling K, van Veen JA. 2003. Nutrients decrease pyrrolizidine alkaloids concentrations in *Senecio jacobaea*. *New Phytologist* 158:175-181.

Hol WHG. 2011. The effect of nutrients on pyrrolizidine alkaloids in *Senecio* plants and their interactions with herbivores and pathogens. *Phytochemisty Reviews* 10:119-126.

Isah T. 2019. Stress and defense responses in plant secondary metabolites production. *Biological research* 52:39.

Jones CG, Firn RD. 1991. On the evolution of plant secondary chemical diversity. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 333:273-280.

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

Joosten L, Mulder PPJ, Klinkhamer PGL, van Veen JA. 2009. Soil-borne microorganisms and soil-type affect pyrrolizidine alkaloids in *Jacobaea vulgaris*. *Plant and Soil* 325:133-143.

Joshi J, Vrieling K. 2005. The enemy release and EICA hypothesis revisited: incorporating the fundamental difference between specialist and generalist herbivores. *Ecology Letters* 8:704-714.

Kessler A, Halischke R. 2007. Specificity and complexity: the impact of herbivore-induced plant response on arthropod community structure. *Current Opinion in Plant Biology* 10:409-414.

Kessler A, Kalske A. 2018. Plant secondary metabolite diversity and species interactions. *Annual Review 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.

Kostenko O, van de Voorde TFJ, Mulder PPJ, van der Putten, Bezemer TM. 2012. Legacy effects of aboveground-belowground interactions. *Ecology Letters* 15:813-821.

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

Leckie BM, D'Ambrosio DA, Chappell TM, Halitschke R, De Jong DM, Kessler A, Kennedy GG, Mutschler MA. 2016. Differential and synergistic functionality of acylsugars in suppressing oviposition by insect herbivores. *PLOS ONE* 11:e0153345.

Li L, Chang Z, Pan Z, Fu Z, Wang X. 2008. Modes of heme binding and substrate access for cytochrome P450 CYP74A revealed by crystal structures of allene oxide synthase. *Proceedings of the National Academy of Sciences of the United States of America* 105:13883-13888.

Liu X, Vrieling K, Klinkhamer PGL. 2017. Interactions between plant metabolites affect herbivores: a study with pyrrolizidine alkaloids and chlorogenic acid. *Frontiers in Plant Science* 8:903.

Macel M, Klinkhamer PGL, Vrieling K, van der Meijden E. 2002. Diversity of pyrrolizidine alkaloids in *Senecio* species does not affect the specialist herbivore *Tyria jacobaea*. *Oecologia* 133:541-550.

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

Maron JL, Agrawal AA, Schemske DW. 2019. Plant-herbivore coevolution and plant speciation. Ecology 100:e02704.

Mint Evolutionary Genomics Consortium. 2018. Phylogenomic mining of the mints reveals multiple mechanisms contributing to the evolution of chemical diversity in Lamiaceae. *Molecular Plant* 11:1084-1096.

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.

Mulder PPJ, López P, Castelari M, Bodi D, Ronczka S, Preiss-Weigert A, These A. 2018. Occurrence of pyrrolizidine alkaloids in animal-and plant-derived food: results of a survey across Europe. *Food Additives & Contaminants: Part A* 35:118-133.

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

Nelson D. 2009. The cytochrome p450 homepage. *Humman Genomics* 4:59-65.

Nelson DR. 2004. Cytochrome P450 Nomenclature, 2004. In Phillips IR, Shephard EA (ed) Methods in molecular biology: cytochrome P450 protocols methods in molecular biology. 2nd edition. Totowa: Humana Press Inc. Vol 320.

Nützmann HW, Huang A, Osbourn A. 2016. Plant metabolic clusters - from genetics to genomics. *New Phytologist* 211:771-789.

Ober D, Hartmann T. 1999. Homospermidine synthase, the first pathway-specific enzyme of pyrrolizidine alkaloid biosynthesis, evolved from deoxyhypusine synthase. *Proceedings of the National Academy of Sciences of the United States of America* 96:14777-14782.

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.

Pelser PB, van den Hof K, Gravendeel B, van der Meijden R. 2004. The systematic value of morphological characters in *Senecio* Sect. *Jacobaea* (Asteraceae) as compared to DNA sequences. *Systematic Botany* 29:790-805.

Pompon D, Louerat B, Bronine A, Urban P. 1996. Yeast expression of animal and plant P450s in optimized redox environments. *Methods in Enzymology* 272:51-64.

Rasmann S, Agrawal AA. 2009. Plant defense against herbivory: progress in identifying synergism redundancy, and antagonism between resistance traits. *Current Opinion in Plant Biology* 12:473-478.

Renault H, Bassard JE, Hamberger B, Werck-Reichhart. 2014. Cytochrome P450-mediated metabolic engineering: current progress and future challenges. *Current Opinion in Plant Biology* 19:27-34.

Richards LA, Dyer LA, Forister ML, Smilanich AM, Dodson CD, Leonard MD, Jeffrey CS. 2015. Phytochemical diversity drives plant-insect community diversity. *Proceedings of the National Academy of Sciences of the United States of America* 112:10973-10978.

Sander H, Hartmann T. 1989. Site of synthesis, metabolism and translocation of senecionine *N*-oxide in cultured roots of *Senecio erucifolius*. *Plant Cell, Tissue and Organ Culture* 18:19-31.

Schuler MA, Werch-Reichhart D. 2003. Functional genomics of P450s. *Annual Review of Plant Biology* 54:629-667.

Sezutsu H, Le Goff G, Feyereisen R. 2013. Origins of P450 diversity. *Philosophical Transactions of the Royal Society B* 368:20120428.

Soldaat LL, Boutin JP, Derridj S. 1996. Species-specific composition of free amino acids on the leaf surface of four *Senecio* species. *Journal of Chemical Ecology* 22:1-12.

Toppel G, Witte L, Riebesehl B, von Borstel K, Hartmann T. 1987. Alkaloid patterns and biosynthetic capacity of root cultures from some pyrrolizidine alkaloid producing *Senecio* species. *Plant Cell Reports* 6:466-469.

Urban P, Cullin C, Pompon D. 1990. Maximizing the expression of mammalian cytochrome P-450 monooxygenase activities in yeast cells. *Biochimie* 72:463-472.

Vrieling K, de Boer N. 1999. Host-plant choice and larval growth in the cinnabar moth: do pyrrolizidine alkaloids play a role? *Entomologia Experimentalis et Applicata* 91:251-257.

Vrieling K, de Vos H, van Wijk CAM. 1993. Genetic analysis of the concentration of pyrrolizidine alkaloids in *Senecio jacobaea*. *Phytochemistry* 32:1141-1144.

Vrieling K, van Wijk CAM. 1994. Cost assessment of the production of pyrrolizidine alkaloids in ragwort (*Senecio jacobaea* L.). *Oecologia* 97:541-546.

Werck-Reichhart D, Feyereisen R. 2000. Cytochrome P450: a success story. *Genome Biology* 1:reviews3003.

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

Witte L, Ehmke H, Hartmann T. 1990. Interspecific flow of pyrrolizidine alkaloids: from plants via aphids to ladybirds. *Naturwissenschaften* 77:540-543.

Witte L, Ernst L, Adam H, Hartmann T. 1992. Chemotypes of two pyrrolizidine alkaloid-containing *Senecio* species. *Phytochemistry* 31:559-565.

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

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

Chapter 2

The evolution of pyrrolizidine alkaloid diversity among and within *Jacobaea* species

Yangan Chen, Patrick P.J. Mulder, Onno Schaap, Johan Memelink, Peter G.L. Klinkhamer, Klaas Vrieling

The evolution of pyrrolizidine alkaloid diversity among and within *Jacobaea* species

Abstract

Plants produce a large amount of secondary metabolites (SMs) showing considerable inter- and intraspecific diversity with respect to concentration and composition as a strategy to cope with environmental stresses. The question how this chemical diversity comes about and why it is maintained in nature remains largely unresolved. Understanding the mechanisms underlying chemical diversity has the potential to shed light on the evolution of plant defenses against herbivores and pathogens. Pyrrolizidine alkaloids (PAs) are a typical class of SMs with high diversity. *Jacobaea* (formerly named *Senecio*) species produce over 80 different PAs and the presence and concentration are believed to be species-specific. We performed a qualitative and quantitative analysis of 80 PAs with liquid chromatography-tandem mass spectrometry (LC-MS/MS) of leaves from 17 *Jacobaea* species including different individuals and populations grown under controlled conditions in a climate chamber. We observed large inter- and intraspecific variation in both PA concentration and composition, which were both species-specific. Senecionine-like PAs, jacobine-like PAs, and otosenine-like PAs were identified as the main PAs driving the diversification of *Jacobaea* PA profiles. In addition, we sequenced 11 chloroplast regions and three nuclear DNA genes to reconstruct the molecular phylogeny of the 17 *Jacobaea* species. By tracing the evolutionary history of each PA using a maximum likelihood phylogeny as a roadmap, we found mainly accidental distributions of these individual PAs. We also found no strong evidence for phylogenetic signals using two different measures (Blomberg's K and Pagel's λ). Nine PAs out of 80 showed a significant phylogenetic signal for Pagel's λ statistics only. These results broadly confirm the results of earlier studies on PAs showing a rather random distribution among *Jacobaea* species. Given the common intraspecific PA diversity found in *Jacobaea* species, we assume that this high PA diversity is the result of regulation of PA expression in plants as a life strategy in response to different biological needs rather than the result of gains and losses of particular PA-related genes during evolution.

Keywords

Secondary metabolite diversity, principal component analysis, hierarchical cluster analysis, Spearman correlation test, ancestral state reconstruction, phylogenetic signal

Introduction

Plant secondary metabolites (SMs) are pervasive in the plant kingdom functioning mainly as defense and/or signaling compounds (Wink, 2003; Moore *et al*., 2014). So far more than 200,000 SMs have been isolated and identified (Kessler and Kalske, 2018), being assigned to different compound classes including alkaloids, flavonoids, terpenoids, and glucosinolates. Within a given taxon, a single class of SM is commonly dominant. Within such a class usually a few major compounds are accompanied by several derivatives and minor related compounds (Wink, 2003). For example, in *Arabidopsis thaliana* 34 different glucosinolates have been identified (Kliebenstein *et al*., 2001), and in the Lamiaceae 147 terpenoids have been found (Mint Evolutionary Genomics Consortium, 2018). Besides the structural diversity, SMs often show remarkable quantitative variation. This is well documented for the abovementioned example of glucosinolates which showed about 20-fold difference in total concentrations among leaves of different ecotypes (Kliebenstein *et al*., 2001). As yet, due to their gigantic number and striking diversity, it is still an ongoing challenge to understand how this SM diversity comes about, and why such a large diversity is maintained in nature (Moore *et al*., 2014).

Aiming at untangling mechanisms behind SM diversity, researchers have put much effort in studying the distribution patterns of SMs under particular phylogenetic frameworks (Wink, 2003; Wink 2008; Pelser *et al*., 2005; Courtois *et al*., 2015; Maldonado *et al*., 2017; Mint Evolutionary Genomics Consortium, 2018). Often, different classes of compounds emerge in an almost mutually exclusive manner in different taxa (Wink, 2003). For instance, glucosinolates are major SMs near-universally in the Brassicaceae, the Capparidaceae and the Caricaceae (Moore *et al*., 2014), while benzylisoquinoline alkaloids occur mainly in the Papaveraceae, the Ranunculaceae, the Berberidaceae and the Menispermaceae (Ziegler and Facchini, 2008). Nevertheless, on a closer look within each chemical class individual compounds can vary in presence and/or quantity in inconsistent ways with or without phylogenetic signals through a clade, although members within the clade often share similar SMs (Pelser *et al*., 2005; Maldonado *et al*., 2017; Mint Evolutionary Genomics Consortium, 2018). This erratic distribution of particular SMs has puzzled botanists and ecologists for some time and has hindered the understanding of evolutionary origins of SM diversity. More dissections of SM diversity in a given chemical class under a particular phylogenetic context are needed.

Pyrrolizidine alkaloids (PAs) are a class of SMs with typical high diversity (Hartmann, 1996). More than 400 PAs existing as monoesters and open chain or macrocyclic diesters have been found in ca. 600 angiosperm species (Chou and Fu, 2006), of which more than 95% belong to the Asteraceae, the Boraginaceae, the Fabaceae and the Orchidaceae (Hartmann, 1999; Lange *et al*., 2011). PAs of the macrocyclic senecionine type are especially diverse with more than 100 structures, which are characteristic for PA containing species of the tribe Senecioneae of the Asteraceae (e.g. the genus *Senecio*) (Langel *et al*., 2011). In *Senecio*,

senecionine *N*-oxide synthesized in roots (Hartmann and Toppel, 1987; Toppel *et al*., 1987) was identified as the backbone structure of most PAs (Hartmann and Dierich, 1998). Senecionine *N*-oxide is transported via the phloem to shoots, where PA diversification takes place (Hartmann *et al*., 1989), resulting in species-specific or ecotype-specific PA bouquets (Hartmann and Dierich, 1998). PAs are present in plants as *N*-oxides and/or tertiary amines (free bases) (Wiedenfeld *et al*., 2008; Mulder *et al*., 2018), and in most cases PAs are synthesized and stored in their polar salt-like *N*-oxide form (Langel *et al*., 2011).

The *Jacobaea* species (26 species and formerly a part of *Senecio* species) all produce PAs (Pelser *et al*., 2005; Langel *et al*., 2011) but the composition and concentration appear to be species-specific (Soldaat *et al*., 1996; Hartmann and Dierich, 1998; Langel *et al*., 2011). Besides interspecific variation, a large variety of PA profiles within species was found. Different chemotypes of *J. vulgaris* and *J. erucifolia* are good examples of this intraspecific variation (Witte *et al*., 1992; Macel *et al*., 2004). The species in the genus *Jacobaea* have often been used to explore the evolutionary basis of SM diversity (Vrieling *et al*., 1993; Hartmann and Dierich, 1998; Macel *et al*., 2004; Cheng *et al*., 2011). Pelser *et al*. (2005) constructed the evolutionary history of qualitative PA variation (presence/absence of PAs) of herbarium samples of 24 *Jacobaea* species using GC-MS. Interestingly they found only weak phylogenetic signals as PA distribution appeared to occur largely incidentally within the whole clade. These authors suggested that PAs evolve and disappear rapidly during evolution.

In this study we wanted to verify or falsify the results of these authors by using more rigid methods. Pelser *et al*. (2005) were restricted by the use of herbarium specimens rather than fresh plant samples. This may have led to unwanted structural transformations and breakdown of particular PAs (Sander and Hartmann, 1989). Furthermore the herbarium specimens were field-collected and not grown under controlled circumstances and material from several specimens had to be combined due to the lack of enough material of single specimens. In the present study, we used a more sensitive analytical method and performed both quantitative and qualitative analyses of 80 PAs with LC-MS/MS of 8 to 10 week old leaves from 17 *Jacobaea* species represented by multiple populations and individuals grown under controlled conditions. In addition we amplified and sequenced 11 chloroplast DNA regions and three nuclear markers to fully resolve the molecular phylogeny of *Jacobaea* species. The resulting phylogenetic tree was used as a 'roadmap' to trace the evolutionary histories and to explore phylogenetic signals present within PA distributions.

Materials and methods

Analytical standards

Formic acid (analytical grade) and ammonium carbonate (analytical grade) were obtained from Sigma-Aldrich, Zwijndrecht, The Netherlands). Acetonitrile (LC-MS grade) and methanol (LC-MS grade) were obtained from Actu-all, Oss, The Netherlands). Twenty-seven PA

analytical standards were available, which were sourced from Phytoplan (Heidelberg, Germany), except for heliotrine (Valence, France), usaramine (BOC Sciences, Shirley, NY, USA) and florosenine (PRISNA, Leiden, The Netherlands). Usaramine *N*-oxide was in-house synthesized by reaction of usaramine with 30% hydroperoxide in methanol/water. Acetylerucifoline and acetylseneciphylline were prepared in-house from respectively, erucifoline and seneciphylline, by acetylation with acetic anhydride and dimethylaminopyridine in toluene. See Appendix 1 for a full list of PA standards used in this study. Stocks in methanol (100 mg/L) of the individual standards were prepared. Aliquots (1 mL) of these stock solutions (except for heliotrine) were combined to prepare a mixed standard solution in methanol (4 mg/mL).

Plant material

Seeds were obtained from botanical gardens or commercial seed companies including 40 accessions (Table S1) representing 17 of the 26 *Jacobaea* species. The seventeen species are evenly distributed throughout the three main clades (i.e. *Incani* s.l. group, *J. vulgaris* s.l. group, and *J. paludosa* group) according to the phylogeny inferred by Pelser *et al.* (2004). Each *Jacobaea* species contains one to three populations with 4-10 (average 6.6) individuals per population (Table S1). *Senecio vulgaris* which belongs to *Senecio* sect. *Senecio* was used as the outgroup for phylogenetic analyses only.

Plant growth and harvest

Seeds were germinated on the surface of wet potting soil covered by plastic bags and the seedlings were transferred in to 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). Plants were kept in a climate room (humidity 70%, 16h/8h light/dark cycle at 20 °C). Two to five fully grown vegetative leaves or early stem leaves were harvested after eight or ten weeks. Leaves were ground into fine powder in liquid nitrogen. A part of the powder was stored at -80 °C until DNA extraction, and the rest was freeze-dried and stored at -20 °C until PA extraction.

Preparation of extracts

Ten milligrams of powdered plant material was extracted with 1 mL of 0.2% formic acid which contained 1 μg/mL of heliotrine as an internal standard. After being shaken for 30 min on a tumbling machine, the extract was centrifuged at 13,000 rpm for 10 min and the supernatant was filtered through a 0.2 μm nylon membrane (Acrodisc 13 mm syringe filter, Pall Life Sciences, Ann Arbor, MI, USA). The filtered solutions were further diluted with milliQ water, depending on the expected concentrations in the extracts (see below). The additional dilution factor ranged from no dilution in the case of *J. cannabifolia*, to 10-fold (*J. adonidifolia, J. erucifolia*), 25-fold (*J. abrotanifolia, J. analoga, J. carniolia, J. leucophylla, J. maritima, J.*

paludosa, J. uniflora), up to 50-fold (*J. alpina, J. aquatica, J arnautorum, J. gnaphalioides, J. incana, J. subalpina, J. vulgaris*).

LC-MS/MS analysis

Analysis of PAs was performed on a LC-MS/MS system consisting of a Waters Acquity UPLC coupled to a Xevo TQ-S tandem mass spectrometer (Waters, Milford, MA, USA), running in positive electrospray mode. Chromatographic separation was achieved on an Acquity BEH C18 analytical column, 150 × 2.1 mm, 1.7 μm particle size (Waters, Milford, MA, USA). Eluent A consisted of water containing 10 mM ammonium carbonate (pH 9.0) and acetonitrile was used as eluent B. The gradient elution was performed as follows: 0.0 min 100% A/0% B, 0.1 min 95% A/5% B, 3.0 min 90% A/10% B, 7.0 min 76% A/24% B, 9.0 min 70% A/30% B, 12.0 min 30% A/70% B, and 12.1-15.0 min 100% A/0% B. The column was kept at 50ºC and a flow rate of 400 μL/min was applied. Two μL sample extract was injected. For each analyte at least two selected precursor to product ion MRM transitions were measured. Cone energy was 40 V and collision energy settings were optimized for the individual compounds. Besides the 24 PAs for which an analytical standard was available, the samples were screened for another 54 PAs for which no standards were available (see below). These PAs were included in the analytical method based on mass spectrometric data obtained from the analysis of extracts prepared from test samples of the individual *Jacobeae* species. See Appendix 1 for an overview of the MS/MS transitions used for the complete set of PAs.

Screening for new PA metabolites

Test samples of the *Jacobeae* species were extracted with 0.2% formic acid as described above, but no additional dilution was applied. The extracts were analyzed by running the LC-MS/MS in parent ion scanning mode. Fragment ions typically present in the fragmentation spectra of PAs were selected: ions with *m/z* 94; 118; 120 and 138 for senecionine-, jacobine- and erucifoline-like PAs; ions with *m/z* 122; 150 and 168 for otosenine type PAs and ions with *m/z* 96; 122 and 140 for platyphylline-like PAs. When a parent ion (the protonated molecular ion) was present in two or more scans produced from different fragment ions, this ion was marked as a potential PA. Further analysis of these potential PAs was conducted by measurement of the complete fragmentation spectrum at different collision energies (typically 20-30-40 eV). Compounds that produced fragmentation spectra indicative for PAs were included in the MRM method. Where possible a tentative assignment was made. In case this was not possible, the compound was annotated as either a free base compound or an *N*-oxide, including the protonated molecular mass and the retention time. Distinction between the free base and *N*-oxide form was made by measurement of the extract with and without chemical reduction (by adding a $Na_2S_2O_5$ solution to part of the extract (Joosten *et al*., 2010)). Compounds that upon reduction disappeared in the chromatogram were considered to be PA *N*-oxides, while compounds that remained stable or increased in concentration were considered to be PA free bases. In total 54 PAs were tentatively identified in one or more of the plant extracts and these were included in the final MRM method. The test extracts were also used to determine the

optimal dilution factor before LC-MS/MS analysis of the extracts of each species. Depending on the concentration of the PAs found in the test extracts a dilution factor was estimated that should ensure that the concentration of the most abundant PAs in the extracts should fall within the dynamic range of the mass spectrometric detector.

Quantification

The samples were run in a non-randomized order. Quantification was performed against a range of mixed standard solutions (0-5-10-25-50-100-200 µg/L) of the PAs in a diluted extract of *Tanacetum vulgare* (tansy). The extract of *T. vulgare* material was prepared in the same way as the other extracts (including a 10-fold additional dilution factor) and was used to mimic a PA-free plant extract. The range of mixed standard solutions was injected at the beginning of the series and at the end. The mixed standard solution of 50 µg/L in *T. vulgare* extract was injected every 30-40 samples, to monitor the performance of the system (drift in retention time, changes in detector sensitivity) during the analysis. For each PA the averaged response of two MRM transitions was used for quantification. For those compounds for which no reference standard was available, a semi-quantitative (indicative) value was obtained by comparison with the most closely related analogue (e.g. an isomer). For metabolites with tentative or unknown structures, no closely related standard could be identified. In such cases the concentration was estimated by taking the sum of the two most intense MRM transitions and comparing this with the sum area of a selected reference standard, as indicated in Appendix 1. Data processing was conducted with MassLynx 4.1 software (Waters Corporation, Milford, MA, USA).

Statistical analyses of metabolomics data

We performed analyses, quantitatively and qualitatively, using absolute concentrations, relative concentrations and binary coded presence (1)/absence (0) of all individual PAs (Appendix 1) of all *Jacobaea* plants. Before any statistical analysis the absolute and relative concentrations were log transformed prior to Pareto scaling, and the binary data were also processed with the Pareto scaling method but without prior transformation.

To extract and display the systematic variation of PA profiles, principal component analyses (PCA) were performed in SIMCA 15.0.2 and MetaboAnalyst (Chong *et al*., 2018). To compare similarities between PA profiles of different plant individuals, hierarchical cluster analyses (HCA) calculated with Euclidean distances and the Ward clustering algorithm were conducted using the tool MetaboAnalyst. We also performed Spearman rank correlation tests between all individual PAs using PA information extracted from all *Jacobaea* plants in MetaboAnalyst.

Phylogenetic analyses

DNA was extracted from frozen leaf powder using the Qiagen DNeasy Plant Mini Kit. In total, 11 plastid regions and three nuclear regions were amplified and sequenced. Primers used for amplification and sequencing are listed in Table S2. PCR products were sequenced on both strands for each region for all species whenever possible. Sequences were edited and

assembled in Sequencher® version 5.0 DNA sequence analysis software (Gene Codes, Ann Arbor, MI, USA), and the resulting consensus sequences were aligned for each region using default parameters in MUSCLE implemented in MEGA7 (Kumar *et al*., 2016) followed by manual curation where necessary. Subsequently, individual alignments were concatenated using Sequence Matrix software (Vaidya *et al*., 2011) to generate a combined dataset. All individual alignments and the combined dataset are available in Additional file 1.

Phylogenetic analyses were conducted using maximum likelihood (ML), Bayesian inference (BI) and maximum parsimony (MP) algorithms. The ML tree was obtained through IQ-tree with automatic evolutionary model selection (Nguyen *et al*., 2015; Kalyaanamoorthy *et al*., 2017) on XSEDE through CIPRES Science Gateway (http://www.phylo.org/). Bootstrap (BS) search was conducted using standard nonparametric bootstrap with 1000 replicates. BI was conducted with MrBayes 3.2.6 (Ronquist and Huelsenbeck, 2003) also on XSEDE through CIPRES Science Gateway. The Markov Chain Monte Carlo analyses (Altekar *et al*., 2004) were run for 1,000,000 generations on four chains. Furthermore, MP analyses were carried out using heuristic search in PAUP* 4.0a164 using 1000 replicates of random taxon addition sequence and the bisection-reconnection branch-swapping option. All characters were included in analyses and equally weighted, and gaps were treated as missing values. BS of all combined datasets were performed with 1000 replicates.

Ancestral state reconstruction and phylogenetic signals

For chemical diversity ancestral state reconstruction, we used the Mk1 model (Lewis, 2001) which was developed for discrete morphological data. This model assumes that transition rates of both forward and backward changes are equal/symmetrical. Overall, the 80 PA compounds were coded as binary traits (presence/absence) for each species. As long as at least one individual of a species contained a given PA, we coded the occurrence of this PA as present since it indicated that the species has the potential to produce this PA. Then the evolutionary history of each compound was traced across the ML topology of total plastid and nuclear evidence using ML methods as implemented in Mesquite 3.6 (https://www.mesquiteproject.org/).

Besides the qualitative aspects of PA composition, we also applied phylogenetic studies to quantitative PA data. We used statistics K (Blomberg *et al*., 2003) and λ (Pagel, 1999) to estimate the strength of phylogenetic signals of continuous PA traits (absolute and relative concentrations) for each species. Values of K or λ indicate whether the distribution of a trait is phylogenetically random (K or $\lambda \approx 0$) or non-random (K or $\lambda \approx 1$). The significance (P) of each K estimate was assessed by randomization test with 10,000 trait simulations. The significance (P) of each λ was evaluated with a likelihood ratio test, where the likelihood of observed λ estimated from the tree was compared to the likelihood of $\lambda = 1$. The tests were conducted using the 'phylosig' function in the 'phytools' package. P-values were adjusted for multiple comparisons by sequential Bonferroni methods using the 'p.adjust' function in the 'stats' package. All analyses were conducted in R version 3.5.1.

Results

Pyrrolizidine alkaloid diversity of *Jacobaea* species

We analyzed 80 PAs with LC-MS/MS extracted from the leaves of 17 *Jacobaea* species encompassing different populations and individuals grown in a climate chamber. The 80 PAs were classified into five structural groups, i.e. senecionine-like, erucifoline-like, jacobine-like, platyphylline-like, otosenine-like (Fig. 1), and unknown PAs (Appendix 1). Except for otosenine-like PAs which do not occur as *N*-oxides, the other five groups contain both tertiary (free base) and *N*-oxide forms.

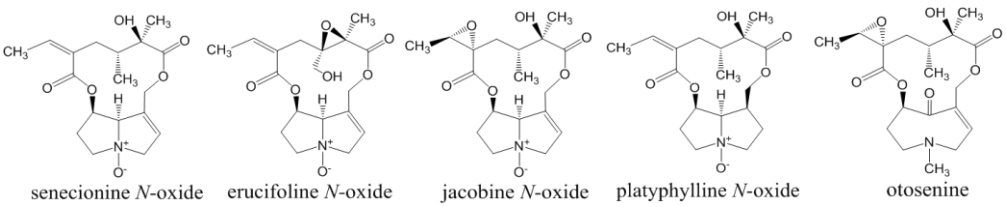

Figure 1. Structural formulas representative for the five different structural groups of PAs.

Population mean trait values have often been used to evaluate traits of plants. The mean concentrations of total PAs, the sum free bases and *N*-oxides, and the sum PA structural groups between different species and different populations were compared (Table S1). The average total PA concentrations in different populations ranged from 6.2 (*J. cannabifolia* 1) to 4301.3 (*J. vulgaris* 2) µg/g dry weight (DW). Even within species, considerable variations of mean total PA concentrations were observed between accessions for some species, such as *J. carniolia* and *J. alpina*. In general, *J. vulgaris s.l.* group contained more accessions with high amounts of PAs. PA free bases were predominant in *J. abrotanifolia*, *J. vulgaris*, *J. maritima* and *J. cannabifolia*, while PA *N*-oxides were dominant in *J. incana* and *J. subalpina*. However, a lack of consistency in the ratios of free bases to *N*-oxides seemed common between different accessions within *Jacobaea* species. Taking *J. aquatica* as an example, in the first accession free bases took up to 78.8%, while in the second accession free bases only accounted for 17.9%. These free base/*N*-oxide ratios were largely influenced by the types of PAs contained by plants. High percentages of jacobine-like or otosenine-like PAs often resulted into high percentages of free bases. Different chemotypes were found for some species including *J. adonidifolia* (erucifoline-type and otosenine-type), *J. aquatica* (senecionine-type and otosenine-type), *J. cannabifolia* (jacobine-type and mixed-type), and *J. paludosa* (jacobine-type and otosenine-type).

To compare differences of PA profiles among and within the *Jacobaea* species more comprehensively, we performed PCA using absolute concentrations, relative concentrations and presence/absence of PA traits of each *Jacobaea* plant. The distribution patterns based on the three aspects of PA traits were highly similar with slight changes of distances (or

dispersion) between observations (Fig. 2A; Fig. S1A; Fig. S1C). Most of the *Incani* s.l. group were separated from the *J. vulgaris* s.l. group based on PC1 and PC2 (Fig. 2A; Fig. S1A; Fig. S1C). The classification resulted mainly from the differences in senecionine-like PAs, jacobine-like PAs and otosenine-like PAs (Fig. 2B; Fig. S1B; Fig. S1D). The *Incani* s.l. group had higher contents of either senecionine-like PAs or otosenine-like PAs while the *J. vulgaris* s.l. group was characterized by more senecionine-like PAs or jacobine-like PAs. Meanwhile, a high degree of overlap between the *J. paludosa* group and both of the other two groups is shown in PCA score plots (Fig. 2A; Fig. S1A; Fig. S1C), clustering mainly below the axis along PC2. This distribution was caused by the lower amount or absence of senecionine-like PAs (Fig. 2B; Fig. S1B; Fig. S1D) in the *J. paludosa* group. On species level, plants were mainly clustered in species-specific ways except that *J. subalpina* and *J. arnautorum* could not be distinguished from each other based on five principal components (Fig. S2-S4). The PA patterns of different populations within some species were different, such as the cases of *J. aquatica* and *J. paludosa* (Fig. 2A; Fig. S1A; Fig. S1C).

We also performed HCA which can give information on closeness between *Jacobaea* plants based on the similarities of their PA bouquets. The clustering results based on the abovementioned three aspects of PA traits (absolute concentration, relative concentration, presence/absence) of all *Jacobaea* plants are shown as heatmaps (Fig. 3 and Fig. S5-S6). To a large extent in all cases the plants appeared to be clustered in species-specific ways, or to be more specific, population-specific ways, although some exceptions were found. Some species were grouped together, like the cluster of *J. subalpina* and *J. arnautorum*, and the cluster of *J. incana*, *J. leucophylla* and *J. uniflora*. Some species showed species-specific PA patterns but intraspecies variation surpassed interspecies differences. For example, the plants of *J. paludosa*, *J. alpina* and *J. aquatica* were always clustered into two different subclusters within the species in population-specific ways. Both *J. adonidifolia* and *J. cannabifolia* had two subclusters based on absolute concentrations or presence/absence of PAs, but had only one cluster by relative concentrations. *J. maritima* had two subclusters based on presence/absence of PA traits, but only one cluster when using either absolute or relative concentrations. Besides the clustering differences within species, the closeness among species varied using different aspects of PA traits. In all cases, the species could be divided into four sets loosely based on their closeness without considering their relative positions within each set: *J. vulgaris* related set (*J. vulgaris*, *J. maritima*, *J. paludosa*, *J. gnaphalioides*, *J. analoga*), *J. aquatica* related set (*J. aquatica*, *J. erucifolia*, *J. alpina*), *J. subapina* related set (*J. subalpina*, *J. arnautorum*, *J. carniolia*, *J. incana*, *J. uniflora*, *J. leucophylla*), and *J. abrotanifolia* related set (*J. abrotanifolia*, *J. adonidifolia*, *J. paludosa*). Nevertheless, the *J. vulgaris* related set was more closely related to the *J. aquatica* related set by absolute concentrations or present/absent PAs than by relative concentrations. And *J. cannabifolia* shifted its position from the *J. vulgaris* related set to the *J. abrotanifolia* related cluster when relative concentrations or binary PA traits were replaced by absolute concentrations.

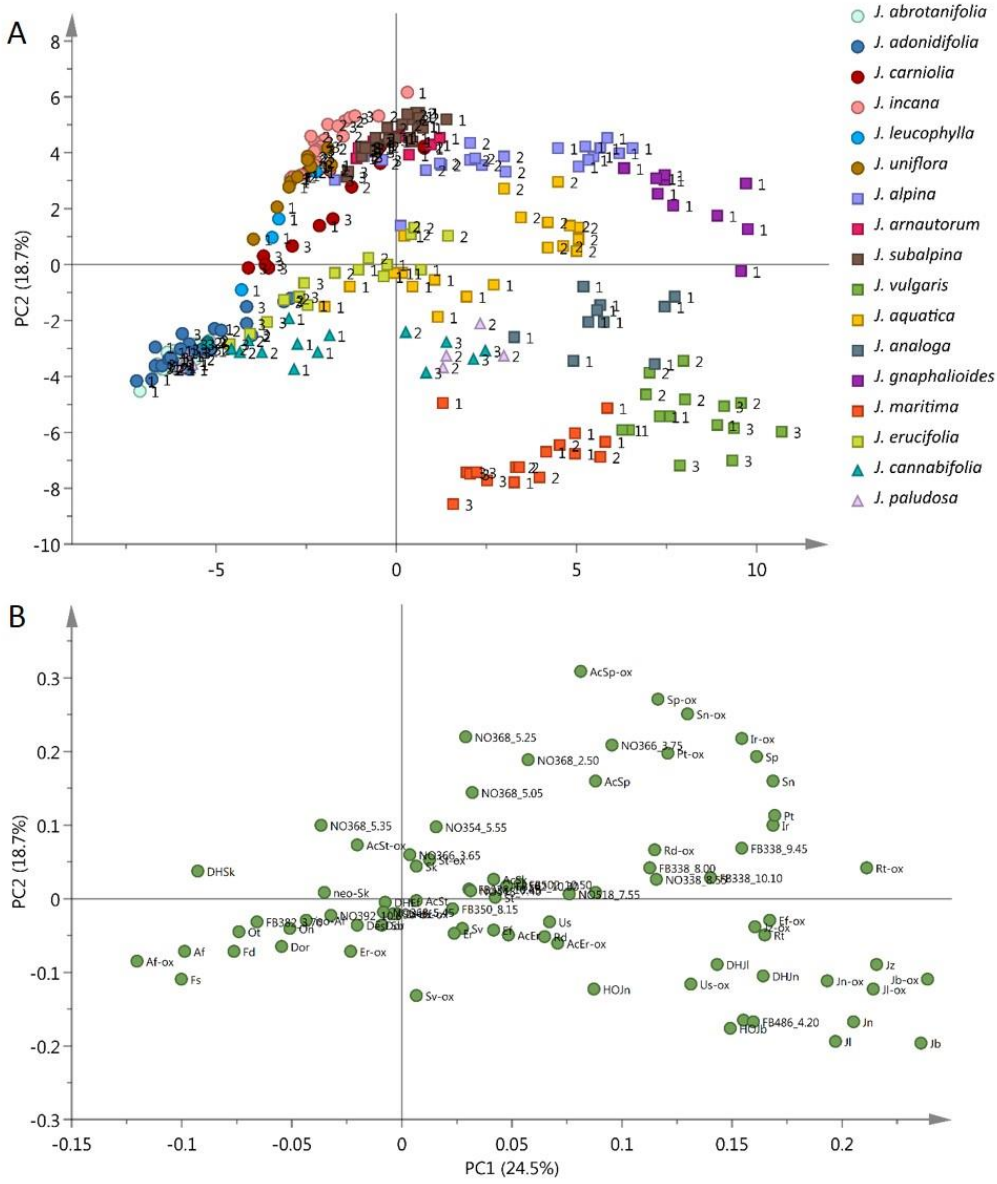

Figure 2. PA concentrations from different individuals and populations of 17 *Jacobaea* species grown in a climate chamber. (**A**) PCA score plot from SIMACA 15.0.2 based on the log-transformed and Pareto-scaled absolute concentrations of 80 PAs. PC1 and PC2 explain 24.5% and 18.7% of the variation, respectively. Each dot represents one plant individual. Different species are coded by different colors as indicated in the legend. Different phylogenetic groups are coded by different shapes: circle (*Incani*-group), square (*J. vulgaris*-group), triangle (*J. paludosa*-group). (**B**) The corresponding PCA loading plot. Each dot represents one PA. Abbreviations of PAs are listed in Appendix 1.

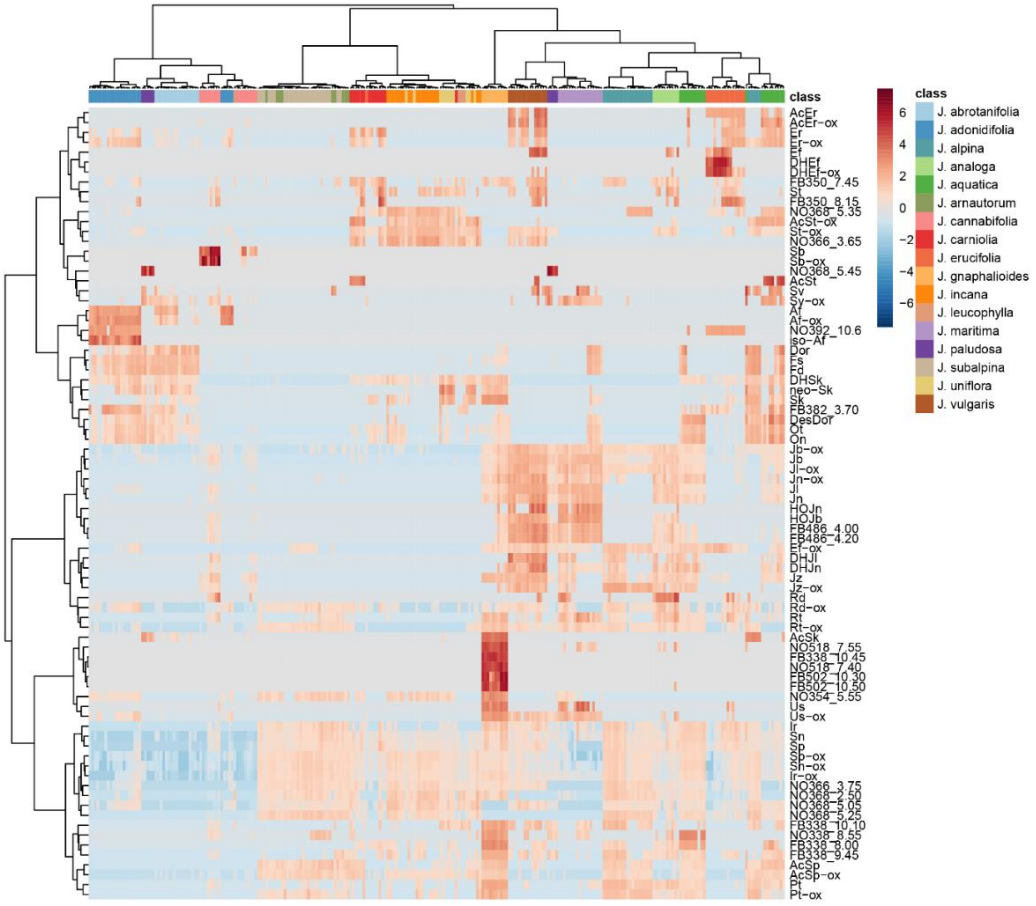

Figure 3. Heatmap representing hierarchical clustering analysis of all individual *Jacobaea* plants based on the absolute concentrations of 80 PAs. The analysis was calculated with Euclidean distances and the Ward clustering algorithm based on the log-transformed and Pareto-scaled absolute concentrations of PAs in the tool MetaboAnalyst. The tree diagram at the top indicates the closeness between different *Jacobaea* plants with regard to PA composition and concentration. Different species are color coded as indicated in the right-most legend. PA name abbreviations are shown in Appendix 1.

Moreover, we evaluated covariations between individual PAs by Spearman rank correlations with the PA information extracted from all *Jacobaea* plants. Based on absolute concentrations, the 80 PAs were roughly clustered into five clusters (Fig. 4). The derived PAs (erucifoline-like, jacobine-like, platyphylline-like, and otosenine-like) were clustered together even though there were some exceptions, whereas basic PAs (senecionine-like PAs) were distributed into different clusters. Jacobine-like PAs, otosenine-like PAs and platyphylline-like PAs were mainly assigned to separate clusters, and erucifoline-like PAs were divided into two clusters. Senecionine-like PAs had scattered distributions across different clusters.

Noticeably, otosenine-like PAs were negatively correlated to jacobine-like PAs, as well as to most of senecionine-like PAs. Similar patterns were found in HCAs of relative concentrations and presence/absence of PAs (Fig. S7-S8).

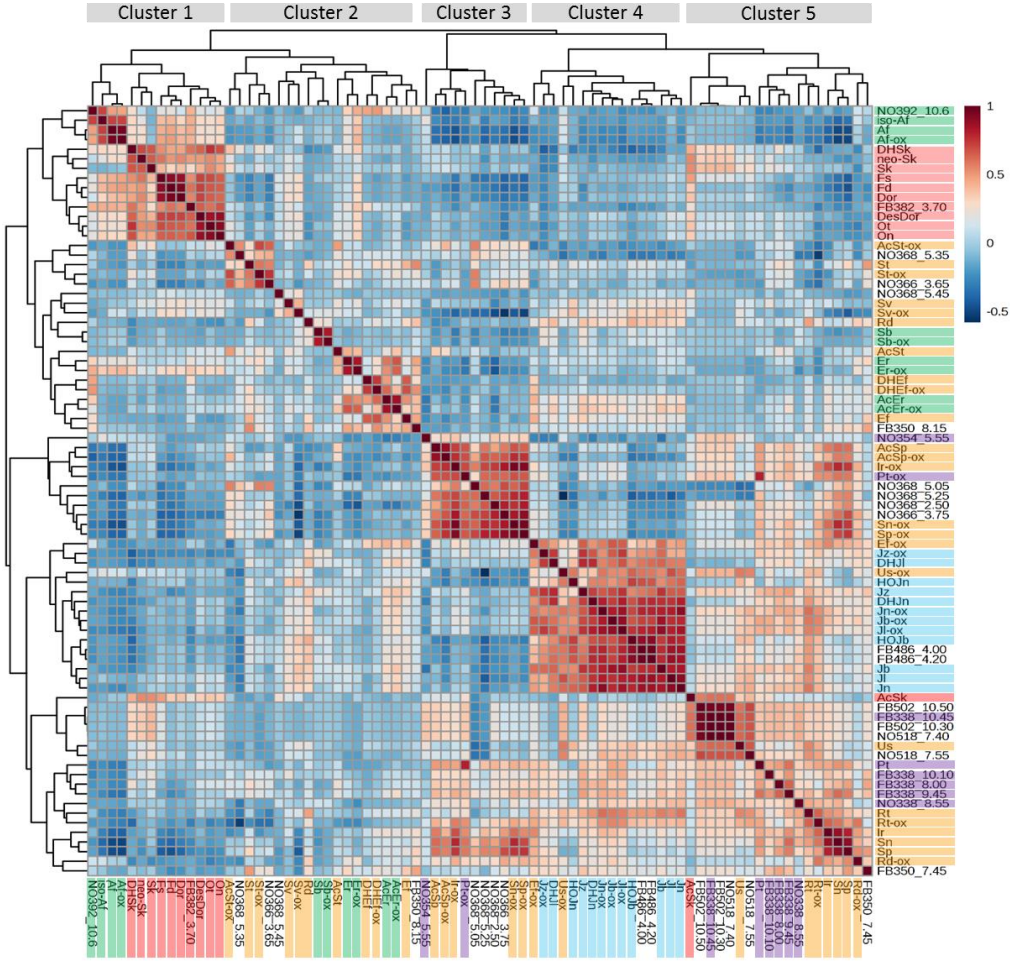

Figure 4. Heatmap representing Spearman rank correlation coefficients between individual PAs based on absolute concentrations of PAs of all *Jacobaea* species. The analysis was calculated with Euclidean distances and the Ward clustering algorithm based on the log-transformed and Pareto-scaled absolute concentrations of PAs in the tool MetaboAnalyst. Names of different known PA structural groups are highlighted with different colors: senecionine-like PAs (orange), erucifoline-like PAs (green), jacobine-like PAs (blue), platyphylline-like PAs (purple), otonesine-like PAs (red). PA name abbreviations are shown in Appendix 1.

Phylogeny of *Jacobaea* species

To obtain "roadmaps" for tracing back the evolutionary origin of PA diversity, we included 17 *Jacobaea* species to reconstruct the phylogeny with *S. vulgaris* as the outgroup in this study. In total, 11 plastid and three nuclear DNA makers were amplified and sequenced (Table S2), which were all included in phylogenetic trees ending up to a total length of 7590 bp (Additional file 1). We used different statistical methods (ML, BI, MP) to recover historical relationships based on the combined plastid and nuclear dataset. The topologies of phylogenetic trees obtained by ML and BI were nearly identical, and were thus represented by the same cladogram, where high BS values coincided with high posterior probabilities (PP) (Fig. 5). The consensus cladogram was largely in agreement with the phylogenetic cladogram based on DNA sequences and morphological data set in the previous study (Pelser *et al*., 2004). The three main clades found earlier were strongly supported: *Incani* s.l. group (*J. abrotanifolia*, *J. adonidifolia*, *J. carniolia*, *J.uniflora* and *J. leucophylla*), *J. vulgaris* s.l. group (*J. alpina*, *J. subalpina*, *J. vulgaris*, *J. analoga*, *J. maritima*, *J. aquatica*, *J. arnautorum*, *J. gnaphalioides* and *J. erucifolia*), and *J. paludosa* group (*J. cannabifolia* and *J. paludosa*). The *Incani* s.l. group was the most basal clade of *Jacobaea* species as a monophyletic clade in ML and BI phylogeny. For the *J. vulgaris* s.l. group, the phylogenetic relationships of seven closely related species (*J. alpina*, *J. analoga*, *J. aquatica*, *J. arnautorum*, *J. maritima*, *J. subalpina*, and *J. vulgaris*) still could not be resolved completely using the ML and BI algorithms based on the DNA regions studied. By using the MP algorithm, the *Incani* s.l. group was a polyphyletic assemblage (Fig. S9). A better resolution of phylogenetic relationship was obtained for the *J. vulgaris* s.l. group. All BS were over 70% except for the placement of *J. analoga* to the clade composed of *J. alpina*, *J. subalpina* and *J. vulgaris*. The clade of *J. aquatica* and *J. arnautorum* was more closely related to *J. gnaphalioides* than the other five species aforementioned. Since the ML algorithm determined the best fit substitution model and showed exactly the same result as the BI algorithm, we used the ML phylogenetic tree in the following steps for ancestral state reconstruction and the check of phylogenetic signals.

Ancestral state reconstruction and phylogenetic signals

We traced the evolutionary history of PA formation (presence/absence) using the abovementioned ML phylogeny based on total plastid and nuclear evidence as the "roadmap". Of the 80 PAs detected by LC-MS/MS, six (senecionine, senecionine *N*-oxide, integerrimine *N*-oxide, seneciphylline, seneciphylline *N*-oxide and riddelliine *N*-oxide) were present in all species while seven (dehydroeruciflorine, senecicannabine, senecicannabine *N*-oxide, iso-adonifoline, and the unidentified PAs FB338 (10.45), NO386 (5.45), FB502 (10.30), and NO518 (7.40)) were unique to a single species. Evolutionary patterns of the remaining PAs were complex, showing irregular presences or absences of individual PAs. Quite often, PAs showed frequent "on/off" changes without noticeable evolutionary direction (Additional file 2). This was well illustrated by the evidence that all three clades of *Jacobaea* species contain both species with or without certain PAs such as erucifoline *N*-oxide (Fig. 6A). On the contrary, some PAs seemed to have originated or been lost only a few times in parallel.

Examples of this pattern included the absence of senkirkine (Fig. 6B) and the presence of adonifoline (Fig. 6C). Unlike the patterns mentioned previously, jaconine *N*-oxide was the only PA identified in our study of which the occurrence showed clear evolutionary direction, as this PA was absent in all species of the *Incani* s.l. clade while present in all species of the other two clades (*J. vulgaris* s.l. clade and *J. paludosa* clade; Fig. 6D).

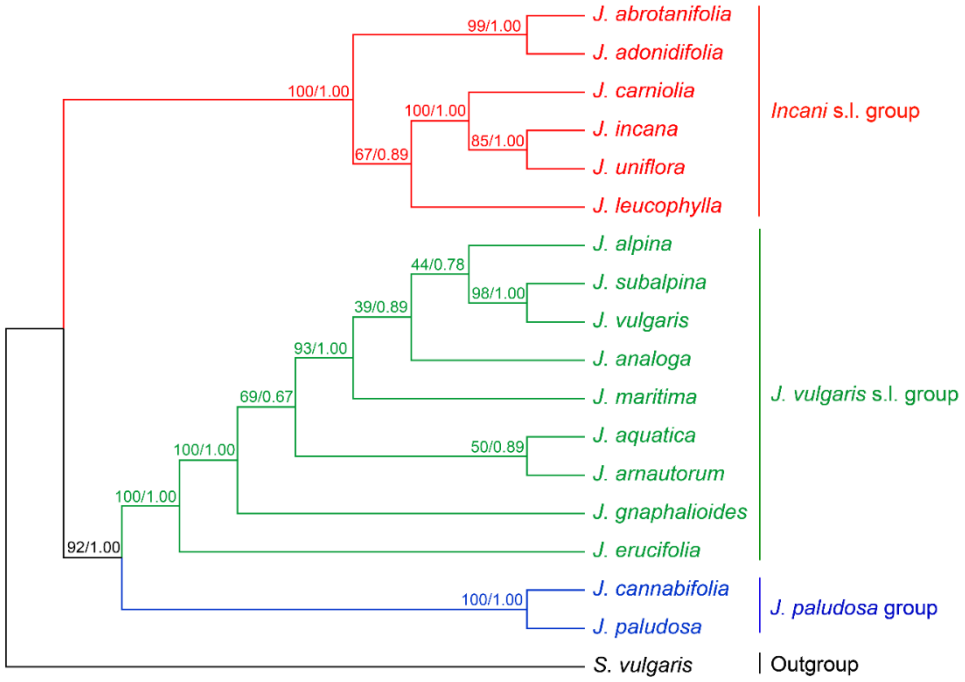

Figure 5. Maximum likelihood (ML) and Bayesian inference (BI) consensus cladogram of 17 *Jacobaea* species inferred from the combined plastid and nuclear dataset. ML bootstrap values and Bayesian posterior probabilities (BS/PP) are indicated above the branches. Different groups are color coded: *Incani* s.l.-group (red), *J. vulgaris* s.l.-group (green), *J. paludosa*-group (blue), and outgroup *S. vulgaris* (black).

Phylogenetic signal is defined as the tendency for related species to resemble each other more than they resemble species drawn at random from the tree (Blomberg and Garland, 2002). We used two quantitative measures, namely Blomberg *et al.*'s (2003) *K* and Pagel's (1999) λ, to measure phylogenetic signals of continuous PA traits (average absolute and relative concentrations) for each species. After sequential Bonferroni adjustment, only three PAs (i.e. dehydroeruciflorine, dehydroeruciflorine *N*-oxide and NO368 (5.45)) showed significant phylogenetic signals ($\lambda \approx 1$; $P < 0.05$) under λ statistics in their absolute concentrations (Additional file 3). Of these three PAs, dehydroeruciflorine was unique to *J. erucifolia* while NO368 (5.45) was unique to *J. paludosa*. Dehydroeruciflorine *N*-oxide was detected in *J. erucifolia* as well as in *J. vulgaris* and *J. analoga* with low amounts only in a few individuals

of the latter two species. For relative concentrations, under λ statistics nine PAs (integerrimine, senecivernine *N*-oxide, eruciflorine, eruciflorine *N*-oxide, dehydroeruciflorine, dehydroeruciflorine *N*-oxide, erucifoline *N*-oxide, acetylerucifoline and NO368 (5.45)) showed significant phylogenetic signals. None of the PAs showed significant phylogenetic signals by *K* statistics either in absolute or relative concentrations (Additional file 3).

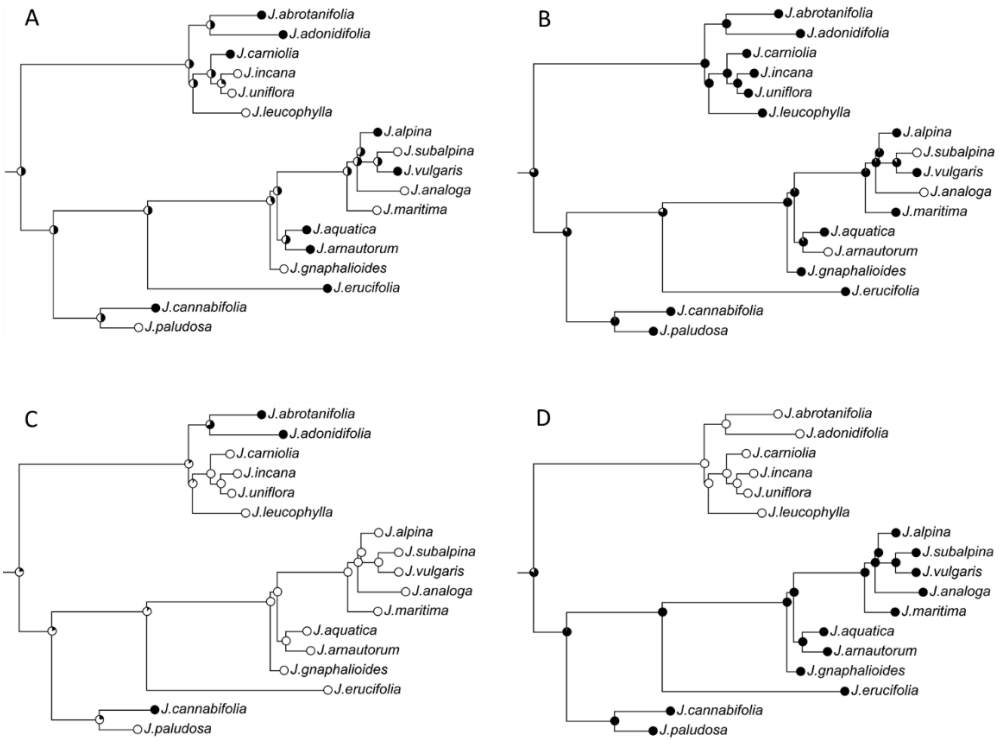

Figure 6. Maximum likelihood ancestral state reconstruction of four PAs. The tree shown is the ML tree from the total plastid and nuclear dataset of 17 *Jacobaea* species. (A) erucifoline *N*-oxide (B) senkirkine (C) adonifoline (D) jaconine *N*-oxide. Character states were binary coded for each species and are shown as small pie charts before taxon names: black (presence) and white (absence). The pie charts shown at individual nodes illustrate the likelihood for the ancestral states.

Discussion

We observed flexible PA profiles among 17 different *Jacobaea* species with respect to both quantitative and qualitative PA variations showing both high inter-/intra-species PA diversity, which supports previous findings (Hartmann and Dierich, 1998; Macel *et al*., 2004; Pelser *et al*., 2005; Langel *et al*., 2011). In total, 80 PAs were detected including both free bases and *N*-oxides in this study, which covered all 26 PAs reduced to the tertiary form except sennecicannabine and deacetyldoronine detected in Pelser *et al*. (2005). In our study, the

average total PA concentration of a species ranged from 32.9 (*J. cannabifolia*) to 3835.7 (*J. gnaphalioides*) µg/g DW. The numbers of PAs varied from 21 (*J. leucophylla*) to 59 (*J. aquatica*) with different relative abundances between different species. Both PA concentrations and compositions were confirmed to be species-specific, although in some cases PA patterns from different populations within species differed from each other and were even surpassing differences between species (Fig. 2; Fig. 3; Fig. S1-S6). However, the taxonomic relationships derived from qualitative and quantitative PA profiles of different *Jacobaea* species are inconsistent with their phylogenetic relationships based on molecular markers. For instance, *J. abrotanifolia* and *J. adonidifolia* were always grouped with the other four species of the *Incani* s.l. clade in different clusters based on their PA patterns (Fig. 3; Fig. S5-6), which is incongruent with the *Incani* s.l. cluster on phylogenetic trees (Fig. 5; Fig. S9). On a closer look, we examined the occurrences and phylogenetic signals of individual PAs using the mean phylogeny of 17 *Jacobaea* species as the "roadmap". By tracing the evolutionary history of PA formation (presence/absence), we found that except the PAs unique to a single species or ubiquitous among all species, PAs appear to distribute incidentally within the genus *Jacobaea*, revealing limited phylogenetic signals, which is in agreement with the findings of Pelser *et al*. (2005). For quantitative PA data, none of the PAs showed significant phylogenetic signals under *K* statistics, while nine of the 80 PAs showed phylogenetic signals based on relative and/or absolute concentrations under λ statistics. These results were similar to those indicated for terpenoids in Lamiaceae, as only 25% (14 out of 57 without a multiple-comparison correction) of the tested monoterpenes showed significant phylogenetic signals for their presence/absence at species level (Mint Evolutionary Genomics Consortium, 2018). This suggests that the distributions of different individual SMs within a chemical class on phylogenetic trees are often random, lacking phylogenetic signals.

Chemical diversity is attributable to the effects of genetic variation, environmental influences and the interaction between these two factors (Moore *et al*., 2014; Kessler and Kalske, 2018). In our study, we grew all the plants under the same condition aiming at minimizing environmental variation. It was demonstrated that the individual PA bouquets were brought about by genetically controlled specific processes in *Senecio* and *Jacobaea* species (Hartmann and Dierich, 1998). Under controlled conditions, 50-100% of the total variation in total PA concentration of *J. vulgaris* is due to genetic differences (Vrieling *et al*., 1993), and PA composition and concentration were genotype-dependent (Macel *et al*., 2004; Cheng *et al*., 2011). However, the phylogenetic distances of *Jacobaea* species were not correlated with differences in their PA bouquets. Even within species different populations which had highly similar DNA sequences showed rather different PA patterns, such as the case of the two populations of *J. paludosa*. This discrepancy between PA profiles and phylogenetic relationships might be due to their maternal effects or gene-environment interactions. According to Kessler and Kalske (2018), organisms interacting with plants can use SM bouquets to find appropriate hosts. With varying SM profiles, plants should lower the chance of attack from herbivores by diverting chemical compositions away from a common host

search pattern used by a potential herbivores. If this process would be stronger within phylogenetically related plants it would decrease the phylogenetic signals of the SMs. Compared to traits such as morphological characters, SMs may be under weak evolutionary constraints due to relatively lower production costs. As long as evolutionary constraints are not limiting the response to selection, even relatively weak selection can lead to adaptive changes, thus resulting in the losses and gains of SMs (Kessler and Kalske, 2018). This rapid evolutionary fine-tuning might be excellent mechanistic basis for plants to cope with multiple selective forces.

The composition of plant SMs are vital in determining the evolutionary success of populations and species (Burow *et al*., 2010). Different SMs may have played different roles during plant evolution and have been exposed to different selective forces. Statistically, jacobine-like PAs, senecionine-like PAs and otosenine-like PAs played more important roles in the classification of different *Jacobaea* species quantitatively and qualitatively in PCA (Fig. 2B; Fig. S1B. Fig. S1D), which suggests that these PAs may be involved in speciation. Senecionine-like PAs have been regarded as biosynthetically basic PAs and can be found in all *Jacobaea* species. The total amount of PAs in plants is controlled by the formation of senecionine *N*-oxide in roots, and the constitutive biosynthesis of senecionine *N*-oxide is genotype dependent (Hartmann and Dierich, 1998). Jacobine-like PAs have a higher percentage of free bases and otosenine-like PAs are only present as free bases, which are regarded as biosynthetically more derived PAs. In general, free base PAs caused a lower survival to insect herbivores compared with *N*-oxides (Liu *et al*., 2017). Cheng *et al*. (2013) found that tertiary amines of jacobine-like PAs and some otosenine-PAs were positively correlated with the oviposition preferences of the specialist herbivore cinnabar moth. Interestingly, most erucifoline-like PAs only had marginal weights in the classification of different species in PCA. Wei *et al*. (2019) used methyl jasmonate to treat *J. vulgaris* and *J. aquatica* mimicking the effects of herbivory, and they found a strong shift from senecionine-like PAs to erucifoline-like PAs in both species. This might reveal that *Jacobaea* species have a similar defense strategy related to erucifoline-like PAs.

At the molecular level, the fact that PA profiles detected in this study were highly plastic among and within *Jacobaea* species, and very often PAs appeared and disappeared on the phylogenetic tree may be due either to convergent evolution of PA specific genes where the ability of plants to produce particular PAs evolved several times, or to differential gene expression where all the plants possess the machinery to produce all PAs but some PA specific genes are not expressed in some species. Given the prevalent intraspecific PA diversity, and the detection of ‘unique’ PAs in more species in other studies, e.g. dehydroeruciflorine was detected in *J. vulgaris* by Carvalho *et al*. (2014), the latter assumption is preferred even though the regulation mechanisms involved in biosynthetic pathways of PAs are unclear. To confirm this hypothesis, we need to understand the molecular basis of PA biosynthesis. Genes involved in SM biosynthesis have often evolved from genes involved in primary metabolism by gene duplication with successive diversification (Ober, 2010; Moore *et al*., 2014). Many of these

genes involved in SM pathways belong to large gene families (Kessler and Kalske, 2018), such as terpene synthases (Tholl, 2006) and cytochrome P450s (Bak *et al*., 2006; Frey *et al*., 2009). Further investigation of the diversity of large gene families may allow insights into the evolution of SM pathways that coordinate the respective enzymes.

In conclusion, we analyzed PA profiles of 17 *Jacobaea* species including different individuals and populations quantitatively and qualitatively. Both PA concentrations and compositions were confirmed to be species-/population-specific. The PAs driving the classification may implicate their important roles in ecological processes of different species. By tracing the occurrence and evaluating the phylogenetic signal of each PA trait, we found that PAs were more incidentally distributed along the phylogeny with limited phylogenetic signals. The PA diversity among and within *Jacobaea* species is more likely the result of differential expression of PA biosynthesis genes as a life strategy in response to different biological needs rather than the result of gains and losses of particular PA biosynthesis genes during evolution.

Acknowledgements

Yangan Chen thanks the China Scholarship Council (CSC) for financial support. We thank Botanical Garden of TU Braunschweig, Botanical Garden-Insitute FEB RAS, Botanischer Garten des Institutes für Botanik der Universität Graz, Botanischer Garten und Botanisches Museum of Freie Universität Berlin, Conservatoire et Jardin botaniques de la Ville de Genève, Giardino Botanico Alpino Rezia, Giardino Botanico Daniela Brescia, Hortus Botanicus Leiden, Hortus Botanicus Tallinnensis, Jardin Botanique Alpin du Lautaret, Jardins botaniques du Grand Nancy et de l'Université de Lorraine, Palace and Botanical Gardens of Balchik, Polish Academy of Sciences Botanical Garden Center for Biological Diversity Conservation in Powsin and Royal Botanic Gardens Kew for donating seeds. We thank Karin van der Veen-van Wijk, Maria Franco Berriel and Rick Hennevelt for their technical assistance, Dr. Young Hae Choi for his kind support for the SIMCA software, and Martine Huberty for her discussion about statistical analyses.

References

Altekar G, Dwarkadas S, Huelsenbeck JP, Ronquist F. 2004. Parallel Metropolis-coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics* 20:407-415.

Bak S, Paquette SM, Morant M, Vinther Moranth A, Saito S, Bjarnholt N, Zagrobelny M, Jorgensen K, Osmani S, *et al.* 2006. Cyanogenic glycosides: a case study for evolution and application of cytochromes P450. *Phytochemistry Reviews* 5:309-329.

Blomberg SP, Garland Jr T. 2002. Tempo and mode in evolution: Phylogenetic inertia, adaptation and comparative methods. *Journal of Evolutionary Biology* 15:899 -910.

Blomberg SP, Garland T, Ives AR. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57:717-745.

Burow M, Halkier BA, Kliebenstein DJ. 2010. Regulatory networks of glucosinolates shape *Arabidopsis thaliana* fitness. *Current Opinion in Plant Biology* 13:348-353.

Carvalho S, Macel M, Mulder PPJ, Skidmore A, van der Putten WH. 2014. Chemical variation in *Jacobaea vulgaris* is influenced by the interaction of season and vegetation successional stage. *Phytochemistry* 99:86-94.

Cheng D, Hirk 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.

Cheng D, van der Meijden E, Mulder PPJ, Vrieling K, Klinkhamer PGL. 2013. Pyrrolizidine alkaloid composition influences cinnabar moth oviposition preferences in *Jacobaea* hybrids. *Journal of Chemical Ecology* 39:430-437.

Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, Wishart DS, Xia J. 2018. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Research* 46:W486-494.

Chou MW, Fu PP. 2006. Formation of DHP-derived DNA adducts in vivo from dietary supplements and Chinese herbal plant extracts containing carcinogenic pyrrolizidine alkaloids. *National Center for Toxicological Research* 22:321-327.

Courtois EA, Dexter KG, Paine CET, Stien D, Engel J, Baraloto C, Chave J. 2015. Evolutionary patterns of volatile terpene emissions across 202 tropical tree species. *Ecology and Evolution* 6, 2854-2864.

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.

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. Site 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.

Hartmann T. 1996. Diversity and variability of plant secondary metabolism: a mechanistic view. *Entomologia Experimentalis et Applicata* 80:177-188.

Hartmann T. 1999. Chemical ecology of pyrrolizidine alkaloids. *Planta* 207:483-495.

Joosten L, Mulder PPJ, Vrieling K, van Veen JA, Klinkhamer PGL. 2010. The analysis of pyrrolizidine alkaloids in *Jacobaea vulgaris*; a comparison of extraction and detection methods. *Phytochemical Analysis* 21:197-204.

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 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.

Lewis PO. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* 50:913-925.

Liu X, Klinkhamer PGL, Vrieling K. 2017. The effect of structurally related metabolites on insect herbivores: a case study on pyrrolizidine alkaloids and western flower thrips. *Phytochemistry* 138:93-103.

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

Maldonado C, Barnes CJ, Cornett C, Holmfred E, Hansen SH, Persson C, Antonelli A, Rønsted N. 2017. Phylogeny predicts the quantity of antimalarial alkaloids within the iconic yellow cinchona bark (Rubiaceae: *Cinchona calisaya*). *Frontiers in Plant Science* 8:391.

Mint Evolutionary Genomics Consortium. 2018. Phylogenomic mining of the mints reveals multiple mechanisms contributing to the evolution of chemical diversity in Lamiaceae. *Molecular Plant* 11:1084-1096.

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.

Mulder PPJ, López P, Castelari M, Bodi D, Ronczka S, Preiss-Weigert A, These A. 2018. Occurrence of pyrrolizidine alkaloids in animal-and plant-derived food: results of a survey across Europe. *Food Additives & Contaminants: Part A* 35:118-133.

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.

Pagel M. 1999. Inferring the historical patterns of biological evolution. *Nature* 401:877-884.

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.

Ronquist F, Huelsenbeck JP. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572-1574.

Sander H, Hartmann T. 1989. Site of synthesis, metabolism and translocation of senecionine *N*-oxide in cultured roots of *Senecio erucifolius*. *Plant Cell, Tissue and Organ Culture* 18:19-31.

Soldaat LL, Boutin JP, Derridj S. 1996. Species-specific composition of free amino acids on the leaf surface of four *Senecio* species. *Journal of Chemical Ecology* 22:1-12.

Tholl D. 2006. Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. *Current Opinion in Plant Biology* 9:297-304.

Toppel G, Witte L, Riebesehl B, von Borstel K, Hartmann T. 1987. Alkaloid patterns and biosynthetic capacity of root cultures from some pyrrolizidine alkaloid producing *Senecio* species. *Plant Cell Reports* 6:466-469.

Vaidya G, Lohman DJ, Meier R. 2011. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics* 27:171-180.

Vrieling K, de vos H, van wijk CAM. 1993. Genetic analysis of the concentration of pyrrolizidine alkaloids in *Senecio jacobaea*. *Phytochemistry* 32:1141-1144.

Wei X, Vrieling K, Mulder PPJ, Klinkhamer PGL. 2019. Methyl jasmonate changes the composition and distribution rather than the concentration of defence compounds: a study on pyrrolizidine alkaloids. *Journal of Chemical Ecology* 45:136-145.

Wiedenfeld H, Roeder E, Bourauel T, Edgar J. 2008. Pyrrolizidine alkaloids: structure and toxicity. Bonn: V&R unipress GmbH.

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

Wink M. 2008. Plant secondary metabolism: diversity, function and it evolution. *Natural Product Communications* 3:1205-1216.

Witte L, Ernst L, Adam H, Hartmann T. 1992. Chemotypes of two pyrrolizidine alkaloid-containing *Senecio* species. *Phytochemistry* 31:559-565.

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

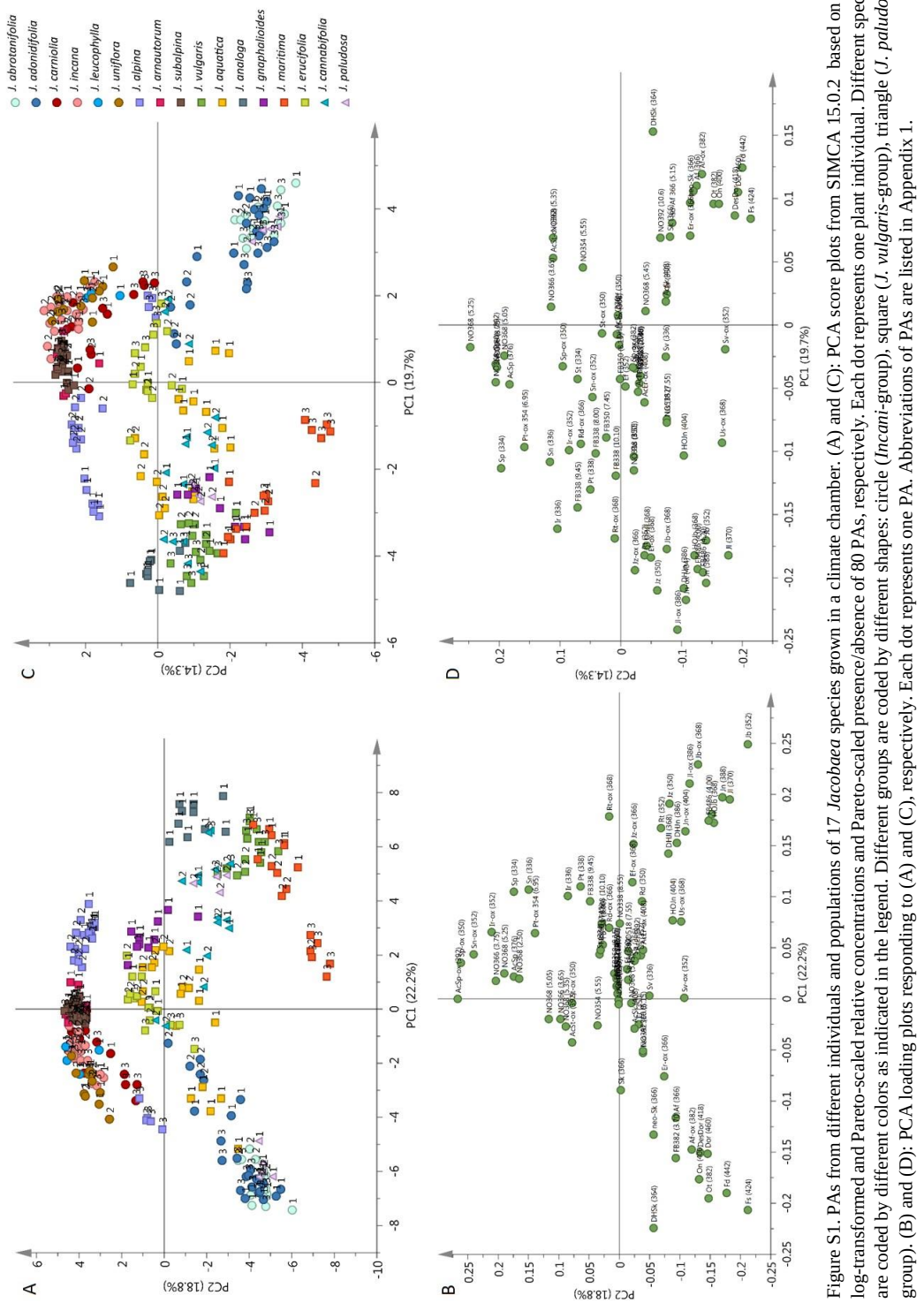

Figure S1. PAs from different individuals and populations of 17 *Jacobaea* species grown in a climate chamber. (A) and (C): PCA score plots from SIMCA 15.0.2 based on the log-transformed and Pareto-scaled relative concentrations and Pareto-scaled presence/absence of 80 PAs, respectively. Each dot represents one plant individual. Different species are coded by different colors as indicated in the legend. Different groups are coded by different shapes: circle (*Incani*-group), square (*J. vulgaris*-group), triangle (*J. paludosa*-group). (B) and (D): PCA loading plots responding to (A) and (C), respectively. Each dot represents one PA. Abbreviations of PAs are listed in Appendix 1.

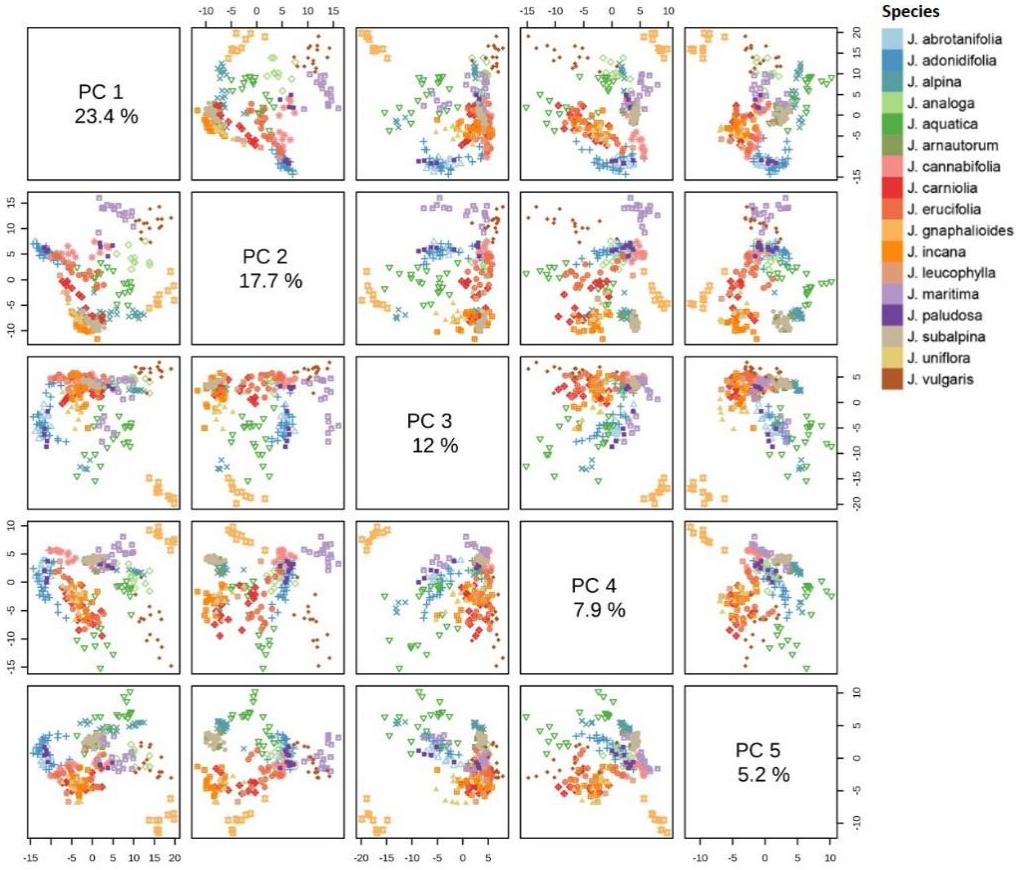

Figure S2. PCA score plots of absolute PA concentration from different individuals and populations of 17 *Jacobaea* species grown in a climate chamber. The plots show the five auto-fit principal components based on the log-transformed and Pareto-scaled absolute concentrations of 80 PAs using the tool MetaboAnalyst. R^2 of each PC is shown in the figure. Each dot represents one plant individual. Different species are coded by different colors as indicated in the legend.

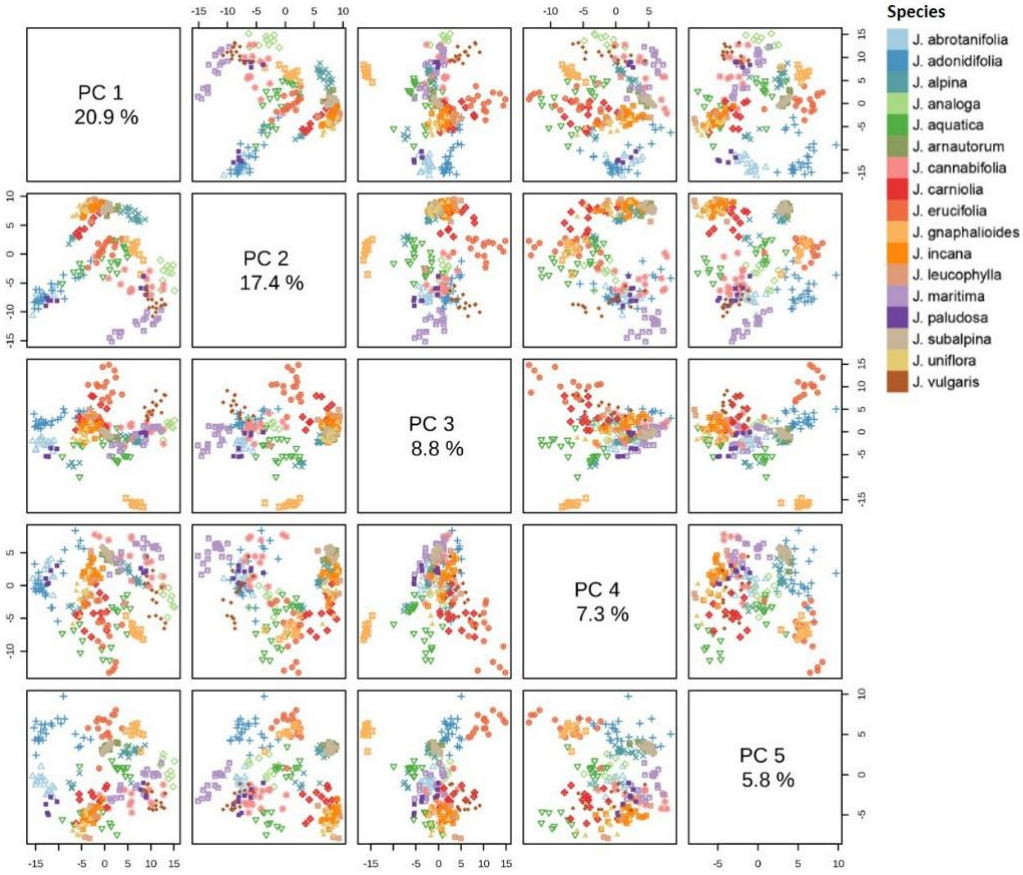

Figure S3. PCA score plots of relative PA concentration from different individuals and populations of 17 *Jacobaea* species grown in a climate chamber. The plots show the five auto-fit principal components based on the log-transformed and Pareto-scaled absolute concentrations of 80 PAs using the tool MetaboAnalyst. R^2 of each PC is shown in the figure. Each dot represents one plant individual. Different species are coded by different colors as indicated in the legend.

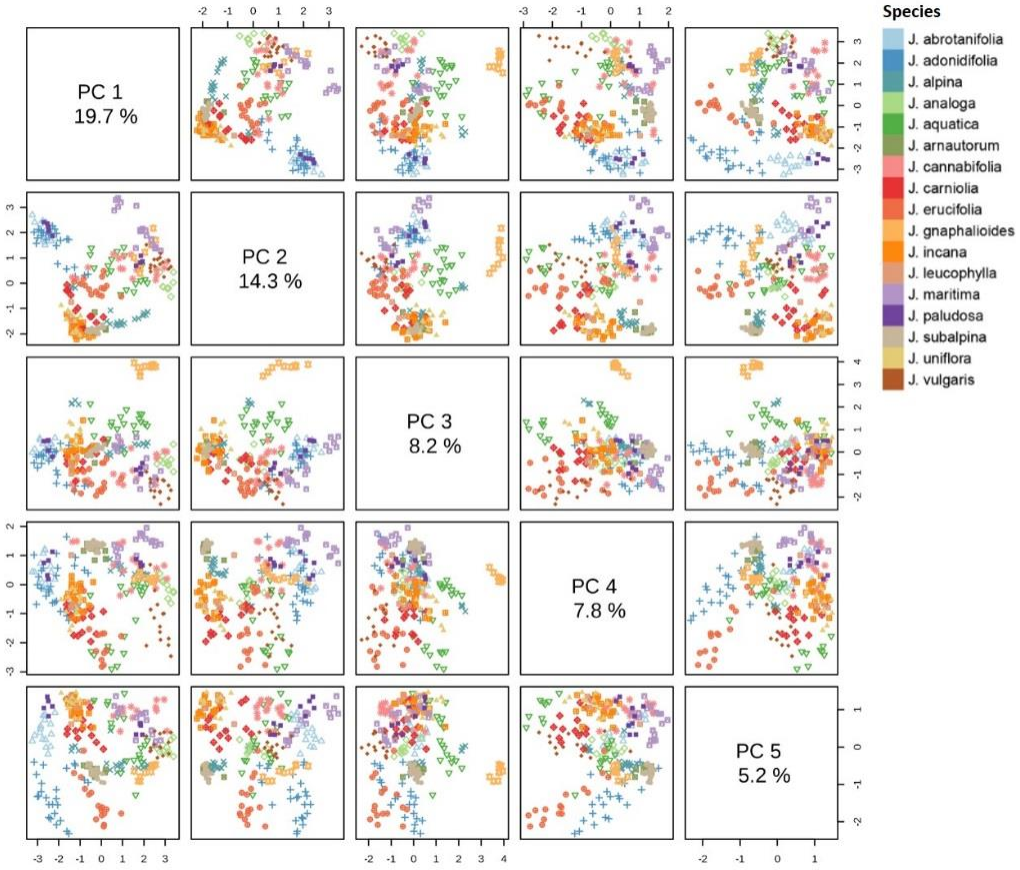

Figure S4. PCA score plots of PAs from different individuals and populations of 17 *Jacobaea* species grown in a climate chamber. The plots show the five auto-fit principal components based on the Pareto-scaled presence/absence of 80 PAs using the tool MetaboAnalyst. R^2 of each PC is shown in the figure. Each dot represents one plant individual. Different species are coded by different colors as indicated in the legend.

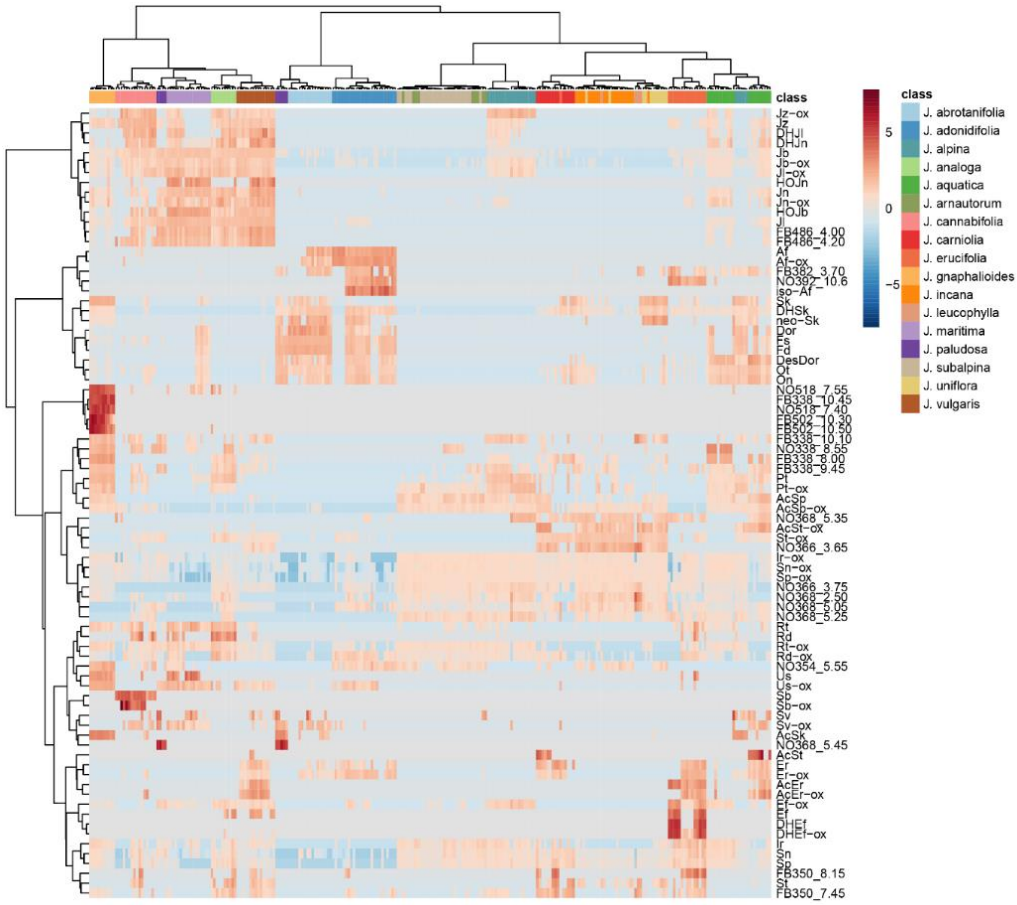

Figure S5. Heatmap representing hierarchical clustering analysis of all individual *Jacobaea* plants based on the relative concentrations of 80 PAs. The analysis was calculated with Euclidean distances and the Ward clustering algorithm based on the log-transformed and Pareto-scaled relative concentrations of PAs in the tool MetaboAnalyst. The tree diagram on the top indicates the closeness between different *Jacobaea* plants. Different species are color coded as indicated in the right-most legend. PA name abbreviations are shown in Appendix 1.

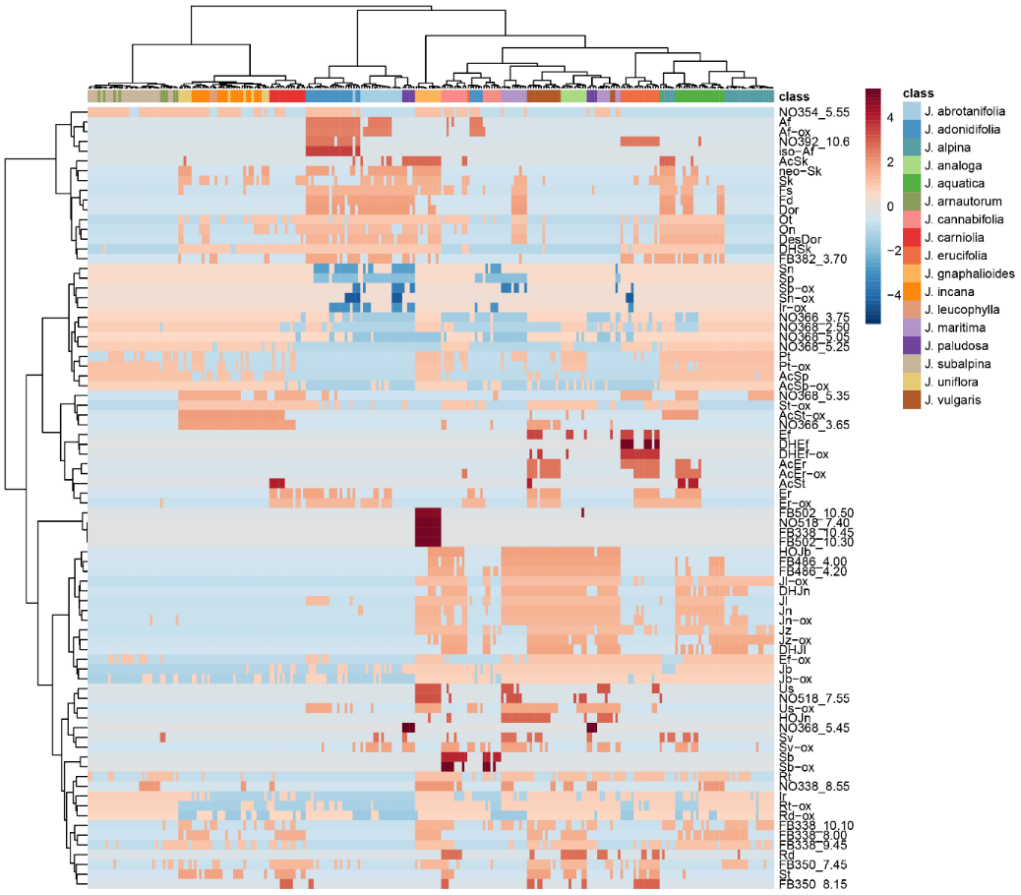

Figure S6. Heatmap representing hierarchical clustering analysis of all individual *Jacobaea* plants based on the presence/absence of 80 PAs. The analysis was calculated with Euclidean distances and the Ward clustering algorithm based on Pareto-scaled presence (1)/absence (0) of PAs in the tool MetaboAnalyst. The tree diagram on the top indicates the closeness between different *Jacobaea* plants. Different species are color coded as indicated in the right-most legend. PA name abbreviations are shown in Appendix 1.

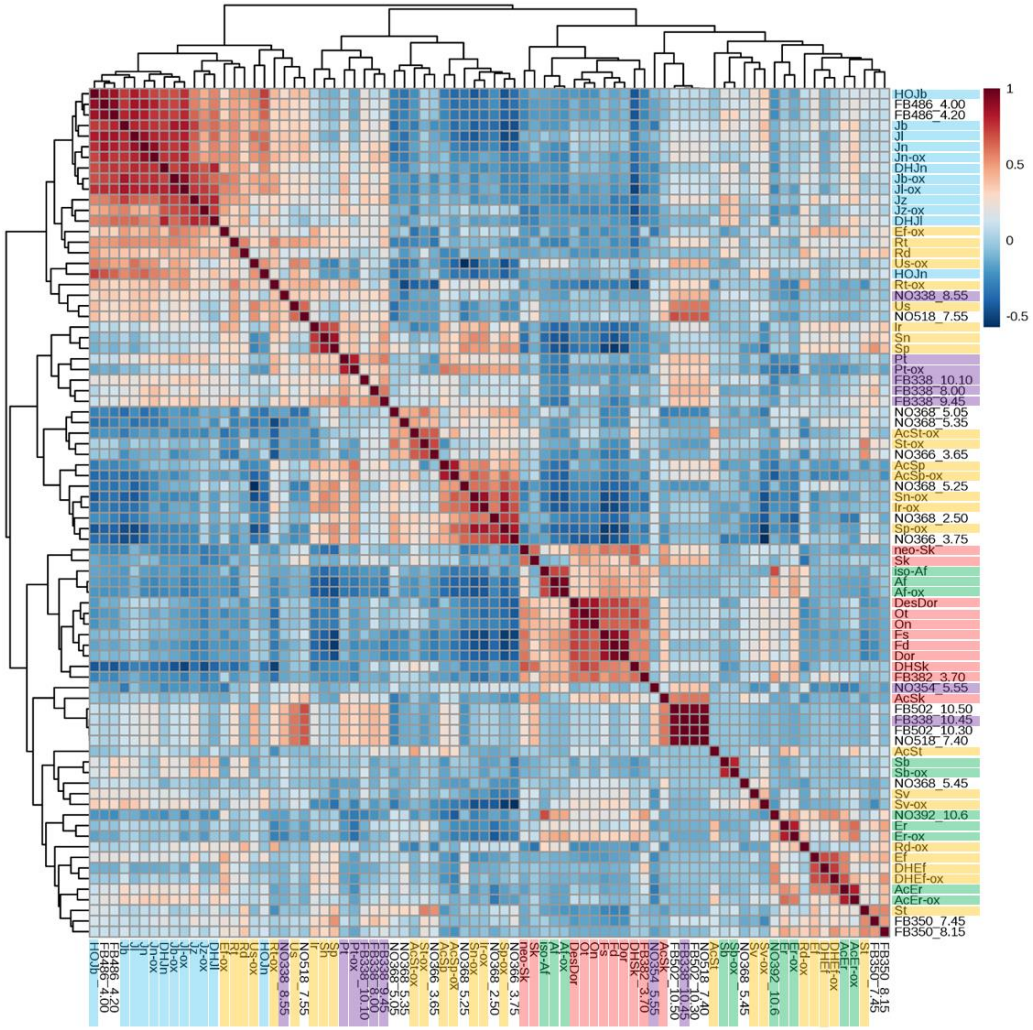

Figure S7. Heatmap representing Spearman rank correlation coefficients between individual PAs based on relative concentrations of PAs of all individual *Jacobaea* plants. The analysis was calculated with Euclidean distances and the Ward clustering algorithm based on the log-transformed and Pareto-scaled relative concentrations of PAs in the tool MetaboAnalyst. Names of different known PA structural groups are highlighted with different colors: senecionine-like PAs (orange), erucifoline-like PAs (green), jacobine-like PAs (blue), platyphylline-like PAs (purple), otonesine-like PAs (red). PA name abbreviations are shown in Appendix 1.

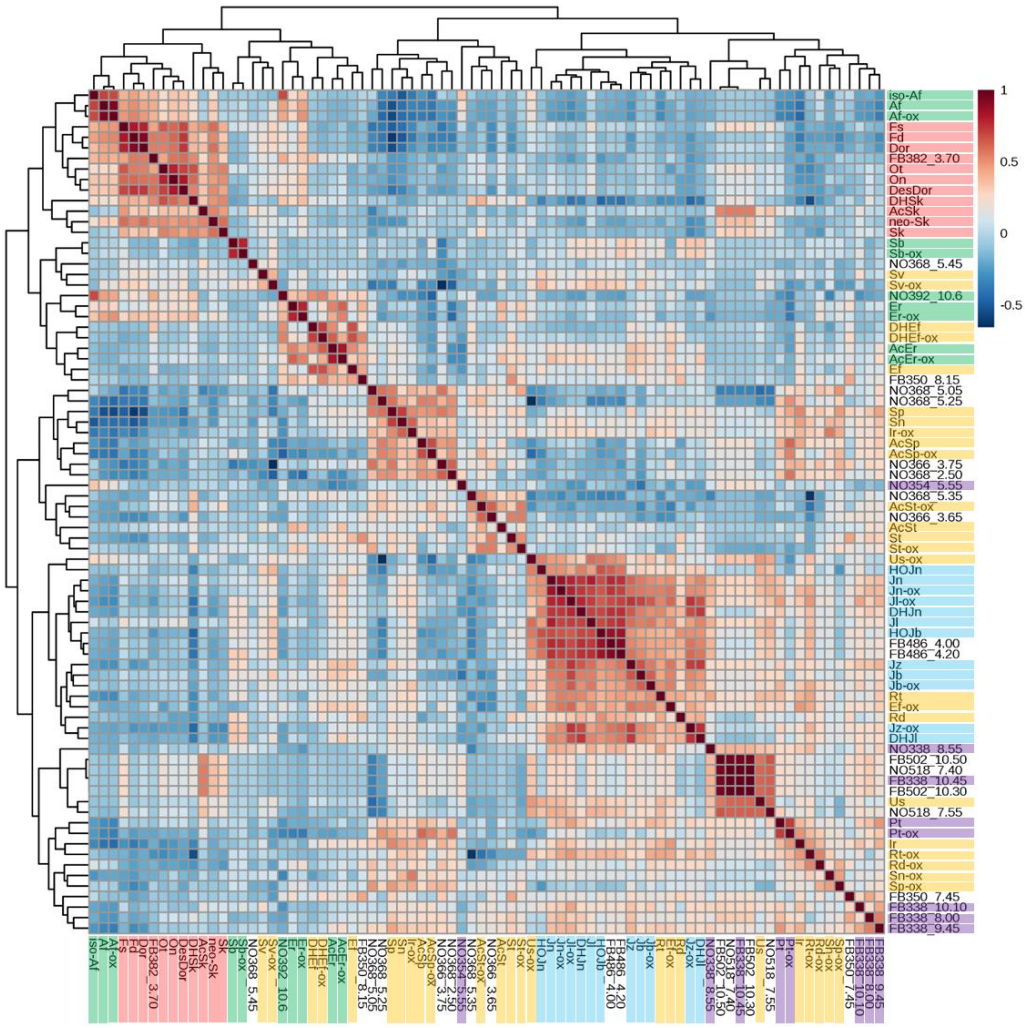

Figure S8. Heatmap representing Spearman rank correlation coefficients between individual PAs based on the presence/absence of PAs of all individual *Jacobaea* plants. The analysis was calculated with Euclidean distances and the Ward clustering algorithm based on Pareto-scaled presence (1)/absence (0) of PAs in the tool MetaboAnalyst. Names of different known PA structural groups are highlighted with different colors: senecionine-like PAs (orange), erucifoline-like PAs (green), jacobine-like PAs (blue), platyphylline-like PAs (purple), otonesine-like PAs (red). PA name abbreviations are shown in Appendix 1.

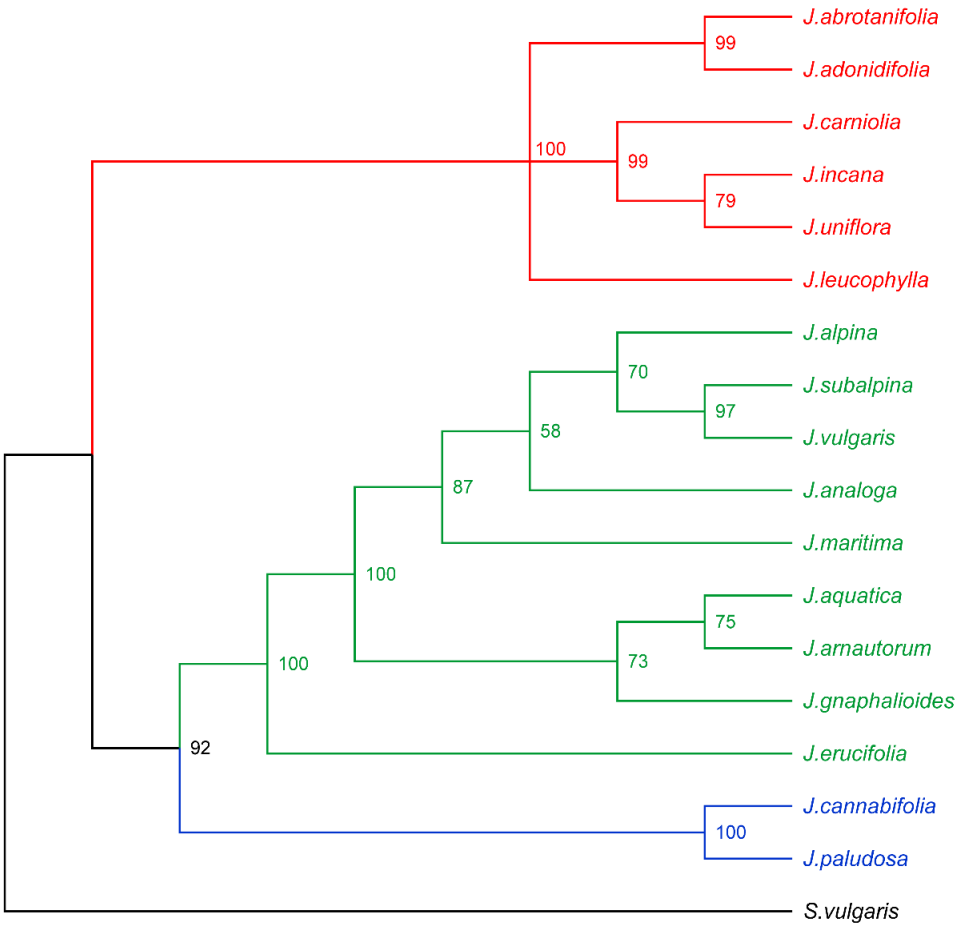

Figure S9. Maximum parsimony (MP) strict consensus cladogram based on the combined plastid and nuclear data of 17 *Jacobaea* species. Bootstrap values are given at each node. Different groups are color coded: *Incani* s.l.-group (red), *J. vulgaris* s.l.-group (green), *J. paludosa*-group (blue), and outgroup *S. vulgaris* (black).

Table S1. Average PA concentrations (μg/g DW) in the leaves of each accession summed over PA groups of 17 *Jacobaea* species.

Clade	Species /Populations	Origins	Individuals	Total PAs		Sum FB		Sum
				Mean	SE	Mean	SE	Mean
Incani s.l. group	*J. abrotanifolia* 1	Austria	5	62.1	7.7	58.4	7.2	3.7
	J. abrotanifolia 2	Switzerland	7	35.5	5.6	35.0	6.2	0.5
	J. abrotanifolia 3	Italy	5	65.1	11.5	64.2	11.5	1.0
	J. adonidifolia 1	Switzerland	10	64.0	12.4	27.9	5.1	36.1
	J. adonidifolia 2	Italy	5	127.1	33.6	37.5	14.5	89.6
	J. adonidifolia 3	France	10	211.3	35.2	132.7	23.4	78.5
	J. carniolia 1	Italy	2	652.0	312.2	252.4	78.3	399.6
	J. carniolia 2	Italy	6	676.8	133.4	48.4	12.2	628.4
	J. carniolia 3	Italy	6	44.8	5.9	19.4	2.0	25.4
	J. incana 1	Alps	6	948.2	402.4	29.9	12.5	918.3
	J. incana 2	Switzerland	8	1153.2	137.6	7.0	1.5	1146.2
	J. incana 3	Switzerland	7	1060.4	157.7	51.9	22.8	1008.5
	J. leucophylla 1	France	4	84.2	50.9	5.4	2.2	78.7
	J. uniflora 1	Alps	6	268.1	75.9	31.4	12.3	236.8
	J. uniflora 2	Switzerland	5	538.0	72.0	202.1	56.0	335.9
J. vulgaris s.l. group	*J. alpina* 1	Alps	9	2129.6	248.5	191.6	15.5	1938.0
	J. alpina 2	France	10	425.4	53.9	28.4	3.7	397.1
	J. alpina 3	Italy	5	2133.8	259.7	1080.2	142.4	1053.6
	J. arnautorum 1	Bulgaria	9	799.9	128.4	36.2	8.3	763.8
	J. subalpina 1	Slovakia	10	1230.8	198.4	52.0	9.2	1178.8
	J. subalpina 2	Austria	9	2125.7	314.1	15.0	1.9	2110.7
	J. subalpina 3	Slovakia	7	672.6	107.3	9.2	1.5	663.4
	J. vulgaris 1	Germany	5	1389.4	118.7	1105.6	103.0	283.8
	J. vulgaris 2	UK	5	4301.3	495.4	2994.9	483.9	1306.4
	J. vulgaris 3	Netherlands	5	3230.4	238.0	2905.0	181.2	325.5
	J. aquatica 1	Germany	10	1129.8	261.2	890.5	276.5	239.3
	J. aquatica 2	UK	10	1412.7	160.3	253.5	35.5	1159.2
	J. analoga 1	India	10	368.1	67.3	172.9	31.6	195.2
	J. gnaphalioides 1	Greece	10	3835.7	306.2	2461.4	126.9	1374.4
	J. maritima 1	Alps	7	516.0	82.1	343.3	57.0	172.7
	J. maritima 2	France	5	1527.0	139.2	1390.8	139.3	136.2
	J. maritima 3	Bulgaria	5	1209.6	95.5	1046.3	120.1	163.2
	J. erucifolia 1	Germany	5	114.7	20.1	49.7	10.5	65.0
	J. erucifolia 2	Netherlands	6	288.8	83.1	32.9	11.7	255.9
	J. erucifolia 3	UK	5	22.3	6.1	5.7	2.2	16.6
J. paludosa group	*J. cannabifolia* 1	Estonia	5	6.2	1.1	4.7	1.1	1.6
	J. cannabifolia 2	Italy	7	9.2	7.2	8.6	7.0	0.6
	J. cannabifolia 3	Russia	4	83.2	21.2	75.9	21.4	7.3
	J. paludosa 1	Poland	5	127.4	36.0	121.6	34.7	5.7
	J. paludosa 2	Germany	4	149.9	26.8	80.2	14.9	69.7

Total = sum of all PAs, FB = free base, NO = *N*-oxide, Sn = senecionine-like, Sp = Seneciphylline-like, Jb = Jacobine-like, Er = erucifoline-like, Pt = platyphylline-like, Ot = otosenine-like and Unk = unknown; SE = standard error.

NO	sum Sn		sum Jb		sum Er		sum Pt		sum Ot		Sum unk	
SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
1.8	1.3	0.4	0.0	0.0	5.5	1.9	0.0	0.0	55.2	7.0	0.0	0.0
0.2	0.4	0.2	0.1	0.1	0.2	0.1	0.0	0.0	34.8	5.7	0.0	0.0
0.2	0.9	0.3	0.1	0.0	1.1	0.3	0.0	0.0	63.1	11.5	0.0	0.0
10.8	1.5	0.4	0.1	0.0	40.8	11.4	0.1	0.0	21.5	5.1	0.1	0.0
24.6	13.1	9.3	0.2	0.1	113.6	29.1	0.1	0.0	0.0	0.0	0.2	0.2
13.2	20.0	8.4	0.3	0.1	65.5	8.0	0.1	0.0	124.4	22.2	1.0	0.4
233.9	518.8	280.3	0.4	0.3	82.0	24.8	12.8	6.9	5.6	1.7	32.4	1.7
133.6	644.0	132.9	0.3	0.1	13.0	3.8	6.9	4.9	1.4	0.8	11.2	2.0
5.9	29.2	5.7	0.0	0.0	5.1	2.4	2.0	1.6	6.8	2.2	1.6	0.4
398.8	906.8	393.9	0.2	0.1	0.0	0.0	12.0	9.4	20.4	12.7	8.9	2.7
136.3	1126.2	136.5	0.1	0.0	0.0	0.0	5.6	2.4	0.4	0.2	20.8	1.7
145.6	991.7	144.6	0.1	0.0	0.0	0.0	7.7	2.2	39.2	22.0	21.8	2.7
50.9	74.4	50.7	0.0	0.0	0.0	0.0	1.2	0.4	2.8	2.2	5.7	1.5
64.5	236.6	64.7	0.0	0.0	0.0	0.0	0.7	0.3	28.6	11.6	2.2	0.4
37.0	332.7	36.9	0.0	0.0	0.0	0.0	2.7	0.8	196.5	56.0	6.1	0.6
235.9	1929.0	236.1	37.4	2.6	0.0	0.0	134.3	10.3	0.0	0.0	29.0	3.9
53.5	365.2	51.2	14.5	2.7	0.0	0.0	29.9	2.8	0.0	0.0	15.8	1.5
190.8	1026.7	187.3	1.5	0.3	1.5	0.4	45.6	5.6	1050.4	142.0	8.1	1.2
121.9	787.4	127.9	0.3	0.1	0.2	0.2	4.7	1.8	0.0	0.0	7.3	0.3
190.7	1220.5	196.9	0.3	0.2	0.0	0.0	2.5	0.4	0.0	0.0	7.4	1.3
312.3	2101.6	312.5	0.1	0.1	0.0	0.0	10.1	5.4	0.0	0.0	13.9	1.8
106.7	663.9	106.5	0.0	0.0	0.0	0.0	1.5	0.5	0.0	0.0	7.2	1.0
62.1	71.5	20.2	1256.0	94.9	33.6	13.7	4.8	1.4	1.7	1.6	21.8	4.1
227.0	195.0	57.0	4053.5	480.7	20.8	6.1	7.7	3.5	0.0	0.0	24.2	5.4
65.3	91.9	23.5	2916.8	188.8	123.0	46.3	4.1	1.1	0.0	0.0	94.7	12.0
47.1	222.1	53.1	13.0	3.4	42.8	10.0	23.5	5.3	826.4	274.7	2.0	0.3
132.6	1124.5	130.0	24.9	2.3	4.3	4.1	89.9	11.1	166.3	31.0	2.8	0.2
61.4	182.2	62.5	154.4	21.3	0.0	0.0	27.2	7.3	0.0	0.0	4.3	0.8
202.4	1246.6	203.8	113.0	30.2	0.0	0.0	1306.9	150.7	1101.8	101.1	67.5	13.6
59.3	80.1	28.1	433.0	70.3	0.0	0.0	0.2	0.0	0.0	0.0	2.8	0.7
42.3	185.8	77.0	1300.4	111.4	0.0	0.0	1.9	0.4	30.0	26.8	8.9	1.7
29.6	29.3	5.2	1087.8	109.3	0.0	0.0	1.1	0.3	85.0	28.2	6.3	1.2
12.6	81.1	14.3	0.5	0.1	27.9	4.7	0.8	0.3	1.8	1.0	2.7	0.5
75.2	219.7	67.7	1.2	0.5	62.5	17.5	1.0	0.5	1.5	0.6	2.9	0.8
4.5	18.9	5.8	0.1	0.0	2.3	0.0	0.0	0.0	0.6	0.3	0.4	0.2
0.3	1.2	0.3	4.2	0.8	0.6	0.4	0.0	0.0	0.1	0.0	0.0	0.0
0.2	3.1	2.7	2.5	2.0	3.3	2.4	0.1	0.1	0.0	0.0	0.1	0.1
0.9	8.9	2.7	47.7	12.8	24.4	8.4	1.7	0.3	0.0	0.0	0.5	0.1
2.2	4.6	2.2	0.1	0.0	0.0	0.0	0.0	0.0	121.4	34.7	1.2	0.3
22.8	51.2	18.7	86.8	13.2	0.0	0.0	8.3	2.9	0.0	0.0	3.6	1.0

Table S2. Primers used for amplification of plastid and nuclear DNA regions in 17 *Jacobaea* species and *S. vulgaris*.

		Forward primer	Reverse primer
plastid regions	psbA-trnH	5'-**TGTAAAACGACGGCCAGT**GTTATGCATGAACGTAATGCTC-3'	5'-**CAGGAAACAGCTATGAC**GCGCATGGTGGATTCACAAATC-3'
	trnK1023	5'-**TGTAAAACGACGGCCAGT**GATTTGGGCCGATTTCTC-3'	5'-**CAGGAAACAGCTATGAC**GCACACGGCTTTCCCTCTG-3'
	trnK39	5'-**TGTAAAACGACGGCCAGT**TGCGGCTAGGATCTTTTACACA-3'	5'-**CAGGAAACAGCTATGAC**TTTTTCAACCCAATCGCTCTTT-3'
	trnT-L	5'-**TGTAAAACGACGGCCAGT**CATTACAAATGCGATGCTCT-3'	5'-**CAGGAAACAGCTATGAC**TCTACCGATTTCGCCATATC-3'
	trnL	5'-**TGTAAAACGACGGCCAGT**CGAAATCGGTAGACGCTACG-3'	5'-**CAGGAAACAGCTATGAC**GGGGATAGAGGGACTTGAAC-3'
	trnL-F	5'-**TGTAAAACGACGGCCAGT**AAAATCGTGAGGGTTCAAGTC-3'	5'-**CAGGAAACAGCTATGAC**GATTTGAACTGGTGACACGAG-3'
	ndhc-atpe1	5'-**TGTAAAACGACGGCCAGT**CGGAGCTAAAAATCCCGAAATGA-3'	5'-**CAGGAAACAGCTATGAC**TGACTCAGAGCACATGGAGC-3'
	ndhc-atpe2	5'-**TGTAAAACGACGGCCAGT**TTGTGCATTGGGCTCTTTCA-3'	5'-**CAGGAAACAGCTATGAC**GCTAGGACACGAGTAGAGGC-3'
	rbcL	5'-**TGTAAAACGACGGCCAGT**ATGTCACCACAAACAGAGACTAAAGC-3'	5'-**CAGGAAACAGCTATGAC**GTAAAATCAAGTCCACCRCG-3'
	rps18	5'-**TGTAAAACGACGGCCAGT**TTGACCTTGAAACAACAACGAT-3'	5'-**CAGGAAACAGCTATGAC**AGAGACAGTTGCTTCTTAATCGTAA-3'
	ndhF	5'-**TGTAAAACGACGGCCAGT**AATAGCTTGGTCTACGGCGGGATT-3'	5'-**CAGGAAACAGCTATGAC**GGAATTCCTAAGAATCCAACGAACAA-3'
nuclear regions	ITS	5'-**TGTAAAACGACGGCCAGT**GGAAGTAAAAGTCGTAACAAGG-3'	5'-**CAGGAAACAGCTATGAC**TCCTCCGCTTATTGATATGC-3'
	ETS	5'-**TGTAAAACGACGGCCAGT**TGACAGATGGCTCTCTTGGA-3'	5'-**CAGGAAACAGCTATGAC**ACTCGGCTCAACTGATTTCTC-3'
	DHS	5'-**TGTAAAACGACGGCCAGT**TGACAGATGGCTCTCTTGGA-3'	5'-**CAGGAAACAGCTATGAC**ACTCGGCTCAACTGATTTCTC-3'

M13 tails used for sequencing are marked in bold.

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,

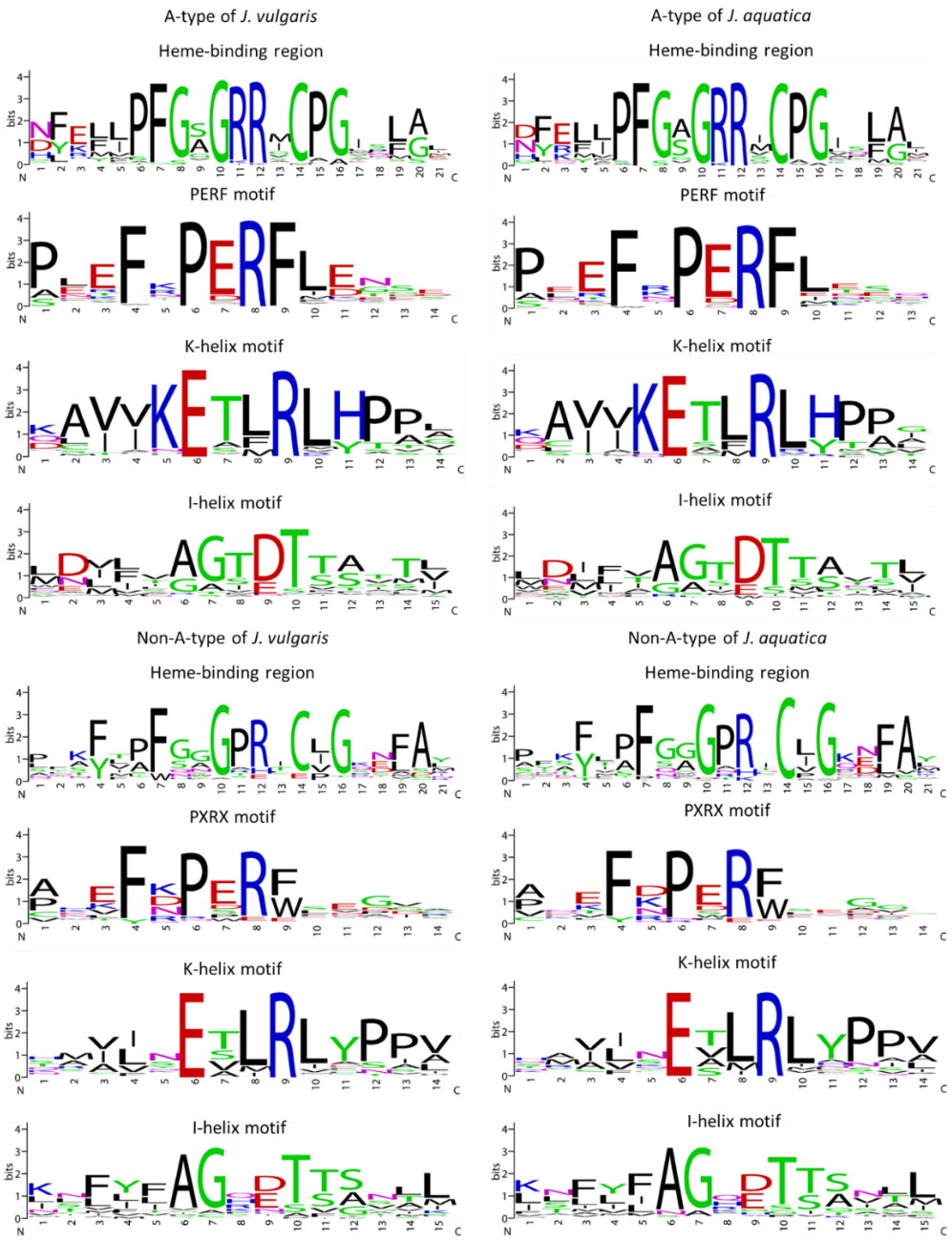

Figure 1. Weblogos of typical conserved motifs identified in the full-length CYP450s divided as A-type (the upper figure) and non-A-type (the lower figure) from *J. vulgaris* (left) and *J. aquatica* (right). The names of the motifs are shown above each logo. The bit score indicates the information content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

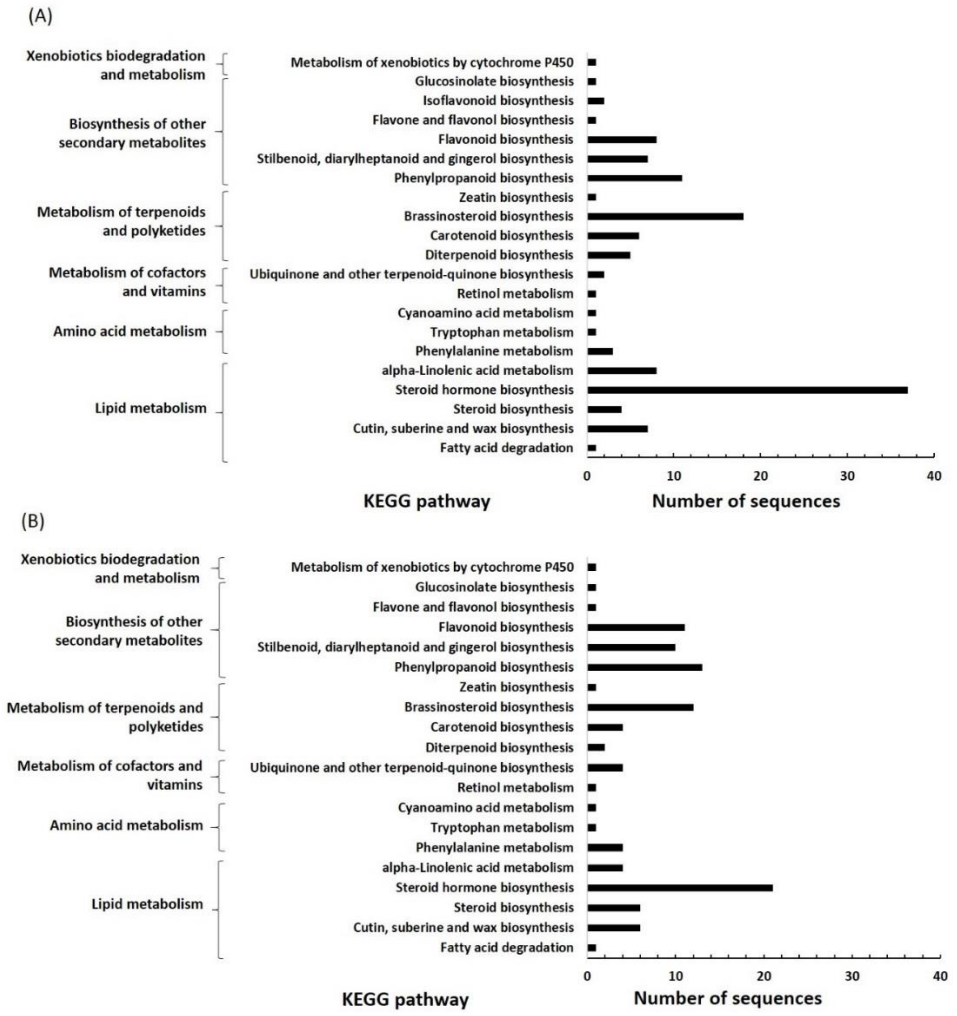

Figure 2. KEGG pathway analysis of predicted CYP450s in two *Jacobaea* species. (A) *J. vulgaris*. (B) *J. aquatica*. The numbers of CYP450 genes involved in the corresponding metabolic processes are shown.

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.*

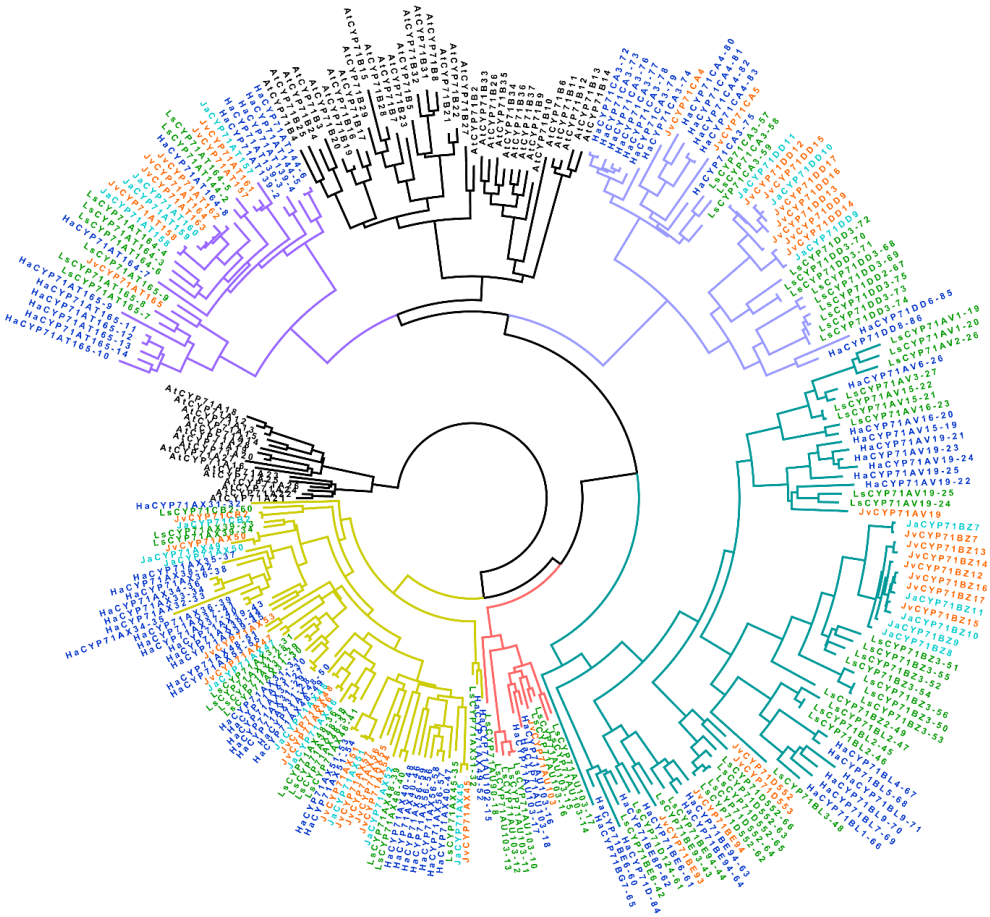

Figure 3. Phylogenetic tree of CYP71 family from five species inferred with the maximum likelihood method. CYP450s are color coded for different species: *J. vulgaris* (orange), *J. aquatica* (light blue), *H. annuus* (dark blue), *L. sativa* (green), *A. thaliana* (black). The branches of the five clades 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.

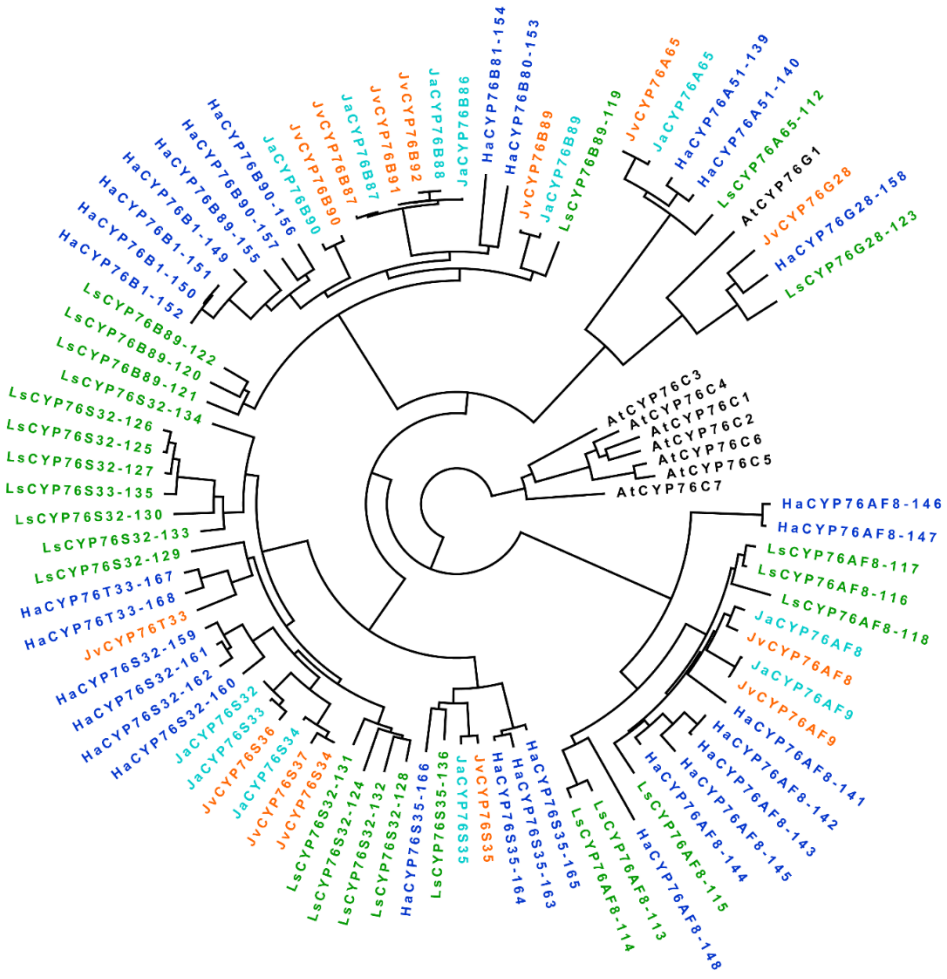

Figure S1. Phylogenetic tree of the CYP76 family from 5 species inferred with the maximum likelihood method. CYP450s are color coded for different species: *J. vulgaris* (orange), *J. aquatica* (light blue), *H. annuus* (dark blue), *L. sativa* (green), *A. thaliana* (black). The names of CYP450s of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

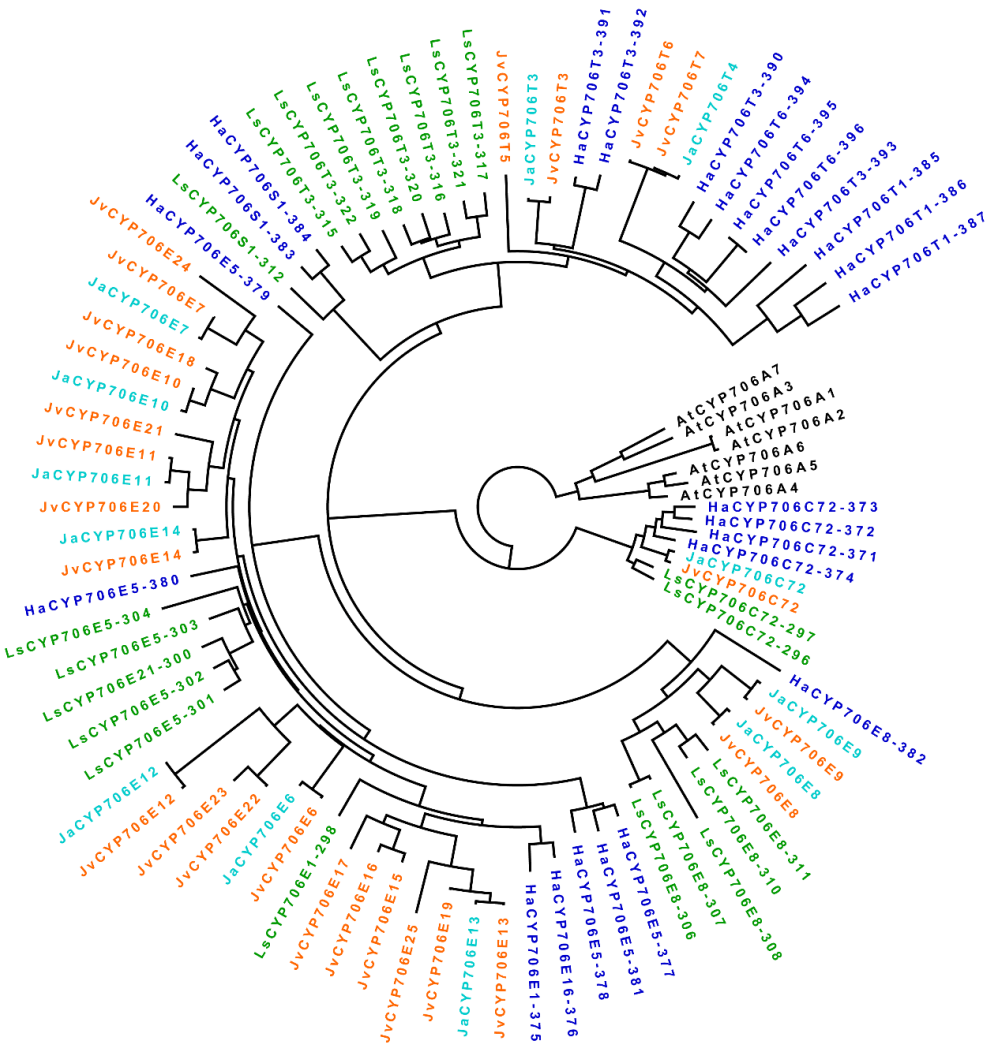

Figure S2. Phylogenetic tree of the CYP706 family from 5 species inferred with the maximum likelihood method. CYP450s are color coded for different species: *J. vulgaris* (orange), *J. aquatica* (light blue), *H. annuus* (dark blue), *L. sativa* (green), *A. thaliana* (black). The names of CYP450s of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

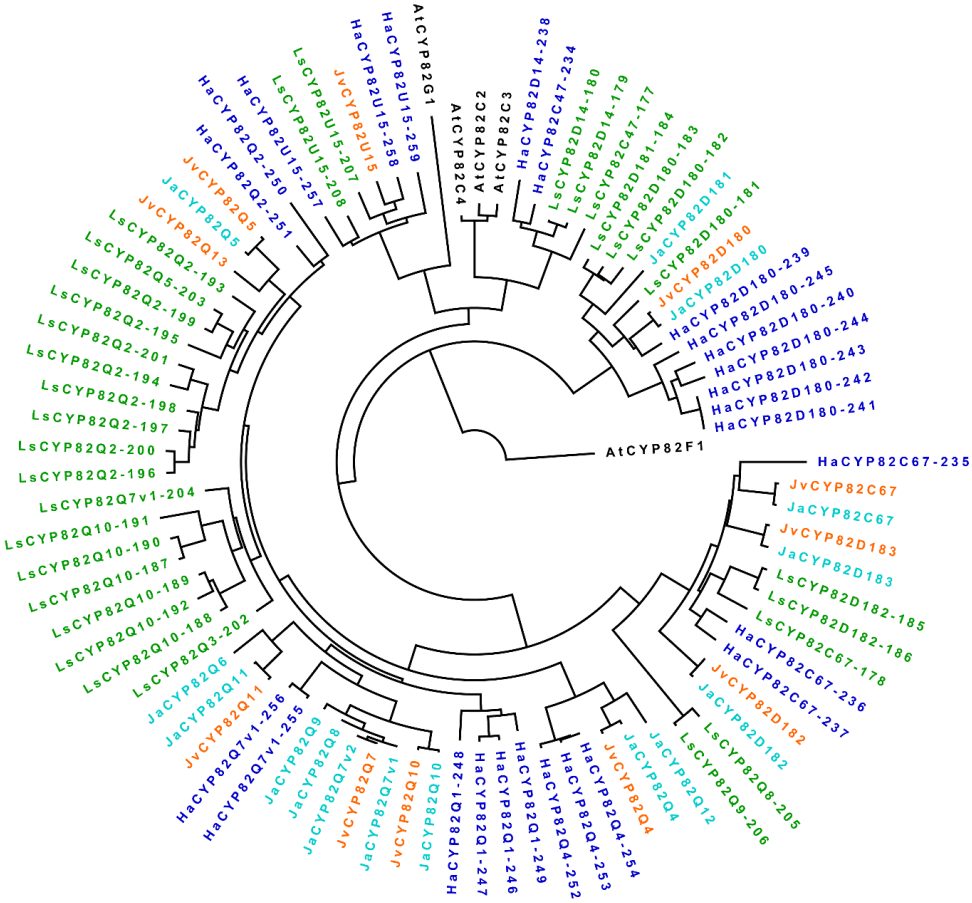

Figure S3. Phylogenetic tree of the CYP82 family from 5 species inferred with the maximum likelihood method. CYP450s are color coded for different species: *J. vulgaris* (orange), *J. aquatica* (light blue), *H. annuus* (dark blue), *L. sativa* (green), *A. thaliana* (black). The names of CYP450s of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

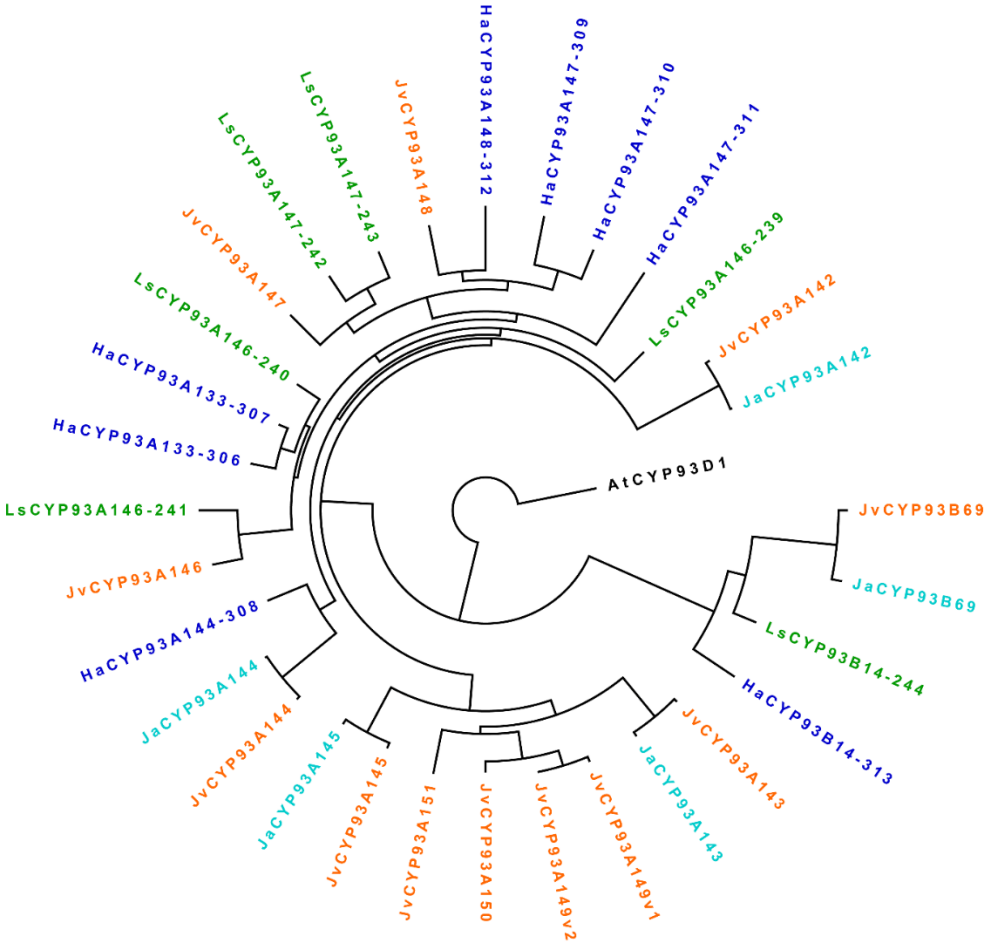

Figure S4. Phylogenetic tree of the CYP93 family from 5 species inferred with the maximum likelihood method. CYP450s are color coded for different species: *J. vulgaris* (orange), *J. aquatica* (light blue), *H. annuus* (dark blue), *L. sativa* (green), *A. thaliana* (black). The names of CYP450s of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

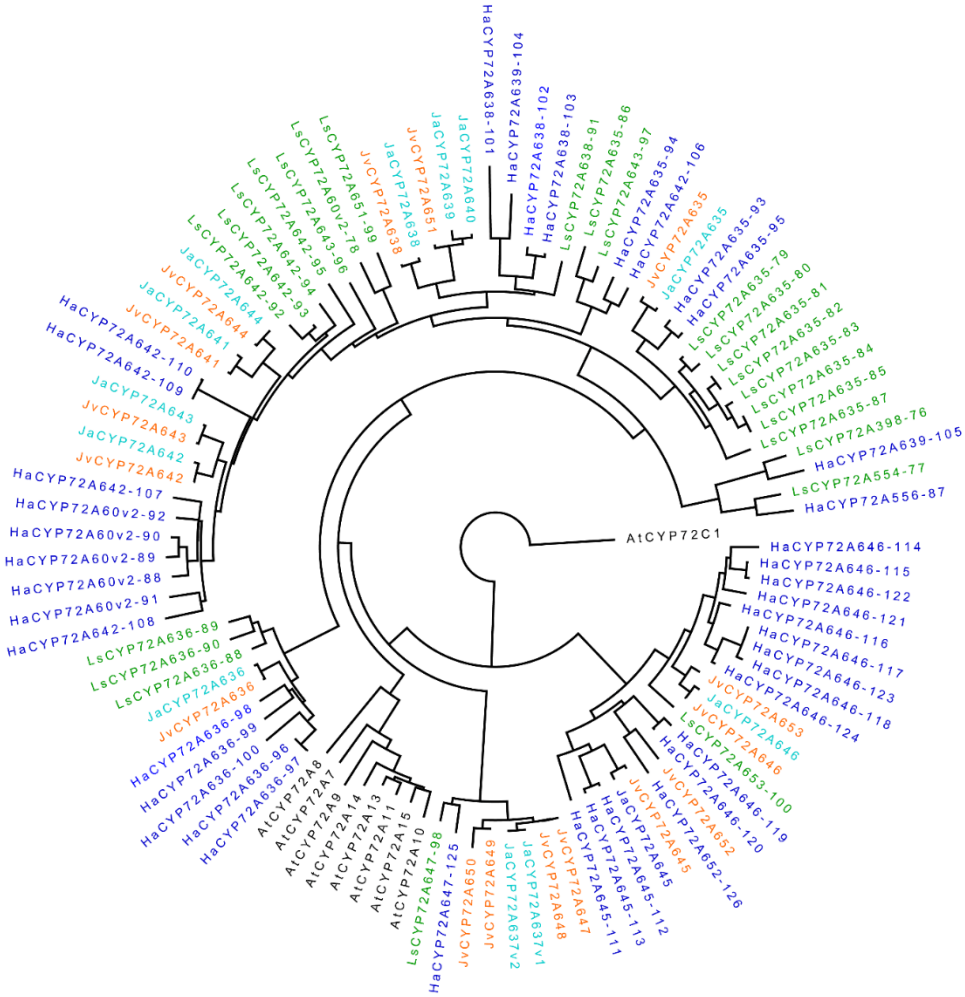

Figure S5. Phylogenetic tree of the CYP72 family from 5 species inferred with the maximum likelihood method. CYP450s are color coded for different species: *J. vulgaris* (orange), *J. aquatica* (light blue), *H. annuus* (dark blue), *L. sativa* (green), *A. thaliana* (black). The names of CYP450s of *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.

Chapter 4

Metabolic and transcriptomic profiling of two *Jacobaea* species and their interspecific hybrids reveals candidate genes involved in the pyrrolizidine alkaloid pathway

Yangan Chen, Patrick P.J. Mulder, Peter G.L. Klinkhamer, Johan Memelink, Klaas Vrieling

Metabolic and transcriptomic profiling of two *Jacobaea* species and their interspecific hybrids reveals candidate genes involved in the pyrrolizidine alkaloid pathway

Abstract

Pyrrolizidine alkaloids (PAs) which are constitutively formed in the plant species containing them provide a powerful defense against herbivores. PAs are a group of secondary metabolites with high diversity. The genes involved in biosynthesis of PAs are not known. To better understand PA metabolism and the origins of its diversity, it is essential to resolve the biosynthesis and the underlying biosynthesis genes. In this study, *Jacobaea* plants were used to study constitutive and induced diversity of the macrocyclic senecionine-type PAs to search for candidate genes for PA biosynthesis. *J. aquatica* and four groups of F_2 hybrids from a cross between *J. vulgaris* and *J. aquatica* with PA contrasts were used to study constitutive PA formation. In addition, a methyl jasmonate treatment on tissue culture plants of *J. vulgaris* was performed to induce changes in PA composition. In total, 44 PAs were detected by LC-MS/MS and PA profiles of different *Jacobaea* samples were compared separately in the constitutive and induced groups by summing up concentrations of PAs containing the same site-specific oxidative modifications, i.e. 15,20-epoxidation, 12,13-epoxidation, 19-hydroxylation, 18-hydroxylation, 13,19-dehydrogenation and 8-oxidation. RNA sequencing was performed separately for the constitutive and induced groups to analyze the expression of cytochrome P450 genes (CYPs) which may be involved in oxidative conversions of PAs. In total, 33 and 27 CYP candidate genes were sieved out for the constitutive and induced PA conversions, respectively. Most of these candidate genes were from the CYP71 clan without known functions. There were 11 CYP subfamilies found both in the constitutive and induced groups, where three subfamilies (CYP72A, CYP706E, CYP82Q) may be responsible for the formation of erucifoline-like PAs which contain both 12,13-epoxidation and 19-hydroxylation.

Keywords

Jacobaea vulgaris, *Jacobaea aquatica*, constitutive defense, induced defense, methyl jasmonate, RNA-seq, cytochrome P450

Introduction

Pyrrolizidine alkaloids (PAs), which are constitutively formed in the plant species containing them provide a powerful defense against herbivores (Hartmann, 1999). PAs are a typical class of SMs with high diversity. More than 400 PAs have been identified from around 6,000 angiosperm species (Chou and Fu, 2006), and are classified into different structural groups. The macrocyclic senecionine type PAs form one of the most diverse group with more than 100 structures found abundantly in the tribe Senecioneae of the family Asteraceae (Langel *et al*., 2011). It is still unclear how PA diversity comes about, and a prerequisite for a better understanding of the origin of this diversity is the elucidation of the structural genes underlying the biosynthetic pathway, as it has been demonstrated that the diversification of PAs are under genetic control (Hartmann and Dierich, 1998; Macel *et al*., 2004).

Most of our current knowledge of PA biosynthetic diversification has come from the studies of the macrocyclic senecionine-type PAs in the Senecioneae (Langel *et al*., 2011). 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 PA biosynthesis (Böttcher *et al*., 1993; Ober and Hartmann, 1999). Senecionine *N*-oxide has been identified as the primary product, which is biosynthesized exclusively in the roots in *Senecio* and *Jacobaea* species (Hartmann and Toppel, 1987; Toppel *et al*., 1987; Hartmann *et al*., 1989; Hartmann and Dierich, 1998). Through the phloem, senecionine *N*-oxide is translocated to the shoots, where the diversification to other PAs occurs (Hartmann *et al*., 1989; Hartmann and Dierich, 1998) mainly via simple one- or two-step reactions including 18-hydroxylation, 13,19-dehydrogenation, 15,20-epoxidation, 12,13-epoxidation, 19-hydroxylation, and a slightly more complex 8-oxidation during the conversion of the retronecine base to the otonecine base (Fig.1; Hartmann and Dierich, 1998; Pelser *et al*., 2005). The enzymes and encoding genes underlying the PA structural diversification in the shoots are not known.

To enable the discovery of the structural genes underlying PA conversions, *Jacobaea* species showing PA contrasts can be used, as different PA patterns are assumed to be the consequences of differential expression of the corresponding genes. A previous study in our lab (Cheng *et al*., 2011) showed that in addition to the PA variation between the parental species *J. vulgaris* and *J. aquatica*, strong PA contrasts in F_2 hybrids exist, indicating that hybridization had a potential role in the evolution of PA diversity in the genus of *Jacobaea*. This hybrid collection can be used for studying the genetic control of constitutive PA formations. In another study (Wei *et al*., 2019) it was found that treatment of tissue culture clones of *J. vulgaris* with methyl jasmonate (MeJA) did not affect total PA concentration but led to a shift in PA composition from senecionine-like PAs to erucifoline-like PAs. Thus, a strategy of correlating metabolic and transcriptomic changes of *J. vulgaris* induced by MeJA is promising for finding structural genes underlying induced PA conversions.

The research described in this study focused on six oxidative modifications (Fig. 1; Fig. S1) of the backbone structure senecionine *N*-oxide, especially on 15,20-epoxidation (mainly found in jacobine-like PAs), 12,13-epoxidaton and 19-hydroxylation (found in erucifoline-like PAs). As these oxidative reactions occur in position-specific and stereoselective manners, cytochrome P450s (CYPs) may be responsible for these reactions (Urlacher and Girhard, 2012; also see Chapter 3). Therefore the CYP gene family was the target of the present study for finding candidate PA biosynthesis genes. *J. vulgaris* (tissue cultures), *J. aquatica* and their F_2 hybrids were grown under controlled conditions aiming at both constitutive and induced contrasts in concentrations of PAs with site-specific modifications. PA abundances were analyzed by LC-MS/MS and gene expression levels by RNA-seq. Different combinations of *J. aquatica* and four F_2 groups with regard to different types of site-specific PA oxidations enabled the identification of CYP candidates underlying constitutive PA diversification. Analysis of tissue culture plants of *J. vulgaris* treated with MeJA allowed the prediction of CYP genes possibly involved in induced PA conversions.

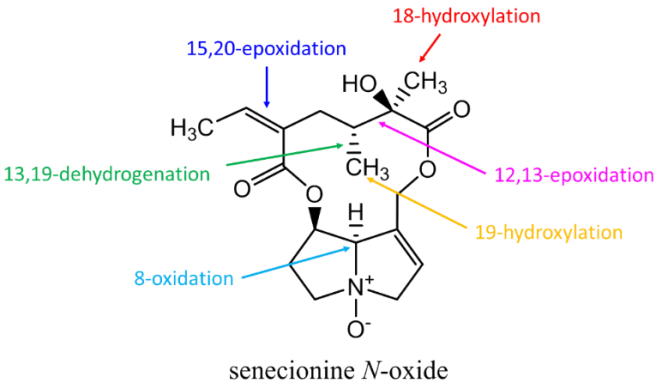

Figure 1. Position-specific and stereoselective oxidative diversifications of senecionine *N*-oxide.

Materials and methods

Plant material

The parental species *J. aquatica* rich in senecionine-like and otosenine-like PAs (Cheng *et al*., 2011) contained higher concentrations of PAs with 13,19-dehydrogenation and 8-oxidation modifications (Fig. S1). Two clones grown from a tissue culture clone (one genotype) of *J. aquatica* originating from a seed collected at the Zwanenwater Reserve (The Netherlands) were combined with five plants (five genotypes) of *J. aquatica* from Avon (England, the UK) constituting the *J. aquatica* sample (Table S1; also see the description for the set *Ja*2 in Chapter 3). The two tissue culture plants were kept in Murashige and Skoog (MS) medium without hormones for two weeks. After rooting cultured plants were transferred to 9×9×10 cm pots

filled with 50% sandy soil (collected from the Meijendel Nature Reserve, The Netherlands), 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). The seeds collected in the UK were germinated on the surface of wet potting soil covered by plastic bags and the seedlings were transferred to pots filled with the soil mixture as indicated above. All plants were kept in a climate room (humidity 70%, light 16 h at 20 °C, dark 8 h at 20 °C) for six weeks. Two to three fully grown leaves were harvested from each individual and were wrapped in aluminum foil and flash frozen in liquid nitrogen. All samples were separately ground into powder with liquid nitrogen, and then identical amounts of powder from each individual sample was pooled for PA analysis and RNA extraction.

Twenty genotypes of F_2 hybrids were selected and grouped into four groups based on previous data (Cheng *et al*., 2011), with the aim to create contrasts in PAs with 15,20-epoxidation, 12,13-epoxidation and 19-hydroxylation (Table S1). The four groups were F_2-1 (high jacobine-like and high erucifoline-like), F_2-2 (high jacobine-like and low erucifoline-like), F_2-3 (low jacobine-like and high erucifoline-like), and F_2-4 (low jacobine-like and low erucifoline-like). A total of eighty five tissue culture individuals from 20 genotypes were kept on MS medium for rooting. After rooting cultured plants were transferred to 9×9×10 cm pots filled with the aforementioned soil mixture. Plants were kept in a climate room for six weeks (humidity 70%, light 16 h at 20 °C, dark 8 h at 20 °C). Two to three fully grown leaves were harvested from each individual and were wrapped in aluminum foil and flash frozen in liquid nitrogen. All the shoots of each plant were ground into powder in liquid nitrogen. Identical amounts of shoot powder from each individual of five genotypes were pooled for each F_2 group for PA analysis and RNA extraction.

In addition, tissue culture plants were used of the other parental species *J. vulgaris* derived from the seed collected from Meijendel for MeJA induction. This setup was mainly to get contrasts in the PAs with 12,13-epoxidation and 19-hydroxylation (found in erucifoline-like PAs; Wei *et al*., 2019). Clonally related issue culture plants of *J. vulgaris* were propagated on MS medium with 0.44 mM benzylaminopurine in a climate room (50% humidity, 16/8 h light/dark cycle at 20 °C). To induce roots 50 cloned plants were transferred to MS medium for two weeks prior to the application of MeJA. One hundred microliters of 4.5 mM MeJA (Sigma-Aldrich) dissolved in 10% ethanol solution were added to the surface of medium, reaching a final concentration of 90 μM after diffusion in each tube, while the same volume of 10% ethanol was added to the control group. The tissue culture plants were harvested at 0, 2, 4, 8 and 16 days after treatment. There were five biological replicates for each treatment at each time point, and these five replicates were pooled after the harvest for their shoots. Pooled samples were ground into powder in liquid nitrogen and stored at -80 °C until RNA extraction, and fractions of each sample were stored at -20 °C until PA extraction.

PA extraction and LC-MS/MS analysis

Ten milligrams of powdered plant material was extracted with 1 mL of 2% formic acid which contained 1 μg/mL of heliotrine as an internal standard. After shaking for 30 minutes on a tumbling machine (Marius Instrumenten, Utrecht, The Netherlands), the extraction mixture was centrifuged at 13,000 rpm for 10 minutes and filtered through a 0.2 μm nylon membrane (Acrodisc 13 mm syringe filter, Pall Life Sciences, Ann Arbor, MI, USA). An aliquot of 25 μL was diluted with 975 μL of 0.05% ammonia and was injected into the LC-MS/MS system.

Analysis of PAs was performed on an LC-MS/MS system consisting of a Waters Acquity UPLC coupled to a Xevo TQ-S tandem mass spectrometer (Waters, Milford, MA, USA), run in positive electrospray mode. Chromatographic separation was achieved on an Acquity BEH C18 analytical column, 150 × 2.1 mm, 1.7 μm particle size (Waters, Milford, MA, USA). Eluent A consisted of water containing 10 mM ammonium carbonate pH 9.0 and acetonitrile was used as eluent B. The gradient elution was performed as follows: 0.0 min 100% A/0% B, 0.1 min 95% A/5% B, 3.0 min 90% A/10% B, 7.0 min 76% A/24% B, 9.0 min 70% A/30% B, 12.0 min 30% A/70% B, and 12.1-15.0 min 100% A/0% B. The column was kept at 50ºC and a flow rate of 400 μL/min was applied. Two μL of the sample extracts was injected.

For each analyte at least two selected precursors to product ion MRM transitions were measured. Cone energy was 40V and collision energy settings were optimized for the individual compounds. Quantification was performed against a range of mixed standard solutions (0-5-10-25-50-100-200 μg/L) of the PAs in a diluted extract of *Tanacetum vulgare* (tansy). The extract of *T. vulgare* material was used to mimic a PA-free plant extract. The range of mixed standard solutions was injected at the beginning of the series and at the end. The mixed standard solution of 50 μg/L in *T. vulgare* extract was injected every 30-40 samples, to monitor the performance of the system (drift in retention times, changes in detector sensitivity) during the analysis. For each PA the averaged response of two MRM transitions was used for quantification. Data processing was conducted with MassLynx 4.1 software (Waters Corporation, Milford, MA, USA).

RNA isolation and transcriptome sequencing

Total RNA was extracted with the NucleoSpin® RNA Plant-Macherey-Nagel kit for seven samples including *J. aquatica* (*Ja*), four F_2 samples and the MeJA-treated tissue-cultured plants of *J. vulgaris* (*Jv*-MeJA) and its control (*Jv*-Control) (except for the four F2 samples, the samples were already described in Chapter 3). RNA integrity 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. 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 strands were then generated after dNTP addition. dUTP instead of dTTP was used to tag the second strand. 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 (New England Biolabs M5505L) to digest the second strand-derived fragments. Yields of pre-amplified libraries were quantified on an Agilent high sensitivity chip. 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 of transcriptomes**

After removal of adapter sequences, the qualities of raw reads obtained from Illumina platforms were checked using FastQC and the bases with low quality (threshold < 30) were removed using trimmomatic via Galaxy (https://usegalaxy.org/) for assembly. The Trinity (Haas *et al*., 2013) program was applied to generate a single assembly for *J. aquatica* and F_2 hybrids (hereafter mentioned as the constitutive transcriptome) based on combining all reads across the five samples (*Ja* and four F_2 samples), and a single assembly for tissue cultures of *J. vulgaris* (hereafter mentioned as the induced transcriptome) by using all reads of both the MeJA-treated sample and its control (also see Chapter 3 for the assembly of the set *Jv*2). The parameters k-mer and minimum assembled contig length were set to 32 and 300 bp, respectively. The quality of each assembly was assessed by aligning reads back to the respective transcriptome by Bowtie2 (Langmead and Salzberg, 2012). After quality assessment, the two transcriptomes were reduced for their redundancies by using the CD-HIT-EST algorithm (version 4.6.8) (Li *et al*., 2006; Fu *et al*., 2012) with nucleotide sequence identity of 100% as the cutoff.

Annotation of transcriptomes and mining of CYPs

The likely coding regions of transcripts in the transcriptomes were predicted by TransDecoder (Version 5.5.0; https://github.com/TransDecoder/TransDecoder/wiki). The putative peptide sequences were blasted against the UniProtKB/Swiss-Prot database for annotation, and were searched for protein domains against the Pfam database by the hidden Markov model (HMM) imbedded in the HMMER program (Version 3.2.1b2; http://hmmer.org). CYP-like contigs were sieved out by hits to the CYP reference database (PF00067) of the Pfam database. These CYP-like contigs were blasted against the CYP database of *J. vulgaris* and *J. aquatica* in house (Chapter 3) for classification.

Differential gene expression and identifying CYP candidates

The program kallisto that relies on pseudoalignments (Bray *et al*., 2016) imbedded in Trinity was used to quantify expression abundance of each transcript for each sample using the respective reference transcriptomes. Two matrices of raw read counts at "gene" level including multiple isoforms were generated separately for the constitutive and induced groups (Table 1). EdgeR (Robinson *et al*., 2010) was subsequently used to identify differentially expressed genes (DEGs). The five samples of the constitutive group, i.e. *Ja*, F_2-1, F_2-2, F_2-3 and F_2-4, were compared in different combinations with regard to different site-specific oxidations.

Specifically, for DEGs related to 15,20-epoxidation, F_2-1, F_2-2 were compared to *Ja*, F_2-3 and F_2-4. For DEGs related to 12,13-epoxidation and 19-hydroxylation, F_2-1, F_2-3 were compared to *Ja*, F_2-2 and F_2-4. For DEGs related to 13,19-dehydrogenation, *Ja* was compared to four F_2 samples, where the dispersion parameter was set to 0.4 since there was only the *Ja* sample (no biological replicate) containing a higher amount of PAs with 13,19-dehydrogenation available. Similarly, only the *Ja* sample containing a higher amount of PAs with 8-oxidation was compared to F_2-1, F_2-3 and F_2-4 setting the dispersion to 0.4 for the responsible DEGs. DEGs of CYP genes were sieved out by *P* value ≤ 0.05 and $\log_2$fold-change (logFC) ≥ 1 at the same time for all site-specific modifications except for 12,13-epoxidation/19-hydroxylation (*P* value ≤ 0.1; $\log_2$fold-change ≥ 1) and 18-hydrogenation (logFC ≥ 1) of the constitutive comparison. In the comparison of induced DEGs between *Jv*-MeJA and *Jv*-Control the dispersion was set to 0.1. CYP candidates were selected for 12,13-epoxidation, 19-hydroxylation and 18-hydroxylation (*P* value ≤ 0.05; logFC ≥ 1).

The expression value of each gene was reported as transcripts per million transcripts (TPM), and the TPM values were cross-sample normalized by the trimmed mean of M-values (TMM) method. Respectively, based on the TMM-normalized expression values of CYP candidates, clustered heatmaps by Euclidean distances of expression abundances after Z-score transformation and log-transformation were drawn for the constitutive and induced groups using the package "pheatmap" (version 1.0.12; https://CRAN.R-project.org/package=pheatmap) in R version 3.5.1.

Results

PA profiles of contrasting groups

In total, 44 PAs were detected and were grouped according to chemical structures (Fig. S1; Table S2). PA profiles were compared based on the summed concentrations of PAs having the same site-specific oxidative modifications.

The constitutive PA profiles were compared across five samples (i.e. *Ja*, F_2-1, F_2-2, F_2-3 and F_2-4) with regard to site-specific PA modifications (Fig. 2). F_2-1 and F_2-2 contained similar amounts of PAs having 15,20-epoxidation (including hydrolysis and chlorolysis of 15,20-epoxide; Fig. S1) which were remarkably higher than amounts in the other three samples (Fig. 2A). The abundance patterns of 12,13-epoxidated and 19-hydroxylated PAs were similarly distributed over the five samples (*Ja*, F_2-1, F_2-2, F_2-3 and F_2-4) as both modifications are found in erucifoline-like PAs (Fig. S1), where F_2-1 and F_2-3 had higher concentrations of 12,13-epoxidated and 19-hydroxylated PAs than F_2-2, F_2-4 and *Ja* (Fig. 2B; Fig. 2C). The total concentrations of 18-hydroxylated PAs were in general low in all five samples, but they were relatively higher in F_2-2 and F_2-4 (Fig. 2D). *Ja* had the highest total concentration of 13,19-dehydrogenated PAs, and the four F_2 groups had similar lower concentrations (Fig. 2E). The conversion from retronecine to otonecine introduces an oxygen at the C8 position for all

otosenine-like PAs. This position-specific modification had a similar abundance pattern as the 13,19-dehydroxylation modification except that F_2-2 had a lower concentration of 8-oxidated PAs compared to the rest three F_2 samples (Fig. 2F).

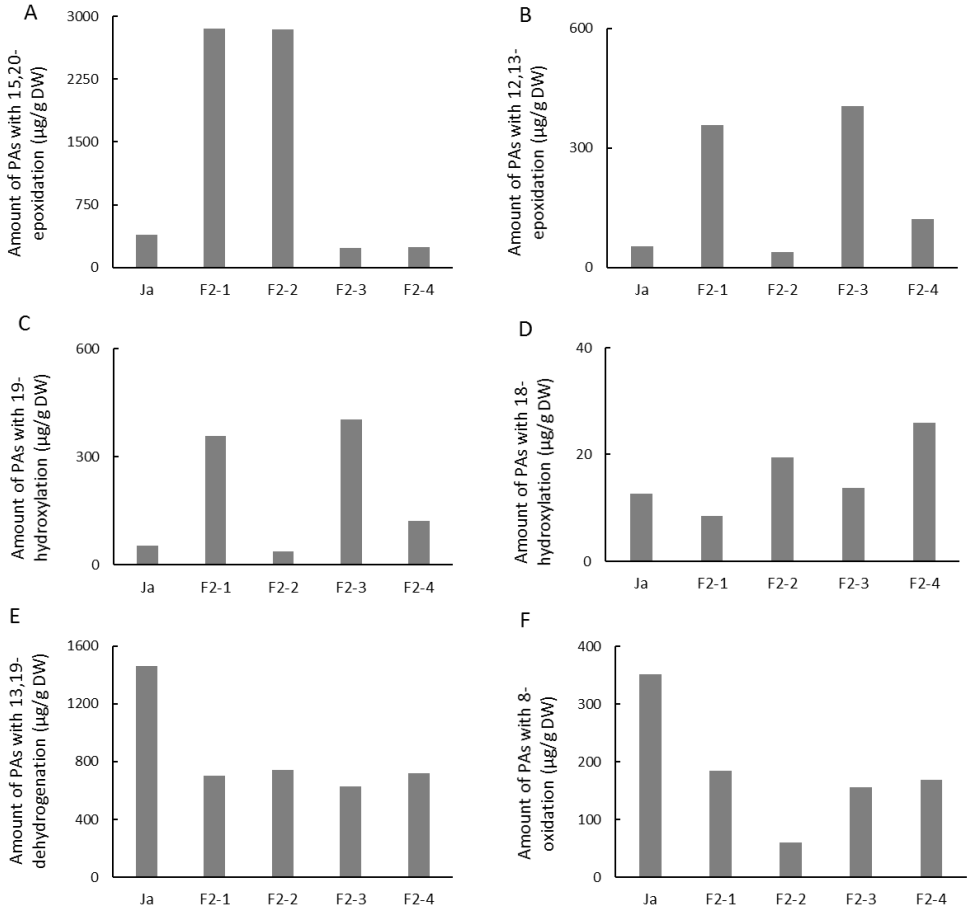

Figure 2. The total concentrations of PAs with the same site-specific oxidative modification derived from the central PA senecionine *N*-oxide in the shoots of *J. aquatica* (*Ja*) and four F_2 groups of a cross between *J. vulgaris* and *J. aquatica*. (A) 15,20-epoxidation (including hydrolysis and chlorolysis of 15,20-epoxide); (B) 12,13-epoxidation (including hydrolysis and chlorolysis of 12,13-epoxide); (C) 19-hydroxylation (including acetylation of 19-hydroxyl); (D) 18-hydroxylation; (E) 13,19-dehydrogenation; (F) 8-oxidation. DW: dry weight.

Tissue culture plants of *J. vulgaris* were treated with MeJA to mimic the effects of herbivore attacks. PA profiles of a time series of tissue culture plants of *J. vulgaris* treated with MeJA or

control solution are shown in Figure 3. Differences between control and induced plants were most pronounced at day 8 after MeJA treatment and therefore this time point was chosen for the gene expression analysis. The concentrations of PAs with 12,13-epoxidation, 19-hydroxylation and 18-hydroxylation showed higher total concentrations in the shoots of MeJA treated tissue culture plants compared to controls at all time points (Fig. 3B-D), although a decrease of PAs with PAs with 18-hydroxylation was found over time (Fig. 3D). The changing patterns of other PA groups (13,19-dehydrogenated, 15,20-epoxidated and 8-oxidated) did not show clear trends (Fig. 3A; Fig. 3E-F).

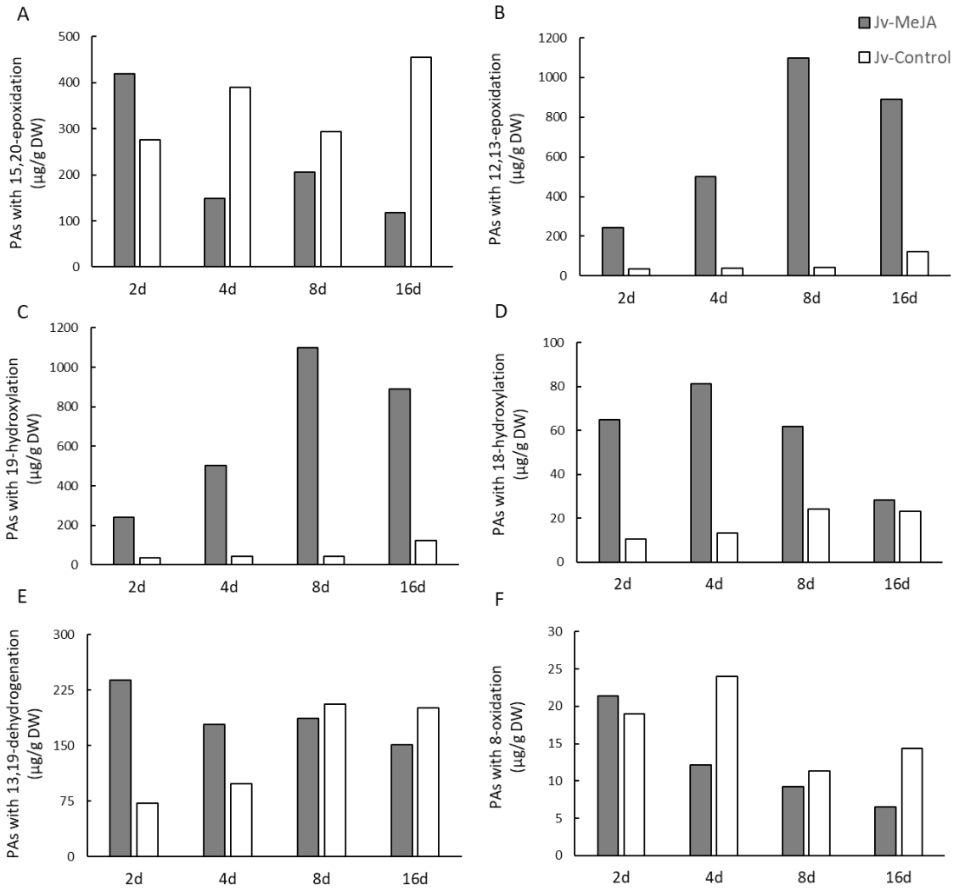

Figure 3. Effects of 90 µM MeJA on PA concentrations in the shoots of tissue culture clones of *J. vulgaris* at different days (d) after treatement. (A) Total amount of PAs with 15,20-epoxidation; (B) total amount of PAs with 12,13-epoxidation; (C) total amount of PAs with 19-hydroxylation; (D) total amount of PAs with or 18-hydroxylation; (E) total amount of PAs with 13,19-dehydrogenation; (F) total amount of PAs with 8-oxidation. DW: dry weight.

De novo assembly of transcriptomes and mining of CYPs

We included seven samples (*Ja*, F_2-1, F_2-2, F_2-3, F_2-4, *Jv*-MeJA, *Jv*-Control) for Illumina sequencing. By combining quality-trimmed reads from different samples, two *de novo* assembled transcriptomes (Table 1) were available as reference transcriptomes for differential gene expression analysis. Both transcriptomes had read mapping rates well above the required value of 70-80% for a good assembly. The constitutive transcriptome contained 393,217 contigs (transcripts) which were assigned to 120,262 genes after redundancy check with 100% identity cutoff. From the constitutive transcriptome, 362 genes corresponding to 1,372 contigs were annotated as CYP-like genes based on Pfam database homologs. The coding regions of the 1,372 contigs were predicted by TransDecoder, resulting to 1,165 coding sequences after removing redundant sequences at the cutoff of 100% identity. Of the 1,165 coding sequences, 277 (24%) had a length of more than 1400 nucleotides (nt) and were regarded as full-length sequences since the lengths of full-length CYPs of *Jacobaea* ranged from 1383 to 1803 nt (Chapter 3). The induced transcriptome had 48,336 genes corresponding to 142,087 contigs after redundancy check at 100% identity cutoff. In the induced transcriptome, 265 CYP-like genes corresponding to 561 CYP-like contigs were found. In total, 502 CYP-like coding sequences without redundancy were found, of which 195 (35%) were longer than 1400 nt and thus considered to be full-length.

Table 1. Overview of relevant numbers associated with *de novo* assemblies of reference transcriptomes and mining of CYPs.

Transcriptomes	Constitutive transcriptome	Induced transcriptome
Sources of reads	*Ja*, F_2-1, F_2-2, F_2-3, F_2-4	*Jv*-MeJA, *Jv*-Control
GC(%)	39.08	39.31
No. of genes	120,262	48,336
No. of transcripts	393,217	142,087
Reads mapped (%)[a]	79.05	83.25
No. of CYP-like genes	362	265
No. of CYP-like contigs	1,372	561
No. of CYP-like cds[b]	1,165	502
No. of CYP-like cds ≥ 1400 nt[c]	277	195

[a]Reads mapped back to the respective transcriptomes concordantly.

[b]The number of coding sequences were reduced for their redundancy at 100% identity cutoff using the CD-HIT-EST algorithm.

[c]nt: nucleotide.

Identification of CYP candidates involved in PA biosynthesis

Differential gene expression analysis was carried out separately for the constitutive (*Ja*, F_2-1, F_2-2, F_2-3 and F_2-4) and induced (*Jv*-MeJA and *Jv*-Control) groups. Candidates were selected based on *P* values, fold changes of expression abundances in combination with elimination of CYPs with known functions such as allene oxide synthase involved in jasmonic acid biosynthesis. Within the constitutive group, different PA contrasts based on different combinations of the five samples enabled the identification of CYP candidates for each site-specific oxidation (Table 2). For 15,20-epoxidation, 13,19-dehydrogenation and 8-oxidation, nine (belonging to subfamilies CYP71CA, CYP701A, CYP90D, CYP706E, CYP82Q, CYP72A, CYP71D, CYP76S and CYP71DD), eight (belonging to subfamilies CYP706E, CYP71AX, CYP706E, CYP706C, CYP71AT, CYP716D, CYP76B and CYP81B) and nine (belonging to subfamilies CYP706E, CYP71AX, CYP706E, CYP716D, CYP706C72, CYP75B, CYP76B, CYP72A and CYP81B) CYP candidate genes were sieved out that satisfied *P* value ≤ 0.05 and logFC ≥ 1 cutoffs, respectively. For 12,13-epoxidation and 19-hydroxylation, eight CYP genes (belonging to subfamilies CYP94D, CYP87A, CYP87A, CYP82Q, CYP707A, CYP706E, CYP706E, CYP72A) were selected as candidates (*P* value ≤ 0.10; logFC ≥ 1 cutoffs). For 18-hydroxylation, nine CYP genes (belonging to subfamilies CYP84A, CYP76G, CYP81B, CYP704A, CYP96A, CYP82D, CYP90D, CYP71AX and CYP71CA) with logFC ≥ 1 cutoff were chosen as candidates although none of these CYPs showed *P* value ≤ 0.05. Ten CYP genes were candidates for two different site-specific oxidations. In total, 33 CYP candidates from clans 71, 72, 85 and 86 were identified for PA conversions, of which 17 CYP genes had putative full-length coding regions (cutoff ≥ 1400 nt). By comparing *Jv*-MeJA to *Jv*-Control, 27 CYPs were sieved out for the combined 12,13-epoxidation, 19-hydroxylation and 18-hydroxylation conversions (*P* value ≤ 0.05; logFC ≥ 1 cutoffs). All candidates belong to the 71 clan except CYP749A and CYP72A that belong to the 72 clan. Only one candidate (belonging to CYP81BG) did not have isoforms with assembled DNA sequences longer than 1400 nt.

In total, 55 CYP genes were selected as candidates for encoding enzymes performing PA conversions as five genes (CYP71CA5, CYP706E12, CYP82D180, CYP706C72, CYP71AT158) were included in both constitutive and induced groups (Table 2). Candidate CYPs for both constitutive and induced PA diversification belong to 11 CYP subfamilies (CYP71AX, CYP71AT, CYP71CA, CYP72A, CYP76B, CYP76S, CYP706C, CYP706E, CYP81B, CYP82D, CYP82Q). Candidate CYPs for both constitutive and induced 12,13-epoxidation and 19-hydroxylation belong to three CYP subfamilies (CYP72A, CYP706E, CYP82Q) (Table 2) and may be of interest for testing their involvement in the formation of erucifoline-like PAs.

Heatmaps were generated to visualize expression patterns of CYP candidate genes based on TMM-normalized expression abundances (Fig. 4). Within the constitutive group, the expression patterns of CYP candidate genes for 15,20-epoxidation (mainly in Cluster 1) were clustered to those for 12,13-epoxidation and 19-hydroxylation (mainly in Cluster 2), followed by those for 18-hydroxylation (mainly in Cluster 3), 13,19-hydroxylation and 8-oxidation

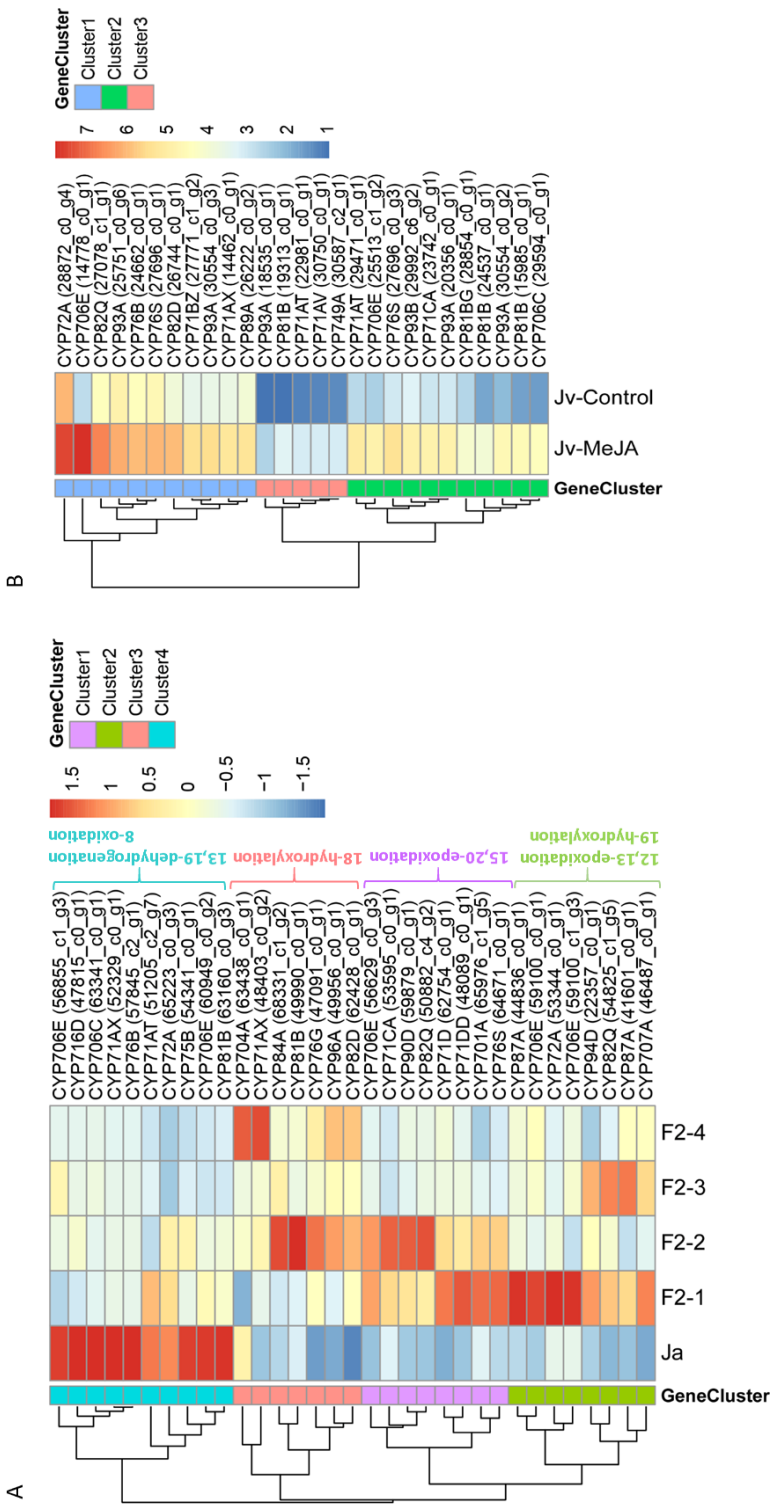

Figure 4. Expression patterns of CYP candidate genes involved in PA biosynthesis in shoots of *Jacobaea* plants. (A) The Z-score transformed heatmap of TMM-normalized expression estimates for 33 CYP candidate genes between *J. aquatica* (*Ja*) and four F_2 samples of a cross between *J. vulgaris* and *J. aquatica*. (B) The log-transformed heatmap of TMM-normalized expression estimates for 27 CYP candidate genes between tissue culture plants of *J. vulgaris* treated with 90 µM MeJA (*Jv*-MeJA) and its control (*Jv*-Control). The hierarchical clustering of CYP genes was conducted based on Euclidean distances between TMM-normalized gene expression value of each CYP gene. Gene names are represented by CYP subfamilies together with codes generated in Trinity (shown in brackets).

Table 2. CYP candidate genes selected for constitutive and induced PA conversions in *Jacobaea* plants.

	Sample contrasts	Site-specific oxidations	Code of CYP candidate gene	logFC of CYP
Constitutive PA conversions	F_2-1, F_2-2 vs *Ja*, F_2-3, F_2-4	15,20-epoxidation (3.3) [a]	JaF2_TRINITY_DN53595_c0_g1[1]	3.7
			JaF2_TRINITY_DN65976_c1_g5	2.3
			JaF2_TRINITY_DN59879_c0_g1[2]	2.2
			JaF2_TRINITY_DN56629_c0_g3	2.0
			JaF2_TRINITY_DN50882_c4_g2	1.4
			JaF2_TRINITY_DN53344_c0_g1[3]	1.4
			JaF2_TRINITY_DN62754_c0_g1	1.4
			JaF2_TRINITY_DN64671_c0_g1	1.3
			JaF2_TRINITY_DN48089_c0_g1	1.0
	F_2-1, F_2-3 vs *Ja*, F_2-2, F_2-4	12,13-epoxidation (2.4) 19-hydroxylation (2.4)	JaF2_TRINITY_DN22357_c0_g1	2.4
			JaF2_TRINITY_DN41601_c0_g1	2.1
			JaF2_TRINITY_DN44836_c0_g1	1.9
			JaF2_TRINITY_DN54825_c1_g5	1.7
			JaF2_TRINITY_DN46487_c0_g1	1.4
			JaF2_TRINITY_DN59100_c0_g1	1.3
			JaF2_TRINITY_DN59100_c1_g3	1.2
			JaF2_TRINITY_DN53344_c0_g1[3]	1.1
	F_2-2, F_2-4 vs *Ja*, F_2-1, F_2-3	18-hydroxylation (2.0)	JaF2_TRINITY_DN68331_c1_g2	1.8
			JaF2_TRINITY_DN47091_c0_g1	1.3
			JaF2_TRINITY_DN49990_c0_g1	1.3
			JaF2_TRINITY_DN63438_c0_g1	1.2
			JaF2_TRINITY_DN49956_c0_g1	1.2
			JaF2_TRINITY_DN62428_c0_g1	1.2
			JaF2_TRINITY_DN59879_c0_g1[2]	1.1
			JaF2_TRINITY_DN48403_c0_g2	1.1
			JaF2_TRINITY_DN53595_c0_g1[1]	1.0
	Ja vs F_2-1, F_2-2, F_2-3, F_2-4	13,19-dehydrogenation (2.1)	JaF2_TRINITY_DN60949_c0_g2[4]	2.8
			JaF2_TRINITY_DN52329_c0_g1[5]	2.7
			JaF2_TRINITY_DN56855_c1_g3[6]	2.3
			JaF2_TRINITY_DN63341_c0_g1[7]	2.2
			JaF2_TRINITY_DN51205_c2_g7	2.2
			JaF2_TRINITY_DN47815_c0_g1[8]	2.1
			JaF2_TRINITY_DN57845_c2_g1[9]	2.0
			JaF2_TRINITY_DN63160_c0_g3[10]	1.8
	Ja vs F_2-1, F_2-3, F_2-4	8-oxidation (2.1)	JaF2_TRINITY_DN60949_c0_g2[4]	2.9
			JaF2_TRINITY_DN52329_c0_g1[5]	2.7
			JaF2_TRINITY_DN56855_c1_g3[6]	2.4
			JaF2_TRINITY_DN47815_c0_g1[8]	2.3
			JaF2_TRINITY_DN63341_c0_g1[7]	2.2

P value	Best hit	No. of isoforms	Identity range of pep (%)	Length range of cds (nt)	CYP
0.003	CYP71CA5 [I]	6	95.9-98.9	726-903	
0.002	CYP701A73	1	97.2	429	
0.039	CYP90D53	2	98.3-98.4	1470-1554	
0.029	CYP706E11	3	98.3-98.7	939-1620	
0.017	CYP82Q13	9	91.5-97.0	360-1560	
0.020	CYP72A636	10	83.7-99.5	342-1659	
0.037	CYP71D552/553	14	94.0-99.5	315-1503	
0.006	CYP76S34/37	9	96.1-100	558-1512	
0.044	CYP71DD9/10/11/15/16/17	13	93.8-100	321-684	
0.086	CYP94D110	2	98.9-100	387-666	
0.086	CYP87A3 [b]	1	76.6	387	
0.060	CYP87A3 [b]	6	72.8	333-1524	
0.077	CYP82Q7	1	96.1	318	
0.098	CYP707A182	1	87.8	279	
0.053	CYP706E7	1	99.2	369	
0.042	CYP706E12 [II]	3	96.8-99.8	948-1551	
0.078	CYP72A636	10	83.7-99.5	342-1659	
0.189	CYP84A117	1	99.00	315	
0.294	CYP76G28	4	98.4-99.0	627-1557	
0.211	CYP81B113	1	95.70	720-1521	
0.322	CYP704A179	1	92.00	315	
0.225	CYP96A156	9	90.9-100	510-1563	
0.302	CYP82D180 [III]	4	97.9-99.6	597-1602	
0.325	CYP90D53	2	96.6-97.9	1470-1554	
0.295	CYP71AX46/53/54	14	62.5-71.1	363-492	
0.462	CYP71CA5 [I]	6	95.9-98.9	726-903	
0.001	CYP706E8	1	96.1	306	
0.003	CYP71AX46	5	91.8-98.6	540-837	
0.011	CYP706E14	1	100.0	384	
0.012	CYP706C72 [IV]	8	93.2-99.8	438-1593	
0.026	CYP71AT158 [V]	1	99.6	819	
0.016	CYP716D63	8	92.4-100	309-1458	
0.029	CYP76B89/90	5	69.5-79.5	483-1470	
0.038	CYP81B108	4	96.3-99.8	813-1506	
0.002	CYP706E8	1	96.1	306	
0.005	CYP71AX46	5	91.8-98.6	540-837	
0.017	CYP706E14	1	100	384	
0.014	CYP716D63	8	92.4-100	309-1458	
0.022	CYP706C72 [IV]	8	93.2-99.8	438-1593	

Table 2 continued

	Sample contrasts	Site-specific oxidations	Code of CYP candidate gene	logFC of CYP
Constitutive PA conversions	*Ja* vs F_2-1, F_2-3, F_2-4		JaF2_TRINITY_DN54341_c0_g1	2.1
			JaF2_TRINITY_DN57845_c2_g1[9]	2.0
			JaF2_TRINITY_DN65223_c0_g3	2.0
			JaF2_TRINITY_DN63160_c0_g3[10]	1.9
Induced PA conversions	*Jv*-MeJA vs *Jv*-Control	12,13-epoxidation (4.7) 19-hydroxylation (4.7) 18-hydroxylation (1.3)	MeJA_TRINITY_DN14778_c0_g1	5.23
			MeJA_TRINITY_DN15985_c0_g1	3.27
			MeJA_TRINITY_DN29594_c0_g1	3.18
			MeJA_TRINITY_DN19313_c0_g1	3.16
			MeJA_TRINITY_DN30554_c0_g2	2.85
			MeJA_TRINITY_DN24537_c0_g1	2.58
			MeJA_TRINITY_DN29471_c0_g1	2.56
			MeJA_TRINITY_DN22981_c0_g1	2.51
			MeJA_TRINITY_DN25513_c1_g2	2.50
			MeJA_TRINITY_DN27696_c0_g3	2.50
			MeJA_TRINITY_DN30750_c0_g1	2.37
			MeJA_TRINITY_DN30587_c2_g1	2.35
			MeJA_TRINITY_DN27078_c1_g1	2.32
			MeJA_TRINITY_DN18535_c0_g1	2.26
			MeJA_TRINITY_DN26744_c0_g1	2.14
			MeJA_TRINITY_DN27771_c1_g2	2.01
			MeJA_TRINITY_DN23742_c0_g1	1.94
			MeJA_TRINITY_DN30554_c0_g3	1.93
			MeJA_TRINITY_DN20356_c0_g1	1.86
			MeJA_TRINITY_DN29992_c6_g2	1.68
			MeJA_TRINITY_DN28854_c0_g1	1.62
			MeJA_TRINITY_DN14462_c0_g1	1.58
			MeJA_TRINITY_DN24662_c0_g1	1.57
			MeJA_TRINITY_DN27696_c0_g1	1.46
			MeJA_TRINITY_DN28872_c0_g4	1.40
			MeJA_TRINITY_DN26222_c0_g2	1.38
			MeJA_TRINITY_DN25751_c0_g6	1.34

Sample contrasts: *Ja* (*J. aquatica*), F_2 (F_2 hybrids from a cross between *J. vulgaris* and *J. aquatica*), *Jv*-MeJA (tissue culture plants of *J. vulgaris* treated with 90 μM MeJA), *Jv*-Control (the control of Jv-MeJA); logFC of CYPs: $\log_2$fold-change of average gene expression levels of CYPs in groups containing PA contrasts; *P* value: the possibility of CYP gene found differentially expressed by chance; Best hit: best hit of a candidate to CYP subfamily against the CYP database of *J. vulgaris* and *J. aquatica* in house (Chapter 3) or the UniProtKB/Swiss-Prot database by the blastp algorithm; Identity range of pep: the range of similarities between isoforms and the reference CYP with best hit in their peptide sequences; Length range of cds: the range of lengths of coding sequences of isoforms in a CYP gene predicted by TransDecoder.

[a] $\log_2$fold-change of average PA concentrations are shown in brackets.

[b] Best hits to CYPs in the UniProtKB/Swiss-Prot database instead of the CYP database of *Jacobaea* species developed in Chapter 3.

[1-10] Ten CYP candidate genes included in two different site-specific oxidations of constitutive PA conversions.

[I-V] Five CYP candidate genes included in both constitutive and induced PA conversions.

P value	Best hit	No. of isoforms	Identity range of pep (%)	Length range of cds (nt)	CYP clan
0.035	CYP75B131	5	99.6-100	294-1572	71
0.038	CYP76B89/90	5	69.5-71.8	483-1470	71
0.035	CYP72A640	1	98.90	315	72
0.042	CYP81B108	4	96.3-99.8	813-1506	71
< 0.001	CYP706E12 [II]	2	100	618-1551	71
< 0.001	CYP81B113	1	100	1524	71
< 0.001	CYP706C72 [IV]	2	99.0-99.4	1152-1575	71
< 0.001	CYP81B112	1	100	1512	71
< 0.001	CYP93A148	2	99.2-100	1512-1524	71
< 0.001	CYP81B109	4	99.2-100	762-1518	71
< 0.001	CYP71AT158 [V]	2	98.9-100	555-1506	71
0.001	CYP71AT165	1	100	1491	71
< 0.001	CYP706E13/19	5	96.9-100	1329-1629	71
< 0.001	CYP76S32/33/36	5	93.5-98.8%	1227-1524	71
0.002	CYP71AV19	2	98.7-99	1437-1506	71
0.002	CYP749A94	4	95.8-99.6	744-1548	72
0.001	CYP82Q5	3	93.9-100	387-1584	71
0.004	CYP93A146	2	99.3-100	927-1527	71
0.002	CYP82D180 [III]	2	95.0-97.9	690-1602	71
0.004	CYP71BZ7	1	99	1503	71
0.005	CYP71CA5 [I]	2	98.8-99.2	1566	71
0.005	CYP93A144	4	98.0-100	867-1524	71
0.007	CYP93A150	2	100	852-1527	71
0.014	CYP93B69	2	100	1542	71
0.019	CYP81BG9	3	97.9-100	576-1170	71
0.021	CYP71AX47	3	99.6-99.8	924-1515	71
0.020	CYP76B90	3	98.9-100	558-1494	71
0.030	CYP76S34/37	8	96.9-100	528-1527	71
0.036	CYP72A651	7	97.2-100	768-1575	72
0.041	CYP89A179	5	98.0-99.0	732-1578	71
0.045	CYP93A147	2	100	546-1569	71

(Cluster 4) (Fig. 4A). The clustering of expression of CYP candidate genes in *Jv*-MeJA and *Jv*-Control can be divided into three clusters roughly, where Cluster 1 had the highest expression while the five CYP candidate genes in Cluster 2 showed lowest expression in both samples (Fig. 4B).

Discussion

PAs are constitutive as well as inducible in plant species containing them. In this study, *J. vulgaris*, *J. aquatica* and their hybrids were chosen as the study system, and separated into constitutive and induced groups. In the constitutive comparison, the presence of different PA patterns in the hybrids and *J. aquatica* provided powerful PA contrasts especially with regard to 15,20-epoxidation, 12,13-epoxidation and 19-hydroxylation. MeJA has been commonly used as an elicitor to mimic the effects of herbivory, resulting in metabolic changes and resistance against herbivores (Chen *et al*., 2006; Largia *et al*., 2015). In this study, tissue culture plants of *J. vulgaris* were treated with MeJA, leading to changes in PA compositions, especially a shift from senecionine-like PAs to erucifoline-like PAs, which in agreement with the results of Wei *et al*. (2019). Also, the results found here are in line with the results of Kostenko *et al*. (2013), where an increase of acetylerucifoline and its *N*-oxide and a decrease of senecionine-like PAs were observed in the shoots of *J. vulgaris* after root herbivory by the wireworm *Agriotes lineraus*. Herbivory resulted in the increase of erucifoline-like PAs and a reduction in senecionine-like PAs. This demonstrates that the PA biosynthesis pathway, especially leading to the production of erucifoline-like PAs, is at least in part controlled by the jasmonic acid signaling pathway.

In this study, gene expression levels of CYPs of plants containing PAs with certain kinds of site-specific functional groups were compared with the levels in plants containing lower amounts of these PAs. Plant CYP enzymes are typically substrate-specific and catalyze highly region-special and stereoselective transformations (Giddings *et al*., 2011). In some cases, a single CYP enzyme can catalyze a few highly related substrates (Giddings *et al*., 2011; Miettinen *et al*., 2017; Hori *et al*., 2018; Dastmalchi *et al*., 2018). It is possible that PAs with the same functional group were catalyzed by a single CYP enzyme or alike enzymes due to their highly similar structures. We analyzed gene expression abundances of CYPs including multiple transcripts (isoforms) based on highly similar sequences classified with the Trinity program since multiple genotypes were included in the samples of the constitutive group giving rise to possible sequence diversity. The number of isoforms within each gene ranged from one to 14 (Table 2). CYPs within a single family or subfamily usually oxidize similar or related compounds (Nelson and Werck-Reichhart, 2011). Nevertheless, it is noted that sometimes even a single amino acid change would can lead to a different function of a CYP enzyme (Schalk and Croteau, 2000).

Most of the CYP candidates selected in this study were from the CYP71 clan, amounting to 70% for the constitutive group and to 93% for the induced group. The predicted evolutionary

youngest clan, the CYP71 clan, has evolved and expanded dramatically recently (Nelson and Werck-Reichhart, 2011; Chapter 3), making it more challenging to predict functions and substrate preferences. The coexpression patterns of CYP candidate genes (Fig. 4) may provide hints of candidates associated in the same biological processes (van Dam *et al*., 2018). However, with lack of information about CYPs involved in PA pathway in the public database, it is not possible to predict functions for the candidates identified in this study.

In summary, in this study metabolic and transcriptomic analyses were performed for seven *Jacobaea* samples separated in constitutive and induced PA groups in order to mine the gene candidates involved in PA biosynthesis. *Jacobaea* samples were differentiated based on the abundances of PAs with six different types of site-specific oxidative modifications, and CYPs were predicted based on the association of their expression levels with concentrations of particular PAs. In total, 55 CYP genes were selected as candidates for encoding enzymes performing PA conversions. For the genes for which full-length coding sequences are not available, RACE techniques can be employed to obtain full-length genes. The hypothesized involvement of candidate enzymes in PA biosynthesis can be verified by heterologous in yeast followed by enzyme assays. Eight of these candidates were tested in this way in the next chapter (Chapter 5) of this thesis.

Acknowledgements

Yangan Chen thanks the China Scholarship Council (CSC) for financial support. Karin van der Veen-van Wijk is thanked for her technical assistance.

References

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

Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* 34:525-527.

Chen H, Jones AD, Howe GA. 2006. Constitutive activation of the jasmonate signaling pathway enhances the production of secondary metabolites in tomato. *FEBS Letters* 580:2540-2546.

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.

Chou MW, Fu PP. 2006. Formation of DHP-derived DNA adducts in vivo from dietary supplements and Chinese herbal plant extracts containing carcinogenic pyrrolizidine alkaloids. *National Center for Toxicological Research* 22:321-327.

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

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.

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.

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. Site 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.

Hartmann T. 1999. Chemical ecology of pyrrolizidine alkaloids. *Planta* 207:483-495.

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.

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

Kostenko O, Mulder PP, Bezemer TM. 2013. Effects of root herbivory on pyrrolizidine alkaloid content and aboveground plant-herbivore-parasitoid interactions in *Jacobaea vulgaris*. *Journal of Chemical Ecology* 39:109-119.

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.

Largia MJV, Pothiraj G, Shilpha J, Ramesh M. 2015. Methyl jasmonate and salicylic acid synergism enhances bacoside A content in shoot cultures of *Bacopa monnieri* (L.). *Plant Cell, Tissue and Organ Culture* 122:9-20.

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.

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

Miettinen K, Pollier J, Buyst D, Arendt P, Csuk R, Sommerwerk S, Moses T, Mertens J, Sonawane PD, Pauwels L, Aharoni A, Martins J, Nelson DR, Goossens A. 2017. The ancient CYP716 family is a major contributor to the diversification of eudicot triterpenoid biosynthesis. *Nature Communications* 8:14153.

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.

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

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.

Schalk M, Croteau R. 2000. A single amino acid substitution (F363I) converts the regiochemistry of the spearmint (-)-limonene hydroxylase from a C6- to a C3-hydroxylase. *Proceedings of the National Academy of Sciences of the United States of America* 97:11948-11953.

Toppel G, Witte L, Riebesehl B, von Borstel K, Hartmann T. 1987. Alkaloid patterns and biosynthetic capacity of root cultures from some pyrrolizidine alkaloid producing *Senecio* species. *Plant Cell Reports* 6:466-469.

Wei X, Vrieling K, Mulder PPJ, Klinkhamer PGL. 2019. Methyl jasmonate changes the composition and distribution rather than the concentration of defence compounds: a study on pyrrolizidine alkaloids. *Journal of Chemical Ecology* 45:136-145.

Ober D, Hartmann T. 1999. Homospermidine synthase, the first pathway-specific enzyme of pyrrolizidine alkaloid biosynthesis, evolved from deoxyhypusine synthase. *Proceedings of the National Academy of Sciences of the United States of America* 96:14777-14782.

Urlacher VB, Girhard M. 2012. Cytochrome P450 monooxygenases: an update on perspectives for synthetic application. *Trends in Biotechnology* 30:26-36.

Robinson MD, McCarthy DJ, Smyth GK. 2010. Edger: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140.

van Dam S, Võsa U, van der Graaf A, Franke L, de Magalhães JP. 2018. Gene co-expression analysis for functional classification and gene-disease predictions. *Briefings in Bioinformatics* 19:575-592.

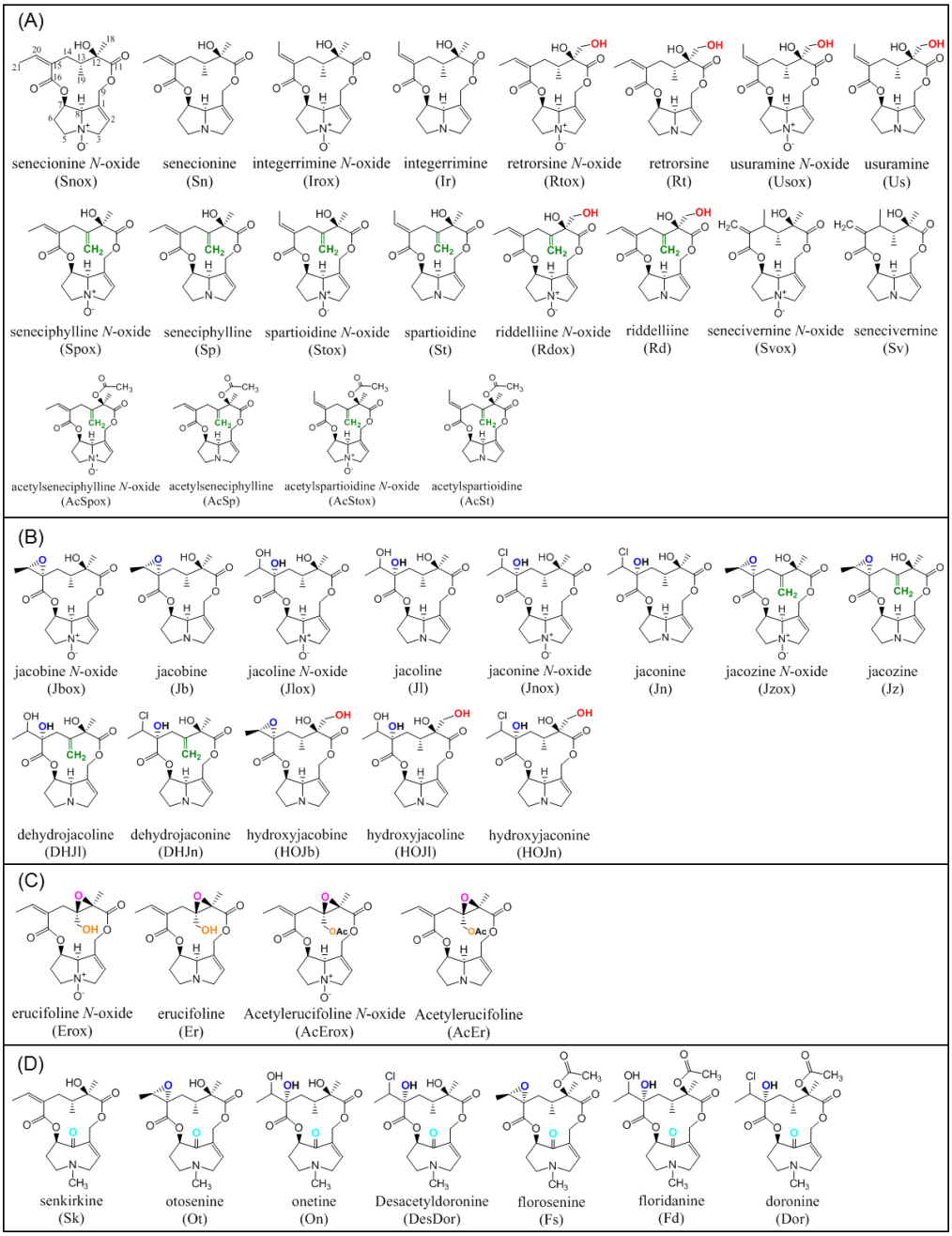

Figure S1. Chemical structures of PAs found in *Jacobaea* plants in this study. (A) Senecionine-like PAs; (B) jacobine-like PAs; (C) erucifoline-like PAs; (D) otosenine-like PAs. Site-specific oxidative modifications were marked by different colors: red = 18-hydroxylation; green = 13,19-dehydrogenation; dark blue = 15,20-epoxidation (including hydrolysis and chlorolysis of 15,20-epoxide); pink = 12,13-epoxidation (including hydrolysis and chlorolysis of 12,13-epoxide); orange = 19-hydroxylation (including acetylation of 19-hydroxyl); light blue = 8-oxidation. Abbreviations of PAs are listed in brackets.

Table S1. Details of samples with PA contrasts used for RNA sequencing analysis.

Samples	PA contrasts[a]	Genotypes	Codes of genotypes	Individuals
J. vulgaris-MeJA (*Jv*-MeJA)	+ Er	1	SJ6	5
J. vulgaris-control (*Jv*-Control)	- Er	1	SJ6	5
J. aquatica (*Ja*)	- Jb, - Er	6	SA6	2
			5 seeds from Avon, England, UK	5
F_2-1	+ Jb, + Er	5	60215	2
			70224	2
			70107	3
			60269	6
			70143	4
F_2-2	+ Jb, - Er	5	70158	2
			60118	5
			60129	8
			70138	5
			60264	5
F_2-3	- Jb, + Er	5	70238	3
			60117	2
			70151	5
			70101	5
			60116	2
F_2-4	- Jb, - Er	5	70108	5
			60260	4
			70154	6
			70160	5
			70116	6

[a] + high concentration; - low concentration; Jb: jacobine-like PAs; Er: erucifoline-like PAs.

Table S2. PA concentrations (μg/g dry weight) in the shoots of seven *Jacobaea* samples detected by LC-MS/MS.

	PAs	*Ja*	F_2-1	F_2-2	F_2-3	F_2-4	*Jv*-MeJA	*Jv*-Control
Senecionine-like PAs	senecionine (336)	29.1	1.8	4.3	14.2	11.5	0.3	3.0
	senecionine *N*-oxide (352)	910.4	55.1	122.	304.8	275.9	4.4	46.1
	integerrimine (336)	4.2	0.9	1.3	1.9	1.6	0.1	0.8
	integerrimine *N*-oxide (352)	196.1	34.1	56.1	71.6	71.3	13.8	17.1
	senecivernine (336)	0.5	0.2	0.2	0.2	0.1	0.2	0.0
	senecivernine *N*-oxide (352)	0.0	0.0	0.0	0.0	0.0	5.1	0.0
	retrorsine (352)	0.4	0.0	0.1	0.3	0.2	0.0	0.5
	retrorsine *N*-oxide (368)	10.6	1.5	3.6	4.7	7.7	6.7	9.4
	usaramine (352)	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	usaramine *N*-oxide (368)	0.1	3.1	12.3	7.6	17.3	7.6	4.6
	seneciphylline (334)	36.8	12.0	14.2	15.4	17.1	0.5	4.6
	seneciphylline *N*-oxide (350)	1252.0	367.2	410.	453.9	577.2	12.7	54.0
	spartioidine (334)	0.4	1.7	1.1	0.6	0.5	0.0	0.4
	spartioidine *N*-oxide (350)	12.9	52.5	29.3	20.7	16.9	7.0	3.8
	riddelliine (350)	0.0	0.0	0.0	0.0	0.0	0.2	0.3
	riddelliine *N*-oxide (366)	1.5	2.0	1.2	1.2	0.8	45.2	6.9
	acetylseneciphylline (376)	4.5	1.5	2.6	3.2	1.9	5.5	21.3
	acetylseneciphylline *N*-oxide	145.6	33.4	84.3	115.7	91.2	78.7	97.9
	acetylspartioidine (376)	0.1	0.2	0.2	0.0	0.0	1.0	0.7
	acetylspartioidine *N*-oxide	1.4	5.8	4.3	2.6	1.2	10.3	2.6
Jacobine-like PAs	jacobine (352)	4.8	588.4	477.	16.7	14.1	71.3	129.0
	jacobine *N*-oxide (368)	23.3	1533.2	181	31.0	23.1	81.3	116.4
	jacoline (370)	0.1	38.5	33.3	0.8	0.7	9.6	12.3
	jacoline *N*-oxide (386)	0.5	40.3	48.7	1.0	0.5	3.8	4.1
	jaconine (388)	1.1	211.5	198.	5.3	3.1	2.4	5.0
	jaconine *N*-oxide (404)	1.3	79.5	112.	1.2	0.9	0.2	0.1
	jacozine (350)	3.7	84.8	41.6	8.0	8.4	5.9	8.3
	jacozine *N*-oxide (366)	5.2	36.6	9.9	12.6	18.3	18.7	4.1
	dehydrojacoline (368)	0.0	12.0	10.6	0.7	0.9	1.1	1.0
	dehydrojaconine (386)	1.0	47.8	29.6	3.9	3.1	0.3	0.3
	HOJb (368)	0.0	1.1	1.1	0.0	0.0	1.7	2.3
	HOJl (386)	0.0	0.0	0.0	0.0	0.0	0.2	0.2
	HOJn (404)	0.0	0.8	1.2	0.0	0.0	0.0	0.0
Erucifoline-like PAs	erucifoline (350)	1.0	15.4	1.2	14.5	3.9	65.2	8.2
	erucifoline *N*-oxide (366)	32.0	209.2	21.2	274.7	71.4	822.7	26.2
	acetylerucifoline (392)	0.3	2.2	0.2	1.8	0.5	6.8	0.8
	acetylerucifoline *N*-oxide	20.2	130.9	15.3	114.1	46.3	202.9	7.7
Otosenine-like PAs	senkirkine (366)	0.8	4.8	0.1	0.0	0.0	0.4	0.5
	otosenine (382)	228.2	119.3	42.8	35.5	93.4	7.9	9.8
	onetine (400)	8.2	6.2	1.9	2.0	4.2	1.0	0.9
	desacetyldoronine (418)	63.2	28.8	10.8	6.0	14.5	0.0	0.1
	florosenine (424)	41.8	16.6	2.9	86.8	43.5	0.0	0.0
	floridanine (442)	1.3	1.1	0.2	3.5	1.9	0.0	0.0
	doronine (460)	7.4	7.1	1.4	21.8	10.9	0.0	0.0
	sum-fb	439.1	1204.8	879.	243.2	236.1	185.9	212.6
	sum-ox	2613.1	2584.3	274	1417.5	1220.	1341.4	406.6
	total	3052.2	3789.2	362	1660.7	1456.	1527.3	619.3

Ja: *J. aquatica*; F_2: F_2 hybrids of *J. vulgaris* and *J. aquatica*; *Jv*-MeJA: tissue culture plants of *J. vulgaris* treated with 90 μM MeJA for eight days; *Jv*-Control: the control group of *Jv*-MeJA; sum-fb: the sum of all PA free bases; sum-ox: the sum of all PA *N*-oxides; total: the sum of all PAs. The numbers in the brackets indicate the precursor mass (m/z) of each PA.

Chapter 5

Tests of cytochrome P450 candidates for the pyrrolizidine alkaloid pathway of *Jacobaea* species using an expression system in yeast

Yangan Chen, Klaas Vrieling, Patrick P.J. Mulder, Peter G.L. Klinkhamer, Erica Wilson, Johan Memelink

Tests of cytochrome P450 candidates for the pyrrolizidine alkaloid pathway of *Jacobaea* species using an expression system in yeast

Abstract

Pyrrolizidine alkaloids (PAs) are a typical class of secondary metabolites with high structural diversity. It is assumed that specific structural diversification of the backbone structure senecionine *N*-oxide to other PAs resulting in unique PA bouquets provides a powerful strategy for plants to cope with biotic and abiotic stresses. Senecionine *N*-oxide undergoes structural transformations including site-specific oxidative modifications and therefore, cytochrome P450s (CYPs) could be involved in PA biosynthesis. Here eight CYP candidates of *Jacobaea* species were tested for their functions by using heterologous expression in yeast and in vitro enzyme assays. The enzyme assays were performed using extracted microsomal membranes and the CYP geraniol 10-hydroxylase was used as a positive control. None of the eight CYP enzymes showed catalytic activities in structural conversion using senecionine/integerrimine, seneciphylline, jacobine, erucifoline or a PA mixture as well as the respective *N*-oxides as substrates. Although this might due to the fact that more optimal reaction conditions are required, most likely these CYPs are not enzymes invoved in the transformation of the tested substrates.

Keywords

yeast, pYeDP60u, microsome, senecionine, intergerrimine, seneciphylline, jacobine, erucifoline

Introduction

Pyrrolizidine alkaloids (PAs) are a characteristic group of plant secondary metabolites (SMs) with a high diversity of chemical structures. Of the senecionine type PAs, senecionine *N*-oxide (Fig. 1) has been identified as the primary product of the biosynthesis of the 12-membered macrocyclic senecionine-type PAs by tracer-feeding experiments (Hartmann and Toppel 1987; Hartmann *et al*., 1989; Hartmann and Dierich, 1998). In shoots of plants of the Senecioneae tribe, senecionine *N*-oxide was found to be diversified into species- or chemotype-specific bouquets of biosynthetically derived PAs including seneciphylline, jacobine and erucifoline (Hartmann and Dierich, 1998). These PAs have been extensively studied with regards to their ecological roles in defense (Macel *et al*., 2005; Nuringtyas *et al*., 2014; Liu *et al*., 2017; Leiss *et al*., 2009; Cheng *et al*., 2011; Wei *et al*., 2015). These studies demonstrated that PAs with different structures have different bioactivities on herbivores. It is assumed that specific structural conversion of senecionine *N*-oxide to other PAs resulting in individual PA bouquets provides a powerful strategy for plants to cope with a dynamic environment (Hartmann and Dierich, 1998). Still an open question is how the PA bouquet is brought about in individual plants. Thus, it is of interest to elucidate the underlying genes of the PA biosynthesis.

Transformation between most of the 12-membered senecionine type PAs can be explained by simple one- or two-step reactions from the primary PA senecionine *N*-oxide (Hartmann and Dierich, 1998; Pelser *et al*., 2005; Langel *et al*., 2011). Senecionine *N*-oxide (Fig. 1) undergoes position-specific and stereoselective structural transformations resulting in the rearrangement of the skeletal structure and in oxidative modifications (Hartmann and Dierich, 1998). The oxidative modifications include 18-hydroxylation, 13,19-dehydrogenation, 15,20-epoxidation, 12,13-epoxidation, 19-hydroxylation and 8-oxidation (see Fig. 1 and Fig. S1 of Chapter 4), which could be catalyzed by cytochrome P450s (CYPs) as CYPs catalyze the oxidative modifications of various substrates using oxygen and NAD(P)H in various SM biosynthesis pathways. Through metabolic and transcriptomic profiling techniques, initial CYP candidates for involvement in PA biosynthesis were selected from *Jacobaea* species (Chapter 4) and their functions remain to be tested.

In order to unequivocally characterize the enzymatic activities of CYPs, heterologous expression of CYP proteins is a critical step. All plant CYPs described so far are bound to membranes, usually to the endoplasmic reticulum (ER). To be active, CYPs need to be coupled with electron-donating proteins, CYP reductases or cytochrome b_5, which are also anchored to the surface of ER. Most commonly, via the NADPH-dependent CYP reductase (CPR) heme-bound O_2 is activated by the successive transfer of two electrons from NADPH (Bak *et al*., 2011), leading to regiospecific and stereospecific oxidative attack of a plethora of substrates. Therefore, yeast is the most often used system for heterologous expression of CYPs. For analysis of plant CYPs the yeast strain WAT11 (Pompon *et al*., 1996) which expresses an NADPH reductase from *Arabidopsis thaliana* instead of the yeast NADPH reductase and the

yeast expression vector pYeDP60 (Urban *et al*., 1990) has been traditionally used (Duan and Schuler, 2006; Miettinen *et al*., 2014; Liu *et al*., 2016; Liu *et al*., 2018). However, the vector pYeDP60 offers a limited selection of restriction sites which complicates the cloning of CYPs. The adaption of the vector to the uracil-excision based (USER^{TM}) cloning technique (Nour-Eldin *et al*., 2006) resulting in pYeDP60u was advocated as more optimal in high-throughput screening of CYPs (Hamann and Møller, 2007).

Here the functional tests of eight CYP candidates available in *Jacobaea vulgaris* were performed. After recombination of selected genes into the pYeDP60u vector, CYP genes were introduced in yeast. Subsequently, microsomes were extracted for *in vitro* enzyme assays using isolated PAs or a PA mixture as substrates, and reaction products were analyzed by LC-MS/MS. Geraniol 10-hydroxylase (G10H; CYP76B6) from *Catharanthus roseus* (Collu *et al*., 2001) was used as a positive control in this study.

Materials and Methods

Plant material and cDNA synthesis

Tissue culture plants of *J. vulgaris* were grown and treated with methyl jasmonate (MeJA) as reported in Chapter 4. The shoots of cultured plants with the MeJA treatment of eight days were ground into fine powder in liquid nitrogen, and were stored at -80 °C. Total RNA was extracted by phenol/chloroform extraction followed by overnight precipitation with ca. 3 M lithium chloride, washed with 70% ethanol, and resuspended in water. First strand cDNA was synthesized using the Thermo Scientific RevertAid First Strand cDNA Synthesis Kit.

Amplification and cloning of CYP genes

Eight CYP candidate genes, i.e. CYP71AT158, CYP71AV19, CYP76S36, CYP706C72, CYP706E12, CYPP81B113, CYP82D180 and CYP82Q5, from different CYP families/subfamilies were selected based on the association of PA abundances with differential gene expression of CYPs between different *Jacobaea* samples (Chapter 4). G10H (CYP76B6) from *C. roseus* was used as a positive control. The full-length coding sequences were amplified by PCR using Phusion® High-Fidelity DNA Polymerase (New England Biolabs) according to manufacturer's instruction (for primers see Table S1). Purified PCR fragments were ligated into the pJET1.2 vector using the Thermo Scientific CloneJET PCR Cloning Kit and introduced into *Escherichia coli* strain XL1-Blue. The resulting plasmids were sent for sequencing of both strands by BaseClear B.V. (The Netherlands).

Preparation of pYeDP60u vector for cloning

The yeast vector pYeDP60u (Hamann and Møller, 2007) was a kind gift from Nicolas Navrot from Strasbourg University. This vector was prepared for cloning according to Nour-Eldin *et al*. (2006) with minor modification. One μg pYeDP60u plasmid was digested with 10 U PacI (New England Biolabs) overnight in a total volume of 50 μL. An additional 5 U PacI was

added in together with 5 U Nt.BbvCI (New England Biolabs), and the digestion was incubated for 2 hours at 37 °C. The linearized vector was purified using the Thermo Scientific GeneJET Gel Extraction Kit.

Expression of CYP genes in yeast

CYP genes were PCR re-amplified using KlearTaq DNA polymerase with the pJET1.2 plasmid constructs as templates. Uracil-containing primers were designed according to Hamann and Møller (2007): forward 5'-GGATTAAU + A + sequence of coding strand of targeted DNA; reverse 5'-GGGTTAAU + optimized stop codon (TAA) + sequence complementary to coding strand of target DNA (Table S1). The PCR fragments were inserted into yeast/*E. coli* shuttle vector pYeDP60u using USER enzyme (New England Biolabs) and introduced in *E. coli* strain XL1-Blue. The genes inserted in the resulting plasmids were re-sequenced with the plasmid primers (forward: 5'-CACGCAAACACAAATACACACAC-3'; reverse: 5'- AAGCACCACCACCAGTAGAG-3'). Subsequently, the obtained constructs were transformed into the *Saccharomyces cerevisiae* WAT11 yeast strain which expresses the *ATR1* CPR from *Arabidopsis thaliana* (Pompon *et al*., 1996) by the improved lithium acetate procedure (Gietz *et al*., 1992) with minor modifications. The transformed cells were plated on SGI plates for autotrophic selection (Pompon *et al*., 1996). Yeast transformed with empty pYeDP60u plasmid and G10H was used as the negative control and the positive control, respectively.

Yeast microsome isolation

Preparation of microsomes from transformed yeasts was carried out according to Liu *et al*. (2016) with minor modifications. Colonies streaked onto SGI plates were pre-cultured in 10 mL liquid SGI medium (20 g/L glucose, 1 g/L bactocasamino acids, 7 g/L yeast nitrogen base without amino acid without ammonium sulfate, 5 g/L ammonium sulfate, 40 mg/L L-tryptophan) at 30 °C for 18 hours in a shaking incubator. Ten milliliters of pre-culture were used to inoculate 200 mL of YPGE medium (10 g/L yeast extract, 10 g/L bactopeptone, 5 g/L glucose, 3% ethanol by volume) and grown for 30 hours at 30 °C at 150 rpm. Induction of gene expression was started by adding 10 mL of 200 g/L galactose and further incubation at 20 °C for 16 hours.

Yeast cells were pelleted at 4,000 rpm for 10 min at 4 °C, washed with TEK buffer (50 mM Tris-HCl pH 7.5, 1 mM EDTA, 100 mM KCl), and resuspended in 2 mL of TES buffer (50 mM Tris-HCl pH 7.5, 1 mM EDTA, 600 mM sorbitol) supplemented with 2-mercaptoethanol and 10 g/L bovine serum albumin, farction V (extraction buffer). The cell resuspensions were homogenized with ca. 5mL of glass beads (450 - 600 μm) by shaking up and down by hand for 5 min in a cold room. Glass beads were washed twice with 10 mL of cold extraction buffer, and lysates were pooled. Cell debris and remaining glass beads were removed from the pooled lysates by centrifugation at 4,000 rpm for 10 min at 4 °C. Supernatants were filtered through Miracloth (Calbiochem®) and microsomes were pelleted by centrifugation at 100,000 g at 4

°C for 1 hour. Pelleted microsomal fractions were resuspended in 1 mL of TEG buffer (50 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 30% (v/v) glycerol) with a Potter-Elvehjem homogenizer. Protein concentrations of microsome preparations were measured by using Bradford reagent (BioRad) in cuvettes on a Nanodrop 2000c spectrophotometer (Thermo Scientific). Microsomal preparations were aliquoted and flash frozen in liquid nitrogen and stored at -80 ℃ until enzyme assays.

***In vitro* enzyme assays**

Enzyme assays were done in a 200 µL final volume of 20 mM potassium phosphate buffer (pH 7.5), containing 200 µg microsomal protein, 500 µM NADPH and 100 µM PA substrate. The mixture was incubated for 30 min at 28 °C and the reaction was terminated by addition of 10 µL of 50% acetic acid. Samples were diluted 10-fold with water and centrifuged at 14,000 rpm for 2 min and the supernatants were passed through a 0.2 µm regenerated cellulose membrane filter (Sartorius). PA substrates (Fig. 1) including a reduced PA extract of leaves of *J. vulgaris*, a mixture containing senecionine and integerrimine, seneciphylline, jacobine, erucifoline (isolated by ExPlant Technologies B.V., The Netherlands) and their corresponding *N*-oxides obtained by incubation with 300 µL of 30% H_2O_2 (Sigma) for 16 hours at 60°C in methanol at a concentration of 5 mg/mL were used for each CYP enzyme of *J. vulgaris* (Table 1). For the positive control G10H geraniol (approx. 98% purity; Sigma) was used as the substrate, and the reaction volume was enlarged to 1 mL final volume as liquid-liquid extraction was required in order to prepare samples for GC-MS which has a low sensitivity. After incubation, the aqueous reaction mixture was extracted three times with 1 mL of ethyl acetate. The ethyl acetate layers were combined and dried using a CentriVap concentrator (Labconco). The residue was redissolved in 100 µL of ethyl acetate before injection into GC-MS. The enzyme assay of the positive control was carried out in duplicate.

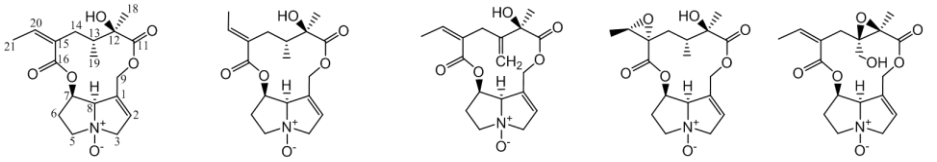

Figure 1. Structures of isolated PAs in *N*-oxide form used as substrates in enzyme assays.

LC-MS/MS analysis of PA conversion

Analysis of reaction mixtures was performed on an LC-MS/MS system consisting of a Waters Acquity UPLC coupled to a Xevo TQ-S tandem mass spectrometer (Waters, Milford, MA, USA), run in positive electrospray mode. Before analysis aliquots (50 µL) of the filtered supernatants were diluted an additional 20 times with water before injection into LC-MS/MS.

Chromatographic separation was achieved on an Acquity BEH C18 analytical column, 150 × 2.1 mm, 1.7 μm particle size (Waters, Milford, MA, USA). Eluent A consisted of water containing 10 mM ammonium carbonate pH 9.0 and acetonitrile was used as eluent B. A gradient elution was performed as follows: 0.0 min 100% A/0% B, 0.1 min 95% A/5% B, 3.0 min 90% A/10% B, 7.0 min 76% A/24% B, 9.0 min 70% A/30% B, 12.0 min 30% A/70% B, 12.1-15.0 min 100% A/0% B. The column was kept at 50ºC and a flow rate of 400 μL/min was applied; 2 μL of the sample extracts was injected.

For each analyte at least two selected precursor to product ion MRM transitions were measured. Cone energy was 40V and collision energy settings were optimized for the individual compounds. Quantification was performed against a range of mixed standard solutions (0-5-10-25-50-100-200 μg/L) of the PAs in a diluted extract of *Tanacetum vulgare* (tansy). The extract of *T. vulgare* material was used to mimic a PA-free plant extract. The range of mixed standard solutions was injected at the beginning of the series and at the end. The mixed standard solution of 50 μg/L in *T. vulgare* extract was injected every 30-40 samples, to monitor the performance of the system (drift in retention times, changes in detector sensitivity) during the analysis. For each PA the averaged response of two MRM transitions was used for quantification. Data processing was conducted with MassLynx 4.1 software (Waters Corporation, Milford, MA, USA).

GC-MS analysis of geraniol conversion

The ethyl acetate solutions were analyzed with a 7890A gas chromatograph equipped with a 7693 automatic sampler and coupled to a 5975C mass single-quadrupole detector (Agilent, Folsom, CA, USA). Samples were separated using a DB-5 GC column (30m×0.25 mm, 0.25 μm film, J&W Science, Folsom, CA, USA) and He (99.9% purity) as a carrier gas at a flow rate of 1 mL/min. The oven temperature was programmed starting at 60 °C and increased to 150 °C at 5 °C/min, then to 240 °C at 15 °C/min. The injector was set at 250 °C and 1 μL of sample was injected in splitless mode. The interface temperature was 280 °C, and the ion source and quadrupole temperature of the mass detector were 230 °C and 150 °C, respectively. Ionization energy in EI mode was 70 eV and peaks were identified by comparison of their ion spectra with the NIST library (version 2008), and by comparison of their retention time and spectra with the standards geraniol and 10-hydroxygeraniol (synthesized by Chiralix B.V., Nijmegen, The Netherlands).

Results

The purpose of this study is to check whether the selected CYP candidates are involved in PA biosynthetic pathway. Aiming to characterize the enzymatic activities of CYPs, heterologous expression of CYP proteins in the yeast expression system was conducted. *In vitro* enzyme

assays were performed using extracted microsomal membranes and different PA substrates, and reaction products were analyzed by LC-MS/MS.

CYP sequences

The initial step of expression of CYPs was to amplify selected CYP sequences from cDNA of the shoots of cultured plants with the MeJA treatment of eight days. The sequence information of CYPs used for amplification were obtained from the *de novo* assembled transcriptome of *J. vulgaris* (Chapter 3; Chapter 4). Six of the eight CYPs showed identical amino acid sequences between Sanger sequencing and RNA-Seq (three CYPs showed identical nucleotide sequences, while the other three had 1-2 different nucleotides). The remaining two CYPs showed two and three different amino acids between two different sequencing approaches.

Geraniol conversion

In order to make sure that the heterologous expression of plant CYPs in yeast and *in vitro* enzyme assays were conducted properly, G10H of *C. roseus* was used as the positive control in our study. In the enzyme assay, geraniol incubated with microsomes expressing CYP76B6 was successfully converted to 10-hydroxygeraniol (Fig. 2A). Most of the geraniol was evaporated during vacuum drying of ethyl acetate extracts (Fig. 2A-B). Metabolic background of enzyme assays (Fig. 2B) as well as background noise of the GC-MS instrument (Fig. 2E) were observed.

Enzyme assays of CYPs from *Jacobaea vulgaris*

None of the eight CYP enzymes selected as candidates for PA biosynthesis showed catalytic activities in structural transformation of PAs using a PA extract, a mixture of senecionine and integerrimine, seneciphylline, jacobine, erucifoline or the respective PAs in *N*-oxide form as substrates (Table 1).

5

Discussion

In the current study, eight CYP candidates (CYP71AT158, CYP71AV19, CYP76S36, CYP706C72, CYP706E12, CYPP81B113, CYP82D180 and CYP82Q5) were chosen from the candidate list obtained in Chapter 4. All amplified CYPs in this study showed identical amino acid sequences to *de novo* assembled sequences of RNA-seq analysis except two (CYP71AT158 and CYP706C72) which showed two and three different amino acids to their virtual sequences. These results provided an evidence for the good quality of *de novo* assembled transcriptome of *J. vulgaris*.

Aiming at characterizing the functions of CYP candidates from *Jacobaea* species to see whether they are involved in PA biosynthesis, we used heterologous expression in yeast and *in vitro* enzyme assays. The successful conversion of geraniol to 10-hydroxygeraniol demonstrated that the yeast expression system followed by *in vitro* enzyme assays can be used

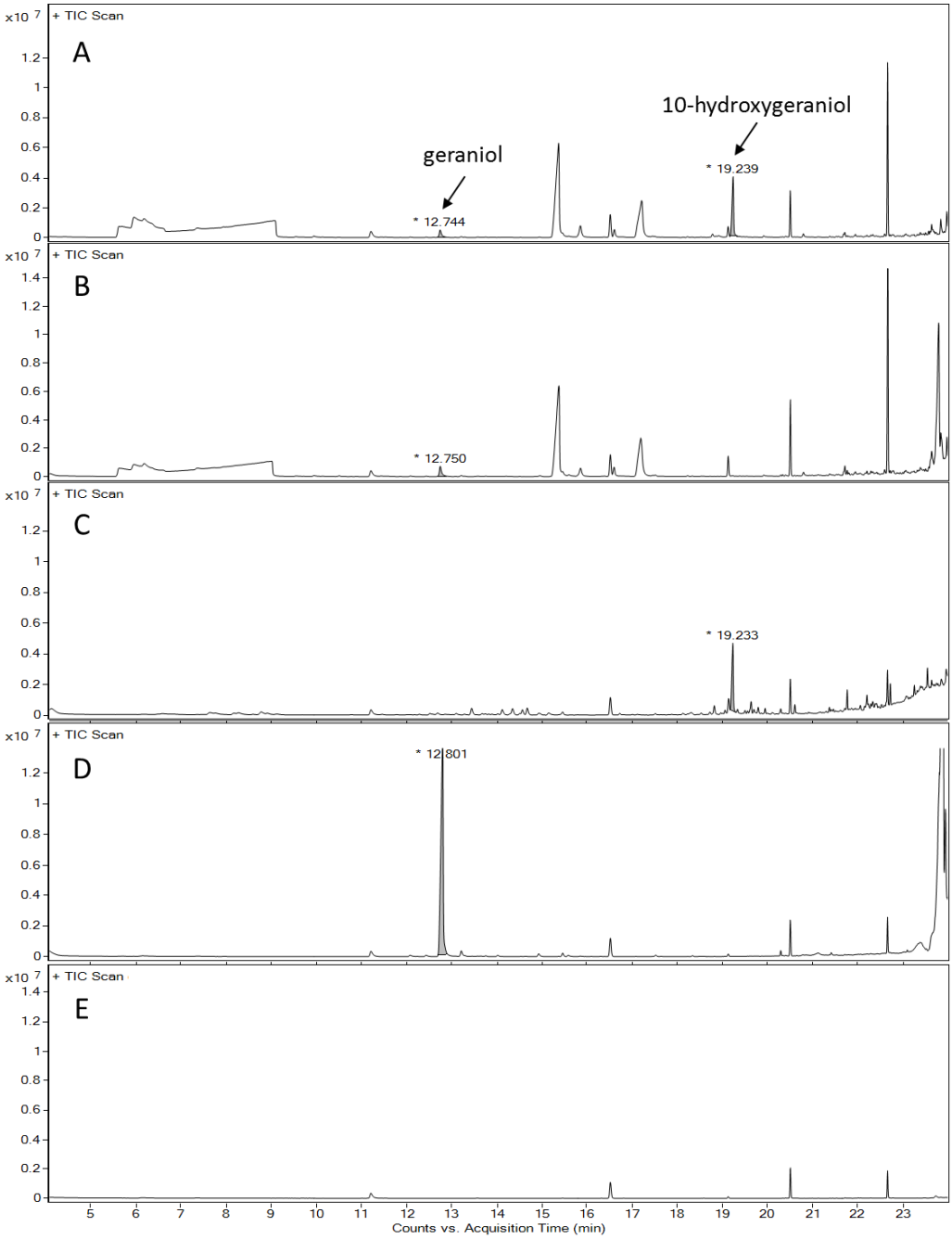

Figure 2. GC-MS analyses of geraniol conversion in enzyme assays with yeast microsomes. (A) Microsomes from yeast expressing G10H; (B) microsomes from yeast expressing the empty pYeDP60u vector; (C) 10-hydroxygeraniol standard; (D) geraniol standard; (E) blank solvent ethyl acetate.

Table 1. Tests of cytochrome P450s (CYPs) for pyrrolizidine alkaloid (PA) conversions by *in vitro* enzyme assays with different PA substrates.

CYP Candidates	PA substrates	Sum FB	Sum Nox	Sn (336)	Sn-ox (352)	Ir (336)	Ir-ox (352)	Sv (336)	Sv-ox (352)	Rt (352)	Rt-ox (368)	Us (352)	Us-ox (368)	Sp (334)
CYP706E12		99.2	0.8	58.6	0.1	34.6	0.2	2.1	0.2	[b]	-	-	-	3.7
CYP81B113		99.5	0.5	59.3	0.1	34.0	0.0	2.3	0.2	-	-	-	-	3.7
CYP71AV19		99.3	0.7	59.3	0.1	33.9	0.2	2.3	0.2	-	-	-	-	3.6
CYP82D180		99.2	0.8	59.2	0.1	34.0	0.2	2.1	0.2	-	-	-	-	3.7
CYP76S36	Sn + Ir	99.2	0.8	59.2	0.1	33.9	0.2	2.2	0.2	-	-	-	-	3.8
CYP71AT158		99.2	0.8	59.4	0.1	33.8	0.2	2.0	0.2	-	-	-	-	3.7
CYP706C72		99.3	0.7	59.3	0.1	34.1	0.2	2.1	0.2	-	-	-	-	3.6
CYP82Q5		99.2	0.8	59.2	0.1	33.7	0.2	2.5	0.2	-	-	-	-	3.6
NC[a]		99.3	0.7	59.4	0.1	34.0	0.2	2.1	0.2	-	-	-	-	3.6
CYP706E12		0.2	99.8	0.2	59.9	0.1	33.8	-	-	-	-	-	-	-
CYP81B113		0.5	99.5	0.3	58.6	0.1	33.8	-	-	-	-	-	-	-
CYP71AV19		0.1	99.9	0.1	59.6	0.0	33.9	-	-	-	-	-	-	-
CYP82D180		0.6	99.4	0.4	59.2	0.2	33.3	-	-	-	-	-	-	-
CYP76S36	Sn-ox + Ir-ox	0.2	99.8	0.2	59.0	0.1	33.8	-	-	-	-	-	-	-
CYP71AT158		0.2	99.8	0.2	59.4	0.1	34.0	-	-	-	-	-	-	-
CYP706C72		0.2	99.8	0.1	59.5	0.1	33.4	-	-	-	-	-	-	-
CYP82Q5		0.3	99.7	0.1	60.1	0.0	33.5	-	-	-	-	-	-	-
NC		0.3	99.7	0.2	59.5	0.1	33.7	-	-	-	-	-	-	-
CYP706E12		99.9	0.1	1.5	-	0.2	-	-	-	-	-	-	-	98.0
CYP81B113		99.9	0.1	1.6	-	0.2	-	-	-	-	-	-	-	98.0
CYP71AV19		99.9	0.1	1.6	-	0.2	-	-	-	-	-	-	-	98.0
CYP82D180		99.9	0.1	1.6	-	0.2	-	-	-	-	-	-	-	98.0
CYP76S36	Sp	99.9	0.1	1.6	-	0.2	-	-	-	-	-	-	-	98.0
CYP71AT158		99.9	0.1	1.6	-	0.2	-	-	-	-	-	-	-	98.0
CYP706C72		99.9	0.1	1.6	-	0.2	-	-	-	-	-	-	-	98.0
CYP82Q5		99.9	0.1	1.6	-	0.2	-	-	-	-	-	-	-	98.0
NC		99.9	0.1	1.6	-	0.2	-	-	-	-	-	-	-	98.0
CYP706E12		0.3	99.7	-	1.8	-	0.2	-	-	-	-	-	-	0.3
CYP81B113		0.6	99.4	-	1.8	-	0.2	-	-	-	-	-	-	0.6
CYP71AV19		0.2	99.8	-	1.8	-	0.2	-	-	-	-	-	-	0.2
CYP82D180		0.8	99.2	-	1.8	-	0.2	-	-	-	-	-	-	0.8
CYP76S36	Spox	0.2	99.8	-	1.7	-	0.2	-	-	-	-	-	-	0.2
CYP71AT158		0.2	99.8	-	1.8	-	0.3	-	-	-	-	-	-	0.2
CYP706C72		0.4	99.6	-	1.8	-	0.2	-	-	-	-	-	-	0.4
CYP82Q5		0.2	99.8	-	1.8	-	0.3	-	-	-	-	-	-	0.2
NC		0.4	99.6	-	1.8	-	0.2	-	-	-	-	-	-	0.4
CYP706E12		99.9	0.1	-	-	-	-	-	-	-	-	-	-	0.1
CYP81B113		99.9	0.1	-	-	-	-	-	-	-	-	-	-	0.1
CYP71AV19		99.9	0.1	-	-	-	-	-	-	-	-	-	-	0.1
CYP82D180		99.9	0.1	-	-	-	-	-	-	-	-	-	-	0.1
CYP76S36	Jb	99.9	0.1	-	-	-	-	-	-	-	-	-	-	0.1
CYP71AT158		99.9	0.1	-	-	-	-	-	-	-	-	-	-	0.1
CYP706C72		99.9	0.1	-	-	-	-	-	-	-	-	-	-	0.1
CYP82Q5		99.9	0.1	-	-	-	-	-	-	-	-	-	-	0.1
NC		99.9	0.1	-	-	-	-	-	-	-	-	-	-	0.1

Sp-ox (350)	St (334)	St-ox (350)	Rd (350)	Rd-ox (366)	Jb (352)	Jb-ox (368)	Jl (370)	Jl-ox (386)	Jn (388)	Jn-ox (404)	Jz (350)	Jz-ox (366)	Er (350)	Er-ox (366)	Un-ox (366)	Ot (382)
-	-	-	-	-	0.1	-	-	-	-	-	-	-	0.1	-	0.2	-
-	-	-	-	-	0.0	-	-	-	-	-	-	-	0.1	-	0.2	-
-	-	-	-	-	0.1	-	-	-	-	-	-	-	0.1	-	0.2	-
-	-	-	-	-	0.1	-	-	-	-	-	-	-	0.1	-	0.2	-
-	-	-	-	-	0.0	-	-	-	-	-	-	-	0.1	-	0.2	-
-	-	-	-	-	0.0	-	-	-	-	-	-	-	0.1	-	0.2	-
-	-	-	-	-	0.0	-	-	-	-	-	-	-	0.1	-	0.2	-
-	-	-	-	-	0.0	-	-	-	-	-	-	-	0.1	-	0.2	-
-	-	-	-	-	0.0	-	-	-	-	-	-	-	0.1	-	0.2	-
2.7	-	-	-	-	-	0.1	-	0.1	-	-	-	-	0.0	0.0	3.2	-
2.7	-	-	-	-	-	0.1	-	0.1	-	-	-	-	0.0	0.9	3.3	-
2.7	-	-	-	-	-	0.1	-	0.1	-	-	-	-	0.0	0.0	3.3	-
2.6	-	-	-	-	-	0.1	-	0.1	-	-	-	-	0.0	0.9	3.2	-
2.7	-	-	-	-	-	0.1	-	0.1	-	-	-	-	0.0	0.8	3.3	-
2.7	-	-	-	-	-	0.1	-	0.1	-	-	-	-	0.0	0.0	3.3	-
2.6	-	-	-	-	-	0.1	-	0.1	-	-	-	-	0.0	0.8	3.2	-
2.6	-	-	-	-	-	0.1	-	0.1	-	-	-	-	0.1	0.0	3.2	-
2.6	-	-	-	-	-	0.1	-	0.1	-	-	-	-	0.0	0.1	3.2	-
0.0	-	0.0	-	-	0.1	-	-	-	-	-	-	-	-	-	-	-
0.0	-	0.0	-	-	0.1	-	-	-	-	-	-	-	-	-	-	-
0.0	-	0.0	-	-	0.1	-	-	-	-	-	-	-	-	-	-	-
0.1	-	0.1	-	-	0.1	-	-	-	-	-	-	-	-	-	-	-
0.1	-	0.0	-	-	0.1	-	-	-	-	-	-	-	-	-	-	-
0.1	-	0.0	-	-	0.1	-	-	-	-	-	-	-	-	-	-	-
0.0	-	0.0	-	-	0.1	-	-	-	-	-	-	-	-	-	-	-
0.0	-	0.0	-	-	0.1	-	-	-	-	-	-	-	-	-	-	-
0.0	-	0.0	-	-	0.1	-	-	-	-	-	-	-	-	-	-	-
95.7	-	0.0	-	1.4	-	0.1	-	-	-	-	0.0	0.1	-	0.1	0.3	-
95.5	-	0.0	-	1.5	-	0.1	-	-	-	-	0.0	0.0	-	0.0	0.3	-
95.9	-	0.0	-	1.4	-	0.1	-	-	-	-	0.0	0.0	-	0.0	0.3	-
95.4	-	0.0	-	1.3	-	0.1	-	-	-	-	0.0	0.1	-	0.0	0.3	-
95.9	-	0.0	-	1.4	-	0.1	-	-	-	-	0.0	0.1	-	0.0	0.3	-
95.8	-	0.0	-	1.4	-	0.1	-	-	-	-	0.0	0.1	-	0.0	0.3	-
95.7	-	0.0	-	1.4	-	0.1	-	-	-	-	0.0	0.1	-	0.0	0.3	-
95.8	-	0.0	-	1.4	-	0.1	-	-	-	-	0.0	0.1	-	0.0	0.3	-
95.7	-	0.0	-	1.3	-	0.1	-	-	-	-	0.0	0.1	-	0.0	0.3	-
-	-	-	-	-	96.3	0.1	0.3	-	0.2	-	2.7	-	0.2	-	-	-
-	-	-	-	-	96.3	0.1	0.3	-	0.2	-	2.7	-	0.2	-	-	-
-	-	-	-	-	96.2	0.1	0.3	-	0.2	-	2.8	-	0.2	-	-	-
-	-	-	-	-	96.1	0.1	0.4	-	0.2	-	2.9	-	0.2	-	-	-
-	-	-	-	-	96.2	0.1	0.3	-	0.2	-	2.8	-	0.2	-	-	-
-	-	-	-	-	96.1	0.1	0.4	-	0.2	-	2.8	-	0.4	-	-	-
-	-	-	-	-	96.2	0.1	0.4	-	0.2	-	2.7	-	0.2	-	-	-
-	-	-	-	-	96.2	0.1	0.4	-	0.2	-	2.8	-	0.2	-	-	-
-	-	-	-	-	96.2	0.1	0.4	-	0.2	-	2.8	-	0.2	-	-	-

CYP706E12		0.5	99.5	-	-	-	-	-	-	-	-	-	-	-
CYP81B113		1.5	98.5	-	-	-	-	-	-	-	-	-	-	-
CYP71AV19		0.3	99.7	-	-	-	-	-	-	-	-	-	-	-
CYP82D180		1.9	98.1	-	-	-	-	-	-	-	-	-	-	-
CYP76S36	Jb-ox	0.4	99.6	-	-	-	-	-	-	-	-	-	-	-
CYP71AT158		0.4	99.6	-	-	-	-	-	-	-	-	-	-	-
CYP706C72		0.5	99.5	-	-	-	-	-	-	-	-	-	-	-
CYP82Q5		0.3	99.7	-	-	-	-	-	-	-	-	-	-	-
NC		1.1	98.9	-	-	-	-	-	-	-	-	-	-	-
CYP706E12		99.8	0.2	-	-	-	-	-	-	0.0	-	0.0	-	-
CYP81B113		99.8	0.2	-	-	-	-	-	-	0.0	-	0.0	-	-
CYP71AV19		99.8	0.2	-	-	-	-	-	-	0.0	-	0.0	-	-
CYP82D180		99.8	0.2	-	-	-	-	-	-	0.0	-	0.0	-	-
CYP76S36	Er	99.8	0.2	-	-	-	-	-	-	0.0	-	0.0	-	-
CYP71AT158		99.8	0.2	-	-	-	-	-	-	0.0	-	0.0	-	-
CYP706C72		99.8	0.2	-	-	-	-	-	-	0.0	-	0.0	-	-
CYP82Q5		99.8	0.2	-	-	-	-	-	-	0.0	-	0.1	-	-
NC		99.8	0.2	-	-	-	-	-	-	0.1	-	0.0	-	-
CYP706E12		0.5	99.5	-	-	-	-	-	-	-	0.2	-	0.0	-
CYP81B113		1.8	98.2	-	-	-	-	-	-	-	0.2	-	0.0	-
CYP71AV19		0.4	99.6	-	-	-	-	-	-	-	0.1	-	0.3	-
CYP82D180		2.3	97.7	-	-	-	-	-	-	-	0.2	-	0.3	-
CYP76S36	Er-ox	0.7	99.3	-	-	-	-	-	-	-	0.1	-	0.3	-
CYP71AT158		0.6	99.4	-	-	-	-	-	-	-	0.2	-	0.3	-
CYP706C72		0.6	99.4	-	-	-	-	-	-	-	0.2	-	0.3	-
CYP82Q5		0.6	99.4	-	-	-	-	-	-	-	0.2	-	0.3	-
NC		1.3	98.7	-	-	-	-	-	-	-	0.2	-	0.3	-
CYP706E12		99.8	0.2	6.4	0.0	2.6	0.0	0.2	-	3.0	-	1.8	-	22.5
CYP81B113		99.8	0.2	6.3	0.0	2.6	0.0	0.2	-	2.9	-	1.7	-	22.4
CYP71AV19		98.9	1.1	6.1	0.0	2.6	0.9	0.2	-	2.9	-	1.8	-	22.5
CYP82D180	A mixture of	99.1	0.9	6.2	0.8	2.6	0.0	0.2	-	2.9	-	1.8	-	22.2
CYP76S36	PA extract	98.2	1.8	6.0	0.8	2.5	0.9	0.2	-	2.9	-	1.8	-	22.1
CYP71AT158		98.1	1.9	6.0	0.8	2.6	0.9	0.2	-	2.8	-	1.7	-	22.5
CYP706C72		98.2	1.8	6.2	0.8	2.7	0.9	0.1	-	2.9	-	1.8	-	22.3
CYP82Q5		98.1	1.9	6.2	0.8	2.6	0.9	0.2	-	2.9	-	1.8	-	21.9
NC		99.0	1.0	6.2	0.8	2.6	0.0	0.2	-	2.9	-	1.8	-	22.5
CYP706E12		0.4	99.6	0.0	9.4	-	4.0	-	-	0.0	3.7	-	3.4	0.0
CYP81B113		1.1	98.9	0.1	9.3	-	3.9	-	-	0.0	3.7	-	3.2	0.2
CYP71AV19		0.3	99.7	0.0	9.6	-	3.9	-	-	0.0	3.7	-	3.3	0.0
CYP82D180	A mixture of	1.3	98.7	0.1	9.4	-	4.0	-	-	0.1	3.7	-	3.3	0.2
CYP76S36	PA extract	0.5	99.5	0.0	9.3	-	4.0	-	-	0.0	3.7	-	3.5	0.1
CYP71AT158	in *N*-oxide form	0.5	99.5	0.0	9.3	-	4.0	-	-	0.0	3.7	-	3.5	0.1
CYP706C72		0.6	99.4	0.0	9.3	-	4.1	-	-	0.0	3.7	-	3.5	0.1
CYP82Q5		0.3	99.7	0.0	9.3	-	3.9	-	-	0.0	3.7	-	3.4	0.0
NC		0.9	99.1	0.1	9.4	-	3.9	-	-	0.0	3.7	-	3.3	0.1

The contents of PAs are shown in their relative abundances (%). Sum FB: the sum of all PA free bases. Sum Nox: the sum of all PA *N*-oxides. Sn: senecionine. Ir: integerrimine. Sv: senecivernine. Rt: retrorsine. Us: usuramine. Sp: seneciphylline. St: spartioidine. Rd: riddelliine. Jb: jacobine. Jl: jacoline. Jn: jaconine. Jz: jacozine. Er: erucifoline. Un: unkown. Ot: otosenine.

[a]NC: negative control using empty vector without CYP insertion.

[b]- percentages of PAs lower than 0.05%.

0.1	-	-	-	0.6	0.5	93.7	-	1.4	-	0.4	-	2.1	-	0.7	0.5	-
0.1	-	-	-	0.6	1.4	93.0	-	1.3	-	0.4	-	2.0	-	0.6	0.5	-
0.1	-	-	-	0.6	0.3	94.1	-	1.4	-	0.4	-	2.0	-	0.7	0.5	-
0.1	-	-	-	0.6	1.9	92.4	-	1.4	-	0.4	-	2.1	-	0.7	0.5	-
0.1	-	-	-	0.6	0.4	93.9	-	1.4	-	0.4	-	2.0	-	0.7	0.5	-
0.1	-	-	-	0.6	0.4	93.9	-	1.4	-	0.4	-	1.9	-	0.7	0.5	-
0.1	-	-	-	0.6	0.5	93.8	-	1.5	-	0.4	-	2.0	-	0.6	0.5	-
0.2	-	-	-	0.6	0.3	93.9	-	1.4	-	0.4	-	2.0	-	0.7	0.5	-
0.1	-	-	-	0.6	1.0	93.3	-	1.4	-	0.4	-	2.0	-	0.6	0.5	-
-	-	-	-	-	2.3	-	-	-	-	-	10.1	-	87.2	0.2	-	-
-	-	-	-	-	2.3	-	-	-	-	-	9.8	-	87.6	0.2	-	-
-	-	-	-	-	2.3	-	-	-	-	-	9.7	-	87.6	0.2	-	-
-	-	-	-	-	2.3	-	-	-	-	-	9.7	-	87.6	0.2	-	-
-	-	-	-	-	2.4	-	-	-	-	-	10.4	-	86.9	0.2	-	-
-	-	-	-	-	2.3	-	-	-	-	-	9.8	-	87.5	0.1	-	-
-	-	-	-	-	2.4	-	-	-	-	-	9.9	-	87.4	0.2	-	-
-	-	-	-	-	2.4	-	-	-	-	-	10.2	-	87.0	0.2	-	-
-	-	-	-	-	2.4	-	-	-	-	-	9.9	-	87.4	0.1	-	-
-	-	-	-	0.1	0.0	5.2	-	0.9	-	0.1	0.0	15.5	0.5	77.2	0.3	-
-	-	-	-	0.0	0.1	6.7	-	0.9	-	0.1	0.3	15.2	1.5	74.7	0.4	-
-	-	-	-	0.1	0.0	5.7	-	0.9	-	0.1	0.1	15.8	0.3	76.3	0.3	-
-	-	-	-	0.1	0.1	5.2	-	0.8	-	0.1	0.3	15.4	1.8	75.3	0.3	-
-	-	-	-	0.0	0.0	5.2	-	0.9	-	0.1	0.1	15.2	0.6	77.1	0.4	-
-	-	-	-	0.0	0.1	5.3	-	1.0	-	0.1	0.1	15.9	0.5	76.2	0.3	-
-	-	-	-	0.1	0.0	5.2	-	1.0	-	0.1	0.1	15.3	0.5	76.9	0.3	-
-	-	-	-	0.0	0.0	5.4	-	0.9	-	0.1	0.2	15.0	0.4	77.2	0.3	-
-	-	-	-	0.1	0.0	5.3	-	1.0	-	0.1	0.2	15.3	1.1	76.1	0.3	-
-	0.6	-	2.9	-	17.4	-	6.8	-	0.2	-	3.0	-	32.2	0.1	-	0.2
-	0.6	-	2.9	-	17.3	-	6.9	-	0.2	-	3.0	-	32.5	0.1	-	0.3
-	0.6	-	2.8	-	17.1	-	6.9	-	0.2	-	2.9	-	32.1	0.1	-	0.2
-	0.6	-	2.9	-	17.4	-	6.9	-	0.2	-	2.8	-	32.0	0.1	-	0.2
-	0.6	-	2.9	-	17.1	-	6.9	-	0.2	-	2.9	-	31.8	0.1	-	0.2
-	0.7	-	2.9	-	17.2	-	6.9	-	0.2	-	2.8	-	31.4	0.1	-	0.2
-	0.6	-	2.8	-	17.2	-	6.9	-	0.2	-	2.8	-	31.4	0.1	-	0.2
-	0.6	-	2.8	-	17.4	-	6.8	-	0.2	-	2.9	-	31.6	0.1	-	0.2
-	0.6	-	2.8	-	17.4	-	7.0	-	0.2	-	3.0	-	31.7	0.1	-	0.2
26.1	-	0.0	-	3.5	0.1	18.7	0.0	8.2	-	0.7	-	2.2	0.1	19.0	0.6	0.2
26.4	-	0.0	-	3.7	0.2	18.0	0.1	8.3	-	0.7	-	2.3	0.3	18.9	0.6	0.2
26.5	-	0.0	-	3.7	0.0	19.2	0.0	8.2	-	0.7	-	1.2	0.0	19.0	0.6	0.2
26.1	-	0.0	-	3.5	0.3	18.3	0.1	8.4	-	0.7	-	2.1	0.3	18.7	0.6	0.2
26.2	-	0.0	-	3.6	0.1	18.7	0.0	8.3	-	0.7	-	2.2	0.1	18.7	0.6	0.2
25.9	-	0.0	-	3.6	0.1	18.6	0.0	8.3	-	0.8	-	2.2	0.1	18.9	0.5	0.2
25.9	-	0.0	-	3.6	0.1	18.5	0.0	8.4	-	0.7	-	2.2	0.1	18.9	0.5	0.2
26.1	-	0.0	-	3.6	0.1	18.5	0.0	8.2	-	0.8	-	2.2	0.1	19.3	0.6	0.2
25.6	-	0.0	-	3.6	0.2	18.6	0.1	8.5	-	0.7	-	2.2	0.2	19.0	0.6	0.2

as the system to test functions of CYP candidates. None of the eight CYPs were found to possess catalytic activities on PA modifications with the selected PA substrates. Most likely the negative results are due to the fact that the CYPs tested in this study were not the enzymes involved in PA biosynthesis. The study finding candidate genes involved in PA biosynthesis (Chapter 4) should be interpreted with caution as long as CYPs responsible for PA diversification have not been identified. As how PA modification is related to CYPs is unknown, defining candidate genes by setting a threshold for the *P*-value and fold change of differential gene expression can be ambiguous. As such, statistical studies do not show causation and are not always sensitive enough to clarify the contribution of CYP abundances on PA concentrations. More candidate genes need to be tested before we can conclude that the candidate approach described in Chapter 4 is not working. Before biochemical characterization, *in silico* enzyme-substrate interaction modelling analysis of binding site features and substrate selectivity might be a potential approach (Raunio *et al*., 2015; Liu *et al*., 2018) to refine the CYP candidate list in later studies. There is a possibility that the genes underlying PA biosynthesis are not CYP genes but other oxidative enzymes such as flavin-dependent monooxygenases and peroxidases (Burton, 2003), which can also be retrieved from transcriptomes described in Chapter 4.

The negative results might be also explained by the lack of efficient expression of CYP enzymes required for catalyzing PA conversions or by suboptimal reaction conditions of enzyme assays for PA conversion. In addition, the PA substrates isolated (Table 1) may contained impurities which might inhibit activities of CYPs. PA conversions might happen but in such low amounts that derived PAs only account for marginal percentages of the total PA concentration and thus is treated as technical variability in LC-MS detection.

Overall, we tested eight CYPs (CYP706E12, CYP81B113, CYP82D180, CYP71AT158, CYP71AV19, CYP706C72, CYP82Q5, CYP76S36) of *J. vulgaris* for their catalytic activities on isolated PAs and PA extracts but found no PA conversion in all combinations. Based on the current results, most likely is that these CYPs are not enzymes involved in PA biosynthesis, at least not with the tested substrates. More CYP candidates are suggested to be tested in further studies.

Acknowledgements

Yangan Chen thanks the China Scholarship Council (CSC) for financial support. We thank Luis Francisco Salomé-Abarca and Hye Kyong Kim for their help in using GC-MS instrument. We thank Blair Berger for her help in using the LC-MS instrument in Leiden University.

References

Burton SG. 2003. Oxidizing enzymes as biocatalysts. *TRENDS in Biotechnology* 21:543-549.

Chen W, Viljoen AM. 2010. Geraniol - A review of a commercially important fragrance material. *South African Journal of Botany* 76:643-651.

Cheng D, Kirk H, Vrieling K, Mulder PPJ, Klinkhamer PGL. The relationship between structurally different pyrrolizidine alkaloids and western flower thrips resistance in F(2) hybrids of *Jacobaea vulgaris* and *Jacobaea aquatica*. *Journal of Chemical Ecology* 37:1071-1080.

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.

D Gietz, A St Jean, R A Woods, and Schiest RH. 1992. Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Research* 20:1425.

Duan H, Schuler MA. 2006. Heterologous expression and strategies for encapsulation of membrane-localized plant P450s. *Phytochemistry Reviews* 5:507-523.

Hamann T, Møller BL. 2007. Improved cloning and expression of cytochrome P450s and cytochrome P450 reductase in yeast. *Protein Expression And Purification* 56:121-127.

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.

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

Leiss KA, Choi YH, Abdel-Farid IB, Verpoorte R, Klinkhamer PGL. 2009. NMR metabolomics of thrips (*Frankliniella occidentalis*) resistance in *Senecio* hybrids. *Journal of Chemical Ecology* 35:219-229.

Liu Q, Beyraghdar Kashkooli A, Manzano D, Pateraki I, Richard L, Kolkman P, Lucas MF, Guallar V, de Vos RCH, Franssen MCR, van der Krol A, Bouwmeester H. 2018. Kauniolide synthase is a P450 with unusual hydroxylation and cyclization-elimination activity. *Nature Communications* 9:4657.

Liu X, Klinkhamer PGL, Vrieling K. 2017. The effect of structurally related metabolites on insect herbivores: A case study on pyrrolizidine alkaloids and western flower thrips. *Phytochemistry* 138:93-103.

Liu Z, Tavares R, Forsythe ES, André F, Lugan R, Jonasson G, Boutet-Mercey S, Tohge T, Beilstein MA, Werck-Reichhart D, Renault H. 2016 Evolutionary interplay between sister cytochrome P450 genes shapes plasticity in plant metabolism. *Nature Communications* 7:13026.

Macel M, Bruinsma M, Dijkstra SM, Ooijendijk T, Niemeyer HM, Klinkhamer PGL. 2005. Differences in effects of pyrrolizidine alkaloids on five generalist insect herbivore species *Journal of Chemical Ecology* 31:1493-1508.

Miettinen K, Dong L, Navrot N, Schneider T, Burlat V, Pollier J, Woittiez L, van der Krol S, Lugan R, Ilc T, Verpoorte R, Oksman-Caldentey KM, Martinoia E, Bouwmeester H, Goossens A, Memelink J, Werck-Reichhart D. 2014. The seco-iridoid pathway from Catharanthus roseus. *Nature Communications* 5: 3606.

Nour-Eldin HH, Hansen BG, Nørholm MHH, Jensen JK, Halkier BA. 2006 Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. *Nucleic Acids Research* 34:e122.

Nuringtyas TR, Verpoorte R, Klinkhamer PG, van Oers MM, Leiss KA. 2014. Toxicity of pyrrolizidine alkaloids to *Spodoptera exigua* using insect cell lines and injection bioassays. *Journal of Chemical Ecology* 40:609-616.

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.

Pompon D, Louerat B, Bronine A, Urban P. 1996. Yeast expression of animal and plant P450s in optimized redox environments. *Methods in Enzymology* 272:51-64.

Raunio H, Kuusisto M, Juvonen RO, Pentikäinen OT. 2015. Modeling of interactions between xenobiotics and cytochrome P450 (CYP) enzymes. *Frontiers in Pharmacology* 6:123.

Urban P, Cullin C, Pompon D. 1990. Maximizing the expression of mammalian cytochrome P-450 monooxygenase activities in yeast cells. *Biochimie* 72:463-472.

Wei X, Vrieling K, Mulder PPJ, Klinkhamer PGL. 2015. Testing the generalist-specialist dilemma: the role of pyrrolizidine alkaloids in resistance to invertebrate herbivores in *Jacobaea* species. *Journal of Chemical Ecology* 41:159-167.

Table S1. Primers used for amplification of CYP genes

	Length of coding region in nt	Normal primer sets	Uracil-containing primer sets
CYP706E12	1551	F: 5'-ATGATATCAAAGCTCAACCATAAGTTTC-3'	F: 5'-**GGATTAAUA**ATGATATCAAAGCTCAACCATAAGTTTCTTG-3'
		R: 5'-TCACATTCATAGAGGGTTGCATC-3'	R: 5'-**GGGTTAAU**TTAACATTCATAGAGGGTTGCATCAG-3'
CYP81B113	1524	F: 5'-ATGGAAACTTTTTATGTGATTCTCTCCATC-3'	F: 5'-**GGATTAAUA**ATGGAAACTTTTTATGTGATTCTCTCCATC-3'
		R: 5'-TCACAACTGAGATAACAATTCCATCATC-3'	R: 5'-**GGGTTAAU**TTACAACTGAGATAACAATTCCATCATC-3'
CYP71AT158	1506	F: 5'-ATGACATTTACATTTTTACTTCTTCTTCTCTCTTTC-3'	F: 5'-**GGATTAAUA**ATGACATTTACATTTTTACTTCTTCTTCTCTCTTTC-3'
		R: 5'-TCAGTACAAATAAGCTAAAAGGCAAAGCTC-3'	R: 5'-**GGGTTAAU**TTAGTACAAATAAGCTAAAAGGCAAAGCTC-3'
CYP82D180	1602	F: 5'-ATGAATATCCAAAGAAAAATTGATCTTC-3'	F: 5'-**GGATTAAUA**ATGAATATCCAAAGAAAAATTGATCTTCTAATGG-3'
		R: 5'-TCACATCGTAGTAGAAACCATTCG-3'	R: 5'-**GGGTTAAU**TTACATCGTAGTAGAAACCATTCGCG-3'
CYP71AV19	1503	F: 5'-ATGGAGCTTCAATCTCCTTTTTCTCTC-3'	F: 5'-**GGATTAAUA**ATGGAGCTTCAATCTCCTTTTTCTCTC-3'
		R: 5'-TTATTGTTGAGGCTTGTAATGGCTTGG-3'	R: 5'-**GGGTTAAU**TTATTGTTGAGGCTTGTAATGGCTTGG-3'
CYP706C72	1575	F: 5'-ATGTCGAACCTAGTCTTCACAGCC-3'	F: 5'-**GGATTAAUA**ATGTCGAACCTAGTCTTCACAGCC-3'
		R: 5'-TCACTCAAACAACGTTGCGTGT-3'	R: 5'-**GGGTTAAU**TTACTCAAACAACGTTGCGTGTGTC-3'
CYP82Q5	1584	F: 5'-ATGAATTTCTATCTCTCATCTTCAACATTTG-3'	F: 5'-**GGATTAAUA**ATGATGAATTTCTATCTCTCATCTTCAACATTTG-3'
		R: 5'-TCAAGCAACATGATACATATTAGAGGATAAAC-3'	R: 5'-**GGGTTAAU**TTAAGCAACATGATACATATTAGAGGATAAACG-3'
CYP76S36	1494	F: 5'-ATGGATAACACCACTTTCCTTCTCTTAT-3'	F: 5'-G**GATTAAUA**ATGGATAACACCACTTTCCTTCTCTTATTTTCG-3'
		R: 5'-TCAAAGACGGATTGGAATAGCCTT-3'	R: 5'-**GGGTTAAU**TTAAAGACGGATTGGAATAGCCTTGAG-3'
CYP76B6	1482	F: 5'-ATGGATTACCTTACCATAATATTAACTTTACT-3'	F: 5'-**GGATTAAUA**ATGGATTACCTTACCATAATATTAACTTTACT-3'
		R: 5'-TTAAAGGGTGCTTGGTACAGC-3'	R: 5'-**GGGTTAAU**TTAAAGGGTGCTTGGTACAGC-3'

The USER compatible linkers are marked in bold. All linkers at the 5' end of forward primers are identical, so are all linkers of reverse primers.

Chapter 6

Summary and conclusions

Summary and conclusions

Plants produce an astonishing variety of secondary metabolites (SMs) which are thought to play vital roles in the fitness of plants through ecological interactions (Wink, 2003; Moore *et al*., 2014). SMs enable plants to deal with antagonists (e.g. herbivores, pathogens, neighboring plants) and mutualists (e.g. pollinators, predators/parasitoids against herbivores, mycorrhizal fungi, rhizobium and other beneficial bacteria) as well as abiotic factors (e.g. UV light, drought, frost) (Kessler and Halitschke, 2007). The most characteristic features of SMs are their striking chemical diversity and inter- or intraspecific variation. So far, an estimated 200,000 SMs have been isolated and identified from plants (Kessler and Kalske, 2018). SMs are assigned to different compound classes such as glucosinolates, alkaloids, terpernoids and flavonoids. Within each chemical class, both qualitative and quantitative SM diversity are common in plants. For instance, in *Arabidopsis thaliana* 34 different glucosinolates have been identified, showing 20-fold difference in total concentration in leaves of different ecotypes (Kliebenstein *et al.*, 2001). As yet, it is poorly understood how SM diversity comes about and why it is maintained in nature (Moore *et al.*, 2014).

In this thesis we focus on SMs involved in plant protection against herbivores. Aiming at a better understanding of mechanisms behind SM diversity, researchers have put much effort in studying the distribution patterns of SMs under particular phylogenetic frameworks (Wink, 2003; Wink, 2008; Pelser *et al.*, 2005; Mint Evolutionary Genomics Consortium, 2018). At the level of a particular class of compounds (e.g. quinolizidine alkaloids, glucosinolates) there is a strong phylogenetic signal although some exceptions exist. This suggests that once the ability to produce the basic structures of these classes of compounds has evolved it is conserved through evolutionary time and not often convergent evolution takes place. In contrast the distributions of different chemical modifications of these structures on phylogenetic trees are often random. That is to say particular SMs within a class lack phylogenetic signals and phylogenetic distances are not correlated with differences in SM bouquets (Pelser *et al*., 2005; Mint Evolutionary Genomics Consortium, 2018; Chapter 2). This suggests that such modifications are not conserved and can rapidly appear or disappear on an evolutionary time scale. An evolutionary explanation would be that the basic structures that helps to protect plants against a broad set of herbivores comes about once and is conserved and that further evolutionary fine-tuning is dependent on the selective effects of the specific set of generalist and specialist herbivores each plant species is confronted with at a particular moment in evolutionary time. At the molecular level, the fact that new basic structures do not rapidly evolve in many plant families suggests that the enzymatic changes needed to create them are relatively complex. The further modification of newly evolved basic structures involves relative simple chemical reactions such as oxidation that are carried out by more general enzymes that are already present in the plant. The multiple functions of such an enzyme are subsequently reduced to single function by duplication of the underlying gene allowing optimizing the catalyzation for the new function (Hughes, 1994; DePristo, 2007). Examples

are the large gene families of cytochrome P450s (CYPs) of which the members show large similarity and that rapidly evolve (Bak *et al*., 2006; Frey *et al*., 2009; Chapter 3). Apparently this enables fast evolution, loss of phylogentic signal and rapid evolutionary fine-tuning of their efficacy. The process of evolutionary tinkering is most likely in addition speeded up by genetic changes in the expression of CYPs. The investigation of the diversity of both SM profiles and the CYP family in plants may allow insights into the evolution of SM pathways that coordinate the respective enzymes.

The pyrrolizidine alkaloids (PAs) of *Jacobaea* were used as the model system in this thesis. PAs are a class of SMs with typical great diversity that are constitutively formed in plants containing them and are thought to mediate the interactions between plants and herbivores (Hartmann, 1999). The *Jacobaea* species (26 species and formerly a part of *Senecio* species, Asteraceae) all produce PAs (Pelser *et al*., 2005; Langel *et al*., 2011) but the composition and concentration are often species-specific (Soldaat *et al*., 1996; Hartmann and Dierich, 1998; Langel *et al*., 2011). Nevertheless, individual PAs seem to have random distributions on the phylogenetic tree lacking phylogenetic signals (Pelser *et al*., 2005). Two hypotheses can be used to explain the random occurrences of PAs in *Jacobaea* species, i.e. (i) convergent evolution where the ability of plant species to produce particular PAs evolved several times, (ii) differential gene regulation of the PA pathway where all the plant species possess the machinery to produce those PAs but do not all express them. In order to figure out which hypothesis is correct, the occurrence and expression of genes underlying SM pathways need to be investigated.

So far, more than 400 PAs have been found (Chou and Fu, 2006), of which more than 100 PAs are macrocyclic senecionine-type (Hartmann and Witte, 1995; Langel *et al*., 2011). Most of our current knowledge of PA biosynthetic diversification has come from the studies of the 12-membered macrocyclic senecionine-type PAs in the Senecioneae (Langel *et al*., 2011). The primary product senecionine *N*-oxide synthesized in roots (Hartmann and Toppel, 1987; Hartmann *et al*., 1988; Hartmann *et al*., 1989) undergoes structural transformations in a position-specific and stereoselective manner resulting in the rearrangement of the skeletal structure and oxidative modifications thereof in shoots (Hartmann and Dierich, 1998; Pelser *et al*., 2005). The enzymes responsible for these processes have not been identified. CYPs catalyze a wide range of regiospecific, stereospecific and irreversible oxidation steps in plant SM biosynthesis (Renault *et al*., 2014), thus playing an important role in the evolution of chemical diversity. Given the common oxygenated site-specific modifications of PAs derived from the primary PA senecionine *N*-oxide, CYPs are good candidates for involvement in PA biosynthesis.

In this thesis the following questions were proposed: what are the distribution patterns of PAs in *Jacobaea* species and are these patterns related to their phylogenetic relationships? How does this PA diversity come about? Are CYPs involved in PA biosynthesis? Following these questions, the experimental chapters of this thesis can be divided into two sections: (i)

the evolution of PA diversity among and within *Jacobaea* species (Chapter 2), (ii) a candidate gene approach targeting CYPs for involvement in PA diversity (Chapter 3-5).

1. The evolution of PA diversity among and within *Jacobaea* species

In chapter 2, aiming to understand the mechanism behind chemical diversity, PA patterns of *Jacobaea* species were studied in a phylogenetic context. The presences and concentrations of 80 PAs in eight to ten week old leaves of 17 *Jacobaea* species including different individuals and populations grown in a controlled chamber were analyzed by LC-MS/MS. A great diversity of PA profiles was observed with the numbers of PAs ranging from 21 to 59 per species, and with the total PA concentrations ranging from 32.9 to 3835.7 µg/g dry weight. These profiles were largely confirmed to be species-specific both qualitatively and quantitatively, which is in line with previous findings (Soldaat *et al*., 1996; Hartmann and Dierich, 1998; Langel *et al*., 2011). Jacobine-like PAs, senecionine-like PAs and otosenine-like PAs contributed more to the classification of different *Jacobaea* species than other PA structural groups. PA patterns from different populations within some *Jacobaea* species (e.g. *J. alpina* and *J. paludosa*) differed from each other even surpassing differences between species. The phylogeny of 17 *Jacobaea* species were reconstructed using 11 chloroplast regions and three nuclear DNA genes to trace the evolution of PA diversity at species level. With ancestral state reconstruction of the occurrences of individual PAs, complex evolutionary patterns for almost all PAs were found showing that the occurrence of virtually all PAs evolved several times, which is in agreement with the findings of Pelser *et al*. (2005). Two different measures, Blomberg's *K* (Blomberg, 2003) and Pagel's λ (Pagel, 1999), were used to evaluate the correlations between quantitative PA traits and phylogenetic relationships, and significant phylogenetic signals for nine out of 80 PAs only under λ statistics were found. Given the common intraspecific PA diversity found in *Jacobaea* species (this thesis; Witte *et al*., 1992; Macel *et al*., 2004), it can be assumed that this high PA diversity is due to the regulation of PA biosynthesis genes in plants as a life strategy to meet their different biological needs rather than the evolutionary gains and losses of particular PA biosynthesis genes.

2. A candidate gene approach targeting CYPs for involvement in PA diversity

Testing our hypothesis that PA diversity in *Jacobaea* species is the result of the regulation of PA biosynthesis genes requires the genes underlying PA biosynthesis. CYPs have formed a large gene family in plants and have often been found to be involved in SM pathways as oxidative enzymes. As diversification of PAs is often through site-specific oxidation (Hartmann and Dierich, 1998; Pelser *et al*., 2005), CYPs are likely to be involved in PA biosynthesis. Yet, no public CYP database of *Jacobaea* species is available at the start of the studies described in this thesis. Therefore, a systematical study of CYPs was performed with regard to their diversity and evolution in *J. vulgaris* and *J. aquatica* as described in Chapter 3. In total, 221 (classified into eight clans and 38 families) and 157 (classified into eight clans and 35 families) full-length CYPs were retrieved from *de novo* assembled transcriptomes of *J.*

vulgaris and *J. aquatica*, respectively. Based on KEGG annotation, the CYPs assigned as being SM metabolic pathway enzymes were all from the CYP71 clan but no CYPs were assigned as being involved in alkaloid pathways. This does not necessarily mean that they are not involved in alkaloid biosynthesis since the current KEGG database does not contain information about PA biosynthetic enzymes. Phylogenetic analyses of the six largest CYP families (CYP71, CYP76, CYP706, CYP82, CYP93 and CYP72) from the two *Jacobaea* species, two other members of the Asteraceae, *Helianthus annuus* and *Lactuca sativa* and the outgroup of *Arabidopsis thaliana* were performed. The phylogenetic trees showed strong lineage-specific expansion of CYPs, suggesting that the evolution of CYPs has been very fast even within the Asteraceae family. Only CYPs of the closely related species *J. vulgaris* and *J. aquatica* were found often in pairs, confirming a close relationship in evolutionary history of these two species. The studies described in Chapter 3 provide a CYP database for future exploration of their functions, including possible involvement in PA biosynthesis and PA diversity.

Chapter 4 describes an attempt to identify candidate CYPs that may be involved in PA biosynthesis based on the association between metabolic and transcriptomic profiles between different *Jacobaea* samples grown under controlled conditions. *J. aquatica* and four groups of F_2 hybrids of a cross between *J. vulgaris* and *J. aquatica* with PA contrasts were used for constitutive PA contrasts, especially in jacobine-like and erucifoline-like PAs. A methyl jasmonate treatment on tissue culture plants of *J. vulgaris* were performed to induce increase of erucifoline-like PAs. In total, 44 PAs were detected by LC-MS/MS and PA profiles of different *Jacobaea* samples were compared separately in the constitutive and induced groups by summing up concentrations of PAs containing the same site-specific oxidative modifications which might be catalyzed by a CYP enzyme: 15,20-epoxidation, 12,13-epoxidation or 19-hydroxylation, 18-hydroxylation, 13,19-dehydrogenation and 8-oxidation. RNA sequencing was performed separately for the constitutive and induced groups to analyze the expression of CYPs which may be involved in PA oxidative conversions. In total, 33 and 27 CYP candidate genes were sieved out for the constitutive and induced PA conversions, respectively. Most of these candidate enzymes were from the CYP71 clan without known functions. There were 11 CYP subfamilies found both in the constitutive and induced groups, where three subfamilies (CYP72A, CYP706E, CYP82Q) may be responsible for the formation of erucifoline-like PAs which contain both 12,13-epoxidation and 19-hydroxylation.

Chapter 5 describes functional tests of eight CYP candidates for involvement in the PA biosynthesis pathway using heterologous expression in yeast and *in vitro* enzyme assays using microsomal membrane preparations presumably containing the expressed CYPs. None of the eight CYP enzyme preparations showed PA conversion with the selected PA substrates (senecionine/integerrimine, seneciphylline, jacobine, erucifoline or a PA mixture as well as the respective *N*-oxides) based on *in vitro* enzyme assays using extracted microsomal membranes. The reasons of the negative results might be lack of efficient expression, and/or requirement for more optimal reaction conditions or better detection method. More likely is

that the tested eight CYPs are not enzymes responsible for PA biosynthesis, or at least do not act on the tested substrates.

3. Discussion and conclusions

Jacobaea species have been frequently used to study their PA diversity (Vrieling *et al*., 1993; Hartmann and Dierich, 1998; Macel *et al*., 2004; Cheng *et al*., 2011). Compared with previous studies, we studied both qualitative and quantitative PA variation at more levels including among species, among populations and among individuals at the same time. Our results revealed stronger evidence that PA patterns of *Jacobaea* species are indeed species-specific both in concentrations and compositions, though for some species PA profiles also differed largely between populations. It is the first study showing that the distribution pattern of concentrations among and within *Jacobaea* species was highly similar to that of compositions, implying the same or closely related mechanisms behind quantitative and qualitative variation of PAs.

The occurrence of individual PAs was studied at species level and limited phylogenetic signals were found. A reasonable assumption is that all *Jacobaea* species possess the machinery for PA production and that differences in PA occurrence are due to differential gene regulation. Based on the fact that PA diversity is largely created by oxidation creations, CYPs, a major class of oxidative enzymes in plants, were chosen as an entry for identification of PA biosynthesis genes. Based on current scientific literature, the curated CYP database developed in this thesis is the first in *Jacobaea* species. A gene-to-metabolite approach was used to identify CYP candidates possibly involved in PA biosynthesis and the functions of eight candidate CYP enzymes was checked but this did not lead to any indication for their involvement in PA biosynthesis. The majority of candidate gene identified in Chapter 4 still needs to be tested. There is a possibility that the enzymes underlying PA biosynthesis are not CYPs but other oxidative enzymes such as flavin-dependent monooxygenases or peroxidases (Burton, 2003), for which gene-to-metabolite correlations can also be retrieved from the data described in Chapter 4.

References

Blomberg SP, Garland T, Ives AR. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57:717-745.

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

Burton SG. 2003. Oxidizing enzymes as biocatalysts. *Trends in Biotechnology* 21:543-549.

Cheng D, Hirk 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.

Chou MW, Fu PP. 2006. Formation of DHP-derived DNA adducts in vivo from dietary supplements and Chinese herbal plant extracts containing carcinogenic pyrrolizidine alkaloids. *National Center for Toxicological Research* 22:321-327.

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. Site of synthesis, translocation and accumulation of pyrrolizidine alkaloid *N*-oxides in *Senecio vulgaris* L. *Planta* 177:98-107.

Hartmann T, Sander H, Adolph R, and Toppel G. 1988. Metabolic links between the biosynthesis of pyrrolizidine alkaloids and polyamines in root cultures of *Senecio vulgaris*. *Planta* 175:82-90.

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.

Hartmann T, Witte L. 1995. Chemistry, biology and chemoecology of the pyrrolizidine alkaloids. In Pelletier SW (ed) Alkaloids: chemical and biological perspective. Oxford: Pergamon Vol. 9, pp.155-233.

Hartmann T. 1999. Chemical ecology of pyrrolizidine alkaloids. *Planta* 207:483-495.

Kessler A, Halischke R. 2007. Specificity and complexity: the impact of herbivore-induced plant response on arthropod community structure. *Current Opinion in Plant Biology* 10:409-414.

Kessler A, Kalske A. 2018. Plant secondary metabolite diversity and species interactions. *Annual Review 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.

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

Liu Q, Beyraghdar Kashkooli A, Manzano D, Pateraki I, Richard L, Kolkman P, Lucas MF, Guallar V, de Vos RCH, Franssen MCR, van der Krol A, Bouwmeester H. 2018. Kauniolide synthase is a P450 with unusual hydroxylation and cyclization-elimination activity. *Nature Communications* 9:4657.

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

Mint Evolutionary Genomics Consortium. 2018. Phylogenomic mining of the mints reveals multiple mechanisms contributing to the evolution of chemical diversity in Lamiaceae. *Molecluar Plant* 11:1084-1096.

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.

Pagel M. 1999. Inferring the historical patterns of biological evolution. *Nature* 401:877-884.

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.

Raunio H, Kuusisto M, Juvonen RO, Pentikäinen OT. 2015. Modeling of interactions between xenobiotics and cytochrome P450 (CYP) enzymes. *Frontiers in Pharmacology* 6:123.

Renault H, Bassard JE, Hamberger B, Werck-Reichhart D. 2014. Cytochrome P450-mediated metabolic engineering: current progress and future challenges. *Current Opinion in Plant Biology* 19:27-34.

Soldaat LL, Boutin JP, Derridj S. 1996. Species-specific composition of free amino acids on the leaf surface of four *Senecio* species. *Journal of Chemical Ecology* 22: 1-12.

Vrieling K, de Vos H, van Wijk CAM. 1993. Genetic analysis of the concentrations of pyrrolizidine alkaloids in *Senecio jacobaea*. *Phytochemistry* 32:1141-1144.

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

Wink M. 2008. Plant secondary metabolism: diversity, function and it evolution. *Natural Product Communications* 3:1205-1216.

Witte L, Ernst L, Adam H, Hartmann T. 1992. Chemotypes of two pyrrolizidine alkaloid-containing *Senecio* species. *Phytochemistry* 31:559-565.

Hughes AL. 1994. The evolution of fuctionally novel proteins after gene duplication. *Proceedings of the Royal Society of London, Series B: Biological Sciences* 256:119-124.

DePristo MA. 2007. The subtle benefits of being promiscuous: adaptive evolution potentiated by enzyme promiscuity. *HFSP Journal* 1:94-98.

Bak S, Paquette SM, Morant M, Vinther Moranth A, Saito S, Bjarnholt N, Zagrobelny M, Jorgensen K, Osmani S, *et al*. 2006. Cyanogenic glycosides: a case study for evolution and application of cytochromes P450. *Phytochemistry Reviews* 5:309-329.

Frey M, Schullehner K, Dick R, Fiesselmann A, Gierl A. 2009. Benzoxazinoid biosynthesis, a model for evolution of secondary metabolic pathways in plants. *Phytochemistry* 70:1645-165

Appendix 1. The full list of PAs detected in *Jacobaea* species by LC-MS/MS.

Abbreviation Code	Name	Structural group	standard for (semi)-quantification
Sn (336)[c]	Senecionine	Senecionine	Senecionine
Sn-ox (352)	Senecionine *N*-oxide	Senecionine	Senecionine *N*-oxide
Ir (336)	Integerrimine	Senecionine	Integerrimine
Ir-ox (352)	Integerrimine *N*-oxide	Senecionine	Integerrimine *N*-oxide
Sv (336)	Senecivernine	Senecionine	Senecivernine
Sv-ox (352)	Senecivernine *N*-oxide	Senecionine	Senecivernine *N*-oxide
Rt (352)	Retrorsine	Senecionine	Retrorsine
Rt-ox (368)	Retrorsine *N*-oxide	Senecionine	Retrorsine *N*-oxide
Us (352)	Usaramine	Senecionine	Usaramine
Us-ox (368)	Usaramine *N*-oxide	Senecionine	Usaramine *N*-oxide
Ef (352)	Eruciflorine*	Senecionine	Retrorsine
Ef-ox (368)	Eruciflorine *N*-oxide*	Senecionine	Retrorsine *N*-oxide
Sp (334)	Seneciphylline	Senecionine	Seneciphylline
Sp-ox (350)	Seneciphylline *N*-oxide	Senecionine	Seneciphylline *N*-oxide
St (334)	Spartioidine	Senecionine	Seneciphylline
St-ox (350)	Spartioidine *N*-oxide	Senecionine	Seneciphylline *N*-oxide
Rd (350)	Riddelliine	Senecionine	Riddelliine
Rd-ox (366)	Riddelliine *N*-oxide	Senecionine	Riddelliine *N*-oxide
DHEf (350)	Dehydroeruciflorine*	Senecionine	Riddelliine
DHEf-ox (366)	Dehydroeruciflorine *N*-oxide*	Senecionine	Riddelliine *N*-oxide
AcSp (376)	Acetylseneciphylline	Senecionine	Acetylseneciphylline
AcSp-ox (392)	Acetylseneciphylline *N*-oxide	Senecionine	Seneciphylline *N*-oxide
AcSt (376)	Acetylspartioidine	Senecionine	Acetylseneciphylline
AcSt-ox (392)	Acetylspartioidine *N*-oxide	Senecionine	Seneciphylline *N*-oxide
Jz (350)	Jacozine	Jacobine	Erucifoline
Jz-ox (366)	Jacozine *N*-oxide	Jacobine	Erucifoline *N*-oxide
Jb (352)	Jacobine	Jacobine	Jacobine
Jb-ox (368)	Jacobine *N*-oxide	Jacobine	Jacobine *N*-oxide
DHJl (368)	Dehydrojacoline	Jacobine	Jacoline
HOJb (368)	Hydroxyjacobine	Jacobine	Jacobine
Jl (370)	Jacoline	Jacobine	Jacoline
Jl-ox (386)	Jacoline *N*-oxide	Jacobine	Jacoline *N*-oxide
DHJn (386)	Dehydrojaconine	Jacobine	Jaconine
Jn (388)	Jaconine	Jacobine	Jaconine
Jn-ox (404)	Jaconine *N*-oxide	Jacobine	Jacoline *N*-oxide
HOJn (404)	Hydroxyjaconine	Jacobine	Jaconine
Er (350)	Erucifoline	Erucifoline	Erucifoline
Er-ox (366)	Erucifoline N-oxide	Erucifoline	Erucifoline N-oxide
Sb (366)	Senecicannabine*	Erucifoline	Erucifoline
Sb-ox (382)	Senecicannabine N-oxide*	Erucifoline	Erucifoline N-oxide
Af (366)	Adonifoline	Erucifoline	Erucifoline

Precursor mass (m/z)	Product ion 1 (m/z)	CE[a] (eV)	Product ion 2 (m/z)	CE (eV)	RT[b] (min)
336.2	94	40	120	30	10.32
352.2	94	40	120	30	6.70
336.2	94	40	120	30	10.03
352.2	94	40	120	30	6.51
336.2	94	40	120	30	10.49
352.2	94	40	120	30	6.75
352.2	94	40	120	30	8.57
368.2	94	40	120	30	5.38
352.2	94	40	120	30	8.29
368.2	94	40	120	30	5.21
352.2	94	40	120	30	6.60
368.2	94	40	120	30	3.80
334.2	120	30	138	30	9.31
350.2	94	40	120	30	5.82
334.2	120	30	138	30	9.06
350.2	94	40	120	30	5.71
350.2	94	40	120	30	7.79
366.2	94	40	118	30	4.46
350.2	94	40	120	30	5.92
366.2	94	40	118	30	3.00
376.2	120	30	138	30	11.60
392.2	94	40	118	30	9.14
376.2	120	30	138	30	11.50
392.2	94	40	118	30	9.03
350.2	94	40	120	30	6.86
366.2	94	40	118	30	3.36
352.2	120	30	155	30	7.71
368.2	120	30	296	25	4.35
368.2	94	40	120	30	4.65
368.2	94	40	296	25	6.28
370.2	94	40	138	30	5.42
386.2	94	40	120	30	2.66
386.2	94	40	120	30	7.70
388.2	94	40	120	30	8.85
404.2	94	40	120	35	4.98
404.2	94	40	120	35	7.43
350.2	94	40	120	30	7.31
366.2	94	40	118	30	3.23
366.2	94	40	120	30	5.98
382.2	94	40	122	30	2.43
366.2	94	40	338	25	5.59

Abbreviation Code	Name	Structural group	standard for (semi)-quantification
Af-ox (382)	Adonifoline *N*-oxide	Erucifoline	Erucifoline *N*-oxide
iso-Af 366 (5.15)	Iso-adonifoline*	Erucifoline	Erucifoline
NO392 (10.6)	Des-HO-acetylerucifoline *N*-oxide*	Erucifoline	Erucifoline *N*-oxide
AcEr (392)	Acetylerucifoline	Erucifoline	Acetylerucifoline
AcEr-ox (408)	Acetylerucifoline *N*-oxide	Erucifoline	Erucifoline *N*-oxide
NO338 (8.55)	des-HO-platyphylline *N*-oxide isomer*	Platyphylline	Senecionine *N*-oxide
Pt (338)	Platyphylline	Platyphylline	Senecionine
Pt-ox (354)	Platyphylline *N*-oxide	Platyphylline	Senecionine *N*-oxide
FB338 (8.00)	Platiphylline isomer*	Platyphylline	Senecionine
FB338 (9.45)	Platiphylline isomer*	Platyphylline	Senecionine
FB338 (10.10)	Platiphylline isomer*	Platyphylline	Senecionine
FB338 (10.45)	Platiphylline isomer*	Platyphylline	Senecionine
NO354 (5.55)	Platiphylline *N*-oxide isomer*	Platyphylline	Senecionine
DHSk (364)	Dehydrosenkirkine	Senkirkine	Senkirkine
Sk (366)	Senkirkine	Senkirkine	Senkirkine
neo-Sk (366)	Neosenkirkine	Senkirkine	Senkirkine
Ot (382)	Otosenine	Senkirkine	Otosenine
FB382 (3.70)	Otosenine isomer	Senkirkine	Otosenine
On (400)	Onetine	Senkirkine	Otosenine
AcSk (408)	Acetylsenkirkine	Senkirkine	Florosenine
DesDor (418)	Desacetyldoronine	Senkirkine	Florosenine
Fs (424)	Florosenine	Senkirkine	Florosenine
Fd (442)	Floridanine	Senkirkine	Florosenine
Dor (460)	Doronine	Senkirkine	Florosenine
FB350 (7.45)	Unknown FB 350	Unknown	Erucifoline
FB350 (8.15)	Unknown FB 350	Unknown	Erucifoline
NO366 (3.65)	Unknown FB 350 *N*-oxide	Unknown	Erucifoline *N*-oxide
NO366 (3.75)	Unknown FB 350 *N*-oxide	Unknown	Erucifoline *N*-oxide
NO368 (2.50)	Unknown FB 352 *N*-oxide	Unknown	Jacobine *N*-oxide
NO368 (5.05)	Unknown FB 352 *N*-oxide	Unknown	Jacobine *N*-oxide
NO368 (5.25)	Unknown FB 352 *N*-oxide	Unknown	Jacobine *N*-oxide
NO368 (5.35)	Unknown FB 352 *N*-oxide	Unknown	Jacobine *N*-oxide
NO368 (5.45)	Unknown FB 352 *N*-oxide	Unknown	Jacobine *N*-oxide
FB486 (4.00)	Unknown FB 486	Unknown	Senecionine
FB486 (4.20)	Unknown FB 486	Unknown	Senecionine
FB502 (10.30)	Unknown FB 502	Unknown	Senecionine
FB502 (10.50)	Unknown FB 502	Unknown	Senecionine
NO518 (7.40)	Unknown FB 502 N-oxide	Unknown	Senecionine N-oxide
NO518 (7.55)	Unknown FB 502 N-oxide	Unknown	Senecionine N-oxide

[a]CE: collision energy; [b]RT: retention time; [c]Precursor mass of known PAs or retention time of unknown PAs shown in brackets; *tentative assignment

Precursor mass (m/z)	Product ion 1 (m/z)	CE[a] (eV)	Product ion 2 (m/z)	CE (eV)	RT[b] (min)
382.2	94	40	354	25	3.41
366.2	94	40	338	25	5.15
392.2	94	40	120	30	10.60
392.2	94	40	120	30	10.60
408.2	94	40	120	30	6.90
338.2	122	25	140	25	8.55
338.2	122	25	140	25	9.75
354.2	120	35	138	30	6.95
338.2	122	25	140	25	8.00
338.2	122	25	140	25	9.45
338.2	122	25	140	25	10.10
338.2	122	25	140	25	10.45
354.2	120	35	138	30	5.55
364.2	122	30	168	30	6.39
366.2	122	30	168	25	7.14
366.2	122	30	168	25	6.95
382.2	122	30	168	25	4.60
382.2	122	30	168	25	3.71
400.2	122	30	168	30	2.58
408.2	94	40	122	30	10.62
418.2	122	30	168	30	5.69
424.2	122	35	168	30	8.43
442.2	122	30	168	30	6.48
460.2	122	35	168	30	9.36
350.2	120	30	138	30	7.45
350.2	120	30	138	30	8.15
366.2	94	40	118	30	3.65
366.2	94	40	118	30	3.75
368.2	94	40	118	30	2.50
368.2	94	40	118	30	5.05
368.2	94	40	118	30	5.25
368.2	94	40	118	30	5.35
368.2	94	40	118	30	5.45
486.2	94	40	120	35	4.00
486.2	94	40	120	35	4.20
502.2	120	25	334	30	10.30
502.2	120	25	334	30	10.50
518.2	118	30	350	30	7.40
518.2	118	30	350	30	7.55

Nederlandse samenvatting

Planten produceren een verbazingwekkende verscheidenheid aan secundaire metabolieten (SMs), kleine organische moleculen waarvan wordt gedacht dat ze een vitale rol spelen in de ecologische interacties van een plant met z'n omgeving (Wink, 2003; Moore *et al*., 2014). SMs zijn van belang voor de relaties van planten met antagonisten zoals herbivoren, pathogenen en naburige planten en met mutualisten zoals bestuivers, predatoren, parasitoïden van herbivoren, mycorrhiza en nuttige bacteriën. Verder zijn SMs betrokken bij de bescherming tegen abiotische factoren zoals UV-licht, droogte en vorst (Kessler and Halitschke, 2007). De meest karakteristieke kenmerken van SMs zijn hun grote chemische diversiteit en de grote variatie binnen en tussen soorten. Momenteel zijn er circa 200.000 SMs geïdentificeerd uit planten (Kessler en Kalske, 2018). SMs komen voor als klassen van chemische verwante stoffen zoals bv. glucosinolaten, alkaloïden, terpenoïden en flavonoïden. Alle chemische klassen van SMs vertonen in vrijwel alle planten zowel kwalitatieve als kwantitatieve variatie. In *Arabidopsis thaliana* zijn bijvoorbeeld 34 verschillende glucosinolaten geïdentificeerd, met een 20-voudig verschil in totale concentratie in de bladeren tussen verschillende ecotypes (Kliebenstein *et al*., 2001). Vooralsnog is er weinig kennis over hoe SM diversiteit tot stand komt in planten en waarom deze variatie gehandhaafd blijft in de natuur (Moore *et al*., 2014).

Dit proefschrift beschrijft studies aan SMs die betrokken zijn bij de afweer van planten tegen herbivoren. Om een beter inzicht te krijgen in SM diversiteit, hebben verschillende wetenschappers de verspreidingspatronen van bepaalde SM klassen over de takken van fylogenetische bomen van plantenfamilies bestudeerd (Wink, 2003; Wink, 2008; Pelser *et al*., 2005; Mint Evolutionary Genomics Consortium, 2018). Op het niveau van een bepaalde klasse van SMs (bv. quinolizidine alkaloïden, pyrrolizidine alkaloïden of glucosinolaten) zijn er nagenoeg altijd sterke fylogenetische signalen aanwezig. Dit suggereert dat wanneer binnen een fylogenetische tak de basisstructuren van een bepaalde klasse van SMs zijn ontstaan, deze worden behouden en dat convergente evolutie zeldzaam is. Daarentegen is het voorkomen van verschillende chemische modificaties van deze basisstructuren binnen een fylogenetische tak vaak willekeurig. Dat wil zeggen dat bepaalde SMs binnen een klasse geen fylogenetische signalen laten zien en dat de fylogenetische afstanden tussen soorten niet gecorreleerd zijn met verschillen in SM samenstellingen tussen die soorten binnen een bepaalde klasse van SMs (Pelser *et al*., 2005; Mint Evolutionary Genomics Consortium, 2018; Hoofdstuk 2). Dit suggereert dat zulke modificaties binnen een klasse van SMs op een evolutionaire tijdschaal snel kunnen verdwijnen of opnieuw kunnen ontstaan. Een evolutionaire verklaring zou kunnen zijn dat de basisstructuren die helpen planten te beschermen tegen een breed scala van herbivoren op een bepaald moment zijn geëvolueerd en dat verdere evolutionaire afstemming afhankelijk is van de selectieve effecten van de specifieke set van generalistische en gespecialiseerde herbivoren waarmee elke plantensoort wordt geconfronteerd op een bepaald moment op een evolutionaire tijdschaal. Op moleculair niveau suggereert het feit dat nieuwe klassen van SMs in veel plantenfamilies niet snel ontstaan dat de enzymatische aanpassingen

die nodig zijn om de basisstructuren te maken relatief complex zijn. De verdere diversificatie van een nieuw geëvolueerde klasse van SMs omvatten relatief eenvoudige chemische reacties zoals oxidaties en methylaties die door meer algemene enzymen die al in de plant aanwezig zijn kunnen worden uitgevoerd. De meerdere functies van een dergelijk enzym kunnen vervolgens gereduceerd worden tot een enkele functie door duplicatie van het onderliggende gen. Eén kopie kan dan voor de oude functie blijven zorgen terwijl de andere kopie geoptimaliseerd kan worden voor de nieuwe functie (Hughes, 1994; DePristo, 2007). Een voorbeeld is de grote gen familie van de cytochroom P450s (CYPs) waarvan de leden grote overeenkomst vertonen en snel dupliceren en evolueren (Bak *et al*., 2006; Frey *et al*., 2009; Hoofdstuk 3). Blijkbaar leidt een dergelijke snelle evolutie tot een snelle aanpassing van hun functie maar daardoor ook tot een verlies van een fylogenetisch signaal in de resulterende SMs. Het lijkt zeer waarschijnlijk dat het proces van evolutionair knutselen bovendien wordt versneld door genetische veranderingen in de expressie van CYPs. Het onderzoeken van de diversiteit van zowel SM profielen als van de variatie in de CYP familie in planten kan inzicht verschaffen in de evolutie van SM biosynthese routes.

De pyrrolizidine alkaloïden (PAs) van het geslacht *Jacobaea* zijn in dit proefschrift als modelsysteem gebruikt. PAs zijn een klasse van SMs met een grote diversiteit. PAs worden constitutief gevormd in planten en er wordt verondersteld dat ze een rol spelen in de interacties tussen planten en herbivoren (Hartmann, 1999). De soorten binnen het geslacht *Jacobaea* (26 soorten die voorheen deel uitmaakten van het geslacht *Senecio* in de familie Asteraceae) produceren allemaal PAs (Pelser *et al*., 2005; Langel *et al*., 2011) maar de samenstelling en concentratie zijn vaak soort-specifiek (Soldaat *et al*., 1996; Hartmann en Dierich, 1998; Langel *et al*., 2011). Niettemin lijken individuele PAs willekeurig verdeeld te zijn over de fylogenetische boom van het *Jacobaea* geslacht (Pelser *et al*., 2005). Twee hypothesen kunnen het willekeurige voorkomen van de verschillende PAs in *Jacobaea* soorten verklaren. Dit zijn (i) convergente evolutie waarbij het vermogen van plantensoorten om bepaalde PAs te produceren verschillende keren is geëvolueerd, en (ii) differentiële genregulatie van PA biosynthese enzymen, waarbij alle plantensoorten over de enzymen beschikken om alle verschillende PAs te produceren, maar deze niet in alle soorten in gelijke mate tot expressie komen. Om er achter te komen welke hypothese juist is, is het noodzakelijk om de genen die ten grondslag liggen aan de PA biosyntheseroute en hun expressie te onderzoeken.

Tot nu toe zijn er meer dan 400 PAs geïsoleerd (Chou en Fu, 2006), waarvan er meer dan 100 van het senecionine-type zijn (Hartmann en Witte, 1995; Langel *et al*., 2011). Het merendeel van onze huidige kennis van de diversiteit van PAs is gebaseerd op studies aan senecionine-type PAs met een 12-ledige macrocyclische ring die voorkomen in de onderfamilie *Senecioneae* (Langel *et al*., 2011). De precursor senecionine *N*-oxide wordt gemaakt in de wortels (Hartmann en Toppel, 1987; Hartmann *et al*., 1988; Hartmann *et al*., 1989) waarna het naar de bovengrondse delen van de plant wordt getransporteerd alwaar positie-specifieke en stereoselectieve transformaties plaatsvinden die resulteren in bijvoorbeeld de herschikking van de ringstructuur en/of verschillende oxidatieve modificaties

(Hartmann en Dierich, 1998; Pelser *et al*., 2005). De enzymen die verantwoordelijk zijn voor deze omzettingen zijn nog niet geïdentificeerd. CYPs katalyseren een breed scala van positie-specifieke en stereospecifieke oxidatiestappen in de SM biosynthese in planten (Renault *et al*., 2014) en spelen daarom een belangrijke rol in de evolutie van chemische diversiteit. Het veelvuldig voorkomen van epoxide-, hydroxyl- en dubbel gebonden zuurstof-groepen in PAs suggereert dat CYPs zeer waarschijnlijk betrokken zijn bij de PA biosynthese.

In dit proefschrift worden de volgende vragen gesteld: Wat zijn de distributiepatronen van PAs in *Jacobaea* soorten en zijn deze patronen gerelateerd aan de fylogenetische afstand tussen deze soorten? Hoe komt de waargenomen PA diversiteit tot stand? Zijn CYPs betrokken bij de PA biosynthese? Naar aanleiding van deze vragen kunnen de experimentele hoofdstukken van dit proefschrift in twee groepen worden verdeeld: (i) de evolutie van de diversiteit in PAs binnen en tussen *Jacobaea* soorten (Hoofdstuk 2), (ii) het gericht onderzoeken of CYPs een rol spelen in de PA diversiteit door middel van een kandidaat-gen benadering (Hoofdstuk 3-5).

(i) De evolutie van PA diversiteit binnen en tussen *Jacobaea* soorten

Studies beschreven in hoofdstuk 2 zijn gericht op een beter begrip van het mechanisme achter de chemische diversiteit van de PAs van *Jacobaea* soorten vanuit een fylogenetische context. De aanwezigheid en concentraties van 80 PAs in acht tot tien weken oude bladeren van 17 *Jacobaea* soorten, met monsters van verschillende individuen en uit verschillende populaties die zijn opgekweekt in een klimaatkamer, zijn geanalyseerd met LC-MS/MS. Een grote diversiteit aan PA profielen is hierbij waargenomen waarbij het aantal PAs varieerde van 21 tot 59 per soort en met totale PA concentraties die varieerden tussen 32,9 tot 3835,7 μg/g drooggewicht. Zowel kwalitatieve als kwantitatieve PA profielen bleken soort-specifiek te zijn en deze bevestigen de resultaten van eerdere onderzoeken (Soldaat *et al*., 1996; Hartmann and Dierich, 1998; Langel *et al*., 2011). Jacobine-achtige, senecionine-achtige en otosenine-achtige PAs droegen het meest bij aan de soort-specifieke PA verschillen van de *Jacobaea* soorten. De verschillen tussen de PA samenstellingen van verschillende populaties binnen sommige *Jacobaea* soorten (bijvoorbeeld *J. alpina* en *J. paludosa*) waren groter dan de verschillen tussen de PA samenstellingen van sommige soorten. Om de evolutie van PA diversiteit op soortniveau te begrijpen is de moleculaire fylogenie van de 17 *Jacobaea* soorten gereconstrueerd met behulp van 11 chloroplast en drie nucleaire DNA markers. Door het voorkomen van de individuele PAs op de takken van de gereconstrueerde fylogenetische boom te plaatsen werd duidelijk dat het voorkomen van vrijwel alle PAs meerdere keren is geëvolueerd. Dit is in overeenstemming met eerdere bevindingen van Pelser *et al*. (2005). Twee verschillende maten, Blomberg's K (Blomberg, 2003) en Pagel's λ (Pagel, 1999), zijn gebruikt om de correlaties tussen kwantitatieve PA eigenschappen en fylogenetische relaties te evalueren. Significante fylogenetische signalen zijn alleen onder de λ statistieken gevonden voor negen van de 80 PAs. Gezien de algemene intraspecifieke PA diversiteit gevonden in *Jacobaea* soorten (dit proefschrift; Witte *et al*., 1992; Macel *et al*., 2004), kan worden

aangenomen dat deze hoge PA diversiteit waarschijnlijk te wijten is aan de regulatie van PA biosynthesegenen in planten als levensstrategie om te voldoen aan hun verschillende biologische behoeften, in plaats van aan de evolutionaire winsten en verliezen van bepaalde PA biosynthesegenen.

(ii) Een kandidaat-gen benadering gericht op CYPs voor hun betrokkenheid bij het tot stand komen van de PA diversiteit

Om de hypothese, dat de PA diversiteit in *Jacobaea* soorten het resultaat is van de regulatie van PA biosynthese genen, te testen, is kennis nodig over de genen betrokken bij PA biosynthese. CYPs vormen grote gen families binnen soorten en spelen vaak een rol in de biosyntheseroutes van SMs. De diversificatie van PAs komt vaak tot stand door positie-specifieke oxidaties (Hartmann and Dierich, 1998; Pelser *et al*., 2005), wat het waarschijnlijk maakt dat CYPs betrokken zijn bij de PA biosynthese. Er was echter nog geen CYP database van *Jacobaea* soorten beschikbaar aan het begin van deze studie. Daarom is een systematische studie van CYPs uitgevoerd met betrekking tot hun diversiteit en evolutie in *J. vulgaris* en *J. aquatica* zoals beschreven in hoofdstuk 3. In totaal zijn respectievelijk 221 (verdeeld over acht clans en 38 families) en 157 (verdeeld over acht clans en 35 families) CYPs van volledige lengte geïdentificeerd uit *de novo* geassembleerde transcriptomen van respectievelijk *J. vulgaris* en *J. aquatica*. De CYPs die op basis van de KEGG database zijn aangewezen als mogelijk betrokken bij SM biosynthese behoren allemaal tot de CYP71 clan. Geen van de gevonden CYPs werd geannoteerd als mogelijk betrokken bij alkaloïd biosynthese. Dit betekent niet noodzakelijk dat CYPs niet betrokken zijn bij de biosynthese van PAs, omdat de huidige KEGG database geen informatie bevat over enzymen die betrokken zijn bij PA biosynthese. Een fylogenetische analyse van de zes grootste CYP families (CYP71, CYP76, CYP706, CYP82, CYP93 en CYP72) van de twee *Jacobaea* soorten, twee andere leden van de familie van de Asteraceae, *Helianthus annuus* en *Lactuca sativa*, en de outgroup *A. thaliana* is uitgevoerd. De fylogenetische bomen vertoonden een sterke soort-specifieke expansie van CYPs, wat suggereert dat de evolutie van CYPs zeer snel is geweest, zelfs binnen de Asteraceae familie. Alleen CYPs van de nauw verwante soorten *J. vulgaris* en *J. aquatica* werden vaak in paren in een terminale clade gevonden, wat een nauwe relatie in de evolutionaire geschiedenis van deze twee soorten bevestigt. De studies beschreven in hoofdstuk 3 hebben geresulteerd in een CYP database voor toekomstige verkenning van hun functies, inclusief mogelijke betrokkenheid bij PA biosynthese en diversiteit.

Hoofdstuk 4 beschrijft een poging tot het identificeren van kandidaat CYPs, welke een rol zouden kunnen spelen in de PA biosynthese, gebaseerd op de associatie tussen PA en CYP transcript profielen van de twee *Jacobaea* soorten, opgegroeid onder gecontroleerde omstandigheden. De contrasten in constitutief aanwezige jacobine-achtige en erucifoline-achtige PAs tussen *J. aquatica* en vier groepen van F_2 hybriden van een kruising tussen *J. vulgaris* en *J. aquatica* zijn gebruikt om kandidaat CYPs te selecteren. Daarnaast zijn de contrasten in profielen gebruikt tussen in weefselkweek opgegroeide *J. vulgaris* planten die

behandeld zijn met methyl-jasmonaat (MeJA) om een toename in erucifoline-achtige PAs te induceren en niet-behandelde planten. In totaal zijn 44 PAs gedetecteerd door middel van LC-MS/MS en PA profielen van de verschillende *Jacobaea* monsters zijn onderling vergeleken in de constitutieve en geïnduceerde PA groepen door middel van het optellen van de concentraties van PAs met dezelfde positie-specifieke oxidatieve modificaties, welke aangebracht zouden kunnen worden door een CYP enzym, nl. 15,20-epoxidatie, 12,12-epoxidatie, 19-hydroxylatie, 18-hydroxylatie, 13,19-dehydrogenatie en 8-oxidatie. De sequenties van de transcriptomen zijn voor alle groepen en behandelingen apart bepaald om het expressieniveau van de CYP genen te kunnen vergelijken. In totaal zijn er voor de constitutieve en geïnduceerde PA conversies respectievelijk 33 en 27 CYP kandidaten gevonden. De meeste van deze CYP kandidaten met onbekende functie behoren tot de CYP71 clan. Kandidaten behorende tot 11 CYP subfamilies werden zowel in de constitutieve als in de MeJA-geïnduceerde groep gevonden, waarvan kandidaten uit drie subfamilies (CYP72A, CYP706E, CYP82Q) verantwoordelijk zouden kunnen zijn voor de vorming van erucifoline-achtige PAs die zowel 12,13 epoxidatie als 19-hydroxylatie bevatten.

Hoofdstuk 5 beschrijft het functioneel testen van acht CYP kandidaten voor hun betrokkenheid bij de PA biosynthese door middel van heterologe expressie in gist en een *in vitro* enzymtest met microsomale membraanpreparaten die naar verwachting de tot expressie gebrachte CYPs bevatten. Bij geen van de acht CYP enzympreparaten is met de gebruikte PA substraten (senecionine/integerrimine, seneciphylline, jacobine, erucifoline, en een PA mix of hun respectievelijke *N*-oxiden) conversie vastgesteld. Redenen voor deze negatieve resultaten zouden te lage CYP expressieniveaus of suboptimale reactiecondities kunnen zijn. Waarschijnlijker is echter dat de acht geteste CYP enzymen niet verantwoordelijk zijn voor PA biosynthese.

Conclusies

Jacobaea soorten worden vaak gebruikt voor de bestudering van de diversiteit van PAs (Vrieling *et al*., 1993; Hartmann and Dierich, 1998; Macel *et al*., 2004; Cheng *et al*., 2011). Vergeleken met eerdere studies was een andere invalshoek van dit proefschrift onderzoek naar de kwalitatieve en kwantitatieve variatie in PAs tussen soorten, populaties en individuen. De resultaten tonen aan dat de PA samenstellingen van *Jacobaea* soorten inderdaad soort-specifiek zijn, zowel in concentratie als in compositie. Hoewel er voor sommige soorten ook grote verschillen zijn tussen de PA samenstellingen van populaties. Dit is de eerste studie die laat zien dat concentraties en composities van PAs binnen en tussen *Jacobaea* soorten sterk vergelijkbare patronen laten zien, daarmee implicerend dat dezelfde of sterk verwante mechanismen een rol spelen bij het tot stand komen van de variatie in PA concentraties en composities.

De aan- en afwezigheid van individuele PAs is bestudeerd op soortniveau, waarbij slechts voor weinig PAs een fylogenetisch signaal werd gevonden. Het lijkt erop dat alle *Jacobaea* soorten de machinerie voor PA productie bezitten maar dat de verschillen in het voorkomen

van PAs het gevolg zijn van differentiële expressie van de biosynthesegenen. Gebaseerd op het feit dat de diversiteit in PAs met name voortkomt uit oxidatiereacties zijn CYPs, een belangrijke klasse van oxidatieve enzymen in planten, gekozen als mogelijke kandidaten voor PA biosynthese genen. De in dit proefschrift beschreven CYP databank is de eerste voor *Jacobaea* soorten. Een gen-tot-metaboliet aanpak is gebruikt om CYP kandidaat genen, welke mogelijk betrokken zouden kunnen zijn bij de PA biosynthese, te identificeren. Vervolgens zijn de functies van acht kandidaat CYP enzymen getest. Hierbij is echter geen indicatie gevonden dat zij een rol zouden spelen bij de biosynthese van PAs. De meerderheid van de in hoofdstuk 4 geïdentificeerde kandidaat genen moeten echter nog functioneel getest worden. Het is verder van belang om in het oog te houden dat de enzymen betrokken bij PA biosynthese niet noodzakelijkerwijs CYPs zijn, maar ook zouden kunnen behoren tot andere klassen van oxidatieve enzymen zoals peroxidases of flavine-afhankelijke mono-oxygenases (Burton, 2003). Omdat deze enzymklassen ook karakteristieke aminozuursequentiemotieven hebben, kunnen de betreffende gen families en kandidaat genen via gen-tot-metaboliet correlaties ook verkregen worden uit de sequentiedata en de metabolietgegevens beschreven in hoofdstuk 4.

Curriculum Vitae

Yangan Chen was born in Zhangping, Fujian Province in China on May 8, 1987 (the 11th day of the 4th month according to the lunar calendar of that year). She started her undergraduate study in Wuhan University majoring in pharmaceutical sciences in 2006 and got her BSc degree in June, 2010. Later on in August 2010, she started her master study at the University of Macau, resulting in the thesis entitled "Phytochemical analysis of *Microctis folium*, *Puerariae lobatae* Radix and *Puerariae thomsonii* Radix". She obtained the degree of Master of Science in Chinese Medicinal Science in August 2013. From September 2013 to August 2014, she worked as a research assistant at the Beijing University of Chinese Medicine. After that, she received financial support from the China Scholarship Council for her PhD research at the Institute of Biology, Leiden University. In October 2014, she started her PhD project in the group of Plant Ecology under the supervision of Prof.dr. Peter Klinkhamer and Dr. Klaas Vrieling, and in the group of Plant Cell Physiology under the supervision of Prof.dr. Johan Memelink. Her work about the evolution of chemical diversity in plants using pyrrolizidine alkaloids in *Jacobaea* species as the study system is described in this thesis.

Publication list

- **Chen Y**, Mulder PPJ, Schaap O, Memelink J, Klinkhamer PGL, Vrieling K. (2019) The evolution of pyrrolizidine alkaloid diversity among and within *Jacobaea* species. (submitted)
- **Chen Y**, Klinkhamer PGL, Memelink J, Vrieling K. (2019) Diversity and evolution of cytochrome P450s of *Jacobaea vulgaris* and *Jacobaea aquatica*. (submitted and under revision)
- **Chen Y***, Song Y*, Wang Y, Yuan Y, Huang X, Ye W, Wang Y, Zhang Q. (2014) Metabolic differentiations of *Pueraria lobata* and *Pueraria thomsonii* using ^{1}H NMR spectroscopy and multivariate statistical analysis. *Journal of Pharmaceutical and Biomedical Analysis* 93:51-58. (*Co-first author)
- **Chen Y***, Li P*, Li P, Yan R, Zhang X, Wang Y, Zhang X, Ye W, Zhang Q. (2013) α-Glucosidase inhibitory effect and HPLC-DAD analysis of *Microctis Folium*. *Molecules* 18:4221-4232. (*Co-first author)
- Song Y, Jing W, **Chen Y**, Yuan Y, Yan R, Wang Y. (2014) ^{1}H nuclear magnetic resonance based-metabolomic characterization of Peucedani Radix and simultaneous determination of praeruptorin A and praeruptorin B. *Journal of Pharmaceutical and Biomedical Analysis* 93:86-94.

Patent

- Li J, Zhang Q, Tu P, Sun Q, Wang J, Zhao Y, Zhang J, **Chen Y**. (2017) The methods of isolation of aescin from *Semen aesculi*. (in application for a patent in China: CN 106589045 A)