Title
Engineering microbial production of terpenoids

Author
Wong, Jeff

Publication Date
2017

Peer reviewed|Thesis/dissertation

Engineering microbial production of terpenoids

by

Jeff Wong

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Plant and Microbial Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jay D. Keasling, Co-chair

Professor Henrik V. Scheller, Co-chair

Professor Krishna Niyogi

Professor John Dueber

Fall 2017

Engineering microbial production of terpenoids

by Jeff Wong

Abstract

Engineering microbial production of terpenoids

by

Jeff Wong

Doctor of Philosophy in Plant and Microbial Biology

University of California, Berkeley

Professors Jay D. Keasling and Henrik V. Scheller, Co-chairs

Terpenoids are a broad class of natural products that have applications in medicine as well as commodity chemicals and biofuels. However, many terpenoids are sourced from field grown plants, which suffer from inconsistent yield and composition. Additionally, many terpenoid biosynthetic pathways are poorly understood or completely uncharacterized due to limitations regarding genetic resources and heterologous expression. Recent improvements in synthetic biology such as DNA synthesis, targeted genome editing, and next generation sequencing have enabled accelerated research and development in this field. In this work I will describe my contributions to the characterization of two plant-derived medicinal terpenoid pathways: the prostratin and valerenic acid biosynthetic pathways.

Chapter 1 begins with a discussion of the challenges and successes associated with engineering terpenoid production in yeast as well as other commonly engineered microbes. This chapter covers their applications in medicine, biofuels and other commodity chemicals.

Chapter 2 presents my work in developing transcriptomic resources for identifying P450s and other candidate genes involved in the biosynthesis of prostratin. This work also functionally characterized the *Jatropha curcas* casbene synthase, as well as ~40 P450s from two Euphorbiaceae species, three of which showed activity on the casbene intermediate. Unfortunately, another group published on the activity of two of these P450s before we could. This study shows the utility of coexpression in the discovery of terpenoid biosynthetic genes.

Chapter 3 develops a high-titer yeast chassis for the production of jolkinol C, an important intermediate in the biosynthesis of prostratin and related compounds. The P450s identified in Chapter 2 were optimized to produce >1 g/L total of triply-oxidized casbanes, and 800mg/L of the target product, jolkinol C. Additionally, this study clarified the role of a short chain alcohol dehydrogenase (ADH) in the pathway; while not necessary for jolkinol C formation, the ADH improved the product formation, suggesting it is important *in planta*, as this gene is found in the jolkinol C biosynthetic gene cluster of *J. curcas*.

Chapter 4 elucidates the final unidentified step in the biosynthetic pathway of a sesquiterpenoid drug, valerenic acid. Specifically, the P450 VoCYP71DJ, identified from coexpression analysis

with the preceding step valerena-4,7(11)-diene synthase, produces oxidized valerena-4,7(11)-diene. Interestingly, the expression of an alcohol dehydrogenase and aldehyde dehydrogenase with VoCYP71DJ1 were necessary to produce valerenic acid in yeast.

The final chapter summarizes this work, as well as suggesting future efforts to further enable the engineering of terpenoid biosynthesis in microbes.

For my fiancée’s support and love

TABLE OF CONTENTS

1. Chapter 1 — Introduction

Including material from published work: Wong, J., Rios-Solis, L., Keasling, J. D. "Microbial Production of Isoprenoids." *Consequences of Microbial Interactions with Hydrocarbons, Oils, and Lipids: Production of Fuels and Chemicals* (2017): 1-24.

1.1. Abstract: an overview of progress in the field of metabolic engineering of microbes for the production of terpenoids

Terpenoids (isoprenoids) are among the most diverse groups of compounds synthesized by biological systems; it has been estimated that there are approximately 30–50,000 known isoprenoids. Isoprenoids are important in maintaining membrane fluidity, electron transport, protein prenylation, cellular and organismal development, and in controlling pests. Many isoprenoids have found applications in fragrances and essential oils, pharmaceuticals, specialty and commodity chemicals and most recently as biofuels. To make all of these applications of isoprenoids possible, their production in microbial hosts is essential. Recently, there has been much progress in producing these complex hydrocarbons in both *Escherichia coli* and *Saccharomyces cerevisiae*. In this chapter, we review recent progress in this area.

1.2. Introduction

Isoprenoids are among the most diverse groups of compounds synthesized by biological systems; it has been estimated that there are approximately 30–50,000 known isoprenoids, which includes sterols, carotenoids and quinines (Beller et al. 2015; McCaskill and Croteau 1997). Isoprenoids are important in maintaining membrane fluidity, electron transport, protein prenylation, cellular and organismal development, as fragrances and essential oils, antibacterial and antifungal agents as well as high value pharmaceuticals and fuel alternatives (McCaskill and Croteau 1997). Terpenes are a classified by the number of isoprene (C5) units that they contain. Hemiterpenoids (C5) such as isopentenols, monoterpenes (C10), such as menthol and camphor, and sesquiterpenes (C15), such as zingiberene (ginger), are the major constituents of herbs and spices. Other sesquiterpenes, diterpenes (C20)are pheromones, defensive agents, and signal transduction agents (Fraga 2005; McGarvey and Croteau 1995; Wang et al. 2013). Higher molecular weight isoprenoids stabilize membranes (cholesterol and other C30 compounds) and serve as photoreceptive agents (carotenoids and other C40 compounds).

1.3. Isoprenoids as Pharmaceuticals

Many terpenoids have been found to exhibit potent biological activity, with several of them in development or in use therapeutically. The anti-malarial drug artemisinin and the anti-cancer agent Taxol serve to illustrate the clinical importance of terpenoids, respectively (Kirby and Keasling 2009). Artemisinin, a sesquiterpenoid extracted from sweet wormwood (*Artemisia annua*), is one of the few anti-malarial drugs for which *Plasmodium* has no known resistance and, as such, has been hailed as a miracle drug for its ability to cure malaria (Cui and Su 2009). Taxol, a diterpene extracted from the Pacific Yew, is extremely effective in the treatment of

certain cancers (ovarian, breast, lung and neck, bladder and cervix, melanoma, and Kaposi's sarcoma) (Jennewein and Croteau 2001; Skeel and Khleif 2011). Unfortunately, the complexity of the molecule precludes commercial total chemical synthesis (Engels et al. 2008). Taxol is currently produced either by semisynthesis from 10-deacetylbaccatin III extracted from the needles of *Taxus* species, or by extraction from plant cell suspension cultures grown with elicitors to improve production (Cusidó et al. 1999).

A range of medicinal diterpenoid compounds (*i.e.,* phorbol esters and the related casbanes, lathyranes, jatrophanes, and ingenanes) are solely produced in Euphorbiaceae and Thymelaceae species from casbene and neocembrene diterpene backbones (Vasas and Hohmann 2014; Kirby et al. 2010). These diterpenoids have gained interest due to unique anti-cancer, anti-HIV, vascular-relaxing, neuro-protective, anti-inflammatory, or immune-modulatory activities (Blumberg 1988; Halaweish et al. 2002; Jiao et al. 2009; Srivalli and Lakshmi 2012; Vasas and Hohmann 2014). Recently, the U.S. FDA approved the use of a related diterpenoid, ingenol mebutate, for the treatment of the premalignant skin condition actinic keratosis (Vasas and Hohmann 2014). Prostratin, a phorbol ester found in *Homalanthus nutans* (Euphorbiaceae), is being developed as an adjuvant therapy to clear latent viral reservoirs, the primary obstacle to eradication of HIV (Wang et al. 2015). A related diterpene, resiniferatoxin, has potential utility in the treatment of chronic pain such as that resulting from osteosarcoma and osteoarthritis (Marwick 2005).

The monoterpene limonene and related derivatives are believed to inhibit farnesylation of the growth promoting protein RAS, and therefore inhibit malignant cell proliferation (Gelb et al. 1995; Gould 1997; Hohl 1996). Additionally, monoterpene indole alkaloids such as vinblastine, vincristine, and camptothecin are commonly used in chemotherapeutic treatments and the pathways of these drugs are currently being researched for utilization in microbial production (Góngora-Castillo et al. 2012; O'Connor and Maresh 2006). The ability to produce terpenoid drugs in microbes could significantly reduce their production costs, reduce pressure on unsustainable plant-derived sources and increase their chances of reaching clinical trials and the market.

1.4 Isoprenoids as Biofuels

There has also been a great deal of interest in terpenes as potential biofuels, because the chemical structures of isoprenoids provide many beneficial aspects as alternative fuels. This includes the common methyl branching which lowers the freezing point significantly, as well as its usual cyclic nature which increases energy density (Beller et al. 2015). This has led to the recent production and testing of several isoprenoids as potential alternative fuels (Harvey et al. 2010; Mack et al. 2014; Yang et al. 2010). Esters of isoprenoid alcohols (C5, C10, and C15) have potential to be used as replacements for petroleum-based diesels (Horton et al. 2003; Singh et al. 2008).

Isoprenoid biosynthesis pathways provide additional routes to C5 hemiterpene alcohols (namely isopentenol and isopentanol) which have higher energy contents than ethanol and high octane

numbers (Yang et al. 2010), which supports their use as gasoline replacements and as anti-knock additives (Cann and Liao 2010; Hull et al. 2006). Recently, due to their favorable energy content, the alcohols 3-methyl-3- and 3-methyl-2-butenol were shown to function as ideal antiknock additives in spark ignition engines (Mack et al. 2014).

In the case of monoterpenes, limonene as well as its fully hydrogenated form, limonane, are considered promising jet-fuel replacements which are been evaluated as a jet-fuel additives (Chuck and Donnelly 2014; Tracy et al. 2009). Pinene dimers have also been shown to have a volumetric heating value similar to that of the tactical fuel JP-10 (Harvey et al. 2010).

Hydroxylated acyclic monoterpenoids such as linalool, geraniol (or its oxidized form geranial) are also potential biofuels which have been found to be soluble in fossil fuels and combusted in a steady manner up to a monoterpene content of 65% and 20% (wt/wt) respectively (Hellier et al. 2013). Using linalool as substrate, a ruthenium (Ru)-based olefin metathesis reaction has been developed for the synthesis of 1-methylcyclopent-2-enol, a promising precursor for the synthesis of methylcyclopentadiene dimer, which can be converted to the high-density missile fuel RJ-4 (Meylemans et al. 2011).

The hydrogenated products of acyclic monoterpenes such as myrcene and ocimene are also considered good biofuel replacements (Tracy et al. 2009). High density renewable fuels have also been successfully synthesized through selective dimerization of α-pinene, camphene, limonene, and crude turpentine (Meylemans et al. 2012).

Diesel is composed of linear, branched, and cyclic alkanes with an average carbon length of 16 which is close to the 15 carbons of sesquiterpenes. The higher branching degree of sesquiterpene gives them the benefit of more stability under high pressure, lower freezing point, reduced premature ignition and increased octane number, although it also lowers the cetane number (Peralta-Yahya et al. 2011). Sesquiterpenes farnesane, farnesol, bisabolene and its hydrogenated form bisabolane have been recognized to have great potential as second generation biofuels (Rude and Schirmer 2009). Farnesane and bisabolane have cetane numbers of 58 and 52 which are within the range for diesel fuels (40-60) (Peralta-Yahya et al. 2011; Wang, Kim, et al. 2011). The carbon ring of bisabolane confers it more energy density per volume of fuel than farsenane, however the latter has a better cetane number and it has been the closest to commercialization (Renninger and McPhee 2008).

Amyris, in collaboration with the french oil company TOTAL and the Brazilian airline GOL have reach industry approval to use farnesane as a jet fuel replacement in a commercial flight. In addition, Amyris has also patented sesquiterpene amorphane to be used as a jet-fuel replacement (Renninger and McPhee 2008).

1.5. Synthesis of Terpene Olefins

The primary building block (C5 unit) for the synthesis of isoprenoids is isopentenyl pyrophosphate (IPP) which is synthesized via two different pathways: mevalonate pathway and non- mevalonate or 1-deoxyxylulose-5-phosphate (DXP) pathway (Fig. 1.1). The mevalonate

pathway is found primarily in eukaryotes and archaea (but has been found in a few prokaryotes), whereas the non-mevalonate pathway is found primarily in prokaryotes and in plastids of photosynthetic eukaryotes (Eisenreich et al. 1998; Lange et al. 2000). The mevalonate pathway produces IPP from three molecules of acetyl-CoA (Hampton et al. 1996; Hampton and Bhakta 1997; Liao et al. 2016; Meigs et al. 1996; Meigs and Simoni 1997; Szkopińska et al. 2000). The genes for the enzymes in the mevalonate-dependent pathway have been cloned from a number of organisms (Bischoff and Rodwell 1996; Bochar et al. 1997; Campbell et al. 1998; Chambliss et al. 1996; Hahn et al. 1996; Hahn and Poulter 1995; Phulara et al. 2016; Potter and Miziorko 1997; Potter et al. 1997; Takahashi et al. 1999; Toth and Huwyler 1996; Weaver et al. 2015).

The more recently elucidated DXP (or MEP) pathway produces IPP and dimethylallyl pyrophosphate (DMAPP) from pyruvate and glyceraldehyde-3-phosphate (Altincicek, A. K. Kollas, et al. 2001; Altincicek, A.-K. Kollas, et al. 2001) (Narciso Campos et al. 2001; N Campos et al. 2001; Cunningham et al. 2000; Hahn et al. 1999); (Rodríguez-Concepción et al. 2001; Rohdich et al. 1999; Rohdich et al. 2002; Rohmer et al. 1993; Schwender et al. 1996; Sprenger et al. 1997; Takahashi et al. 1998). Synergy between both pathways has been observed in a *E. coli* engineered strain with both the DXP and the MVA pathway, where it was shown that overexpression of the dual pathway resulted in a 4.8-fold and 1.5-fold increase in the flux through the MEP and MVA pathway respectively *(Yang et al. 2016)*.

DMAPP acts as a primer for the sequential additions of IPP by the isoprenyl pyrophosphate synthases to form C10 geranyl pyrophosphate (GPP), C15 farnesyl pyrophosphate (FPP), C20 geranylgeranyl pyrophosphate (GGPP), and larger isoprenyl pyrophosphates (Fig. 1.1). Several prenyltransferases have been characterized, and their genes cloned. These prenyltransferases are fairly selective for the chain condensation lengths and stereochemistries of their substrates and products (Tarshis et al. 1994; Tarshis et al. 1996; Wang and Ohnuma 1999; Wang 2000; Leipoldt et al. 2015).

Terpene olefins are synthesized from the linear terpene pyrophosphate esters GPP, FPP, and GGPP. Cyclization of GPP by terpene cyclases (or synthases) forms the monoterpenes, cyclization of FPP forms the sesquiterpenes, and cyclization of GGPP forms the diterpenes (Dickschat 2016).

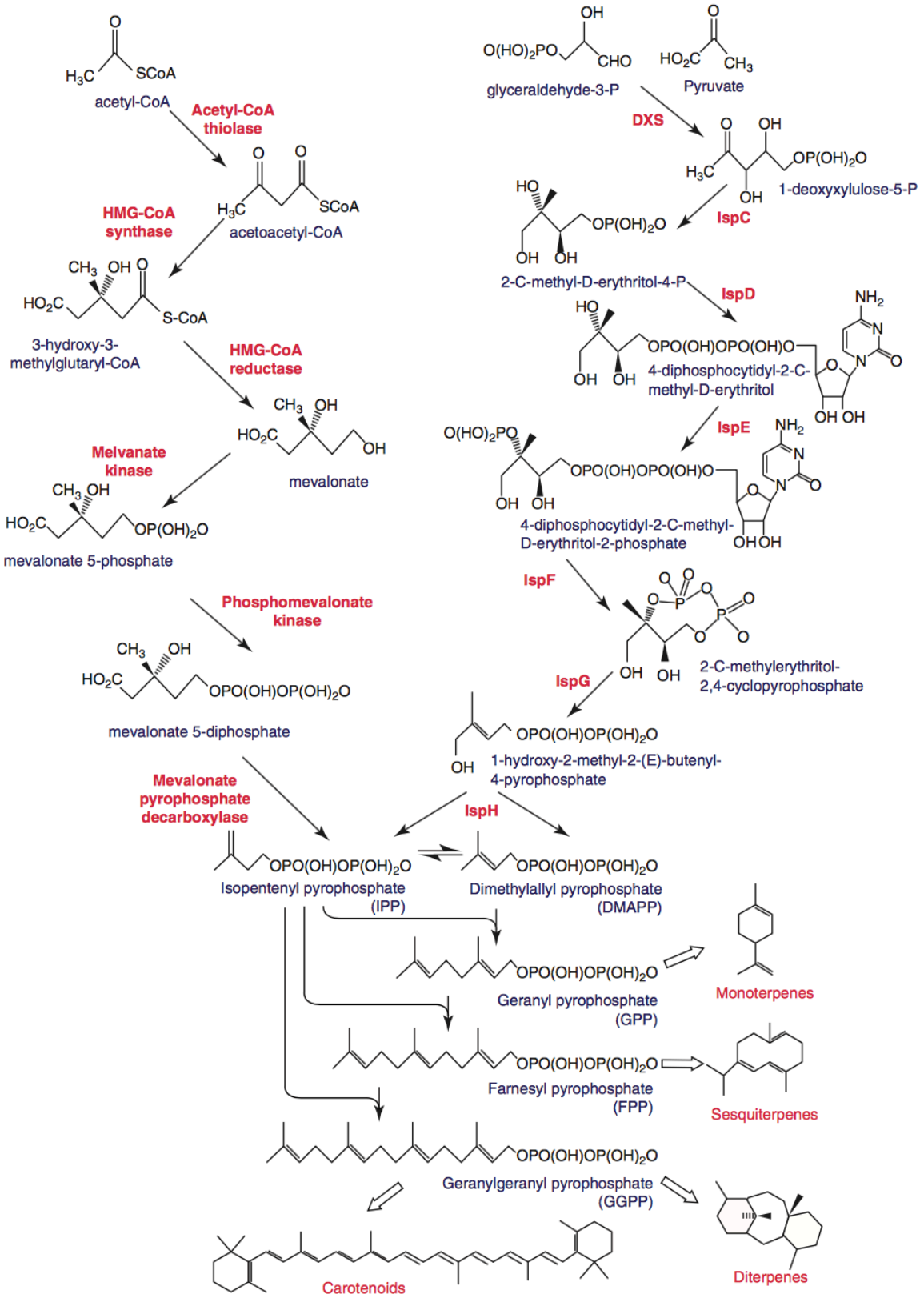


Fig. 1.1. Terpene biosynthetic pathways. Left: DXP pathway. Right: Mevalonate pathway.

The terpene cyclases are similar to the prenyl transferases in structure and reaction mechanism. However, they differ from prenyl transferases in two respects: (1) terpene cyclases catalyze intramolecular reactions, whereas prenyltransferases catalyze intermolecular reactions; (2) different terpene cyclases can take the same substrate and catalyze the formation of very different products (Chappell 1995). While some of the terpene cyclases form one or a few products, there are some cyclases that produce a variety of products from a single substrate type; for example, the δ-selinene and γ-humulene synthases of *Abies grandis* (grand fir) produce 34 and 52 total sesquiterpenes, respectively (Steele et al. 1998).

In general, the condensation and cyclization steps are independently catalyzed. Nevertheless, exemptions like enzymes that comprise a C-terminal isoprenyl transferase domain and an N-terminal cyclase domain have been discovered which can catalyze both condensation and cyclization reactions (Ozaki et al. 2014; Chiba et al. 2013)

1.6. Metabolic Engineering of Isoprenoid Production

Due to the ease of genetic manipulation and industrial fermentation of *E. coli* and *S. cerevisiae*, most isoprenoid metabolic engineering efforts have focused on these hosts, although metabolic engineering in other host organisms for isoprenoid production has been promising (Melillo et al. 2013; Wriessnegger et al. 2014; Zhan et al. 2014). *E. coli* and *S. cerevisiae* have been engineered to produce high levels of mono-, sesqui-, and diterpenes for both biofuel and medicinal applications (Alonso-Gutierrez et al. 2013; George, Thompson, et al. 2015; Meadows et al. 2016; Paddon et al. 2013; Trikka et al. 2015).

1.7. Production of Terpenes in *E. coli*

E. coli was used to functionally express several native and codon optimized plant terpene synthases and confirmed the production of the corresponding terpenes: carene and myrcene synthases (monoterpene synthases); δ-cadinene, epi-aristolochene, vetispiradiene, germacrene C, selinene, γ-humulene, amorphadiene, and epi-cedrol synthases (sesquiterpene synthases); and casbene, ent-kaurene, and abietadiene synthases (diterpene synthases) (Kim et al. 2015; Martin et al. 2001; Reiling et al. 2004). Originally most of these terpenes were produced at levels of less than 1 μg L^{-1} (when using the native DXP pathway). This was approximately 500-fold lower than lycopene production in a similar system.

To address the difficulties expressing plant terpene synthases in *E. coli*, the two sesquiterpenes of *E. coli* codon optimized genes of amorphadiene synthase (ADS) and epi-cedrol synthase (EPC) were synthesized. The cDNA clones for both genes had been previously obtained from *Artemisia annua (Mercke et al. 1999; Mercke et al. 2000)*, and were cloned into the high-copy expression vector pTrc99A.

Expression of the synthetic genes increased both enzyme production and product formation. The production of amorphadiene was 13-fold higher than that of δ-cadinene when using endogenous FPP. An SDS-PAGE analysis of cells expressing the ADS gene showed a visible band near the predicted gene size of 64 KDa; the protein produced from the native plant gene was not evident

on a gel, even when the rare codon tRNA plasmids were included. The synthetic genes were co-expressed with the genes for the latter steps of the mevalonate pathway (MBIS) (Fig 1.2); with 20 mM mevalonate in the culture medium, *E. coli* produced greater than 1.7 mg L^{-1} amorphadiene. When the synthetic amorphadiene synthase gene was co-expressed with the full mevalonate pathway (Fig 1.2), *E. coli* produced 10 mg L^{-1} amorphadiene (Kizer et al. 2008). To prevent the loss of the volatile amorphadiene, dodecane was added to the culture medium. In this two-phase cultivation, we were able to produce nearly 1 g L^{-1} amorphadiene, with the amorphadiene accumulating in the organic phase (dodecane) (Newman et al. 2006). This system (engineered host plus production process) represented a production level approximately 1,000,000-fold higher than the previous best native plant gene system or the system based on the native *E. coli* isoprenoid biosynthetic (DXP) pathway.

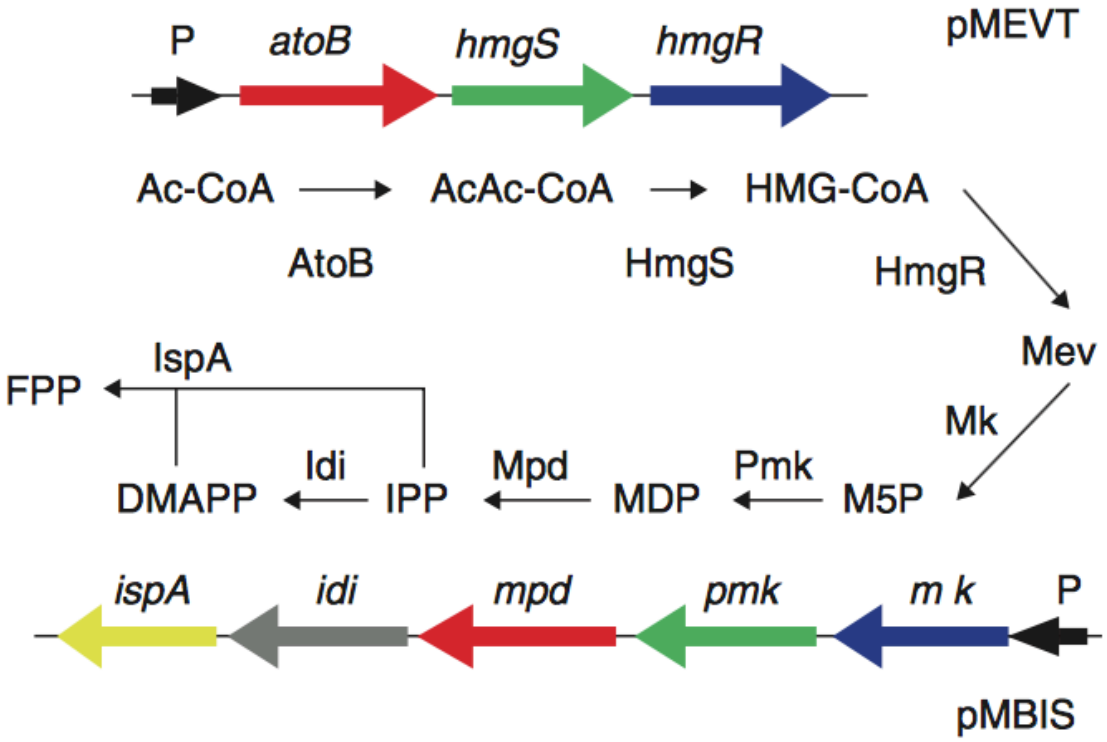


Fig. 1.2. Constructs used to express the mevalonate pathway in E. coli. The gene names are shown above each ORF. P, promoter; Ac-CoA, acetyl CoA; AcAc-CoA, acetoacetyl-CoA; HMG-CoA, hydroxymethylglutaryl-CoA; Mev, mevalonate; M5P, mevalonate-5-phosphate; MDP, mevalonate-5-pyrophosphate; IPP, isopentenyl pyrophosphate; FPP, farnesyl pyrophosphate.

The system was further improved by the identification of two rate-limiting enzymes (mevalonate kinase (MK) and amorphadiene synthase (ADS). By optimizing promoter strength to balance expression of the encoding genes, the pathway bottlenecks were alleviated and improved production fivefold (Anthony et al. 2009). Using gene variants of HMGS and HMGR derived from *Staphylococcus aureus* (mvaS and mvaA, respectively) as well as developing a high-density fermentation process, amorphadiene production was further improved to yield up 27.4 g

L^{-1} in a 2L bioreactor (Tsuruta et al. 2009).

High-level production of mono and diterpene synthases in *E. coli* requires the introduction of either GPP or GGPP production capabilities along with monoterpene or diterpene synthases. In addition to the genes involved in production of IPP and DMAPP, polyprenyltransferases have also been cloned and successfully expressed. To produce monoterpenes and diterpenes, *E. coli* FPP synthase gene (ispA) has been mutated at the fourth and fifth positions N-terminal to the first aspartate rich domain using site-directed mutagenesis to transform the enzyme to either a GPP synthase or a GGPP synthase, respectively, similar to the research done on the FPP synthase of *B. stearothermophilus* (Narita et al. 1999; Ohnuma et al. 1996). The products of both mutant enzymes were verified by *in vitro* enzyme assays and by the *in vivo* production of monoterpenes and diterpenes in engineered *E. coli* (Reiling et al. 2004). The addition of these genes to the operons encoding the IPP biosynthetic pathway created a host organism capable of overproducing precursors to monoterpenes, sesquiterpenes, diterpenes, and carotenoids. The engineered GGPP precursor generating system (GGPP synthase made from the native *E. col*i FPP synthase) was evaluated for the production of the diterpenes casbene and ent-kaurene. Coupling the high-flux mevalonate pathway described previously for amorphadiene production to the GGP or GGPP synthase system allowed for the production of various isoprenoids bisabolene, farnesene, limonene, pinene, casbene or sabinene among others (Alonso-Gutierrez et al. 2013; Wang, Yoon, et al. 2011; Sarria et al. 2014) (Zhang et al. 2014). Further engineering of the MVA *E. coli* platform guided by proteomics and metabolomics using GPP synthase from *Arabidopsis thaliana* and limonene synthase from *Mentha spicata* allowed the production of up to 650 mg L^{-1} of limonene (Alonso-Gutierrez et al. 2015). Using the same optimization strategy, a titer of 1.15 g L^{-1} of bisabolene was achieved using shake-flasks by using a codon optimized bisabolene synthase from *Abies grandis* (Alonso-Gutierrez et al. 2015; Peralta-Yahya et al. 2011).

The synthesis of monoterpenes pinene and myrcene in *E. coli* using the MVA pathway has also recently been achieved up to titers of 32 and 58 mg L^{-1} respectively (Kim et al. 2015; Sarria et al. 2014; Yang et al. 2013). The lower titer of pinene in comparison to limonene suggested that pinene synthase is probably the bottleneck , considering that both pathways shared the same MVA pathway and precursors (Beller et al. 2015).

E. coli was used as a microbial system for the heterologous synthesis of β-phellandrene, Heterologous co-expression of the MVA pathway was required to confer β-phellandrene production up to 3 mg per g dcw (Formighieri and Melis 2014). In the case of diterpenes, the major advances in heterologous production of the terpene olefin in *E. coli* have been achieved for taxadiene, which is the precursor of the anticancer drug Taxol (Paclitaxel). In this case, (Ajikumar et al. 2011) selected the native DXP pathway instead of the MVA pathway, and using a multivariate-modular approach allowed the production of taxadiene up to 1 g L^{-1} culture in a 2 L bioreactor, by using a previously identified truncated version of *Taxus brevifolia* taxadiene synthase (Huang et al. 2001). In this work, in order to optimize the production, the pathway was divided into an upstream module comprising the DXP pathway leading to IPP synthesis, and a

downstream module integrated by the *Taxus* geranylgeranyl pyrophosphate and taxadiene synthase. Both modules were combined with different promoters and gene copy numbers in order to identify the optimum combination for high diterpene production.

1.8. Production of Terpenes in *S. cerevisiae*

Yeast is a particularly attractive host because of its robustness in industrial fermentation, genetic stability, and greater capacity for P450 expression (relative to *E. coli*) (Chang et al. 2007; Ro et al. 2006). Using yeast as a heterologous host is now simplified due to recent advancements in markerless genome engineering using Cas9, which have significantly reduced the time to produce the necessary modifications for an engineered yeast strain (Horwitz et al. 2015; Lee et al. 2015). The development of yeast strains for artemisinin production has shown the potential of using yeast as a heterologous host for terpene production, even surpassing terpene titers in *E. coli* cell culture (Paddon et al. 2013). Terpenoid pathway engineering in *S. cerevisiae* has largely focused on the production of the C15-based sesquiterpenes whose titers have reached >10 g L^{-1} (in fed-batch bioreactors, >1 g/L in laboratory shake flasks) (Westfall et al. 2012). Development of monoterpenoid and diterpenoid production in yeast has lagged behind, with fewer studies and lower titers (<20mg L^{-1} and <1 g L^{-1} respectively) (Ignea et al. 2014; Trikka et al. 2015). For the Taxol pathway, the first enzyme, taxadiene synthase, has been functionally expressed in *E. coli* to produce >300 mg L^{-1} taxadiene (Ajikumar et al. 2010). However, only three of the subsequent enzymes could be functionally expressed, and, until recently, only in a co-cultured *S. cerevisiae* strain since P450 expression in *E. coli* has proven problematic as these enzymes are typically membrane bound and require a P450 reductase to provide electrons for the oxygenation reactions (Biggs et al. 2016; Zhou et al. 2015). Functional expression of taxadiene synthase in yeast has also been problematic; titers reported in the literature remain low (<10 mg L^{-1}) hampering further pathway identification and heterologous reconstitution (Dejong et al. 2006; Engels et al. 2008) (approximately 70 mg/L taxadiene has been reported (Ding et al. 2014); however the authors' NMR and mass spectra do not match those previously published, leading us to discount this claim). For prostratin and other phorbols, functional expression of casbene synthase and subsequent enzymes in yeast has also proved challenging (Kirby et al. 2010; Luo et al. 2016).

The most successful instances of terpene production in *S. cerevisiae* rely on several approaches. Many groups have independently shown that HMGR is the rate limiting enzyme of the isoprenoid pathway; this is due to negative feedback of the HMGR encoded enzyme by pathway products, resulting in degradation of the enzyme (Hampton and Garza 2009). A truncated version of HMGR (tHMGR) was found to improve sesquiterpene titers, as this modification removes the endogenous regulation of this enzyme in yeast (Ro et al. 2006). Additionally, upregulation of every enzyme in the mevalonate pathway has been shown to improve titer by diverting flux of acetyl-coa to isoprenoid production (Paddon et al. 2013). Recently, Amyris showed that modifying yeast central metabolism can produce significantly higher titers (Meadows et al. 2016). Removal of the PDH-bypass enzymes and replacement with heterologous enzymes for non-oxidative glycolysis to reduce carbon lost as CO_2, conserve ATP and reducing equivalents resulted in production of farnesene >15% by volume. Expressing multiple copies of a terpene

synthase has been shown to improve final terpene titer. Fusions of terpene synthases with the preceding enzyme in the pathway has been used with varying levels of success in sesqui- and di-terpene pathways (Albertsen et al. 2011; Ignea et al. 2014; Ignea et al. 2015). An N-terminal fusion of yeast native Erg20p FPP synthase with patchoulol synthase resulted in a strain with a doubled patchoulol titer compared to strains with an unfused pathoulol synthase (Albertsen et al. 2011). Fusion proteins have been used with great success in yeast diterpene production. Ignea and colleagues fused a mutant GGPP-producing Erg20p, Erg20F96Cp, to the C-terminus of the type II 8OH-CPP synthase from *Cistus creticus*, resulting in a titer of 400mg L^{-1} of sclareol (Ignea et al. 2015). They coexpressed this fusion enzyme with other type I terpene synthases, producing increased titers of these terpene products, cis-abienol and abietadiene, as well. Ignea and colleagues also developed a mutant Erg20p producing GPP, Erg20F96W-N127Wp (Ignea et al. 2014). When fused to the N-terminus of sabinene synthase and combined with erg20 and erg9 downregulation, they reported a final titer of 17mg L^{-1}, or 340-fold increase in sabinene relative to the starting strain. Trikka and colleagues reported the highest titer of any diterpenoid in yeast, 750mg/L, of sclareol in shake flasks (Trikka et al. 2015). This titer was achieved by using a carotenogenic screen to identify gene deletion mutants with improved diterpene yield. Codon optimization has also provided mixed results for the production of terpenes in yeast (Andersen-Ranberg et al. 2016; Westfall et al. 2012); however, some of the highest diterpene titers have resulted from a combinatorial expression screen of codon optimized type 1 and 2 diterpene synthases, producing nearly 400 mg L^{-1} of various diterpenes (Andersen-Ranberg et al. 2016).

1.9. Functionalization of Terpenes

Most terpenes of medical importance are functionalized in at least one position on the hydrocarbon backbone. Functionalization, which may include glycosylation, acetylation, hydroxylation, benzoylation, and additional ring closures of the terpene backbone, expands further the diversity of possible terpenes. Relatively few of the enzymes responsible for terpene functionalization have been purified and/or characterized and only a fraction of the genes encoding these proteins have been cloned. To date, the most extensive characterization has been done for the biosynthesis of the diterpenes taxol (Ajikumar et al. 2011) and gibberellin. Most functionally characterized terpene modifying enzymes to date have been P450s. These are the largest class of modifying enzymes by far, representing 1% of all encoded genes in plants (Hamberger and Bak 2013).

1.9.1 Monoterpenes

The monoterpenes menthol and carvone (from *Mentha x piperita* and *M. spicata*, respectively) are both derived from limonene but differ in the position of oxygenation. The cytochrome P450 hydroxylases (limonene-3-hydroxylase and limonene-6-hydroxylase) that form these two products from limonene were cloned from mint trichome cDNA libraries, expressed in *E. coli* and S. cerevisiae, and characterized *in vitro* (Haudenschild et al. 2000; Lupien et al. 1995; Schalk and Croteau 2000; Wüst and Croteau 2002). Co-expression of a NADPH-cytochrome P450 reductase gene in *S. cerevisiae* or addition of the purified NADPH- cytochrome P450

reductase to *E. coli* lysate reconstituted the enzyme activity *in vitro*.

Croteau and colleagues introduced the carvone biosynthetic pathway into *E. coli* (Carter et al. 2003). They demonstrated that the regiospecific cytochrome P450 limonene hydroxylase (Haudenschild et al. 2000) and menthofuran synthase (a cytochrome P450 monooxygenase) (Bertea et al. 2001) from mint could be functionally expressed in *E. coli* and *S. cerevisiae*. By introducing into *E. coli* a GPP synthase, limonene synthase, cytochrome P450 limonene-6-hydroxylase, and carveol dehydrogenase, they were able to produce 5 mg L^{-1} of (-)-limonene. Unfortunately, most of limonene was excreted into the medium so that the final product was not detectable. Only when limonene was fed to cells was carvone detectable. Again, understanding the factors limiting P450 function could significantly impact the amount of final product formed.

In addition to limonene hydroxylases, a geraniol 10-hydroxylase has been cloned from *Catharanthus roseus* and expressed in *S. cerevisiae* for *in vitro* analysis (Collu et al. 2001). It was shown that a single amino acid substitution could convert a C6-hydroxylase to a C3-hydroxylase (Schalk and Croteau 2000), indicating that plant P450s have great potential for protein engineering. In the case of functionalization of limonene, a second cassette with the genes of cytochrome P450 from *Mycobacterium* sp. and a ferredoxin reductase was added to the *E. coli* strain encoding the MVA pathway, in combination with a geranyl diphosphate synthase from *Abies grandis* and a limonene synthase from *Mentha spicata*, allowing the production of up to 100 mg L^{-1} of perillyl alcohol (Alonso-Gutierrez et al. 2013). Recently three new cytochrome P450 monooxygenases were isolated and expressed in *E. coli* demonstrating the hydroxylation of 1,8-cineole (eucalyptol) yield the products putatively identified as (1*S*)-2α-hydroxy-1,8-cineole or (1*R*)-6α-hydroxy-1,8-cineole (Unterweger et al. 2016).

The hydroxylation of 1,8-cineole to (1R)-6β-hydroxy-1,8-cineole using *Pseudomonas putida* as a host in combination with a cytochrome P450 monooxygenase its native redox partner cindoxin from *Citrobacter braakii* allowed a production of 13 g L^{-1} in a bioreactor (substrate 1,8-cineole was added exogenously) (Mi et al. 2016). This highlighted the ability of *P. putida* to resist various toxic compounds, including several monoterpenes which have been found toxic in *E. coli* (Chubukov et al. 2015). The expansion of the enzyme toolbox for modification of monoterpenes will be critical to create starting materials to discover new functionalizing enzymes as well as been used as hydroxylated derivatives in a range of industrial applications.

1.9.2. Sesquiterpenes

5-epi-aristolochene-1,3-dihydroxylase (CYP71D20), one of the first sesquiterpene modifying P450s discovered, was found responsible for oxidizing 5-epi-aristolochene twice forming capsidiol in *Nicotiana tabacum* by coexpression with an *Arabidopsis thaliana* NADPH-cytochrome P450 reductase in *S. cerevisiae* (Ralston et al. 2001). Perhaps the most well-known sesquiterpene modifying P450, amorphadiene oxidase (CYP71AV1), catalyzes three successive oxidations, using an amorphadiene substrate and forming artemisinic acid (Ro et al. 2006). More recently identified, CYP71BA1 forms 8-hydroxy-α-humulene from α-humulene in the zerumbone biosynthetic pathway in *Zingiber zerumbet (Yu et al. 2011)*. In grape, VvSTO2 of the

CYP71BE family forms (-)-rotundone via the oxidation of α-guaiene (Takase et al. 2016).

Many promiscuous sesquiterpene hydroxylases have been identified; Cankar and coworkers found a P450 (CYP71AV8) from chicory that converts (+)-valencene to nootkatone, germacrene A to germacra-1(10),4,11(13)-trien-12-oic acid, and amorphadiene to artemisinic acid (Cankar et al. 2011). Premnaspirodiene oxygenase (CYP71D55) from *Hyoscyamus muticus* oxidizes premnaspirodiene, 5-epi-aristolochene, valencene, and 5-epi-eremophilene (Takahashi et al. 2007). (+)-δ-cadinene-8-hydroxylase (CYP706B1), involved in the hydroxylation of (+)-δ-cadinene in the gossypol biosynthetic pathway, was found to oxidize (-)-α-cubebene, (-)-α-muurolene, α-humulene, and (-)-δ-cadinene and has been expressed in S. cerevisiae (Luo et al. 2001; Wang and Essenberg 2010). Other sesquiterpene modifying enzymes have been found, though they are few. A short chain alcohol dehydrogenase was found to form zerumbone from (Okamoto et al. 2011). A pathway in *E. coli* and *S. cerevisiae* was developed for the production of artemisinic acid, which can be converted directly into the anti-malarial drug artemisinin in two chemical steps (Fig. 1.3) (Roth and Acton 1989). Addition of an amorphadiene oxidase (AMO) and associated redox partners from *A. annua* enabled production of artemisinic acid (Chang et al. 2007; Ro et al. 2006). Although the native gene (nAMO) had no detectable in vivo or in vitro activity, codon-optimization coupled with N-terminal transmembrane domain engineering generated two constructs that were competent to carry out the first oxidation step *in vivo* to generate the alcohol congener of artemisinic acid at low levels (0.18–0.45 mg L^{-1}). Use of the redox partners with those from *A. annua* increased productivity 12-fold to 5.6 mg L^{-1} of alcohol. Finally, use of the most appropriate promoters and expression vector allowed much higher *in vivo* productivity of fully-oxidized artemisinic acid. Further coexpression of an alcohol dehydrogenase and aldehyde dehydrogenase in *S. cerevisiae* improved production of artemisinic acid from amorphadiene (Westfall et al. 2012).

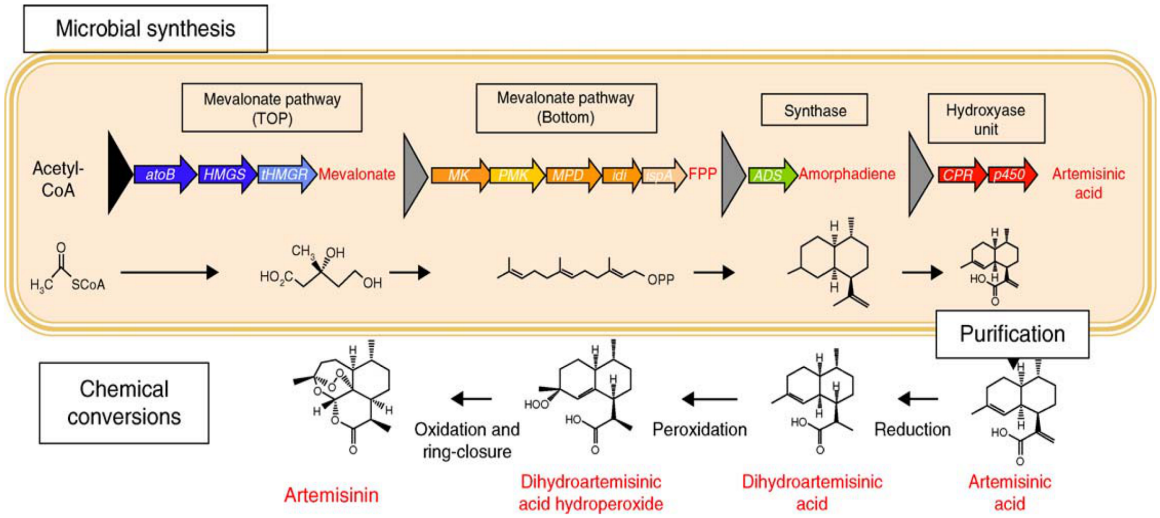


Fig. 1.3. Semi-synthesis strategy for producing artemisinin. The genes encoding the mevalonate-based FPP biosynthetic pathway were introduced from Saccharomyces cerevisiae and Escherichia coli into E. coli. The genes encoding the enzymes in the pathway were organized into two operons to allow for easy optimization. The MevT operon contains genes responsible for transforming three acetyl-CoA's into mevalonate, and the MBIS operon contains genes responsible for transforming mevalonate to FPP. In addition, the genes encoding the amorphadiene synthase, oxidase, and redox partners were also introduced into the host strain. The microbial strain produces artemisinic acid, which can be transformed to artemisinin using established chemistry.

1.9.3. Diterpenes

To date, most diterpene modifying enzymes have been discovered in pathways related to labdane-type diterpenoids and taxane-type diterpenoids (Hamberger et al. 2011; Ignea et al. 2016; Zerbe et al. 2013). Hamberger and colleagues found P450s in the CYP720B family that oxidize diterpene resin acids producing anti-herbivory diterpenes in Sitka spruce trees (Hamberger et al. 2011). Additional steps of the carnosic acid biosynthesis pathway have been discovered; Ignea and coworkers identified four P450s, CYP76AH24, CYP71BE52, CYP76AK6, and CYP76AK8, that oxidize intermediates in carnosic acid pathway by screening using heterologous expression in yeast (Ignea et al. 2016). In ingenane and phorbol ester biosynthesis, two groups found JcCYP71D495 and ElCYP71D445 oxidize casbene to 9-ketocasbene, while JcCYP726A35, JcCYP726A20, and ElCYP726A27 form the unusual ring closure seen in the intermediate jolkinol using heterologous expression in *N. benthamiana* and in yeast microsomes (King et al. 2016; Luo et al. 2016).

The biosynthetic pathway to Taxol comprises approximately twenty steps, starting with the cyclization of the prenyl diphosphate geranylgeranyl pyrophosphate (GGPP) to the diterpene taxa-4(5),11(12)-diene, followed by multiple hydroxylations, two acylations, two benzoylations, oxetane ring formation, and side chain addition to generate the final product. Three taxadiene

hydroxylating cytochrome P450s (taxadiene 5α-hydroxylase and taxane 10β-hydroxylase and 13α-hydroxylase) had been characterized by the early 2000s (Jennewein and Croteau 2001; Walker and Croteau 2001; Hefner et al. 1996; Schoendorf et al. 2001); two of these hydroxylases, taxadiene 5α-hydroxylase and taxane 10*β*- hydroxylase, have been functionally expressed in *Saccharomyces cerevisiae* as a microbial consortium with *E. coli* providing the taxadiene precursor (Zhou et al. 2015). More recently, six additional taxoid hydroxylases were characterized, with taxane 2α-hydroxylase and 7*β*-hydroxylase directly contributing to taxol biosynthesis (Kaspera and Croteau 2006). In addition, Walker and colleagues identified two acyltransferases and a benzoyltransferase in the taxol biosynthetic pathway, with taxadien-5α-ol acetyl-transferase expressed in yeast (Walker and Croteau 2001; Zhou et al. 2015). While many enzyme candidates have been found, the specific order in which the enzymes act is still unknown. Biochemical characterization of many enzymes has proved difficult because of the lack of intermediates with which to test them. Functional co-expression of such isolated enzymes presents the possibility of fermentation-based production of complex and highly substituted natural products. However, significant challenges remain in determining all of the factors that must be co-expressed with the hydroxylases to produce the decorated terpene in sufficient quantities to examine subsequent steps in the pathway and eventually to produce the desired end product, in this case taxol. Based on the hydroxylation pattern of previously related taxoid compounds, it has been proposed that the hydroxylation reaction order could be C5, C10, C2, C9, C13 followed by C7 and C1 (Croteau et al. 2006). The hydroxyl groups of C2, C5 and C10 are then acetylated, and C2 further benzoylated. Late stage Taxol intermediates such as 1-deacetylbaccatin III require the oxidation of the hydroxyl group at C9 and the addition of the side chain β-phenylalanoyl at C13 (Croteau et al. 2006).

1.10. Production of Isoprenoids in Other Hosts

The past two decades have witnessed a significant effort to develop and optimize isoprenoid production in common hosts such as *Escherichia coli* or *S. cerevisiae*. Recently, many studies have focused in genetically modifying alternate microorganisms in order to develop novel hosts for isoprenoid production platforms.

Corynebacterium glutamicum has been engineered to produce pinene by co-expressing a geranyl diphosphate synthases and pinene synthases from *Pinus taeda* and *Abies grandis* in combination with an over-expressed native 1-deoxy-d-xylulose-5-phosphate synthase and isopentenyl diphosphate isomerase (Kang et al. 2014). This resulted in a pinene production of 27 μg g^{-1} cell dry weight.

The DXP pathway of *Streptomyces venezuelae* was engineered and combined with a codon optimized bisabolene synthase from *A. grandis* , allowing the production of up to 10 mg L^{-1} of bisabolene (Phelan et al. 2015). These titers still represent an approximate order of magnitude less of bisabolene production compared to *S. cerevisiae* (Peralta-Yahya et al. 2011), but titers are similars to pinene production in *E. coli (Sarria et al. 2014)*.

Genetic engineering of the cyanobacterium *Synechocystis* allowed the production of isoprene

from CO_2 and H_2O (Lindberg et al. 2010). This was achieved by heterologous expression of the codon optimized *Pueraria montana* isoprene synthase in *Synechocystis,* enabling photosynthetic isoprene production of up 50 μg per g dry cell weight per day. *Synechocystis sp.* was also used to express a β-caryophyllene synthase gene from *Artemisia annua,* allowing the synthesis of β-caryophyllene up to 46 mg L^{-1} *(Reinsvold et al. 2011)* . *Synechocystis sp.* has also been engineered to produce limonene by overexpressing three native enzymes of the DXP pathway (*dxs*, crtE and ipi) in combination with a codon optimized *Schizonepeta tenuifolia* limonene synthase (Kiyota et al. 2014). The engineered strain allowed the photosynthetic production of limonene during 300 h reaching an overall production of 1 mg L^{-1}.

Cyanobacterium *Synechococcus sp* has been engineered to produce titers of 4 mg L^{-1} limonene and 0.6 mg L^{-1} α-bisabolene through heterologous expression of the *Mentha spicatal* limonene synthase or the *Abies grandis* (E)-α-bisabolene synthase genes respectively (Davies et al. 2014). This highlighted *Synechococcus sp*. as a promising platform for terpenoid biosynthesis.

Aspergillus nidulans has also been shown to be capable of heterologous terpene production like amorphadiene by encoding an *Artemisia annua* amorpha-4,11-diene synthase (Lubertozzi and Keasling 2008). Using the same host, (Bromann et al. 2016) showed that by overexpression of the *Fusarium fujikuroi ent*-kaurene synthase and *Citrus unshiu* gamma-terpinene synthase, the authors were able to produce *ent*-kaurene and gamma-terpinene production, respectively. Nevertheless, in both cases the production titers of terpenes in *A. nidulans* was relatively low.

1.11. Production of Terpene-Based Biofuels

Recent work suggests the isoprenoid pathway may be utilized to produce the proposed gasoline additives isopentanol and isoamylacetate (Hull et al. 2006). For example, a pyrophosphatase isolated from *Bacillus subtilis* has been demonstrated to dephosphorylate IPP to form isopentenol (Withers et al. 2007). The acetylation of isopentanol by an engineered *E. coli* has also been demonstrated (Horton et al. 2003; Singh et al. 2008). Additionally, saturated or lightly (mono- or di-) unsaturated mono- and sesquiterpenes may find use as diesel and jet fuels. The ability to modify terpene cyclases for production of novel terpenes will greatly expand the number of potential fuels molecules that can be synthesized microbially (Martin et al. 2001; Yoshikuni et al. 2006).

Biological platforms for the production of bisabolene and farnesene has been successfully developed (Peralta-Yahya et al. 2011; Wang, Yoon, et al. 2011), nevertheless the hydrogenation step of these sesquiterpenes to produce the corresponding biofuels bisabolane and farnesane is still limited in terms of titers (Kung et al. 2014). Therefore, at the current state, an option is to produce isoprenoid biofuels using an hybrid process where the sesquiterpene is produced in a microbial host followed by a chemical step to produce the reduced fuel (George, Alonso-Gutierrez, et al. 2015).

In the case of bisabolene, after optimizing the MVA pathway to improve flux towards FPP and using a codon optimized bisabolene synthase from *Abies grandis,* titers of bisabolene up to 900 mg/L in both *E. coli* and *S. cerevisiae* were achieved (Peralta-Yahya et al. 2011). After

identifying gene deletions that improved isoprenoid production, the production of bisabolene was increased up to 5.2 g/L of bisabolene using a biofermenter (Özaydın et al. 2013). These deletions typically improved mRNA and protein stability of pathway enzymes (i.e. *arp6*, *kex1*), or downregulated competing pathways (i.e. *lac1*, *ser33*, *erg24*).

Farnesene synthase from different sources like *A. annua* (Picaud et al. 2005)), *Picea abies* (Martin et al. 2004), *Zea mays (Köllner et al. 2009)* among others have been cloned and characterized in *E. coli*. Farnesene has been produced in *E. coli* from FPP via heterologous expression of a *Malus x domestica* farnesene synthase reaching 320 mg L^{-1} (Wang, Yoon, et al. 2011). Using the previously described *E. coli* and *S. cerevisiae* production bioplatforms (Pitera et al. 2007; Kizer et al. 2008; Peralta-Yahya et al. 2011) with the optimized MVA pathway, and by recruiting the farnesene synthase from *A. annua* and *P. abies*, the production of farnesene in *E. coli* and *S. cerevisiae* reached up to 1.1 g L^{-1} and 0.72 g L^{-1} respectively. Farnesene is currently been produced by Amyris from sugarcane by using an evolved *S. cerevisiae* PE-2 strain where theoretical mass yield higher than >50 % have been achieved (Chandran et al. 2011; George, Alonso-Gutierrez, et al. 2015). Novel process configurations integrating fermentation and product recovery, cell reuse and low-cost technologies for product separation (Cuellar et al. 2013), in combination with novel synthetic biology tools applied for further strain engineering will likely have a positive impact for the biofuels commercialization.

2. Chapter 2 — Development of Euphorbiaceae genetic resources for elucidation of the prostratin biosynthetic pathway

2.1. Abstract

Euphorbiaceae diterpenoids have been a source for human medicines for millenia. This class of compounds includes well-documented natural product medicines used in the treatment of cancer (ingenol-3-angelate), for latent HIV-1 activation (prostratin, 12-deoxyphorbol 13-phenylacetate), and as analgesics (resiniferatoxin). However, many of the plant species that produce these compounds are threatened in the wild and do not produce significant quantities of the compounds of interest. Additionally, knowledge of these pathways and their heterologous production in microbes is hampered by a lack of molecular, genetic, and genomic resources in the native plant species. We generated large-scale transcriptome sequence and expression profiles for two species of Euphorbiaceae that produce medicinally important diterpenoids: *Jatropha curcas* and *Homalanthus nutans,* a species with no previous sequence data. Additionally, we produced a draft genome assembly for *H. nutans* to aid identification of prostratin biosynthetic gene clusters. Further, we functionally tested a casbene synthase and elicited its expression *in planta* to probe the prostratin biosynthetic pathway using coexpression analysis, and engineered production of casbene in yeast at titers of ~80 mg/L. Using next-generation sequencing technology, we sampled the transcriptomes of these species across a variety of tissue types and conditions. These transcriptomes and expression abundance matrices provide powerful tools for interrogating the diterpenoid pathways in these species and advances our understanding of these pathways for applications in microbial synthesis.

2.2. Introduction

Euphorbiaceae diterpenoids are used widely in traditional medicine across the world, and one of these compounds, ingenol-3-angelate, has received FDA approval (Siller et al. 2009). Despite the importance of these drugs, their biosynthetic pathways are poorly understood, while farming these species and synthetic approaches towards supplying these drugs are costly and inefficient (Tanino et al. 2003; Shimokawa et al. 2007; Wender et al. 2008). The biosynthesis of these compounds originates from the isoprenoid biosynthetic pathway, after which either of the diterpene synthases casbene synthase (CBS) or neocembrene synthase cyclize the universal diterpene precursor geranylgeranyl diphosphate into casbene or neocembrene, respectively (Fig. 2.1). These hydrocarbons are subsequently oxidized to form diverse, medicinally active compounds with one or more ring closures, resulting in lathyranes and jatrophanes (one ring closure) and subsequently jatropholanes, tiglianes, ingenanes, rhamnofolanes, and daphnanes (two ring closures) (Vasas and Hohmann 2014). Studies have suggested common biosynthetic origins of these compounds, but no evidence for this claim exists to date.

Some research has been done on the initial steps of lathyrane and jatrophane biosynthetic pathways. Kirby and colleagues found that CBSs were conserved in several divergent Euphorbiaceae species (Kirby et al. 2010). Further, Kirby and colleagues identified several P450s from *Ricinus communis*, *Euphorbia resinifera*, and *Homalanthus nutans* with high homology to other terpene oxidases, in the CYP71D clade. We coexpressed these P450s with a casbene synthase on high-copy plasmids in yeast to identify casbene oxidases, resulting in the identification of two *R. communis* P450s and one *E. resinifera* P450 that oxidize unknown positions of casbene (work not published). Later, King and colleagues mined the sequenced *R. communis* genome to identify the same P450s, which were classified as RcCYP726A14 and RcCYP726A18, both of which produce 5-hydroxycasbene (King et al. 2014). The activities of these P450s stimulated further discussion that this class of enzymes may be involved in the ring closure of casbene, forming lathyrane intermediates on the way to more advanced Euphorbiaceae diterpenoids.

To elucidate the biosynthesis of medicinally important Euphorbiaceae diterpenoids, we decided to study the tigliane pathways in *Jatropha curcas* and *Homalanthus nutans*. *J. curcas* was selected due to its production of large amounts of casbene-derived tigliane Jatropha factors, multiple genome and transcriptome sequences, as well as its susceptibility to virus-induced gene silencing, an important tool for validating gene functionality *in planta* (Li et al. 2016; Ye et al. 2009; Natarajan et al. 2010; Costa et al. 2010; Natarajan and Parani 2011; Sato et al. 2011; Wang et al. 2013) *H. nutans* is of particular interest because it produces prostratin, an antiviral compound with potential for treating HIV (Johnson et al. 2008). Indeed, the Jatropha factor tiglianes contain the same structure as prostratin and are a proposed intermediate in the pathway to other diterpenoids, including ingenanes, important anticancer compounds (Tanino et al. 2003).

In this study, we functionally tested a *J. curcas* CBS in yeast to confirm its product and elicited the tigliane pathway using pectinase. Next, we constructed transcriptomic resources for *J. curcas* by assembling cDNA libraries of various tissue types and the elicited tissues. Coexpression analysis of the resulting transcript expression profiles showed 967 genes with similar expression profiles to *J. curcas* CBS, many of which were involved in plant defense. We also assembled an *H. nutans* transcriptome across diverse set of tissue types and analyzed the transcriptome for conserved prostratin biosynthetic genes found in another prostratin producing species, *Euphorbia fischeriana*. The resulting data were used to screen P450 candidates from both species, two of which oxidized the C9 position of casbene. This study highlights the importance of gene expression analysis, which can provide a more targeted approach for identifying putative biosynthetic genes for functional testing than other strategies.

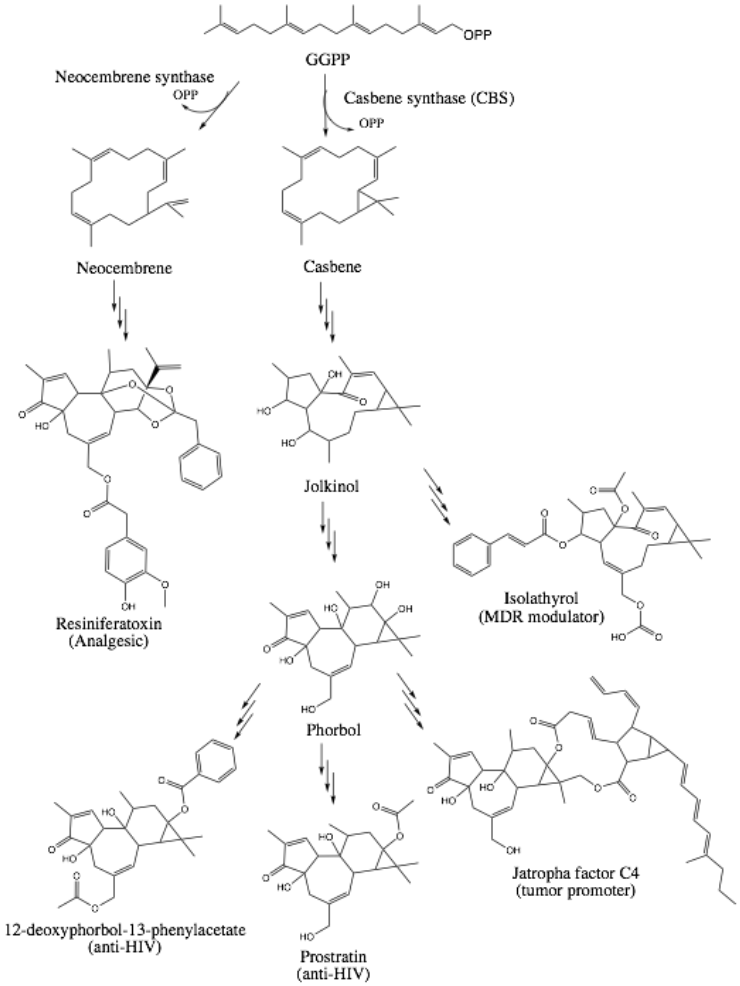


Fig. 2.1. Euphorbiaceae diterpenoid important medicinal derivatives and their predicted pathways. Lathyrane, tigliane and ingenane pathways begin with casbene, which is subsequently oxidized by unknown enzymes to form final products.

2.3. Materials and methods

2.3.1. RNA isolation, CBS elicitation, cDNA library construction, and sequencing

RNA was isolated from *R. communis*, *J. curcas* and *H. nutans* root, mature leaf, young leaf, bark, stem and pectinase elicited tissues using the RNeasy Plant Mini Kit (Qiagen), after freezing and grinding under liquid nitrogen. Young leaves were cut and elicited by soaking in 2000 units of pectinase (Sigma cat. P4716) in 3ml of water on wet paper towels in petri dishes for 20 hours while controls were treated with only water. cDNA libraries were constructed using the Illumina Paired End DNA library construction kit (Illumina, San Diego, CA). Illumina RNA-Seq libraries were prepared with the adapter ligation method (NEBNext Ultra RNA Library Prep Kit for Illumina, New England BioLabs) and barcoded with index oligonucleotides (NEBNext Multiplex Oligos for Illumina, New England BioLabs). All cDNA synthesis for qPCR and cloning was performed using poly(T) primers. A list of biosynthetic genes cloned for

this study are listed in Supplementary figure S2.5. All libraries were sequenced on the Illumina Platform at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 OD018174 Instrumentation Grant. Sequencing reads were aligned to the previously published *J. curcas* reference genome (Sato et al. 2011) with TopHat 2 (version 2.0.13) with a max intron length of 2500 and differential analysis was performed with Cuffdiff (version 2.2.1) with the –frag-bias-correct and the –multi-read-correct flags. Infinite values for differential expression resulting from assigned reads in one condition were assigned a $\log_2$ value of 10. Statistical analysis and plotting was done in R.

2.3.2. Jatropha factor extraction for LCMS-MS analysis

~200mg of elicited and non-elicited *J. curcas* leaves were frozen under liquid nitrogen and ground to a fine powder. The resulting powder was transferred to a container and extracted in methanol by placing in a Branson Ultrasonic Benchtop Cleaner (Model B2510MT) and sonicated for 10 minutes. The resulting solution was pelleted at 4000 rpm for 3 min and the supernatant collected. This process was repeated three times. Combined supernatants were filtered in 10,000 kDa MWCO Amicon Ultra 0.5ml filters (Millipore cat. UFC501096) by spinning at 8000 rpm for 30 min at 4°C. The 10μL of filtered liquid was diluted in 90μL of 50% water (1% formic acid) and 50% acetonitrile immediately before LCMS-MS analysis, and stored at 4°C.

2.3.3. Synthetic oligonucleotides.

Oligonucleotides were commercially synthesized (Integrated DNA Technologies, Inc.).

2.3.4. Yeast strain construction, culture and fermentation conditions, and extraction

The parent *Saccharomyces cerevisiae* strain used for screening CBS was GTy116 {MATa, leu2-3,112::HIS3MX6-GAL1p-ERG19/GAL10p-ERG8;ura3-52::URA3-GAL1p-MvaSA110G/GAL10p-MvaE (codon optimized); his3Δ1::hphMX4-GAL1p-ERG12/GAL10p-IDI1; trp1-289::TRP1_GAL1p-CrtE(*X. dendrorhous*)/GAL10p-ERG20;YPRCdelta15::NatMX-GAL1p-CrtE(opt)/GAL10p-CrtE} previously used by our lab (Reider Apel et al. 2017). The first 51 residues of the JcCBS cDNA were truncated and the resulting sequence was cloned into pRS426-SacII and expressed under control of the GAL1 promoter (Ro et al. 2006). All PCRs used Phusion Hot Start II DNA polymerase (www.thermofisher.com, cat. F549L). The following touchdown PCR cycling conditions were used for all PCRs: 1 cycle of 98°C for 15 sec; 25 cycles of 98°C for 10 sec, 65°C for 30 sec (dropping 1 degree each cycle after the first cycle), 72°C for 30 sec, and then 25 cycles of 98°C for 10 sec, 50°C for 30 sec, 72°C for 30 sec. Transformations were performed via heat-shock using ~ 800 ng plasmid, and 20 min heat shock at 42°C, then plated all cells on selective agarose plates (Gietz and Woods 2002). Colonies were screened by PCR directed at the promoter and front of the coding sequence and one representative colony sequenced. Three biological replicates were analyzed. Selective agar plates used for transformations were purchased from Teknova (www.teknova.com, cat. C3080).

Liquid selective medium used to grow transformants contained 0.2% (w/v) complete supplement mixture (CSM) lacking uracil (www.sunrisescience.com, cat. 1004-100), 0.67% yeast nitrogen base (www.difco.com, cat. 291920), and 2% dextrose. Cultures were grown in glass test tubes for strain maintenance and production runs. Production cultures were cultured for 48 hr in 3 ml of CSM -URA with 1.8% galactose, shaken at 200 rpm, before terpenoid extraction for analysis. All strains were grown at 30°C. Subsequently, cultures were extracted 1:1 with EtOAc spiked with 10 mg/L trans-caryophyllene (Sigma cat. C9653) by shaking for 30 minutes, then spun at 21,952 x g for 1 minute. For GC-MS analysis, the resulting organic phase was removed and transferred to GC vials. A casbene standard containing known concentrations of the internal standard and casbene extracted from a casbene producing strain was used to determine titer.

2.3.5. Casbene synthase expression analysis by qPCR

To determine the relative expression levels of *JcCbs* for the elicitation experiments, RNA samples were extracted as described above and treated with DNaseI (ThermoFisher Scientific RapidOut DNA Removal Kit) to remove genomic DNA contaminations. One-step qRT-PCR was carried out using a SYBR Green-based kit (Invitrogen) on a StepOnePlus Real-Time PCR System (Applied Biosystems). Melting curve analysis was performed according to thermocycler specifications. Relative expression levels of *JcCbs* were analyzed with the ΔΔct method and normalized to expression of the ubiquitin gene using previously described primers (Booranasrisak et al. n.d.). The following primers were designed for amplification of CBS for qPCR analysis: CBS1FWD (CAACCGGCATTTGAATTATTAAG), CBS1RVS (CGAGACGGCACAATGAGTC).

2.3.6. WGCNA for gene candidate analysis of P450 genes from J. curcas and H. nutans

The weighted correlation network analysis (*WGCNA*) R package (Langfelder and Horvath 2008) was used to identify gene co-expression modules from our *J. curcas* pooled tissue and treatment types; Cuffdiff output was loaded into *WGCNA*. The similarity matrix was raised to the power of six to identify co-expression modules. A single block size of 40,000 was used. Sequences and the annotations of genes that coexpressed with *JcCbs* in the turquoise module were selected, and reannotated using HMMER and BLASTP.

2.3.7. OrthoMCL for gene candidate analysis of P450 genes from J. curcas and H. nutans

Orthologs were identified in the proteomes of *A. thaliana* (TAIR: AGI proteins), *E. fischeriana* (Plantrans DB: http://lifecenter.sgst.cn/plantransdb), and *H. nutans* using OrthoMCL (v 1.4) (Li et al. 2003) using an E-value cutoff of 1e-5 and the default parameters. To reduce false grouping of paralogs due to alternative isoforms, a representative transcript defined as the model that produces the longest peptide was used in the orthologous clustering. Gene clusters are displayed in Venn diagram.

2.3.8. Phylogenetic analysis

Phylogenetic analysis was performed using the entire predicted amino acid sequences of *J. curcas* and *H. nutans* P450 family proteins and related *R. communis* and *J. curcas* terpene-modifying P450 proteins from the GenBank database (King et al. 2016; King et al. 2014). Sequence alignments were generated on the basis of comparison of the amino acid sequences using the MAFFT L-INS-i algorithm with default parameters. Alignments for each partition were generated using the default settings (gap opening penalty = 1.53 and offset value = 0.00) (Katoh et al. 2002). A consistent alignment was selected using TrimAl, with the parameter automated1 (Capella-Gutiérrez et al. 2009). Maximum likelihood analyses were conducted with RAxML v.7.2.8 (Stamatakis 2014). Twenty randomized starting trees were generated with which the initial rearrangement setting and the number of distinct rate categories were determined. The best-known likelihood tree was found by performing 1000 repetitions for each of the amino acid datasets. One thousand non-parametric bootstrap replications were then performed using the bootstrap algorithm. The resulting tree was visualized using FigTree (Fig. S2.2). The scale bar of 0.2 indicates a 20% change and each number shown next to the branches is the number of replicate trees in which the related taxa clustered in the bootstrap test.

2.3.9. Agrobacteria tumefaciens strain construction and Nicotiana benthamiana transient expression

Nicotiana benthamiana plants were grown in growth chambers (Percival-Scientific) at 25 °C in 16/8 h light/dark cycles with 60% humidity. Leaves of 4-week-old plants were infiltrated with Agrobacterium tumefaciens strain GV3101 (OD600=1.0) carrying pYB vectors of interest as previously described (Sparkes et al. 2006). Booster plasmids encoding *Coleus forskohlii* 1-deoxy-D-xylulose 5-phosphate synthase (*CfDXS*) and geranylgeranyl pyrophosphate synthase (*CfGGPPS*) were kindly provided by Dr. Johan Andersen-Ranberg and coinfiltrated with appropriate vectors containing other pathway enzymes (Luo et al. 2016). *JcCBS* and candidate P450 genes were expressed in separate plasmids under the 35S-CpMV promoter, constructed using the Gibson assembly method (Gibson et al. 2009). Infiltrated plants were returned to the same growth conditions after infiltration. 200 mg of fresh leaf tissue was collected 4 days post infiltration, weighed out and ground with a bead-beater (Qiagen, TissueLyser) at maximum speed for 5 min, resuspended in hexane, and vortexed for 30 min. The samples were transferred into a 2 ml microtube and spun down at maximum speeds for 1 min at room temperature to clear off the supernatant that was then taken for further analysis. Previously described cloned *R. communis* P450s were used as positive controls for P450 expression (Fig. S2.4). Casbene and resulting oxidized products were analyzed in the supernatant by gas chromatography mass spectrometry.

2.3.10. Metabolite quantification using GCMS and LCMS-MS

An aliquot of the sample (1 µL) was injected into a cyclosil B column (J&W Scientific) operating at a He flow rate of 1 mL/min on GC–MS (GC model 6890, MS model 5973 Inert, Agilent). An initial temperature of 120°C was held for 3 min, followed by ramping to 250°C at a rate of 20°C/min to 250°C, and then held at 250°C for another 3 min. The total flow was set to 8.3 ml/min and helium flow was set to 1 ml/min. A previously described method was used for LCMS-MS analysis of Jatropha factors (Baldini et al. 2014). A 12-O-Tetradecanoylphorbol-13-acetate (Sigma cat. P8139) standard curve was used to quantify Jatropha factors.

2.4. Results and discussion

2.4.1. Functionally testing the J. curcas casbene synthase and eliciting the Jatropha factor pathway

Prior to this study, only three casbene synthases had been functionally tested; these included *R. communis* CBS, *Euphorbia esula* CBS and *Sapium sebiferum* CBS (Kirby et al. 2010). Thus we set out to identify a *J. curcas* CBS that we could use to study additional pathway genes, aiding pathway characterization in this species. Several putative *J. curcas* casbene synthases have been described by Nakano et al., who identified JcCSH as the most likely diterpene synthase candidate for casbene production based on sequence homology (Nakano et al. 2012). Thus, we cloned and expressed JcCSH, hereafter referred to as JcCbs1, on a 2µ plasmid in a GGPP-producing yeast strain to identify casbene synthase from this species. This yeast strain, JWy1, produced ~80 mg/L of casbene.

To produce cDNA libraries for coexpression analysis, *J. curcas* seedlings were harvested at three weeks old. Young leaf, mature leaf, stem, and root tissue types were isolated and. Additionally, we decided to test elicitation of the pathway to further enhance coexpression analysis, as previous studies have shown that casbene synthase is likely involved in plant defense response. Initially, we tested methyl jasmonate (MeJA) as an elicitor of Jatropha factor production, using casbene synthase expression as a proxy for pathway upregulation and measuring Jatropha factor levels using LCMS-MS. Young leaf tissue disks were excised from seedlings and placed in petri dishes with either water or MeJA to elicit the Jatropha factor pathway. Unfortunately, this type of elicitation failed, thus we turned to pectinase as an elicitor, as this fungal enzyme has been shown to elicit casbene synthase expression in *R. communis* (Bruce and West 1982). This resulted in nearly a two-fold increase in Jatropha factor production in the leaves, and an over 800-fold increase in casbene synthase transcript levels after only 20 hours of elicitation (Fig. 2.2). Thus, we pooled pectinase elicited leaf tissues and controls for cDNA libraries along with tissue type cDNA libraries for next generation sequencing.

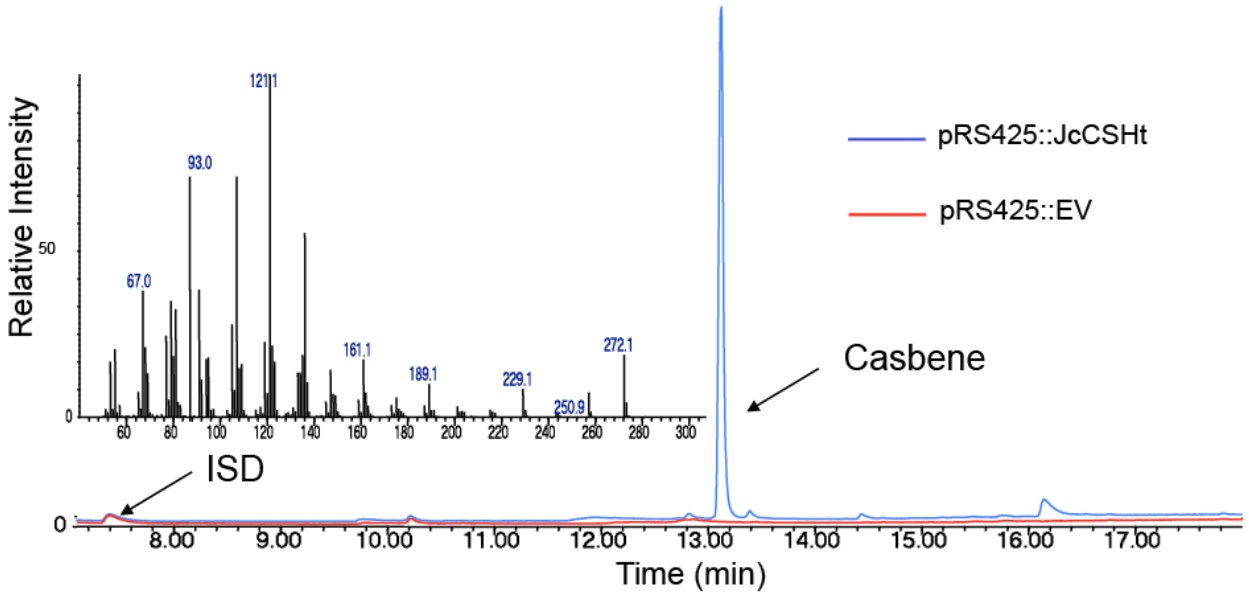


Fig. 2.2. JcCbs1 (JcCSH) produces casbene in engineered yeast. A truncated version of a previously identified putative casbene synthase was coexpressed with *J. curcas* cytochrome P450 reductase (JcCpr1) on a high copy plasmid. Casbene elutes at ~13.2 min. 10 mg/L of trans-caryophyllene was used as internal standard (ISD). Fragmentation pattern of new peak matches the fragmentation of a casbene standard.

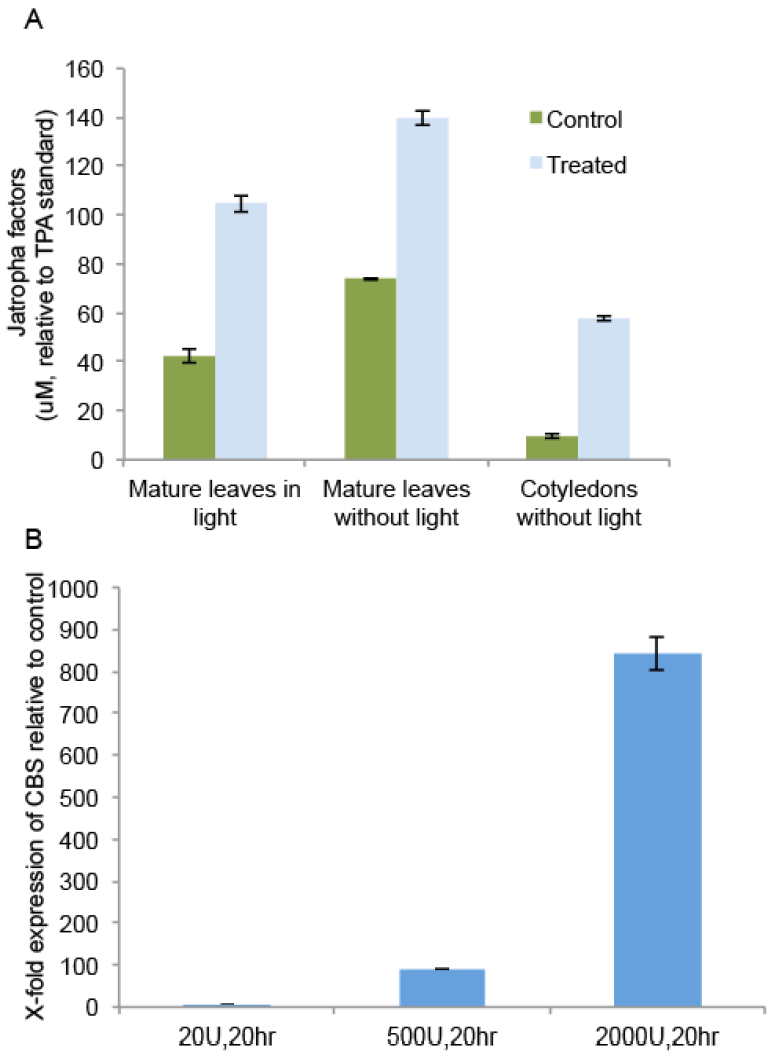


Fig. 2.3. Pectinase elicitation of *J. curcas* young leaves. (A) Pectinase (2000 units, 20hr) treated leaves contained double the concentration of Jatropha factors as water-treated leaves. (B) qPCR shows *JcCBS1* upregulation in pectinase treated leaves relative to water treated controls. Assays were conducted under light and dark conditions to test for optimal induction of *JcCBS1*. Data represent the averages of three biological replicates; error bars show one standard deviation from the mean.

2.4.2. Identifying casbene oxidase candidates in J. curcas

After producing a *J. curcas* transcriptome assembly, we generated a gene expression network using WGCNA (Langfelder and Horvath 2008). This network identified a pool of 967 candidate pathway genes with similar expression to casbene synthase (Fig. 2.4). We noticed that many of the genes in this WGCNA module were related to plant defense; putative flavinoid biosynthetic genes, i.e. CYP82A, genes coexpressed with *JcCBS1*. 32 putative P450s also coexpressed with casbene synthase, as well as three polyketide synthases, and several BAHD acetyltransferases. To test all possible gene candidates, we cloned additional *J. curcas* P450s not identified in the

coexpression analysis, but rather by sequence homology to known *R. communis* casbene oxidizing P450s in the CYP726A clade, as well as those in the CYP71D clade, which are proposed terpene oxidases conserved particularly in Euphorbiaceae species (Zerbe et al. 2013).

2.4.3. Development of a H. nutans transcriptome and identifying additional casbene oxidase candidates

Currently, only three species of plants are known to contain prostratin, namely, *Homalanthus nutans*, *Euphorbia fischeriana*, and *Euphorbia cornigera (Johnson et al. 2008; Miana et al. 1985; Wang et al. 2006)*. Although prostratin has been studied as a resource for latent HIV eradication, the genetic resources for *H. nutans* and *E. cornigera* are completely nonexistent, while only a root transcriptome of *E. fischeriana* exists (Barrero et al. 2011). Our group previously signed an agreement with the Samoan government to microbially produce prostratin. *H. nutans* is an important source for prostratin, a traditional Samoan medicine, with diverse uses from treating hepatitis to circumcision wounds (Wender et al. 2008; Maartens et al. 2014). Thus we sought to identify biosynthetic gene candidates that directly contribute to prostratin formation. We collected tissues of a population of *H. nutans* previously found to produce prostratin in a previous survey (Johnson et al. 2008). We used Illumina sequencing to produce a transcriptome sequence derived from eight samples of mature leaf, young leaf, root and bark cDNA. After assembly of the transcriptome, we identified four P450 candidates with CYP726A homology from the transcriptome (Fig. S2.2). Promisingly, additional analysis using OrthoMCL identified groups of orthologous genes present in the root transcriptome of another prostratin producing species, *Euphorbia fischeriana* (Li et al. 2003) (Fig. 2.5). Indeed, homologs for these P450s appeared to be shared between these species.

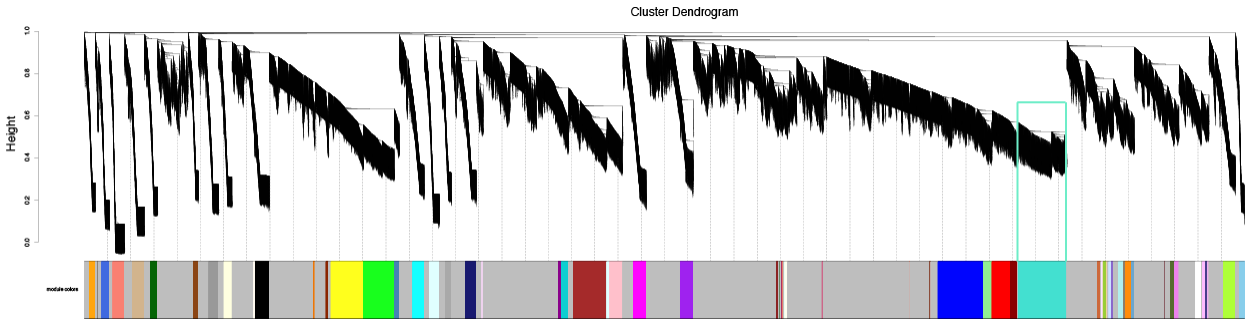


Fig. 2.4. (A) Weighted Gene Correlation Network Analysis (WGCNA) was applied to the 30,854 genes from the *J. curcas* transcriptome. A total of 69 modules were identified, each represented by a color; the algorithm assigned names to modules after unique colors. Genes within each module were strongly correlated with each other across all the samples; 967 genes, including 32 putative P450s, coexpress with *JcCBS1* in the turquoise module, circled in turquoise.

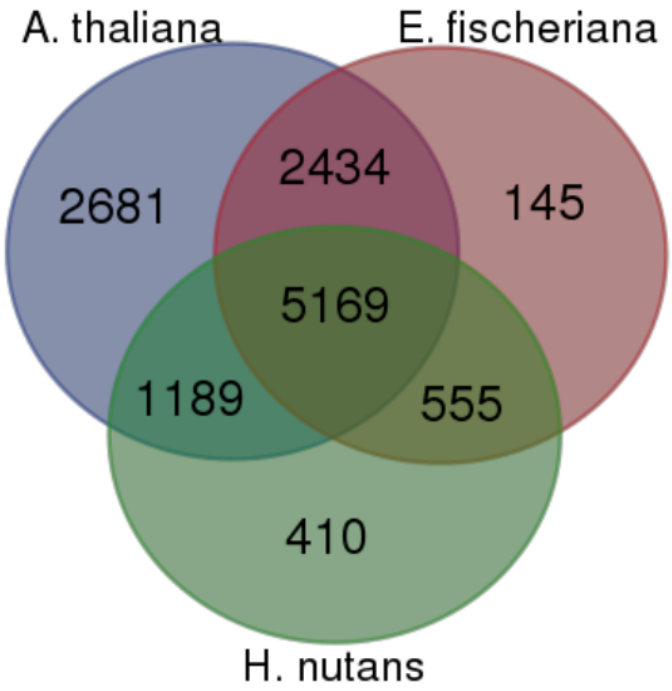


Fig. 2.5. Shared orthologous genes among Euphorbiaceae *H. nutans*, *E. fischeriana*, and *A. thaliana*. Orthologs were identified using OrthoMCL. Genes from different species were considered as orthologs if the shared homology in their deduced amino acid sequences (BlastP, e < $1e10^{-5}$) was more than 50% of the size of the genes being compared. Numbers of gene ortholog clusters shared between species are in black.

2.4.4. Functional testing of P450 candidates reveals three casbene oxidases

Next, we tested several P450 gene candidates for activity in casbene biosynthetic pathways. In general, functional testing of P450s has proven a daunting task in part since P450s make up nearly 1% of a plant's proteome ; microbial heterologous expression is often poor; heterologous hosts lack of organellar membranes P450s are natively targeted ; P450 activity requires co-expression of a suitable cytochrome P450 reductase (CPR) colocalized with the P450; .As a result, active P450 expression is notoriously difficult. Thus, we decided to use *N. benthamiana* as a heterologous host for our studies, as we anticipated the necessity of expressing multiple P450s.

We coexpressed P450s from *H. nutans* and *J. curcas* in *N. benthamiana* with upstream isoprenoid biosynthetic genes *CfDXS*, *CfGGPPS*, and the terpene synthase *JcCBS1*. We saw P450 activity on casbene from three candidates; JcCYP71D495 and JCCYP71D496 produce oxidized forms of casbene that match their previously described mix of 9-keto and 9-hydroxycasbene products (King et al. 2016), while the other candidate, referred to as JcCYP21,

produced hydroxy and keto forms of oxidized casbane with a different fragmentation pattern than the other P450s, albeit at low levels (Fig. 2.6). Although we tested JcCYP726A20, later shown to be a 5,6-casbene oxidase (King et al. 2016), we did not detect production of any oxidized casbene from this P450. Unfortunately, none of the *H. nutans* P450s produced any oxidized casbene.

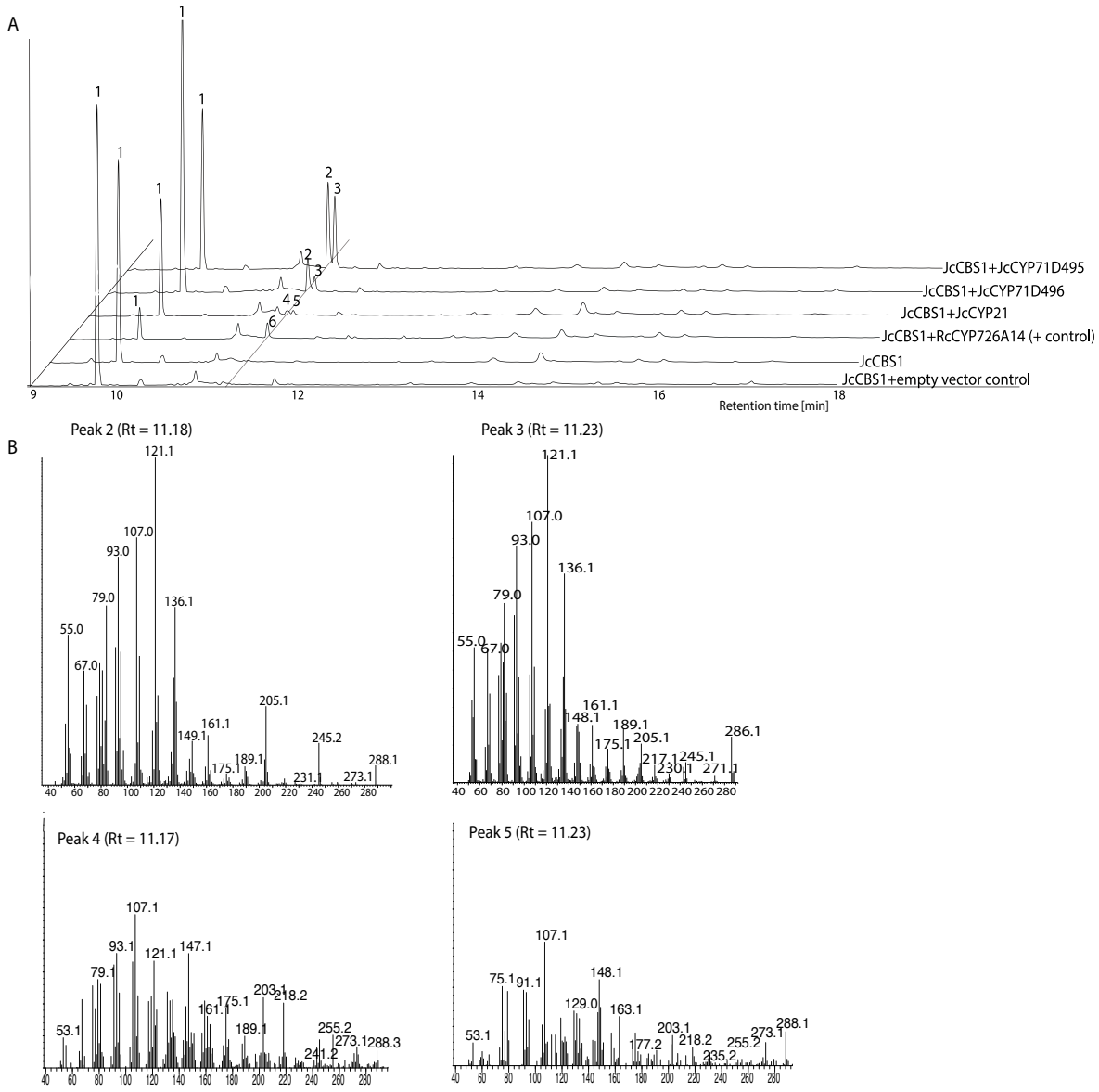


Fig. 2.6. Transiently expressed *J. curcas* P450s form keto- and hydroxycasbene in *N. benthamiana*. (A) GCMS chromatograms of JcCYP71D495 and JcCYP71D496 products show formation of 9-ketocasbene (retention time 17.18) and 9-hydroxycasbene (retention time 17.23) (B) Mass spectra of peaks 2,3,4, and 5. JcCYP21 produces new casbanes singly oxidized at an unknown position. Note, all oxidized casbanes have similar retention times.

2.5. Conclusion

In identifying *J. curcas* casbene synthase, assembling gene expression profiles and transcriptomic resources for *J. curcas* and *H. nutans*, and screening putative casbene oxidases in *N. benthamiana*, we have developed resources for identifying additional gene candidates responsible for medicinal diterpenoid biosynthesis in Euphorbiaceae. Using WGCNA and homology searches allowed us to identify JcCYP71D445, JcCYP71D446 and JcCYP21 as casbene oxidases. Although the P450s we retrieved from the *H. nutans* transcriptome and ortholog analysis did not oxidize casbene, we believe that additional prostratin biosynthetic genes may be shared in the ortholog groups shared between *H. nutans* and *E. fischeriana* identified in this study. The titers of casbene produced by *S. cerevisiae* suggest that this may be the optimal host for screening additional candidate biosynthetic genes.

3. Chapter 3 — High-titer production of lathyrane diterpenoids from sugar by engineered *Saccharomyces cerevisiae*

Including material from submitted work: Wong, J., de Rond, T., d'Espaux, L., van der Horst, C., Dev, I., Rios-Solis, L., Kirby, J., Scheller, H., Keasling, J. D. (2017). High-titer production of lathyrane diterpenoids from sugar by engineered *Saccharomyces cerevisiae*

3.1. Abstract

Euphorbiaceae are an important source of medically important diterpenoids, such as the anticancer drug ingenol-3-angelate and the antiretroviral drug prostratin. However, extraction from the genetically intractable natural producers is often limited by the small quantities produced, while the organic synthesis of terpene-derived drugs is challenging and similarly low-yielding. While transplanting the biosynthetic pathway into a heterologous host has proven successful for some drugs, it has been largely unsuccessful for diterpenoids due to their elaborate biosynthetic pathways and lack of genetic resources and tools for gene discovery. We engineered casbene precursor production in *S. cerevisiae*, verified the ability of six *Euphorbia lathyris* and *Jatropha curcas* cytochrome P450s to oxidize casbene, and optimized the expression of these P450s and an alcohol dehydrogenase to generate jolkinol C, achieving ~800 mg/L of jolkinol C and over 1 g/L total oxidized casbanes, the highest titer of oxidized diterpenes in yeast reported to date. This strain enables the semisynthesis of biologically active jolkinol C derivatives and will be an important tool in the elucidation of the biosynthetic pathways for ingenanes, tiglianes, and lathyranes. These findings demonstrate the ability of *S. cerevisiae* to produce oxidized drug precursors in quantities that are sufficient for drug development and pathway discovery.

3.2. Introduction

A range of medicinal diterpenoid compounds (*e.g.,* casbanes, lathyranes, jatrophanes, tiglianes, and ingenanes) produced solely in Euphorbiaceae and Thymelaceae plants have gained interest due to their unique anti-cancer, anti-HIV, vascular-relaxing, neuro-protective, anti-inflammatory, and immunomodulatory activities (Vasas and Hohmann 2014; Srivalli and Lakshmi 2012; Jiao et al. 2009; Halaweish et al. 2002; Wang et al. 2015). These compounds include ingenol-3-angelate, recently approved by the US FDA for the treatment of the premalignant skin condition actinic keratosis (Siller et al. 2009) (Fig. 3.1b); prostratin, in phase I clinical trials as an adjuvant therapy to clear latent viral reservoirs; and resiniferatoxin, studied for its powerful analgesic effects (Johnson et al. 2008; Payne et al. 2005). Indeed, the demand for medicinal diterpenoids is exemplified by the rise of paclitaxel and other Taxus-derived drugs, which is estimated to be a $2.3 billion industry worldwide (Bishop and Pagiola 2012).

The medicinal properties and unusual structures of lathyrane diterpenoids, a group of tricyclic diterpenes derived from the hydrocarbon casbene, as well as tigliane and ingenane diterpenes have led to many attempts at chemical synthesis and direct purification from plants (Shimokawa et al. 2007; Wender and McDonald 1990; Tanino et al. 2003; Vasas and Hohmann 2014). Prostratin is produced via a semisynthetic route, and ingenol-3-angelate by mass extraction from *Euphorbia peplus* plants. However, chemical synthesis is complicated by the multiple chiral centers in these diterpenoid compounds, while extraction from plant tissues is inefficient and costly, with yields typically in the range of 0.1% to 10% from starting material (Tanino et al. 2003; Wender et al. 2008). The development of a microbial production host for lathyrane diterpenoids could significantly reduce the production cost and increase the availability of these compounds. Such a host could also produce other medically important diterpenoids that are not produced naturally or at levels insufficient to detect.

Yeast is a particularly attractive host for microbial engineering because of its robustness in industrial fermentation and genetic tractability. The development of yeast strains producing >10 g/L of the anti-malarial artemisinic acid has demonstrated the potential of using yeast as a heterologous host for oxidized natural product production(Paddon et al. 2013; Westfall et al. 2012). Development of monoterpenoid (C5) and diterpenoid (C20) production in yeast has lagged behind sesquiterpenoid (C15) production, with fewer studies and lower titers (<20mg/L and <1 g/L, respectively) (Trikka et al. 2015; Ignea et al. 2014; Jongedijk et al. 2015; Amiri et al. 2016; Ignea et al. 2011; Andersen-Ranberg et al. 2016).

The biosynthesis of lathyrane isoprenoids begins with the common five-carbon isoprenoid precursors IPP and DMAPP, which are sequentially condensed to form the universal 20-carbon isoprenoid intermediate geranylgeranyl-pyrophosphate (GGPP). A diterpene synthase found in several Euphorbiaceae, casbene synthase (CBS), cyclizes GGPP to form casbene (Fig. 3.1a) (Kirby et al. 2010). From casbene, the biosynthetic route to lathyrane isoprenoids is poorly understood, but is thought to proceed through intermediates such as jolkinol C *via* cytochrome P450-catalyzed oxidations and possibly a short-chain alcohol dehydrogenase (ADH). Different groups have debated the involvement of the ADH in the formation of jolkinol C; Luo and colleagues found that, in addition to the CYPs required to oxidized casbene, an *Euphorbia lathyris* ADH (ElAdh1) was necessary for jolkinol C formation in *N. benthamiana* and *in vitro* (Luo et al. 2016). In contrast, King and colleagues never mentioned the use of an ADH for the production of jolkinol C, but saw production of jolkinol C from coexpression of casbene synthase and two *Jatropha curcas* CYPs (King et al. 2016).

Reconstituting complex oxidized terpenoid biosynthetic pathways, such as those of lathyranes, in heterologous hosts has posed numerous problems requiring optimizing the expression of a cytochrome P450 reductase (CPR) and the CYP, as well as balancing the redox environment within the cell (Paddon et al. 2013; Renault et al. 2014). Indeed, only about 40% of all plant-derived CYPs tested in yeast express poorly if at all (Renault et al. 2014). It is important to note that yeast has proved to be a better host for CYP expression than prokaryotic hosts such as

Escherichia coli due to the presence of the endoplasmic reticulum in the former organism. A previous study found that despite the coexpression of the pathway enzymes required to produce jolkinol C in *Nicotiana benthamiana* and *in vitro*, coexpression of these enzymes in yeast did not result in jolkinol C formation (Luo et al. 2016). This research illustrates the need to engineer casbene and lathyrane pathways in yeast to aid functional testing of additional pathway enzymes.

The present work describes the establishment of yeast strains producing lathyrane diterpenoids as a platform for producing varied medicinal compounds. We have constructed a strain that synthesizes the lathyrane backbone casbene at high titer. We verified the ability of *E. lathyris* and *J. curcas* P450s to oxidize casbene, and resolved some discrepancies between the two pathways previously proposed (Luo et al. 2016; King et al. 2016). Finally, we optimized the expression of these P450s and an ADH to generate jolkinol C at high titers producing a strain with all pathway genes stably integrated as a chassis for additional gene discovery.

3.3. Methods

3.3.1 Strain construction

The parent *Saccharomyces cerevisiae* strain used for all engineering was GTy116 {MATa,leu2-3,112::HIS3MX6-GAL1p-ERG19/GAL10p-ERG8;ura3-52::URA3-GAL1p-MvaSA110G/GAL10p-MvaE (codon optimized); his3Δ1::hphMX4-GAL1p-ERG12/GAL10p-IDI1;trp1-289::TRP1_GAL1p-CrtE(X.den)/GAL10p-ERG20;YPRCdelta15::NatMX-GAL1p-CrtE(opt)/GAL10p-CrtE} previously used by our lab (Reider Apel et al. 2017). The integration cassettes for all subsequent strains (Table 1, Table S1) were created using the software tools CASdesigner (casdesigner.jbei.org) and DIVA(diva.jbei.org) and integrated using the previously reported, cloning-free methodology via Cas9-aided homologous recombination (Reider Apel et al. 2017). Integration cassettes containing 1-kb flanking homology regions targeting a chosen genomic locus were constructed by PCR amplifying donor DNA fragments using primers generated by CASdesigner, then co-transformed with a Cas9-gRNA plasmid (pCut) targeting the chosen genomic locus. CASdesigner primers provide 30–60 nt of inter-fragment homology allowing 1–5 separate fragments to assemble via homologous recombination *in vivo*. pCuts targeting empty genomic loci (e.g., 208a, 1622b) were available pre-cloned, and pCuts targeting new sites (e.g., for deletions) were assembled in vivo from a linear backbone and a linear PCR fragment containing the new gRNA sequence, as described previously (Reider Apel et al. 2017). The new gRNA sequence for the URA3 locus (Suppl. Table 1) was chosen using DNA2.0 (www.dna20.com/eCommerce/cas9/input). To generate donor DNA fragments, native sequences—e.g., chromosomal homology regions, promoters—were amplified from CEN.PK2-1C genomic DNA, while heterologous sequences—e.g., P450 coding sequences (Fig. S3.8)—were amplified from synthetic gene blocks codon-optimized (for expression in *S. cerevisiae*) and synthesized by Integrated DNA Technologies (www.idtdna.com).

All PCRs used Phusion Hot Start II DNA polymerase (www.thermofisher.com, cat. F549L). The following touchdown PCR cycling conditions were used for all PCRs: 1 cycle of 98°C for 15

sec; 25 cycles of 98°C for 10 sec, 65°C for 30 sec (dropping 1 degree each cycle after the first cycle), 72°C for 30 sec, and then 25 cycles of 98°C for 10 sec, 50°C for 30 sec, 72°C for 30 sec. Transformations were performed via heat-shock using ~ 200 ng pCut, ~ 1 µg donor DNA per sample, and 20 min heat shock at 42°C, then plated all cells on selective agarose plates (Gietz and Woods 2002). For assembling a pCut targeting a new site by homologous recombination, we used 200 ng linear pCut backbone and 500 ng of a 1- kb fragment containing the gRNA sequence, as described (Reider Apel et al. 2017). For multi-site integrations, we used 200 ng total linear pCut backbone, and the same amounts of gRNA fragment and donor DNA for each site as we would have for a single integration. Colonies were screened by PCR directed at the target locus, and for integrations, one representative colony sequenced. Three to four biological replicates were analyzed for each strain.

3.3.2 Development of chimeric JcCYP726A35

To produce a JCCYP726A35 chimera for expression in yeast, the length of the plastidial transit sequence was estimated using ChloroP prediction, at about 132 residues. The ER localizing tag of an ER localized C5ox CYP, ElCYP726A27, was annotated by ChloroP and the front 36 residues were used to replace the ChloroP predicted native plastidial transit sequence of JcCYP726A35 (EMANUELSSON et al. 1999). The resulting chimeric protein was ordered from IDT.

3.3.3. Synthetic genes and oligonucleotides

Oligonucleotides and synthetic genes were commercially synthesized (Integrated DNA Technologies, Inc.). All codon optimized sequences were designed based on the IDT online tool. Sequences of synthetic genes can be found in Figure S3.8.

3.3.4. Culture and fermentation conditions

Selective agar plates used for transformations were purchased from Teknova (www.teknova.com, cat. C3080). Liquid selective medium used to grow transformants contained 0.2% (w/v) complete supplement mixture (CSM) lacking uracil (www.sunrisescience.com, cat. 1004-100), 0.67% yeast nitrogen base (www.difco.com, cat. 291920), and 2% dextrose. Nonselective medium contained 1% yeast extract, 2% peptone (Difco cat. 288620 and 211677, respectively), and either 2% dextrose (YPD) or 2% galactose and 0.2% dextrose (YPG). Nonselective agar YPD plates were purchased from Teknova (cat. Y100). Cultures were grown in plastic 96-deep well plates (www.vwr.com, cat. 29445-166) and glass test tubes for strain maintenance, while 2 ml of media in 24-deep well plastic plates (CWR cat. 89080-534) were used for all production runs. Production cultures were spiked with 50 mg/L trans-caryophyllene (sigma cat. C9653) as an internal standard. Plastic plates were covered with aeraseal film (www.excelscientific.com, cat. BS-25) and shaken at 800 rpm in a Multitron shaker

(www.infors-ht.com, model AJ185). Production runs were cultured for 48 hr in 2 ml of YPG before terpenoid extraction for analysis. Glass tubes were shaken at 200 rpm. All strains were grown at 30°C.

3.3.5. Confocal microscopy

To visualize GFP expression of tagged CBS variants in yeast strains, strains were grown in 5 ml YPD overnight, then back-diluted 1:100 into the same medium and grown 3–6 h at 200 rpm and 30 °C. Then, 1 ml of culture volume was centrifuged at 10,000 rpm on a table-top centrifuge, washed with 1x water, and 1ul of the cell pellet was imaged using a Zeiss LSM 710 confocal system mounted on a Zeiss inverted microscope (www.zeiss.com) with a 63Å~ objective and processed using Zeiss Zen software.

3.3.6. Metabolite quantification using GCMS and HPLC-UV

Yeast cultures grown in 2 ml of YPgal in 24-deep well plastic plates for 48 hours were extracted 1:1 with EtOAc spiked with 50 mg/L trans-caryophyllene by shaking for 30 minutes, then spun at 21,952 x g for 1 minute. For GC-MS analysis, the resulting organic phase was removed and transferred to GC vials. An aliquot of the sample (1 µL) was injected into a cyclosil B column (J&W Scientific) operating at a He flow rate of 1 mL/min on GC–MS (GC model 6890, MS model 5973 Inert, Agilent). An initial temperature of 120°C was held for 3 min, followed by ramping to 250°C at a rate of 20°C/min to 250°C, and then held at 250°C for another 3 min. The total flow was set to 8.3 ml/min and helium flow was set to 1 ml/min. For HPLC analysis, the resulting organic phase was removed and dried down at 54°C in vacuum, and the extract was redissolved in 500 µL of MeOH. 25 µL of the extract was analyzed on an Agilent HPLC 1200 series using a Zorbax Eclipse XDB-C18 column (Agilent, 5-micron, 4.6 mm x 250 mm), kept at 40 C. Mobile phases A and B were water and methanol respectively. A flow of 1ml/min was used. The gradient profile was as follows: 5 min constant at 10% B, a linear gradient from 10% B to 100% B in 10 min, held for 13 min, a second linear gradient from 100% B to 10% B for 1 min, and final step of 10% B maintained for 11 min. The following wavelengths were monitored: 204 nm, 254 nm, 270 nm, and 280 nm, and 270 nm was used for quantification. All production measurements were performed in biological triplicates or quadruplicates. A casbene standard containing known concentrations of the internal standard and casbene extracted from a casbene producing strain was used to determine titer. Farnesol (sigma cat. F203) was used as a standard for singly hydroxylated casbanes, while authentic standards of jolkinol C, 6,9-dihydroxy-5-ketocasbene, and 6-hydroxy-5-ketocasbene were isolated and weighed and verified by H^1 NMR from JWY521 culture to quantify these products of our strains.

3.4. Results

3.4.1. Engineering casbene production in yeast

Past studies have reported casbene production in *S. cerevisiae* by expressing the casbene synthase gene (*CBS*) from plasmids (Kirby et al. 2010; Luo et al. 2016). To date, titers remain under 30 mg/L, hampering development of advanced lathyrane diterpenoid production strains. Previous studies expressing plant-derived diterpene synthases in yeast suggest protein insolubility as a limiting factor in heterologous expression and demonstrated improved activity by tagging with soluble proteins (Reider Apel et al. 2017; Ignea et al. 2015). To develop a strain for high-level production of casbene, we optimized soluble expression of Cbsp using protein tagging strategies. We integrated a truncated version of the gene encoding *Jatropha curcas* Cbsp *(JcCBS1)* alone or with various protein tags into a GGPP-producing strain developed by our group, GTy116 (Reider Apel et al. 2017; Nakano et al. 2012). These tags included a maltose binding protein (MBP, codon-optimized for yeast expression) attached to the N-terminus of JcCbs1p, which led to 15-fold improvement in casbene titers over the untagged JcCbs1p variant (Fig. 3.2a). Two other constructs, green fluorescent protein (GFP) attached to the C-terminus of JcCbs1p, and a yeast codon-optimized *JcCBS1*, both showed similar casbene titers as the untagged, non-codon optimized JcCbs1p. Ignea and colleagues reported an over 200-fold increase in diterpene production when they expressed a mutated GGPP-producing *S. cerevisiae ERG20* variant, *ERG20(F96C)* in a diterpene strain. Fusion of the protein encoded by this gene to the diterpene synthase in their system appeared to improve the titer further, likely due to increased solubility of the terpene synthase (Ignea et al. 2015). We screened the fusion protein JcCbs1-Erg20(F96C)p and saw modest titer improvements. However, a two-tag variant of JcCbs1p with an MBP tag on the N-terminus and an Erg20(F96C)p fusion on the C-terminus outperformed all other JcCbs1p variants, resulting in ~80 mg/L casbene. To examine protein solubility improvement from protein tagging strategies, we developed a 3-tag construct, MBP-JcCbs1-Erg20(F96C)-GFPp, for imaging. This strain showed significantly elevated and disbursed GFP expression within the cell relative to the JcCbs1-GFPp construct, indicating that the MBP tag and Erg20(F96C)p fusion indeed aided expression, possibly due to increased protein stability (Fig. 3.2b). A strain with three additional copies of *CBS* (hereafter referred to as JWY509) led to a final titer of ~160 mg/L casbene, by far the highest titer of casbene in yeast reported to date.

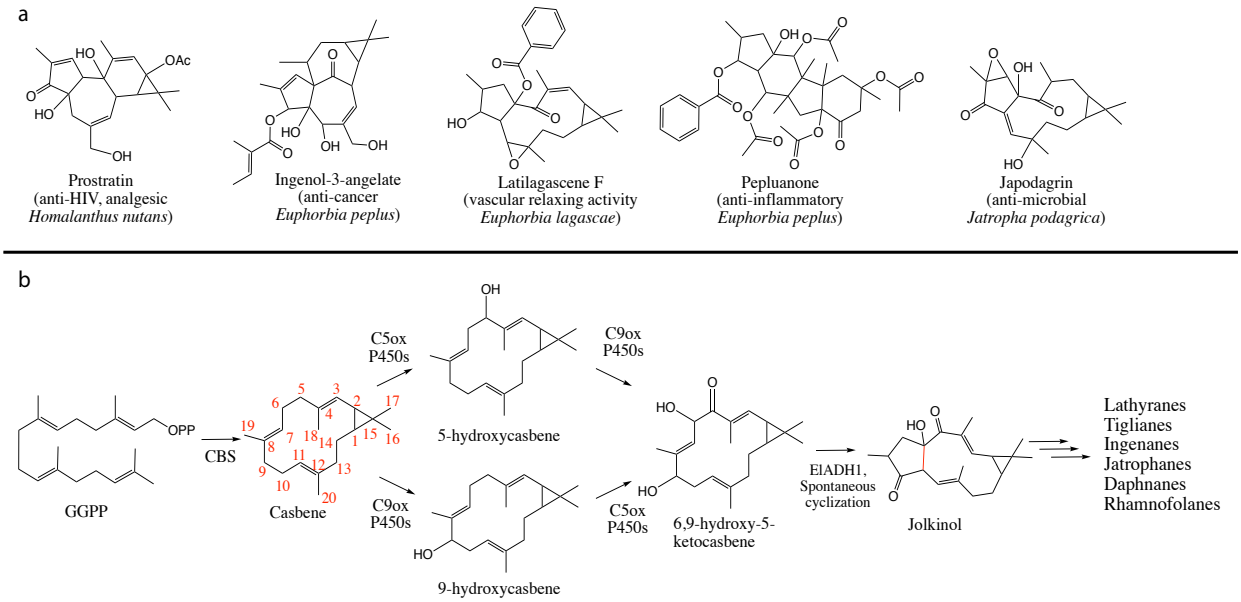


Fig. 3.1. Euphorbiaceae diterpenoid biosynthetic pathways originate from casbene. (a) Structures of bioactive Euphorbiaceae diterpenes. (b) Proposed pathway for the production of medicinal Euphorbiaceae diterpenoids begins with geranylgeranyl diphosphate (GGPP) from the DXP pathway in plants, which is cyclized by casbene synthase (CBS) and subsequently oxidized by numerous P450s. Enzymes required to form jolkinol C in both *J. curcas* and *E. lathyris* have been functionally characterized. C9OX CYP (ElCYP71D445p/JcCYP71D495p) homologs oxidize the C-9 position of casbene, while C5OX CYP (ElCYP726A27p/JcCYP726A35p/JcCYP726A20p) homologs oxidize the C-5 position of casbene. Consequently, both enzymes have been shown to catalyze the hydroxylation of C-6 position of casbene. The formation of a carbonyl at position 5n has been postulated to result in the tautomerization and spontaneous ring closure forming jolkinol C. Hamberger and colleagues have shown the necessity of the alcohol dehydrogenase ElAdh1p to form jolkinol C *in vitro* using *E. lathyris* enzymes (Luo et al. 2016). Additional steps towards production of decorated diterpenes are diverse and unknown.

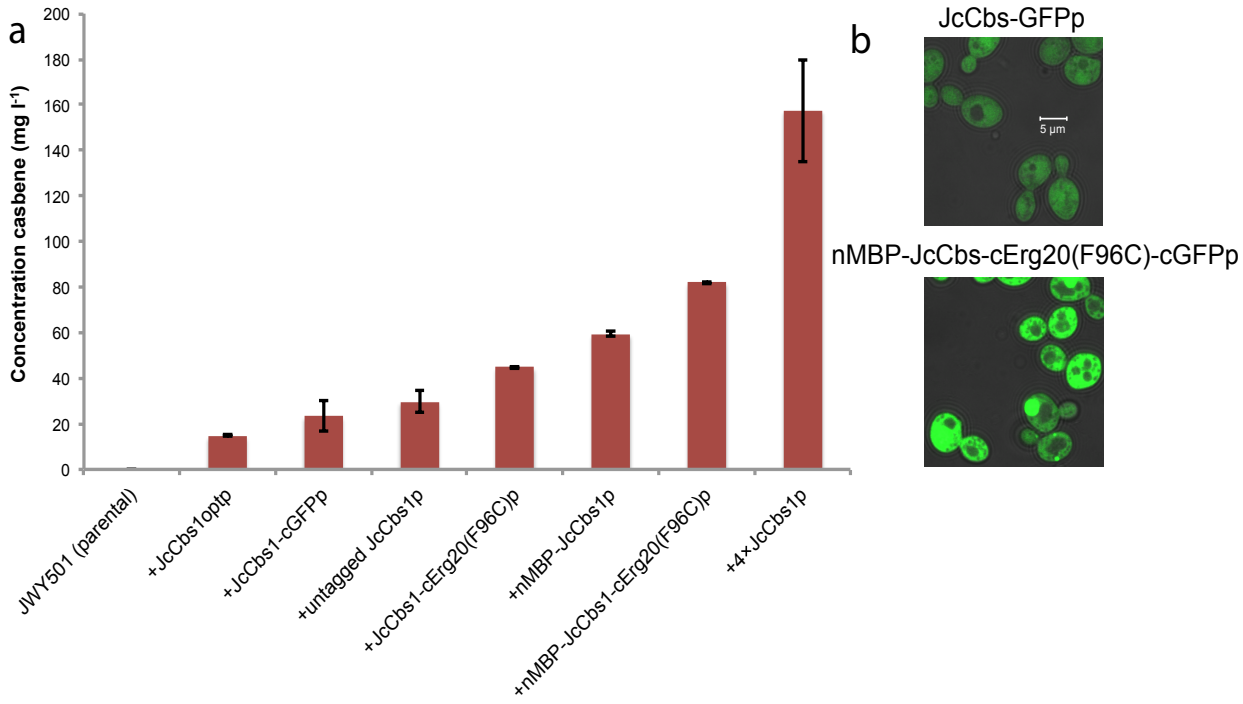


Fig. 3.2. Engineering casbene production in yeast. (a) Different protein tags attached to *J. curcas* casbene synthase (JcCbs1p) improved casbene titer. All *JcCBS1* variants were integrated into the GGPP-producing parent strain, JWY501. Strain JWY509, containing four copies of *JcCBS1*, was used as the parent for all P450 testing in subsequent studies. Data represent the averages of three biological replicates; error bars show one standard deviation from the mean. The lowercase letter in front of the tag name indicates the terminus of CBS to which the tag was attached, ie N- or C-terminus. (b) GFP-tagging experiments with JcCbs1p reveal increased expression and protein stability with solubility tags. Confocal microscopy studies show low expression from a single integrated copy of GFP fused JcCbs1p in yeast (top). A single integrated copy of three tag *JcCBS1* (producing nMBP-JcCbs1-cErg20[F96C]-cGFPp) results in increased expression levels (bottom).

3.4.2. Engineering CYP expression in yeast.

Besides the synthesis of the diterpene casbene, lathyrane biosynthesis involves CYP-catalyzed oxidation at three positions of the casbene hydrocarbon skeleton, as well as the formation of a C-C bond between carbons 6 and 10 (see numbering scheme in Fig. 3.1b). We examined CYPs known and hypothesized to act on casbene in our high-titer casbene-producing yeast strain JWY509. CYPs from *J. curcas* and *E. lathyris* reported to act on C-9 of casbene (hereafter referred to as C9OX CYPs) and ones that act on C-5 and C-6 (hereafter referred to as C5OX CYPs) were cointegrated with a *CPR*—required for channelling electrons from NADPH to the CYP heme domain—found in the transcriptome of *J. curcas*, hereafter referred to as *JcCPR1* (Luo et al. 2016; King et al. 2016). We used JcCpr1p for both the *E. lathyris* and *J. curcas* CYPs because it has been shown that plant P450-CPR interacting domains are highly conserved across

plants species, and previous studies have had success using non-cognate CPRs within Euphorbiaceae for functional testing of casbene C9OX and C5OX CYPs (Luo et al. 2016; Jensen and Møller 2010).

JWY509 cultures expressing C9OX CYPs and JcCpr1p produced 9-hydroxycasbene, 9-ketocasbene, or a mixture of both, as identified by GC-EIMS fragmentation patterns (Fig. 3.3, Fig. S3.10). Additionally, we decided to test the CYP adjacent to the C9OX2 in the published *J. curcas* jolkinol C biosynthetic gene cluster (Fig. S3.1) (King et al. 2016). This enzyme, JcC9OX1p, whose activity was not previously reported, showed low-level activity on casbene at the C-9 position when expressed in our casbene-producing yeast strain. Of the C9OX CYPs tested, the most productive enzyme was the codon-optimized *E. lathyris* ElC9OX1p, which consumed 90% of the casbene relative to the parental strain, while producing more than 44 mg/L of C-9 oxidized casbanes. The most productive enzyme from *J. curcas* was JcC9OX2p (in JWY511), which consumed 70% of the casbene relative to the parental strain, while producing 35 mg/L C-9 oxidized casbene. Interestingly, ElC9OX1p produced primarily 9-hydroxycasbene, while the other C9OX candidates produced a relatively equal mix of the 9-ketocasbene and 9-hydroxycasbene intermediates (Fig. S3.3). Due to ElC9OX1p showing both the highest total C-9 oxidation activity, as well as product specificity for 9-hydroxycasbene—our desired intermediate—we chose this strain for further pathway engineering.

Table 1| Yeast strains and gene nomenclature

Strain	Parent (+ additional genetic changes)	Compound	References
GTY116	*MATa leu2-3,112::His3MX6_P_{GAL1}-ERG19/P_{GAL10}-ERG8 ura3-52::URA3_P_{GAL1}-mvaS(A110G)/P_{GAL10}-mvaE(CO) his3Δ1::hphMX4_P_{GAL1}-ERG12/P_{GAL10}-IDI1 trp1-289::TRP1_P_{GAL1}-crtE(X.den)/P_{GAL10}-ERG20 yprcδ15::natMX_P_{GAL1}-crtE(opt)/P_{GAL10}-crtE*	GGPP	ref. 1
JWY501	GTY116 (*ura3-52* prototrophy removed for use of Cas9 system)	GGPP	This work
JWY509	JWY3 (*ARS1622b::P_{GAL1}-CBS ARS1014a::P_{GAL1}-nMBP-CBS, P_{GAL10}-nGFP-CBS ARS308a::P_{GAL1}-nMBP-CBS-erg20F96c*)	Casbene	This work
JWY510	JWY509 (*ARS911b::P_{GAL10}-JcCPR1, P_{GAL1}-JcC9OX1*)	C-9 oxidized casbanes	This work
JWY511	JWY509 (*ARS911b::P_{GAL10}-JcCPR1, P_{GAL1}-JcC9OX2*)	C-9 oxidized casbanes	This work
JWY512	JWY509 (*ARS911b::P_{GAL10}-JcCPR1, P_{GAL1}-ElC9OX1*)	C-9 oxidized casbanes	This work
JWY513	JWY509 (*ARS911b::P_{GAL10}-JcCPR1, P_{GAL1}-ElC5OX1*)	C-5 oxidized casbanes	This work
JWY514	JWY509 (*ARS911b::P_{GAL10}-JcCPR1, P_{GAL1}-JcC5OX1*)	C-5 oxidized casbanes	This work
JWY515	JWY509 (*ARS911b::P_{GAL10}-JcCPR1, P_{GAL1}-JcC5OX2*)	C-5 oxidized casbanes	This work
JWY516	JWY512 (ARS*1021b::P_{GAL1}-JcC5OX2*)	Jolkinol	This work
JWY517	JWY512 (ARS*1021b::P_{GAL1}-JcC5OX1*)	Mix of C-5/C-9 oxidized c	This work
JWY518	JWY512 (ARS*1021b::P_{GAL1}-ElC5OX1*)	Mix of C-5/C-9 oxidized c	This work
JWY519	JWY511 (*ARS1021b::P_{GAL1}-JcC5OX2*)	Jolkinol	This work
JWY520	JWY516 (*HIS3b::P_{GAL1}-JcADH1*)	Jolkinol	This work
JWY521	JWY519 (*HIS3B::P_{GAL1}-JcADH1*)	Jolkinol	This work

Abbreviation	Published name	References
JcCBS1	*JcCSH*	ref. 11
JcC9OX1	*JcCYP71D496*	ref. 12
JcC9OX2	*JcCYP71D495*	ref. 12
ElC9OX1	*ElCYP71D445*	ref. 13
ElC5OX1	*ElCYP726A20*	ref. 13
JcC5OX1	*JcCYP35a20*	ref. 12
JcC5OX2	*JcCYP726A20*	ref. 12

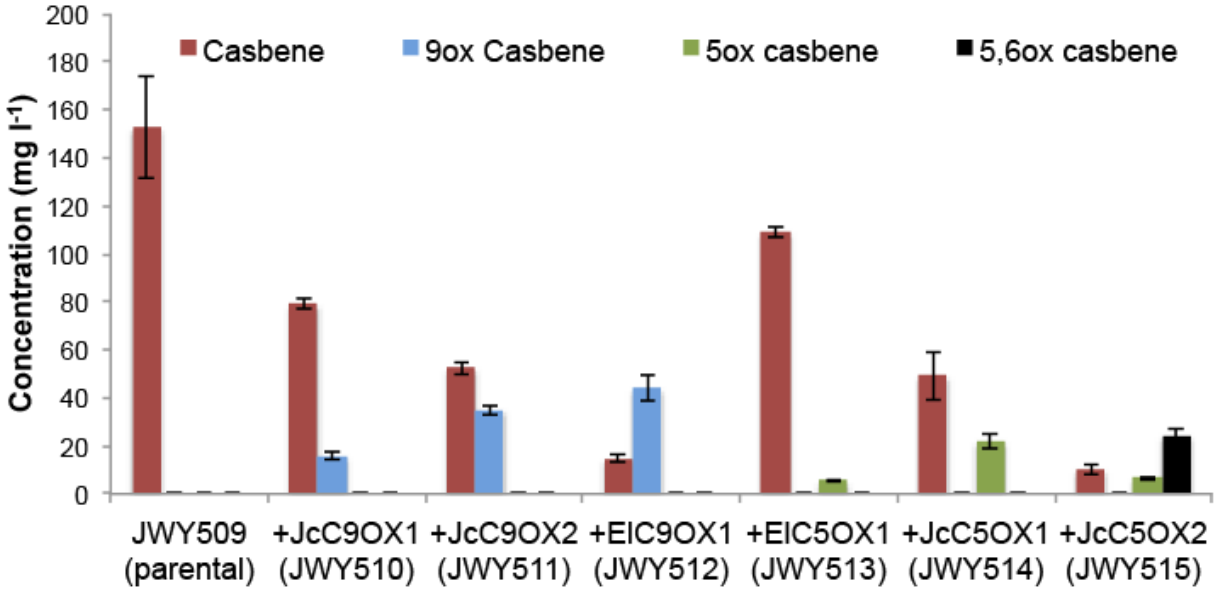


Fig. 3.3. Titers of casbene, singly- and doubly-oxidized casbanes (sum of hydroxy and keto forms), detected in the presence of different C9OX and C5OX CYP variants. All CYP variant constructs were individually cointegrated with the *J. curcas CPR* (*JcCPR1*) into the high titer casbene strain JWY509. The formation of 9-ketocasbene has been shown to be a dead-end product and is likely due to overactivity of C9OX on an accumulating 9-hydroxycasbene, resulting in double oxidation at the C-9 position. We presume that 9-hydroxycasbene will be consumed in the presence of a C5OX CYP (Luo et al. 2016), so we estimated CYP activity by the combined accumulation of C-9 oxidized casbanes. Cultures were extracted after 48 hours and analyzed for oxidized casbane production by GC-MS. ElC9OX1p outperforms all C9OX CYPs in C-9-oxidized casbene production, while JcC5OX2p outperforms all the other C5OX CYPs in production of C-5/C-6 oxidized casbene. Data represent the averages of three replicate cultures; error bars show s.d. One of the C5OX CYPs, JcC5OX1p, has been shown to localize to the plastidial membrane rather than the endoplasmic reticulum when heterologously expressed in *N. benthamiana* (King et al. 2016). To test the expression of this CYP in yeast, we designed a chimeric version of the enzyme by replacing the predicted N-terminus plastidial transit sequence with the beginning of the ER-targeting ElC5OX1p sequence. The resulting chimeric protein was a similar length to the non-plastid localizing C5OX CYPs (EMANUELSSON et al. 1999) (Fig. S3.8). Upon expressing C5OX CYPs in JWY509, we noticed the production of 5-ketocasbene as well as 6-hydroxy-5-ketocasbene from only JcC5OX2p, which is consistent with a previous report using the *N. benthamiana* expression system (King et al. 2016) (Fig. 3.3b). Both other C5OX CYPs produced only 5-ketocasbene at lower titers.

3.4.3. Combining C9OX and C5OX CYPs to produce jolkinol C.

After we found the most productive C9OX and C5OX CYPs, we decided to test our C5OX CYPs in the best C9OX CYP strain, JWY512, for production the desired product jolkinol C. It

was unclear what the products of coexpressing C5OX CYPs in strain JWY512 would be, given that previous studies disagree on this matter (Luo et al. 2016; King et al. 2016). Despite detecting jolkinol C in *N. benthamiana* when expressing *E. lathyris* C9OX CYP and C5OX CYP, coexpressing these enzymes in yeast did not result in the production of jolkinol C, but rather the formation of primarily 5-hydroxy-9-ketocasbene, a purported dead-end product (Luo et al. 2016). This result may be attributed to differences in pH or some intracellular environmental condition responsible for the ring closure in plant cells that is different in yeast cells. We therefore tested C9OX CYP and C5OX CYP combinations in yeast and analyzed product profiles by GC-MS, HPLC, and NMR (Fig. S3.4, S3.5, and S3.6). HPLC-UV analysis was possible since casbanes and lathyranes with α,β-unsaturated carbonyls absorb in the 270-290 nm range (Seip and Hecker 1983). Indeed, when these samples were run on HPLC, we saw the appearance of three major peaks, which NMR analysis showed were 6-hydroxy-5-ketocasbene, 6,9-dihydroxy-5-ketocasbene (proposed to be the final intermediate before jolkinol C in the *J. curcas* pathway) and (in one strain) jolkinol C (Fig. S3.5). The only C5OX CYP that produced detectable jolkinol C when expressed in JWY512 was JcC5OX2p, producing ~360 mg/L jolkinol C. (Fig. 3.4). The chimeric JcC5OX1p did not produce any detectable jolkinol C.

We cointegrated the best C9OX CYP and C5OX CYP from *J. curcas*, surmising that P450s from the same species may cooperate to produce higher levels of jolkinol C. This strain, JWY519, produced a slightly higher level of jolkinol C (400 mg/L) than the cross-species pair (JWY518). Thus, we used this strain for further strain development (Fig. 3.4).

3.4.4. ADH's role in jolkinol C production and strain engineering to improve jolkinol C titer

Previous studies on lathyrane diterpenoid biosynthesis have suggested the involvement of Euphorbiaceae ADHs in the formation of jolkinol C. One study found an *E. lathyris* ADH (ElAdh1p) was required for jolkinol C production *in vitro* and *in planta* (Luo et al. 2016) while another group working on the *J. curcas* pathway reported jolkinol C production in the absence of an ADH, despite a homolog of ElAdh1p being present in a *J. curcas* gene cluster (NW_012124159) containing jolkinol C biosynthetic genes including *JcCBS1*, *C9OX*, and *C5OX* (King et al. 2016). We thus sought to determine the role of ADH in the formation of jolkinol C in yeast.

We proceeded to test native and codon-optimized forms of *ElADH1* and *JcADH1* in our top jolkinol C producers JWY518 and JWY519. We found that all integrated ADHs drastically improved jolkinol C titer by about 2-fold (Fig. 3.3), and increased the levels of the non-ring closed accumulated intermediate, 6,9-dihydroxy-5-ketocasbene. The jolkinol production levels from strains containing the two ADHs from the two species did not differ statistically and neither codon optimization nor protein-tagging of the ADHs further improved titer (Fig. S3.7). While testing different cultivation conditions for jolkinol C production, we noticed a consistent increase in titers of jolkinol C and oxidized casbane intermediates when higher surface are:volume ratios or higher agitation rates were used, likely due to increased oxygen supply improving CYP

activity (Fig. S3.7). This indicates a need for additional strain and process engineering for industrial applications. Our final top producing strain, JWY521, produced 801 mg/L +/- 42mg of jolkinol C. This high titer illustrates the promise of using engineered yeast for further gene discovery of related pathways, circumventing the problem of low yield often associated with gene functional testing using other techniques.

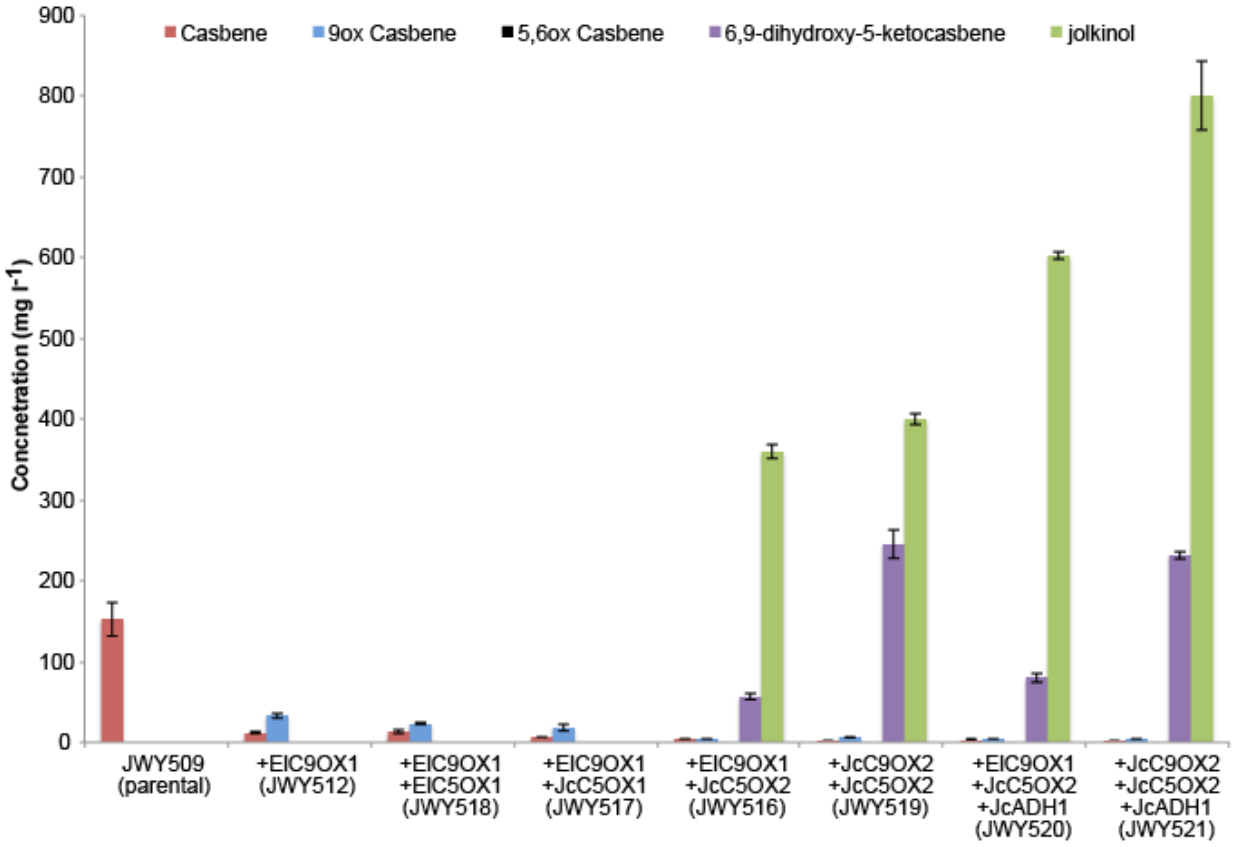


Fig. 3.4. Production of 6,9-hydroxy-5-ketocasbene and jolkinol C with various combinations of C9OX CYP, C5OX CYP, and ADH variants. C5OX CYPs were screened for jolkinol C production when expressed with the most productive C9OX CYP, ElC9OX1p. Only JcC5OX2p formed any detectable level of jolkinol C. Additionally, we integrated this CYP into the best *J. curcas* C9OX CYP parental strain containing JcC9OX2p. It is important to note that ElC5OX1p produced low titers of C-5 oxidized casbanes in JWY509 and produced low levels of 9-keto-5-hydroxycasbene, a purported dead end product, while producing no triply oxidized casbanes nor jolkinol C. All strains were cultured in 2 ml of YPG medium in millititer plates, grown for 48 h, and the culture medium analyzed for casbene (red bars) and C-9 oxidized casbene (blue bars) by GC-MS, while all other compounds were measured by HPLC-UV. Bars represent mean values ±1 s.d. of three replicate cultures. All compounds were quantified using authentic standards purified from high-producing strains.

3.5. Conclusions

Diterpenoid drugs have a long history of success in medicine. Although discovery of biosynthetic pathways can be arduous, we believe that applying the techniques used in this work in yeast can provide a stable biosynthetic platform for accelerated gene discovery for valuable FDA-approved compounds such as ingenane-3-angelate and taxol. These natural products are often difficult to source in quantities necessary for drug clinical trials due to the large number of chiral centers, which make efficient chemical synthesis difficult, and low yields from plant cell culture or farmed plants. We believe that protein tagging strategies, optimizing copy number, and screening libraries of enzymes from different species that perform the same reaction can allow one to find the optimal combinations of terpenoid biosynthetic genes for industrial drug production in yeast strains, particularly for oxygenated compounds.

Our data conclusively show that ElC9OX1p and JcC5OX2p coexpressed alongside an ADH allow for the biosynthesis of the important lathyrane precursor jolkinol C. This pathway illustrates the potential of combinatorial CYP screening and the use of yeast to characterize CYP expression in terpenoid biosynthetic pathways. The final engineered jolkinol C strain (JWY521) produced 800 mg/L jolkinol C. These titers highlight the advantage of yeast as a host over the previously used transient expression in *N. benthamiana* for gene discovery. The extraction of jolkinol C from this yeast chassis greatly simplifies gene discovery of the pathway, while using *N. benthamiana* as an expression host requires harvesting and extracting many plants and performing many downstream isolation steps to produce enough jolkinol C for gene discovery. Additionally, the methods used in the paper are advantageous over previous terpenoid gene discovery methods in yeast, as previous groups have relied on high-copy number plasmids, which suffer from plasmid instability, high levels of expression variability, and the limited number of genes that can be expressed from such a system (Ro et al. 2006; Kirby et al. 2010; Luo et al. 2016).

Because previous studies have suggested yeast codon-optimization of CYPs improved titer of target compounds, we used all codon-optimized CYPs (Paddon et al. 2013; Luo et al. 2016). However, we decided to test whether this trend was consistent in the downstream pathway enzymes, the ADHs. Interestingly, the codon-optimized ADH strains did not outperform their non codon-optimized counterparts. However, we believe that codon optimization can be used as a general strategy for improving enzyme expression in diterpenoid pathways.

The necessity of an ADH for the formation of jolkinol C has been debated. However, our work shows that although the ADHs tested greatly improve jolkinol C production, they are not necessary for production of jolkinol C. These findings resolve the differences seen in King and Luo's experiments. King and colleagues produced jolkinol C by transient expression of CBS, C9OX CYP and C5OX CYP in *N. benthamiana* and did not require the use of an ADH, while Luo and colleagues only saw the formation of jolkinol C in the presence of ElAdh1p *in vitro* and *in planta (King et al. 2016; Luo et al. 2016)*.

The lack of jolkinol C production in the *E. lathyris* system in the absence of an ADH is likely due to the low activity of ElC5OX1p on the C-6 position of the casbene skeleton, even when coexpressed with a C9OX CYP. To this effect, our strain JWY518 produced a doubly oxidized casbane, shown to be 9-keto-5-hydroxycasbene by Luo and colleagues, which is a dead-end product (Fig. S3.3). Based on *in vitro* data reported by Luo and colleagues, we believe that the ADH converts 6,9-dihydroxy-5-ketocasbene to 6-hydroxy-5,9-diketocasbene, which is poised for spontaneous cyclization into jolkinol C. Perhaps, the *J. curcas* CYPs have some ability to effect a second oxidation on the C-9 position of the intermediate 6,9-dihydroxy-5-ketocasbene and thus produce detectable levels of jolkinol C in the absence of an ADH. The dramatic jolkinol C titer improvements in the ADH strains supports the direct involvement of an ADH in jolkinol C biosynthesis.

While the ADH integrations caused a massive increase in jolkinol C titers, it also nearly doubled titers of the precursor 6,9-dihydroxy-5-ketocasbene. We cannot explain the massive total diterpenoid titer increase in the jolkinol C-producing strains relative to the casbene- and singly/doubly oxidized casbane-producing strains. It is possible that the triply oxidized casbanes and lathyranes are less cytotoxic than casbene and singly or doubly hydroxylated casbene.

Our work significantly advances the field of diterpenoid biosynthesis in yeast and specifically for jolkinol C, for which we engineered the biosynthetic steps required for its heterologous production. This jolkinol C chassis strain will significantly simplify discovery of additional genes in the biosynthetic pathways of ingenol-3-angelate and tiglianes such as prostratin. This work represents the highest levels of oxidized diterpenoids produced to date in any microorganism.

4. Chapter 4 — *De novo* synthesis of the sedative valerenic acid in *Saccharomyces cerevisiae*

Including material from submitted work: Wong, J., d'Espaux, L., van der Horst, C., Dev, I., Keasling, J. D. (2017). *De novo* synthesis of the sedative valerenic acid in *Saccharomyces cerevisiae*

4.1. Abstract

Valeriana officinalis (Valerian) root extracts have been used by European and Asian cultures for millennia for their anxiolytic and sedative properties. However, the efficacy of these extracts suffers from variable yields and composition, making these extracts a prime candidate for microbial production. Recently, valerenic acid, a C15 sesquiterpenoid, was identified as the active compound that modulates the $GABA_A$ channel. Although the first committed step, valerena-4,7(11)-diene synthase, has been identified and described, the complete valerenic acid biosynthetic pathway remains to be elucidated. Sequence homology and tissue-specific expression profiles of *V. officinalis* putative P450s led to the discovery of a *V. officinalis* valerena-4,7(11)-diene oxidase, VoCYP71DJ1, which required coexpression with a *V. officinalis* alcohol dehydrogenase and aldehyde dehydrogenase to complete valerenic acid biosynthesis in yeast. Further, we demonstrated the stable integration of all pathway enzymes in yeast, resulting in the production of 140mg/L of valerena-4,7(11)-diene and 4mg/L of valerenic acid in millititer plates. These findings showcase *Saccharomyces cerevisiae*'s potential as an expression platform for facilitating multiply-oxidized medicinal terpenoid pathway discovery, possibly paving the way for scale up and FDA approval of valerenic acid and other active compounds from plant-derived herbal medicines.

4.2. Introduction

The popularity of herbal medicines is exemplified by a $60 billion industry worldwide, and 20% of the total drug market (Kirk and Dunker 2014). Valerian — a medicinal preparation of *Valeriana officinalis* rhizome, root, and stolon extracts — is one such herbal medicine, and has been valued for its anxiolytic and sedative properties for millenia (Houghton 1988; Eadie 2004). Valerian is currently 'generally recognized as safe' by the US Food and Drug Administration, and has been approved as a natural sleep aid in several countries (Kumar 2006). While *V. officinalis* extracts contain many bioactive compounds such as phenolic acids, valepotriates, lignans, flavonoids, amino acids, alkaloids, and tannins, the compound responsible for valerian's well-known activities is the sesquiterpenoid valerenic acid (Hritcu and Cioanca 2016; Trauner et al. 2008; Takemoto et al. 2009) (Fig. 4.1a). Interestingly, valerenic acid has been shown to be produced in only two related genera, *Valeriana* and *Centranthus,* primarily in the root and inflorescence tissues (Hassan et al. 2008). Until recently, the mechanism for valerenic acid activity was unknown. *In vivo* studies in mice showed that valerenic acid allosterically modulates $GABA_A$ receptor activity leading to sedative or anxiolytic effects (Becker et al. 2014; Benke et

al. 2009; Trauner et al. 2008). Interestingly, studies showed that *V. officinalis* root extracts contain another naturally occurring valerenic acid-derived sesquiterpenoid, acetoxyvalerenic acid, that diminishes the anxiolytic effects of valerenic acid by binding to identical sites (Felgentreff et al. 2012; Khom et al. 2007; Benke et al. 2009) (Fig. 4.1a). Indeed, studies have shown that extracts with high valerenic acid:acetoxyvalerenic acid ratios have more pronounced sedative effects (Felgentreff et al. 2012). Effective use of valerian is hampered by inaccurate dosage guidelines and highly variable acetoxyvalerenic acid content, which often makes up a significant amount of crude root extract (Becker et al. 2014; Felgentreff et al. 2012). Indeed, synthesis of valerenic acid is a source of isolated valerenic acid but suffers from low yields (ca. 6%) after many chemical conversion steps starting from (*R*)-pulegone (Kopp et al. 2009). Microbial production of valerenic acid promises isolation of valerian's active compound in the absence of antagonistic acetoxyvalerenic acid, lowers production costs, and does not suffer from inconsistent composition nor variable yields of plant derived valerenic acid — the current source of this herbal medicine.

The biosynthetic pathway for valerenic acid has not been fully elucidated, but is thought to proceed from central carbon metabolism through the mevalonate pathway and farnesyl pyrophosphate (FPP), then cyclized by a sesquiterpene synthase, valerena-4,7(11)-diene synthase (VDS), to form valerena-4,7(11)-diene, and likely decorated by one or more P450s, acyltransferases, and other modifying enzymes. Previous studies have found that the highest concentration of valerenic acid is localized to *V. officinalis* root tissues, and an expression study showed *VDS* transcripts were almost exclusively expressed in the root relative to other tissues (Yeo et al. 2013). Several *Asteraceae* P450s in the CYP71D clade have been shown to oxidize a primary carbon on a sesquiterpene hydrocarbon substrate. These include the *Artemisia annua* AaCYP71AV1 (which converts amorpha-4,11-diene to artemisinic acid, an important precursor to the anti-malarial drug artemisinin), various germacrene A oxidases (GAOs) (which convert germacrene A to germacra-1(10),4,11(13)-trien-12-oic acid, the precursor to the anti-cancer compound costunolide (Nguyen et al. 2010; Ro et al. 2006; Ikezawa et al. 2011), and *Thapsia garganica* TgCYP76AE2 (which catalyzes analogous oxidations on epikunzeaol to form epidihydrocostunolide (Andersen et al. 2017)). Due to the high similarity of other *Asteraceae* sesquiterpenoid and the valerenic acid biosynthetic pathways, previous studies have postulated that a P450 may catalyze conversion of valerenadiene into valerenic acid (Fig. 4.1a).

Thus far, attempts at identifying the full biosynthetic pathway for producing the important drug valerenic acid or reconstituting this activity in a heterologous host have failed (Ricigliano et al. 2016). We decided to use yeast as an expression platform, as our preliminary studies and other studies have encountered difficulties using *N. benthamiana* as a heterologous host for oxidized sesquiterpene production likely due to endogenous activities such as glycosylation (Fig. S4.1) (van Herpen et al. 2010). Here, we engineered a yeast chassis for the production of the valerenic acid precursor valerenadiene at a titer of ~140 mg/L. Then, we used phylogenetic and expression

analysis to identify a root-upregulated *V. officinalis* valerenadiene oxidase, VoCYP71DJ1, and use dehydrogenases to produce a yeast strain capable of generating valerenic acid at 4 mg/L.

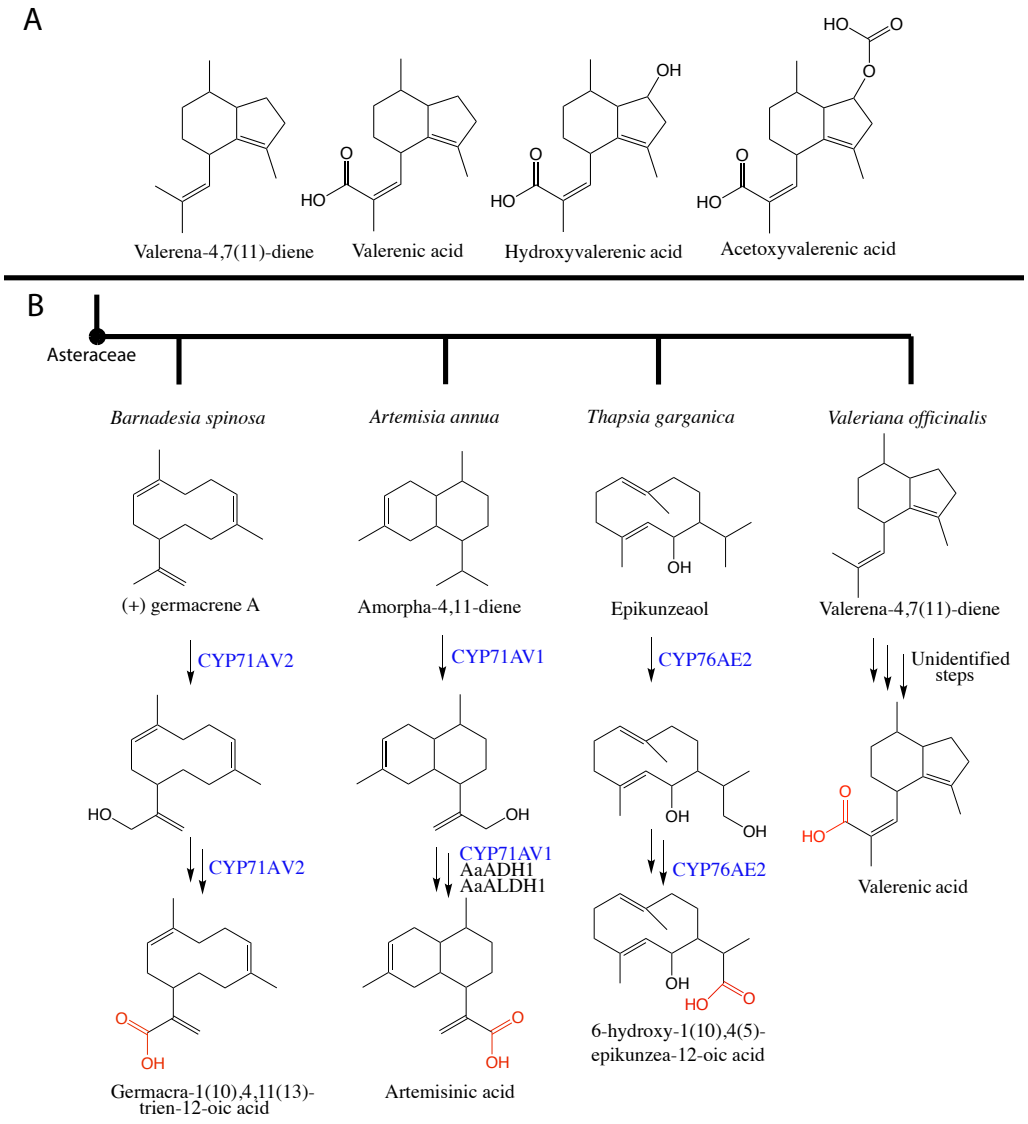


Fig. 4.1. (A) Important valerenic acid derivatives and related compounds. Valerenic acid, hydroxyvalerenic acid and acetoxyvalerenic acid are the three most abundant valerenic acids in *V. officinalis*. Valerenic acid and valerena-4,7(11)-diene are reported to have sedative effects, while acetoxyvalerenic acid has antagonistic activities. (B) P450 activity in *Asteraceae* sesquiterpenoid drug pathways. Sesquiterpenes are produced in all plant species from MEV pathway-derived FPP by sesquiterpene synthases. Subsequently, P450 (CYP) activity is responsible for sesquiterpene acid formation in characterized biosynthetic pathways from various *Asteraceae*. *Artemisia annua* alcohol dehydrogenase (AaADH) and *A. annua* aldehyde dehydrogenase (AaALDH) have been shown to catalyze the conversion of artemisinic alcohol to artemisinic acid alongside P450 activity in the artemisinin pathway. Functionally characterized P450s are shown in blue.

Table 1
Yeast strains used in this study

Strain	Parent (+ additional genetic changes)	References
	erg9::KanMX_P_{CTR3}-ERG9 leu2-3,112::His3MX6_P_{GAL1}-ERG19/P_{GAL10}-ERG8 ura3-52::URA3_P_{GAL1}-mvaS(A110G)/P_{GAL10}-mvaE(CO)	
GTy23	his3Δ1::hphMX4_P_{GAL1}-ERG12/P_{GAL10}-IDI1	This study
JWy601	GTy23 (ura3-52 prototrophy removed for use of Cas9 system)	This study
JWy602	JWy601 (ARS1622b::P_{GAL1}-VDS-GFP)	This study
JWy603	JWy601 (ARS1622b::P_{GAL1}-VDS)	This study
JWy604	JWy601 (ARS1622b::P_{GAL1}-nMBP-VDS)	This study
JWy605	JWy601 (ARS1622b::P_{GAL1}-nMBP-VDS-GFP)	This study
JWy606	JWy601 (ARS1622b::P_{GAL1}-nMBP-VDS-ERG20)	This study
JWy607	JWy601 (ARS1622b::P_{GAL1}-nMBP-VDS-ERG20-GFP)	This study
JWy608	JWy605 (ARS1014a::P_{GAL1}-nMBP-VDS-ERG20 ARS1114a::P_{GAL1}-nMBP-VDS-ERG20 ARS308a::P_{GAL1}-nMBP-VDS-ERG20)	This study
JWy609	JWy608 (ARS911b::P_{GAL10}-AaCPR, P_{GAL1}-VoCYP714A33)	This study
JWy610	JWy608 (ARS911b::P_{GAL10}-AaCPR, P_{GAL1}-VoCYP81Q107)	This study
JWy611	JWy608 (ARS911b::P_{GAL10}-AaCPR, P_{GAL1}-VoCYP71D510)	This study
JWy612	JWy608 (ARS911b::P_{GAL10}-AaCPR, P_{GAL1}-VoCYP71D511)	This study
JWy613	JWy608 (ARS911b::P_{GAL10}-AaCPR, P_{GAL1}-VoCYP71BE87)	This study
JWy614	JWy608 (ARS911b::P_{GAL10}-AaCPR, P_{GAL1}-VoCYP71DJ1)	This study
JWy615	JWy614(ARS1021b::P_{GAL10}-VoADH1, P_{GAL1}-VoALDH1)	This study
JWy616	JWy615 (ARS607b::P_{GAL1}-VoCYP714A33)	This study
JWy617	JWy615 (ARS607b::P_{GAL1}-VoCYP81Q107)	This study
JWy618	JWy615 (ARS607b::P_{GAL1}-VoCYP71D510)	This study
JWy619	JWy615 (ARS607b::P_{GAL1}-VoCYP71D511)	This study
JWy620	JWy615 (ARS607b::P_{GAL1}-VoCYP71BE87)	This study
JWy621	JWy615(ARS511b::P_{GAL1}-nMBP-VoADH1)	This study
JWy622	JWy615(ARS511b::P_{GAL1}-nMBP-VoALDH1)	This study
JWy623	JWy615(ARS511b::P_{GAL1}-nMBP-VDS-ERG20)	This study
JWy624	JWy614(ARS1021b::P_{GAL10}-AaADH1, P_{GAL1}-AaALDH1)	This study
JWy625	JWy608 (ARS911b::P_{GAL10}-VoCPR1, P_{GAL1}-VoCYP71DJ1)	This study
JWy626	JWy626(ARS1021b::P_{GAL10}-AaADH1, P_{GAL1}-AaALDH1)	This study

4.3. Materials and methods

4.3.1. Strain construction

The parent *Saccharomyces cerevisiae* strain used for all engineering was GTy23 {erg9::KanMX_PCTR3-ERG9 leu2-3,112::His3MX6_PGAL1-ERG19/PGAL10-ERG8 ura3-52::URA3_PGAL1-mvaS(A110G)/PGAL10-mvaE(CO) his3Δ1::hphMX4_PGAL1-ERG12/PGAL10-IDI1} previously used by our lab. The integration cassettes for all subsequent strains (Table 1) were created using the software tools CASdesigner (casdesigner.jbei.org) and DIVA (diva.jbei.org) and integrated using the previously reported, cloning-free methodology via Cas9-aided homologous recombination (Reider Apel et al. 2017). Integration cassettes containing 1-kb flanking homology regions targeting a chosen genomic locus were constructed by PCR amplifying donor DNA fragments using primers generated by CASdesigner, then co-transformed with a Cas9-gRNA plasmid (pCut) targeting the chosen genomic locus. CASdesigner primers provide 30–60 nt of inter-fragment homology allowing 1–5 separate fragments to assemble via homologous recombination *in vivo*. pCuts targeting genomic loci were assembled in vivo from a linear backbone and a linear PCR fragment containing the new gRNA sequence, as described previously (Reider Apel et al. 2017). The new gRNA sequence for the URA3 locus was chosen using DNA2.0 (www.dna20.com/eCommerce/cas9/input). To generate donor DNA fragments, native sequences—e.g., chromosomal homology regions, promoters—

were amplified from CEN.PK2-1C genomic DNA, while heterologous sequences—e.g., P450 coding sequences (Fig. S4.9)—were amplified from synthetic gene blocks codon-optimized for expression in *S. cerevisiae*. All genes integrated in this study were expressed under galactose inducible promoters.

All PCRs used Phusion Hot Start II DNA polymerase (www.thermofisher.com, cat. F549L). The following touchdown PCR cycling conditions were used for all PCRs: 1 cycle of 98°C for 15 sec; 25 cycles of 98°C for 10 sec, 65°C for 30 sec (dropping 1 degree each cycle after the first cycle), 72°C for 30 sec, and then 25 cycles of 98°C for 10 sec, 50°C for 30 sec, 72°C for 30 sec. Transformations were performed via heat-shock using ~ 200 ng pCut, ~ 1 µg donor DNA per sample, and 20 min heat shock at 42°C, then plated all cells on selective agarose plates (Gietz and Woods 2002). For assembling a pCut targeting a new site by homologous recombination, we used 200 ng linear pCut backbone and 500 ng of a 1- kb fragment containing the gRNA sequence, as described (Reider Apel et al. 2017). For multi-site integrations, we used 200 ng total linear pCut backbone, and the same amounts of gRNA fragment and donor DNA for each site as we would have for a single integration. Colonies were screened by PCR directed at the target locus, and for integrations, one representative colony was sequenced. Three to four biological replicates were analyzed for each strain.

4.3.2. Synthetic genes and oligonucleotides.

Oligonucleotides and synthetic genes were commercially synthesized (Integrated DNA Technologies, Inc.). All codon optimized sequences were designed based on the IDT online tool. Sequences of synthetic genes can be found in Figure S4.9. P450 ORF predictions for gene synthesis were selected from the publically available transcriptome at Medicinal Plant Genomics Resource (http://medicinalplantgenomics.msu.edu). Previously described synthetic, yeast codon optimized *Artemisia annua* ADH1 (JQ582842.1) and ALDH1 (JQ609276.1) were used in this study (Paddon et al. 2013).

4.3.3. Culture and fermentation conditions

Selective agar plates used for transformations were purchased from Teknova (www.teknova.com, cat. C3080). Liquid selective medium used to grow transformants contained 0.2% (w/v) complete supplement mixture (CSM) lacking uracil (www.sunrisescience.com, cat. 1004-100), 0.67% yeast nitrogen base (www.difco.com, cat. 291920), and 2% dextrose. Nonselective medium contained 1% yeast extract, 2% peptone (Difco cat. 288620 and 211677, respectively), and either 2% dextrose (YPD) or 2% galactose and 0.2% dextrose (YPG). Nonselective agar YPD plates were purchased from Teknova (cat. Y100). Cultures were grown in plastic 96-deep well plates (www.vwr.com, cat. 29445-166) and glass test tubes for strain maintenance, while 2 ml of medium in 24-deep well plastic plates (CWR cat. 89080-534) were used for all production runs. Production cultures were spiked with 100 mg/L trans-caryophyllene (sigma cat. C9653) as an internal standard. Plastic plates were covered with aeraseal film

(www.excelscientific.com, cat. BS-25) and shaken at 800 rpm in a Multitron shaker (www.infors-ht.com, model AJ185). Production runs were cultured for 48 hr in 2 ml of YPG before terpenoid extraction for analysis. Glass tubes were shaken at 200 rpm. All strains were grown at 30°C.

4.3.4. Confocal microscopy

To visualize GFP expression of tagged VDS variants in yeast strains, strains were grown in 5 ml YPD overnight, then back-diluted 1:100 into the same medium and grown 3–6 h at 200 rpm and 30 °C. Then, 1 ml of culture volume was centrifuged at 21,952 x g on a table-top centrifuge, washed with 1x water, and 1ul of the cell pellet was imaged using a Zeiss LSM 710 confocal system mounted on a Zeiss inverted microscope (www.zeiss.com) with a 63Å~ objective and processed using Zeiss Zen software.

4.3.5. Identification of differentially expressed P450s

Previously reported assembled transcripts and expression abundance estimations were retrieved from the Medicinal Plant Genomics Resource (http://medicinalplantgenomics.msu.edu) and Plantrans DB (http://lifecenter.sgst.cn/plantransdb). To select P450s for functional testing, published transcriptomes of *V. officinalis* were mined using a profile hidden Markov model search (HMMER) (Finn et al. 2011). Correlation of expression profiles was calculated by the Pearson product-moment correlation coefficient of log2 FPKM, with all log2 FPKM values less than zero were set to zero. Subsequent heatmaps were generated using Multiple Experiment Viewer Software (MeV) version 4.5 (Howe et al. 2010). DNA sequences identified in this work are deposited in GenBank™ with the following accession numbers: VoCYP71DJ1, JX494699; VoCYP, JX494700; VoCYP, JX494701; VoCYP, JX494702; VoCYP, JX494703; VoCYP, JX494704; VoCPR1, JX494705.

4.3.6. Phylogenetic analysis

Phylogenetic analysis was performed using the entire predicted amino acid sequences of *V. officinalis* P450 family proteins and related terpene-modifying P450 proteins from the GenBank database. Fourteen P450s with homology to *V. officinalis* candidates with known roles in terpenoid biosynthetic pathways were used in the analysis. The following accessions were used: *Valeriana officinalis* VoCYP71D442 (ALU63882.1), *Santalum album* SaCYP76F40 (AHB33947.1), *Santalum album* SaCYP76F39 (AHB33940.1), *Catharanthus roseus* CrCYP76B6 (CAC80883), *Thapsia garganica* TgCYP76AE2 (AQY76213.1), *Lactuca sativa* LsCYP71BL2 (AEI59780.1), *Barnadesia spinosa* BsCYP71AV7 (D5JBX1.1), *Tanacetum cinerariifolium* TcCYP71AV2 (AGO03789.1), *Artemisia annua* AaCYP71AV1 (Q1PS23.1), *Cynara cardunculus var. Scolymus* CcCYP71BL5 (AIA09038.1), *Cynara cardunculus var. Scolymus* CcCYP71AV9 (AIA09035.1), *Cichorium intybus* CiCYP71AV8 (E1B2Z9.1), *Arabidopsis thaliana* AtCYP714A1 (NP_001332750.1), and *Arabidopsis thaliana* AtCYP714A2

(NP_001331286.1). Sequence alignments were generated on the basis of comparison of the amino acid sequences using the MAFFT L-INS-i algorithm with default parameters. Alignments for each partition were generated using the default settings (gap opening penalty = 1.53 and offset value = 0.00) (Katoh et al. 2002). A consistent alignment was selected using TrimAl, with the parameter automated1 (Capella-Gutiérrez et al. 2009). Maximum likelihood analyses were conducted with RAxML v.7.2.8 (Stamatakis 2014). Twenty randomized starting trees were generated with which the initial rearrangement setting and the number of distinct rate categories were determined. The best-known likelihood tree was found by performing 1000 repetitions for each of the amino acid datasets. One thousand non-parametric bootstrap replications were then performed using the bootstrap algorithm. The resulting tree was visualized using FigTree. The scale bar of 0.2 indicates a 20% change and each number shown next to the branches is the number of replicate trees in which the related taxa clustered in the bootstrap test.

4.3.7. Metabolite quantification using GC-MS

Yeast cultures were grown in 2 mL of YPgal with 0.25μM $CuSO_4$ (Sigma cat. 209201) in 24-deep well plastic plates for 48 hours. For GC-MS analysis of valerenadiene, the cultures were overlaid with 0.4mL dodecane (spiked with internal standard), which was pipetted into 1.5mL Eppendorf tubes and spun at 21,952 x g for 1 minute. The resulting organic phase was removed and transferred to GC vials and mixed with EtOAc. For GC-MS analysis of valerenic acid, the cultures were extracted 1:1 with EtOAc spiked with 50 mg/L trans-caryophyllene by shaking with 100μL of glass beads for 8 minutes at a frequency of 28 Hz in a Retsch mixer mill MM 400, then centrifuged at 21,952 x g for 1 minute. The resulting organic phase was removed and dried down at 54°C in vacuum, resuspended in 41 μL EtOAc, 4 μL of 40% v/v tetrabutylammonium hydroxide (TBAH) solution (Sigma cat. 86854). Then, 5 μL of iodomethane (Sigma cat. I8504) was added to the sample, and the mixture was agitated by vortex for 10 s. An aliquot of the sample (1 μL) was injected into a cyclosil B column (J&W Scientific) operating at a He flow rate of 1 mL/min on GC–MS (GC model 6890, MS model 5973 Inert, Agilent). An initial temperature of 120°C was held for 3 min, followed by ramping to 250°C at a rate of 20°C/min, and then held at 250°C for another 3 min. The total flow was set to 8.3 mL/min and helium flow was set to 1 mL/min. All production measurements were performed in biological triplicates or quadruplicates. A caryophyllene standard (Sigma cat. C9653) and a valerenic acid standard (Sigma cat. 51964) containing known concentrations of the internal standard were used to determine titers of valerenadiene and valerenic acid, respectively.

4.4. Results and discussion

4.4.1. Engineering production of valerenadiene precursor in yeast

First, we engineered a yeast strain to produce the valerenic acid precursor, valerenadiene, by expressing chromosomally integrated copies of valerenadiene synthase (*VDS*), which converts FPP to valerenadiene. Previous studies have produced low titers of valerenadiene in yeast (~1 mg/L) from unstable, high-copy plasmids (Yeo et al. 2013; Pyle et al. 2012). We used a

previously developed mevalonate overproducing strain, GTy23, in which all enzymes converting acetyl-CoA to GPP are overexpressed from galactose-inducible promoters. After, we integrated a single copy of *VDS*, but noted low production, about 3 mg/L, consistent with other non-modified sesquiterpene synthases (Fig. 4.2) (Son et al. 2014; Beekwilder et al. 2014). Many studies have shown diterpene and sesquiterpene synthases suffer from cytosolic insolubility or instability when expressed in yeast (Ignea et al. 2015; Reider Apel et al. 2017). Thus, we employed protein tagging strategies to visualize insoluble protein and to improve protein solubility. A single integration of a VDS-GFP tagged variant showed severe protein aggregation (Fig. 4.2b, Fig. S4.2, left). We proceeded to test other recently described protein tag variations. We tested several combinations of an N-terminal MBP tag, a C-terminal GFP tag, and a C-terminal ERG20 tag (ERG20 being the enzyme catalyzing the formation of FPP). The best VDS variant was MBP-VDS-ERG20, which resulted in substantial titer improvements to ~30 mg/L, almost an order of magnitude increase over untagged VDS. Integrating three additional copies of MBP-VDS-ERG20 led to strain JWy608, which produced a titer of ~140 mg/L valerenadiene (Fig. 4.2). This strain was used for all P450 functional testing.

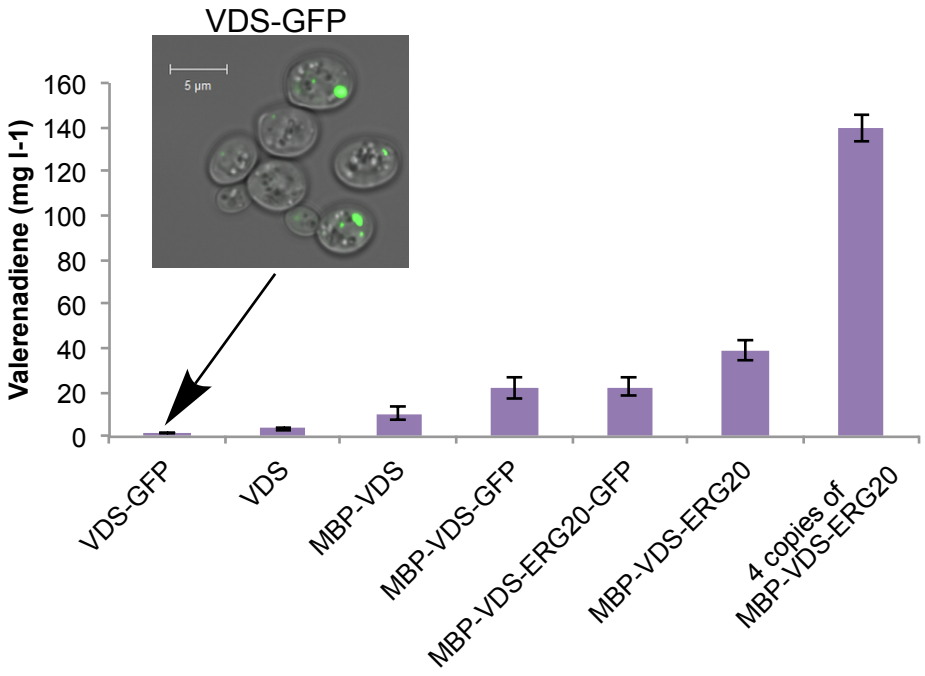


Fig. 4.2. Titer improved with VDS tagging strategies. Confocal microscopy of yeast cells expressing VDS-GFP shows GFP aggregation. ERG20 appeared to improve VDS solubility and provide increased flux through the pathway by increasing the supply of FPP. Cultures were extracted after 48 hours and analyzed for valerenadiene production by GC-MS. MBP-VDS-ERG20 narrowly outperformed both MBP-VDS-GFP and MBP-VDS-ERG20-GFP in production of valerenadiene. This VDS variant was integrated two additional times to produce the high titer valerenadiene strain, JWy608. Data represent the averages of three replicate cultures; error bars show s.d.

4.4.2. Identifying putative valerenadiene oxidase genes from V. officinalis

To select P450s for functional testing, we mined the published transcriptomes of *V. officinalis* using a profile hidden Markov model to search for all P450s (Yeo et al. 2013; Pyle et al. 2012; Finn et al. 2011). Subsequently, we used isoform expression data generated by a previous study to identify likely candidates involved in valerenic acid biosynthesis based on root tissue preferential expression, and their correlation with the expression pattern of *VDS* in the tissue types (Yeo et al. 2013). However, we also identified two P450s, *VoCYP81Q107* and *VoCYP71D510*, that had constitutive expression over all tissue types, similar to the expression of upstream isoprenoid biosynthetic gene homologs for *HMGR*, *IDI*, and *FPPS* (Fig. 4.3). Of the candidates, the expression profile of *VoCYP71DJ1* shared the highest similarity to the expression profile of *VDS*, with no expression in the leaf or callus, low expression in the stem, and the highest expression in the root (Fig. 4.3).

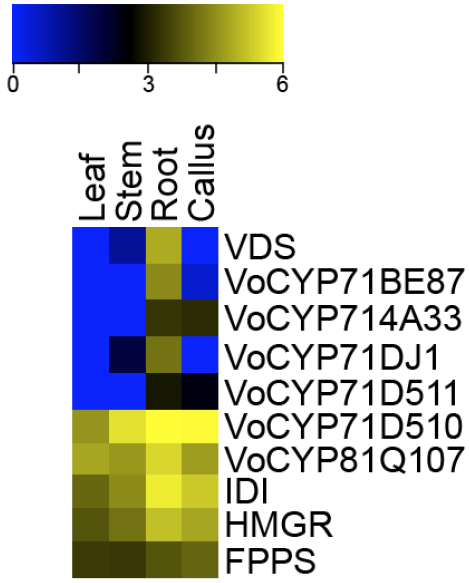


Fig. 4.3. Heat map of candidate P450s involved in valerenic acid biosynthesis upregulated in the root. Sesquiterpene biosynthetic precursor enzymes upstream of valerenadiene are highly expressed in all tissues, while VDS is almost exclusively expressed in root tissue. P450s were selected by high expression in the root of *V. officinalis*; note, VoCYP71DJ1 (valerenadiene oxidase) shares a similar expression profile among tissue types with VDS. Expression values in log2 FPKM (fragments per Kilobase of transcript per million fragments mapped) were used, negative values were set to zero. Expression values shown represent the different developmental tissues.

We made a phylogenetic tree of all putative *V. officinalis* P450s and included other functionally tested *Asteraceae* P450s that oxidize sesquiterpenes (Fig. 4.4). We noticed that all of these upregulated P450s had homology to known terpene modifying enzymes, with the exception of VoCYP81Q107. VoCYP714A33 has homology to gibberellin oxidases (Nomura et al. 2013),

such as AtCYP714A1 and AtCYP714A1, while VoCYP71BE87 shares homology with the *Vitis vinifera* P450 VvCYP71BE5 responsible for the formation of the sesquiterpenoid (-)-rotundone, an important component of wine flavor (Takase et al. 2016). VoCYP81Q107 has homology to P450s involved in the sesamin biosynthetic pathway (Hata et al. 2010). Four of the six identified P450 candidates were classified as CYP71D P450s, consistent with the classification of other *Asteraceae* sesquiterpene oxidases that have been found to catalyze the oxidation of a primary carbon on sesquiterpenes forming the respective acids (Nguyen et al. 2010; Ro et al. 2006; Andersen et al. 2017; Ikezawa et al. 2011; Nelson and Werck-Reichhart 2011). Because amorphadiene oxidase AaCYP71AV1 catalyzes three successive oxidations on amorphadiene, similar to the oxidations seen in our target pathway, we expected high homology of AaCYP71AV1 to our candidates. Of the selected putative *V. officinalis* P450s in the CYP71D family, VoCYP71D510 and VoCYP71D511 share approximately 51% homology with AaCYP71AV1, while VoCYP71DJ1 and VoCYP71BE87 share approximately 44% homology with AaCYP71AV1 at the protein sequence level (Fig. S4.3). VoCYP71D510 and VoCYP71D511 share high homology to VoCYP71D442, previously identified as a possible *V. officinalis* P450 candidate involved in valerenic acid biosynthesis, but this enzyme did not coexpress with *VDS* or upstream enzymes and did not produce oxidized sesquiterpenes (Ricigliano et al. 2016).

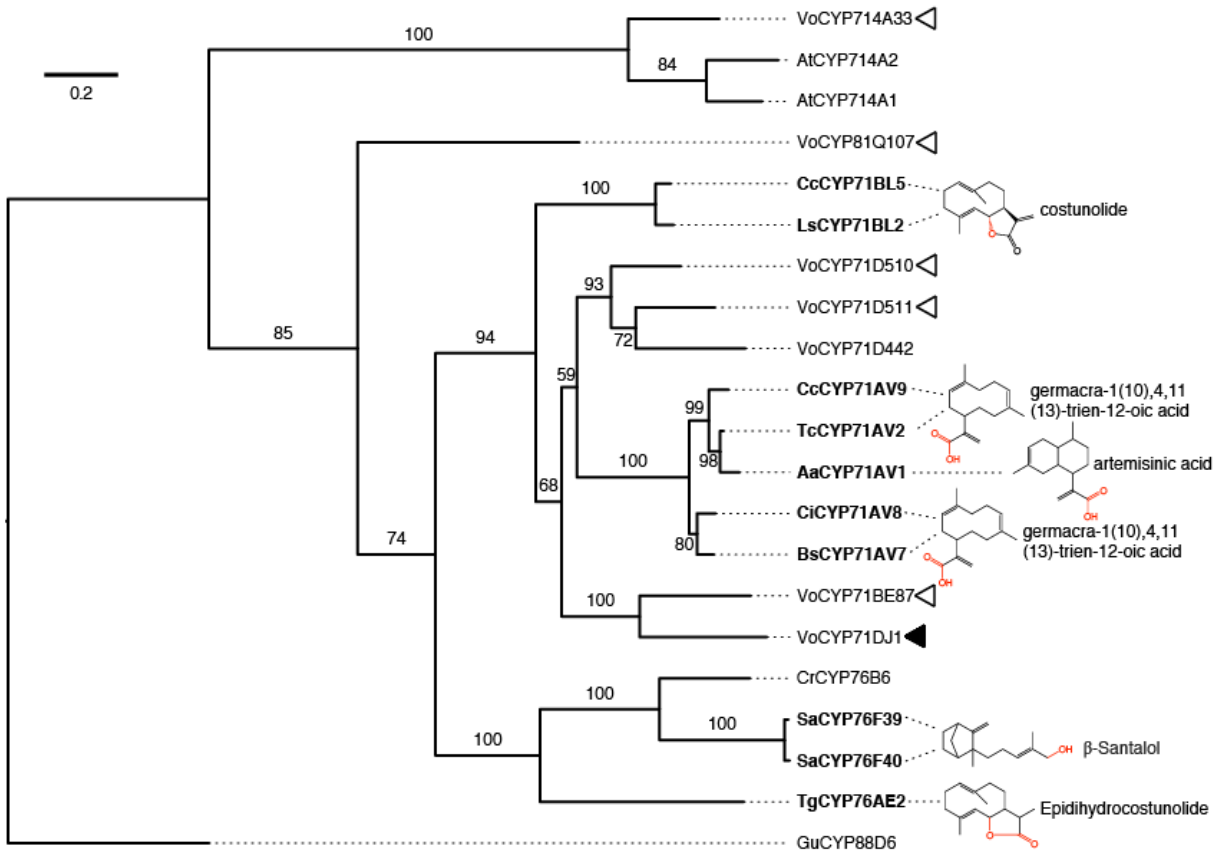


Fig. 4.4. Phylogenetic tree of *V. officinalis* P450s and related terpene-modifying P450s. Note, Germacrene A oxidase P450s from *Asteraceae* species clade with many of the *V. officinalis* candidates.The neighbor-joining tree was generated using MAFFT and RAxML. The numbers indicate the bootstrap value (%) from 100 replications. The scale bar shows the amino acid substitution ratio. GuCYP88D6, *Glycyrrhiza uralensis* β-amyrin-11-oxidase (AB433179), was used as the outgroup. *V. officinalis* CYP proteins identified in this study are marked with arrowheads. The closed arrowhead indicates VoCYP71DJ1 having the activity of valerenic alcohol synthesis and the open arrowheads indicate CYPs incapable of producing oxidized valerenadiene in this study.

4.4.3. Functional identification of V. officinalis P450s acting on valerenadiene

To test the candidate P450s for activity, we cointegrated a single copy of each candidate with *A. annua* cytochrome P450 reductase, *AaCPR1*, into our high valerenadiene producer strain, JWy608. Unfortunately, no valerenic acid was produced by any of the candidate P450s. However, we detected the production of trace amounts of hydroxylated valerenadiene by one of our candidate P450s, VoCYP71DJ1 (Fig. 4.1a, Fig. S4.5). We were surprised to find that none of the other candidates functioned as a valerenadiene oxidase, despite sequence similarity and upregulated expression in the root. To determine if coexpression of a native *V. officinalis* cytochrome P450 reductase with VoCYP71DJ1 would improve valerenic acid biosynthesis over

the *A. annua* AaCPR1, we identified and expressed a *V. officinalis* cytochrome P450 reductase, *VoCPR1*, alongside *VoCYP71DJ1*. However, the addition of VoCPR1 did not improve valerenic acid production relative to the titers seen in the strains expressing *AaCPR1*, thus we used other strategies to improve titer (Fig. S4.7).

4.4.4. Expression of alcohol and aldehyde dehydrogenases forms valerenic acid

Due to the low titer of valerenic acid in our strain, we surmised that another P450 candidate could be responsible for the conversion of valerenic alcohol to valerenic acid. We individually integrated all other P450 candidates into the valerenic acid base strain, JWy615, but failed to see improved valerenic acid titer (Fig. S4.6). Additionally, hydroxyvalerenic acid (Fig. 4.1a), a valerenic acid derivative, was also not detected in these strains expressing additional P450 candidates.

Several *Asteraceae* pathways produce respective sesquiterpenoid acids in two steps, requiring only a synthase and a P450. However, several of these pathways and those of other terpenoids show increased production of the final product by overexpressing alcohol and aldehyde dehydrogenases, including the artemisinic acid, jolkinol C, zerumbone, and germacra-1(10),4,11(13)-trien-12-oic acid pathways (Okamoto et al. 2011; de Kraker et al. 2001; Luo et al. 2016; Paddon et al. 2013).

We mined the *V. officinalis* transcriptome for genes with homology to *A. annua* dehydrogenases and found *VoADH1* and *VoALDH1*. *VoADH1* had poor expression in the root transcriptome, while a *VoALDH1* was highly expressed in the root and stem tissues (Fig. S4.4). *VoADH* and *VoALDH* were cointegrated into our yeast strain expressing *VoCYP71DJ1*, JWy614, resulting in trace amounts of valerenic acid, as determined by the mass spectrum and retention time relative to an authentic valerenic acid standard (Fig. 4.5). To further improve titer, we integrated the previously published *A. annua* dehydrogenases *AaADH* and *AaALDH*, used to improve artemisinic acid titer in yeast, surmising that they might have activity on valerenic alcohol and valerenic aldehyde substrates. Cointegration of *A. annua AaADH1* and *AaALDH1* in JWy614 increased valerenic acid titer (Fig. 4.5); thus, although the VoADH1 and VoALDH1 identified in this study formed valerenic acid in yeast, it is possible that one or both of these enzymes are not actually responsible for the formation of valerenic acid in the native plant, which is consistent with their expression profiles. Our final strain, JWy627, produces 4mg/L of valerenic acid. These results show that the techniques that led to the successful microbial production of artemisinin can be applied to other terpenoid pathways.

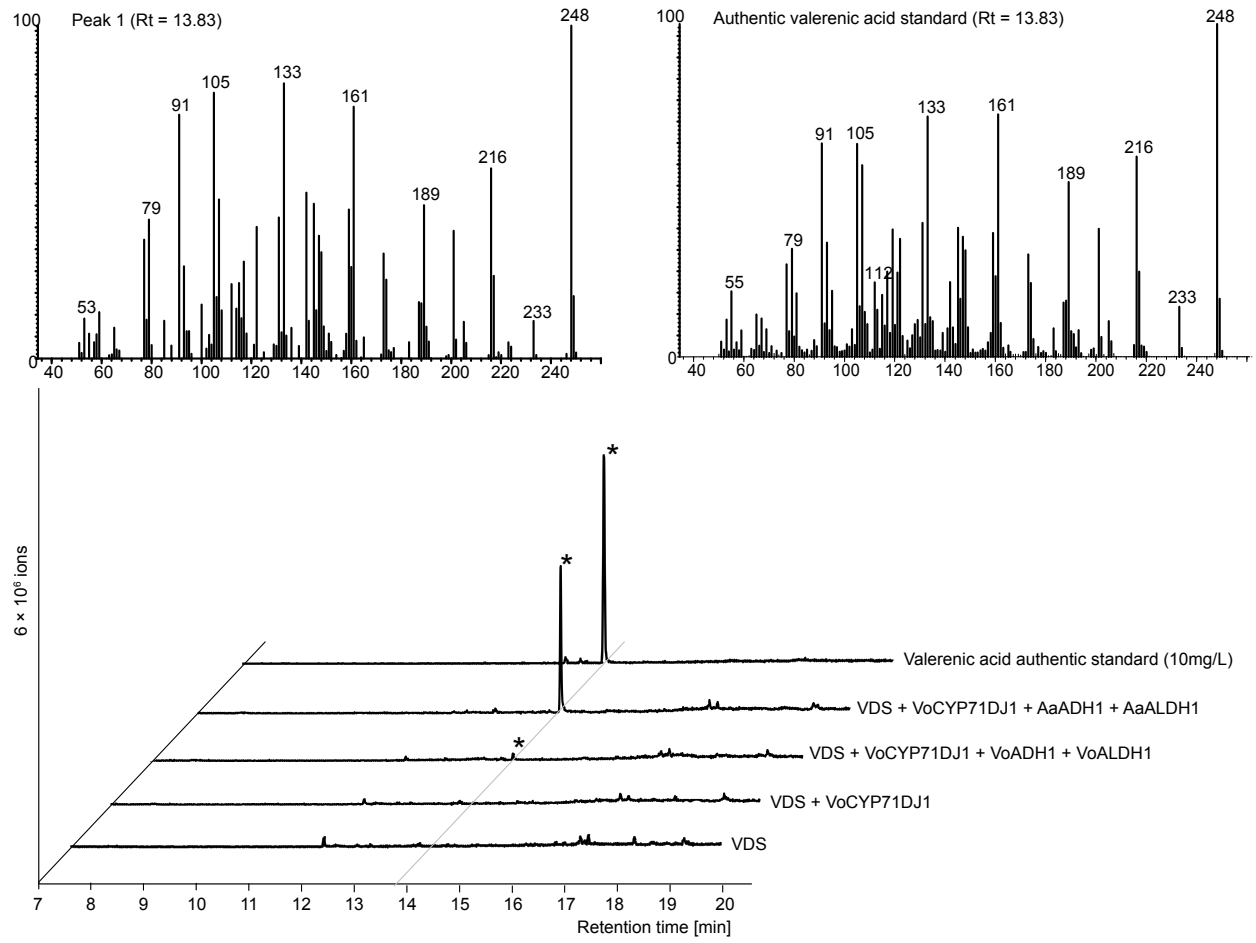


Fig. 4.5. *In vivo* production of valerenic acid by *V. officinalis* P450s and dehydrogenases in yeast. GC analysis identifies sesquiterpenoid products of extracts of yeast cultures with integrated candidate genes.

Coexpression of *A. annua* dehydrogenases improved production of valerenic acid to ~4mg/L. VoCYP71DJ1 was cointegrated with *A. annua* CPR. Ethyl acetate-extractable fractions were derivatized and analysed by GC–MS in extracted ion mode (*m*/*z* 248). Mass spectra of valerenic acid (retention time 13.83 min, detected as methyl ester) relative to authentic standard is shown. Valerenic acid denoted with asterisk.

4.5. Conclusion

We have engineered a yeast chassis for the production of the sedative valerenic acid by engineering the production of the backbone valerenadiene, identified novel P450 oxidizing valerenadiene, and overexpressed these two genes, and completed the valerenic acid biosynthetic pathway in yeast. Valerenadiene and valerenic acid were produced at 140 mg/L and 4 mg/L, respectively. Phylogenetic and expression analyses were necessary to identify a valerenadiene oxidase, VoCYP71DJ1. Further, expression of an ADH and ALDH were required to produce a yeast strain capable of generating valerenic acid. Microbially produced valerenic acid may allow

for more accurate studies of this drug, as plant derived material contains many bioactive compounds, including acetoxyvalerenic acid, a compound with antagonistic effects of valerenic acid. The gene testing strategy used in this study could prove valuable for gene discovery in other medicinally important *Asteraceae* sesquiterpenoid pathways. These findings also illustrate that closely related P450s have been fine tuned by evolutionary pressure for specialized metabolism. Additional strain engineering will be necessary to improve valerenic acid titers for industrial applications.

5. Conclusions and Research Needs

Because of the sheer number of isoprenoids, they hold potential in both the fields of medicine and sustainable biofuels. Microbial production of isoprenoids offers a potentially more sustainable, higher yielding alternative to chemical synthesis or extraction from the natural host. The ability to decorate the terpene backbone with oxygen in the correct position through the functional expression of terpene oxidases opens up the possibility of producing complex drugs and other important molecules. While remarkable progress has been made in improving titer of terpenes in microbial production, there is still a need for pathway discovery, particularly in the area of P450s and other modifying enzymes, to broaden the varieties of useful terpenoids produced in microbes. Additionally, the work above illustrates that P450 expression in yeast still requires significant optimization to improve flux through these steps while maintaining the yeast cell's health.

6. References

Ajikumar, P.K., Xiao, W.-H., Keith E. J. Tyo, Wang, Y., Simeon, F., Leonard, E., Mucha, O., Phon, T.H., Pfeifer, B. and Stephanopoulos, G. 2011. ISOPRENOID PATHWAY OPTIMIZATION FOR TAXOL PRECURSOR OVERPRODUCTION IN ESCHERICHIA COLI - The Science and Applications of Synthetic and Systems Biology - NCBI Bookshelf. . Available at: https://www.ncbi.nlm.nih.gov/books/NBK84464/.

Ajikumar, P.K., Xiao, W.-H., Tyo, K.E.J., Wang, Y., Simeon, F., Leonard, E., Mucha, O., Phon, T.H., Pfeifer, B. and Stephanopoulos, G. 2010. Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. *Science* 330(6000), pp. 70–74.

Albertsen, L., Chen, Y., Bach, L.S., Rattleff, S., Maury, J., Brix, S., Nielsen, J. and Mortensen, U.H. 2011. Diversion of flux toward sesquiterpene production in Saccharomyces cerevisiae by fusion of host and heterologous enzymes. *Applied and Environmental Microbiology* 77(3), pp. 1033–1040.

Alonso-Gutierrez, J., Chan, R., Batth, T.S., Adams, P.D., Keasling, J.D., Petzold, C.J. and Lee, T.S. 2013. Metabolic engineering of Escherichia coli for limonene and perillyl alcohol production. *Metabolic Engineering* 19, pp. 33–41.

Alonso-Gutierrez, J., Kim, E.-M., Batth, T.S., Cho, N., Hu, Q., Chan, L.J.G., Petzold, C.J., Hillson, N.J., Adams, P.D., Keasling, J.D., Garcia Martin, H. and Lee, T.S. 2015. Principal component analysis of proteomics (PCAP) as a tool to direct metabolic engineering. *Metabolic Engineering* 28, pp. 123–133.

Altincicek, B., Kollas, A.-K., Eberl, M., Wiesner, J., Sanderbrand, S., Hintz, M., Beck, E. and Jomaa, H. 2001. *LytB* , a novel gene of the 2-*C* -methyl-D -erythritol 4-phosphate pathway of isoprenoid biosynthesis in*Escherichia coli*. *FEBS Letters* 499(1–2), pp. 37–40.

Altincicek, B., Kollas, A.K., Sanderbrand, S., Wiesner, J., Hintz, M., Beck, E. and Jomaa, H.

2001. GcpE is involved in the 2-C-methyl-D-erythritol 4-phosphate pathway of isoprenoid biosynthesis in Escherichia coli. *Journal of Bacteriology* 183(8), pp. 2411–2416.

Amiri, P., Shahpiri, A., Asadollahi, M.A., Momenbeik, F. and Partow, S. 2016. Metabolic engineering of Saccharomyces cerevisiae for linalool production. *Biotechnology Letters* 38(3), pp. 503–508.

Andersen, T.B., Martinez-Swatson, K.A., Rasmussen, S.A., Boughton, B.A., Jørgensen, K., Andersen-Ranberg, J., Nyberg, N., Christensen, S.B. and Simonsen, H.T. 2017. Localization and in-Vivo Characterization of Thapsia garganica CYP76AE2 Indicates a Role in Thapsigargin Biosynthesis. *Plant Physiology* 174(1), pp. 56–72.

Andersen-Ranberg, J., Kongstad, K.T., Nielsen, M.T., Jensen, N.B., Pateraki, I., Bach, S.S., Hamberger, B., Zerbe, P., Staerk, D., Bohlmann, J., Møller, B.L. and Hamberger, B. 2016. Expanding the Landscape of Diterpene Structural Diversity through Stereochemically Controlled Combinatorial Biosynthesis. *Angewandte Chemie* 55(6), pp. 2142–2146.

Anthony, J.R., Anthony, L.C., Nowroozi, F., Kwon, G., Newman, J.D. and Keasling, J.D. 2009. Optimization of the mevalonate-based isoprenoid biosynthetic pathway in Escherichia coli for production of the anti-malarial drug precursor amorpha-4,11-diene. *Metabolic Engineering* 11(1), pp. 13–19.

Baldini, M., Ferfuia, C., Bortolomeazzi, R., Verardo, G., Pascali, J., Piasentier, E. and Franceschi, L. 2014. Determination of phorbol esters in seeds and leaves of Jatropha curcas and in animal tissue by high-performance liquid chromatography tandem mass spectrometry. *Industrial Crops and Products* 59, pp. 268–276.

Barrero, R.A., Chapman, B., Yang, Y., Moolhuijzen, P., Keeble-Gagnère, G., Zhang, N., Tang, Q., Bellgard, M.I. and Qiu, D. 2011. De novo assembly of Euphorbia fischeriana root transcriptome identifies prostratin pathway related genes. *BMC Genomics* 12, p. 600.

Becker, A., Felgentreff, F., Schröder, H., Meier, B. and Brattström, A. 2014. The anxiolytic effects of a Valerian extract is based on valerenic acid. *BMC complementary and alternative medicine* 14, p. 267.

Beekwilder, J., van Houwelingen, A., Cankar, K., van Dijk, A.D.J., de Jong, R.M., Stoopen, G., Bouwmeester, H., Achkar, J., Sonke, T. and Bosch, D. 2014. Valencene synthase from the heartwood of Nootka cypress (Callitropsis nootkatensis) for biotechnological production of valencene. *Plant Biotechnology Journal* 12(2), pp. 174–182.

Beller, H.R., Lee, T.S. and Katz, L. 2015. Natural products as biofuels and bio-based chemicals: fatty acids and isoprenoids. *Natural Product Reports* 32(10), pp. 1508–1526.

Benke, D., Barberis, A., Kopp, S., Altmann, K.-H., Schubiger, M., Vogt, K.E., Rudolph, U. and Möhler, H. 2009. GABA A receptors as in vivo substrate for the anxiolytic action of valerenic acid, a major constituent of valerian root extracts. *Neuropharmacology* 56(1), pp. 174–181.

Bertea, C.M., Schalk, M., Karp, F., Maffei, M. and Croteau, R. 2001. Demonstration that menthofuran synthase of mint (Mentha) is a cytochrome P450 monooxygenase: cloning, functional expression, and characterization of the responsible gene. *Archives of Biochemistry and Biophysics* 390(2), pp. 279–286.

Biggs, B.W., Lim, C.G., Sagliani, K., Shankar, S., Stephanopoulos, G., De Mey, M. and Ajikumar, P.K. 2016. Overcoming heterologous protein interdependency to optimize P450-mediated Taxol precursor synthesis in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* 113(12), pp. 3209–3214.

Bischoff, K.M. and Rodwell, V.W. 1996. 3-Hydroxy-3-methylglutaryl-coenzyme A reductase from Haloferax volcanii: purification, characterization, and expression in Escherichia coli. *Journal of Bacteriology* 178(1), pp. 19–23.

Bishop, J. and Pagiola, S. eds. 2012. *Selling Forest Environmental Services: Market-Based Mechanisms for Conservation and Development*. Taylor & Francis.

Blumberg, P.M. 1988. Protein kinase C as the receptor for the phorbol ester tumor promoters: sixth Rhoads memorial award lecture. *Cancer Research* 48(1), pp. 1–8.

Bochar, D.A., Brown, J.R., Doolittle, W.F., Klenk, H.P., Lam, W., Schenk, M.E., Stauffacher, C.V. and Rodwell, V.W. 1997. 3-hydroxy-3-methylglutaryl coenzyme A reductase of Sulfolobus solfataricus: DNA sequence, phylogeny, expression in Escherichia coli of the hmgA gene, and purification and kinetic characterization of the gene product. *Journal of Bacteriology* 179(11), pp. 3632–3638.

Booranasrisak, T., Phaonakrop, N., Jaresitthikunchai, J., Virunanon, C., Roytrakul, S. and Chulalaksananukul, W. Proteomic evaluation of free fatty acid biosynthesis in Jatropha curcas L. (physic nut) kernel development. *African Journal of Biotechnology*. Available at: https://www.ajol.info/index.php/ajb/article/view/131746.

Bromann, K., Toivari, M., Viljanen, K., Ruohonen, L. and Nakari-Setälä, T. 2016. Engineering Aspergillus nidulans for heterologous ent-kaurene and gamma-terpinene production. *Applied Microbiology and Biotechnology* 100(14), pp. 6345–6359.

Bruce, R.J. and West, C.A. 1982. Elicitation of casbene synthetase activity in castor bean : THE ROLE OF PECTIC FRAGMENTS OF THE PLANT CELL WALL IN ELICITATION BY A FUNGAL ENDOPOLYGALACTURONASE. *Plant Physiology* 69(5), pp. 1181–1188.

Campbell, M., Hahn, F.M., Poulter, D.C. and Leustek, T. 1998. Analysis of the isopentenyl diphosphate isomerase gene family from Arabidopsis thaliana. *Plant Molecular Biology* 36(2), pp. 323–328. Available at: http://link.springer.com/article/10.1023/A:1005935516274.

Campos, N., Rodríguez-Concepción, M., Sauret-Güeto, S., Gallego, F., Lois, L.M. and Boronat, A. 2001. Escherichia coli engineered to synthesize isopentenyl diphosphate and dimethylallyl diphosphate from mevalonate: a novel system for the genetic analysis of the 2-C-methyl-d-erythritol 4-phosphate pathway for isoprenoid biosynthesis. *The Biochemical Journal* 353(Pt 1),

pp. 59–67.

Campos, N., Rodríguez-Concepción, M., Seemann, M., Rohmer, M. and Boronat, A. 2001. Identification of*gcpE* as a novel gene of the 2-*C* -methyl-D -erythritol 4-phosphate pathway for isoprenoid biosynthesis in*Escherichia coli*. *FEBS Letters* 488(3), pp. 170–173.

Cankar, K., van Houwelingen, A., Bosch, D., Sonke, T., Bouwmeester, H. and Beekwilder, J. 2011. A chicory cytochrome P450 mono-oxygenase CYP71AV8 for the oxidation of (+)-valencene. *FEBS Letters* 585(1), pp. 178–182.

Cann, A.F. and Liao, J.C. 2010. Pentanol isomer synthesis in engineered microorganisms. *Applied Microbiology and Biotechnology* 85(4), pp. 893–899.

Capella-Gutiérrez, S., Silla-Martínez, J.M. and Gabaldón, T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15), pp. 1972–1973.

Carter, O.A., Peters, R.J. and Croteau, R. 2003. Monoterpene biosynthesis pathway construction in Escherichia coli. *Phytochemistry* 64(2), pp. 425–433.

Chambliss, K.L., Slaughter, C.A., Schreiner, R., Hoffmann, G.F. and Gibson, K.M. 1996. Molecular cloning of human phosphomevalonate kinase and identification of a consensus peroxisomal targeting sequence. *The Journal of Biological Chemistry* 271(29), pp. 17330–17334.

Chandran, S.S., Kealey, J.T. and Reeves, C.D. 2011. Microbial production of isoprenoids. *Process Biochemistry* 46(9), pp. 1703–1710.

Chang, M.C.Y., Eachus, R.A., Trieu, W., Ro, D.-K. and Keasling, J.D. 2007. Engineering Escherichia coli for production of functionalized terpenoids using plant P450s. *Nature Chemical Biology* 3(5), pp. 274–277.

Chappell, J. 1995. Biochemistry and Molecular Biology of the Isoprenoid Biosynthetic Pathway in Plants. *Annual review of plant physiology and plant molecular biology* 46(1), pp. 521–547.

Chiba, R., Minami, A., Gomi, K. and Oikawa, H. 2013. Identification of ophiobolin F synthase by a genome mining approach: a sesterterpene synthase from Aspergillus clavatus. *Organic Letters* 15(3), pp. 594–597.

Chubukov, V., Mingardon, F., Schackwitz, W., Baidoo, E.E.K., Alonso-Gutierrez, J., Hu, Q., Lee, T.S., Keasling, J.D. and Mukhopadhyay, A. 2015. Acute Limonene Toxicity in Escherichia coli Is Caused by Limonene Hydroperoxide and Alleviated by a Point Mutation in Alkyl Hydroperoxidase AhpC. *Applied and Environmental Microbiology* 81(14), pp. 4690–4696.

Chuck, C.J. and Donnelly, J. 2014. The compatibility of potential bioderived fuels with Jet A-1 aviation kerosene. *Applied energy* 118, pp. 83–91.

Collu, G., Unver, N., Peltenburg-Looman, A.M.G., van der Heijden, R., Verpoorte, R. and Memelink, J. 2001. Geraniol 10-hydroxylase[1] , a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis. *FEBS Letters* 508(2), pp. 215–220.

Costa, G.G.L., Cardoso, K.C., Del Bem, L.E.V., Lima, A.C., Cunha, M.A.S., de Campos-Leite, L., Vicentini, R., Papes, F., Moreira, R.C., Yunes, J.A., Campos, F.A.P. and Da Silva, M.J. 2010. Transcriptome analysis of the oil-rich seed of the bioenergy crop Jatropha curcas L. *BMC Genomics* 11, p. 462.

Croteau, R., Ketchum, R.E.B., Long, R.M., Kaspera, R. and Wildung, M.R. 2006. Taxol biosynthesis and molecular genetics. *Phytochemistry reviews : proceedings of the Phytochemical Society of Europe* 5(1), pp. 75–97.

Cuellar, M.C., Heijnen, J.J. and van der Wielen, L.A.M. 2013. Large-scale production of diesel-like biofuels - process design as an inherent part of microorganism development. *Biotechnology Journal* 8(6), pp. 682–689.

Cui, L. and Su, X. 2009. Discovery, mechanisms of action and combination therapy of artemisinin. *Expert Review of Anti-Infective Therapy* 7(8), pp. 999–1013.

Cunningham, F.X., Lafond, T.P. and Gantt, E. 2000. Evidence of a role for LytB in the nonmevalonate pathway of isoprenoid biosynthesis. *Journal of Bacteriology* 182(20), pp. 5841–5848.

Cusidó, R.M., Palazón, J., Navia-Osorio, A., Mallol, A., Bonfill, M., Morales, C. and Piñol, M.T. 1999. Production of Taxol® and baccatin III by a selected Taxus baccata callus line and its derived cell suspension culture. *Plant Science* 146(2), pp. 101–107.

Davies, F.K., Work, V.H., Beliaev, A.S. and Posewitz, M.C. 2014. Engineering Limonene and Bisabolene Production in Wild Type and a Glycogen-Deficient Mutant of Synechococcus sp. PCC 7002. *Frontiers in bioengineering and biotechnology* 2, p. 21.

Dejong, J.M., Liu, Y., Bollon, A.P., Long, R.M., Jennewein, S., Williams, D. and Croteau, R.B. 2006. Genetic engineering of taxol biosynthetic genes in Saccharomyces cerevisiae. *Biotechnology and Bioengineering* 93(2), pp. 212–224.

Dickschat, J.S. 2016. Bacterial terpene cyclases. *Natural Product Reports* 33(1), pp. 87–110.

Ding, M.-Z., Yan, H.-F., Li, L.-F., Zhai, F., Shang, L.-Q., Yin, Z. and Yuan, Y.-J. 2014. Biosynthesis of Taxadiene in Saccharomyces cerevisiae : selection of geranylgeranyl diphosphate synthase directed by a computer-aided docking strategy. *Plos One* 9(10), p. e109348.

Eadie, M.J. 2004. Could valerian have been the first anticonvulsant? *Epilepsia* 45(11), pp. 1338–1343.

Eisenreich, W., Schwarz, M., Cartayrade, A., Arigoni, D., Zenk, M.H. and Bacher, A. 1998. The deoxyxylulose phosphate pathway of terpenoid biosynthesis in plants and microorganisms. *Chemistry & Biology* 5(9), pp. 221–233.

EMANUELSSON, O., NIELSEN, H. and HEIJNE, G.V. 1999. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites | Protein Science |

Cambridge Core. *Protein Science*. Available at: https://www.cambridge.org/core/journals/protein-science/article/chlorop-a-neural-network-based-method-for-predicting-chloroplast-transit-peptides-and-their-cleavage-sites/A4CFAC2436726045600012D2F8C22DF0.

Engels, B., Dahm, P. and Jennewein, S. 2008. Metabolic engineering of taxadiene biosynthesis in yeast as a first step towards Taxol (Paclitaxel) production. *Metabolic Engineering* 10(3–4), pp. 201–206.

Felgentreff, F., Becker, A., Meier, B. and Brattström, A. 2012. Valerian extract characterized by high valerenic acid and low acetoxy valerenic acid contents demonstrates anxiolytic activity. *Phytomedicine: International Journal of Phytotherapy and Phytopharmacology* 19(13), pp. 1216–1222.

Finn, R.D., Clements, J. and Eddy, S.R. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research* 39(Web Server issue), pp. W29-37.

Formighieri, C. and Melis, A. 2014. Carbon partitioning to the terpenoid biosynthetic pathway enables heterologous β-phellandrene production in Escherichia coli cultures. *Archives of Microbiology* 196(12), pp. 853–861.

Fraga, B.M. 2005. Natural sesquiterpenoids. *Natural Product Reports* 22(4), pp. 465–486.

Gelb, M.H., Tamanoi, F., Yokoyama, K., Ghomashchi, F., Esson, K. and Gould, M.N. 1995. The inhibition of protein prenyltransferases by oxygenated metabolites of limonene and perillyl alcohol. *Cancer Letters* 91(2), pp. 169–175.

George, K.W., Alonso-Gutierrez, J., Keasling, J.D. and Lee, T.S. 2015. Isoprenoid drugs, biofuels, and chemicals--artemisinin, farnesene, and beyond. *Advances in Biochemical Engineering/Biotechnology* 148, pp. 355–389.

George, K.W., Thompson, M.G., Kang, A., Baidoo, E., Wang, G., Chan, L.J.G., Adams, P.D., Petzold, C.J., Keasling, J.D. and Lee, T.S. 2015. Metabolic engineering for the high-yield production of isoprenoid-based C_5 alcohols in E. coli. *Scientific reports* 5, p. 11128.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A. and Smith, H.O. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* 6(5), pp. 343–345.

Gietz, R.D. and Woods, R.A. 2002. Screening for protein-protein interactions in the yeast two-hybrid system. *Methods in Molecular Biology* 185, pp. 471–486.

Góngora-Castillo, E., Childs, K.L., Fedewa, G., Hamilton, J.P., Liscombe, D.K., Magallanes-Lundback, M., Mandadi, K.K., Nims, E., Runguphan, W., Vaillancourt, B., Varbanova-Herde, M., Dellapenna, D., McKnight, T.D., O'Connor, S. and Buell, C.R. 2012. Development of transcriptomic resources for interrogating the biosynthesis of monoterpene indole alkaloids in medicinal plant species. *Plos One* 7(12), p. e52506.

Gould, M.N. 1997. Cancer chemoprevention and therapy by monoterpenes. *Environmental Health Perspectives* 105(4), pp. 977–979.

Hahn, F.M., Hurlburt, A.P. and Poulter, C.D. 1999. Escherichia coli open reading frame 696 is idi, a nonessential gene encoding isopentenyl diphosphate isomerase. *Journal of Bacteriology* 181(15), pp. 4499–4504.

Hahn, F.M. and Poulter, C.D. 1995. Isolation of Schizosaccharomyces pombe isopentenyl diphosphate isomerase cDNA clones by complementation and synthesis of the enzyme in Escherichia coli. *The Journal of Biological Chemistry* 270(19), pp. 11298–11303.

Hahn, F.M., Xuan, J.W., Chambers, A.F. and Poulter, C.D. 1996. Human isopentenyl diphosphate: dimethylallyl diphosphate isomerase: overproduction, purification, and characterization. *Archives of Biochemistry and Biophysics* 332(1), pp. 30–34.

Halaweish, F.T., Kronberg, S., Hubert, M.B. and Rice, J.A. 2002. Toxic and aversive diterpenes of Euphorbia esula. *Journal of Chemical Ecology* 28(8), pp. 1599–1611.

Hamberger, B. and Bak, S. 2013. Plant P450s as versatile drivers for evolution of species-specific chemical diversity. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 368(1612), p. 20120426.

Hamberger, B., Ohnishi, T., Hamberger, B., Séguin, A. and Bohlmann, J. 2011. Evolution of diterpene metabolism: Sitka spruce CYP720B4 catalyzes multiple oxidations in resin acid biosynthesis of conifer defense against insects. *Plant Physiology* 157(4), pp. 1677–1695.

Hampton, R., Dimster-Denk, D. and Rine, J. 1996. The biology of HMG-CoA reductase: the pros of contra-regulation. *Trends in Biochemical Sciences* 21(4), pp. 140–145.

Hampton, R.Y. and Bhakta, H. 1997. Ubiquitin-mediated regulation of 3-hydroxy-3-methylglutaryl-CoA reductase. *Proceedings of the National Academy of Sciences of the United States of America* 94(24), pp. 12944–12948.

Hampton, R.Y. and Garza, R.M. 2009. Protein quality control as a strategy for cellular regulation: lessons from ubiquitin-mediated regulation of the sterol pathway. *Chemical Reviews* 109(4), pp. 1561–1574.

Harvey, B.G., Wright, M.E. and Quintana, R.L. 2010. High-Density Renewable Fuels Based on the Selective Dimerization of Pinenes. *Energy & fuels : an American Chemical Society journal* 24(1), pp. 267–273.

Hassan, E., Tayebeh, R., Tousi Samaneh, E., Sadr Zeinalabedin, B., Vahid, N. and Mehdi, Z. 2008. Quantification of Valerenic Acid and its Derivatives in Some Species of Valeriana L. and Centranthus longiflorus Stev. *Asian Journal of Plant Sciences* 7(2), pp. 195–200.

Hata, N., Hayashi, Y., Okazawa, A., Ono, E., Satake, H. and Kobayashi, A. 2010. Comparison of sesamin contents and CYP81Q1 gene expressions in aboveground vegetative organs between two Japanese sesame (Sesamum indicum L.) varieties differing in seed sesamin contents. *Plant*

Science 178(6), pp. 510–516.

Haudenschild, C., Schalk, M., Karp, F. and Croteau, R. 2000. Functional expression of regiospecific cytochrome P450 limonene hydroxylases from mint (Mentha spp.) in Escherichia coli and saccharomyces cerevisiae. *Archives of Biochemistry and Biophysics* 379(1), pp. 127–136.

Hefner, J., Rubenstein, S.M., Ketchum, R.E.B., Gibson, D.M., Williams, R.M. and Croteau, R. 1996. Cytochrome P450-catalyzed hydroxylation of taxa-4(5),11(12)-diene to taxa-4(20),11(12)-dien-5a-o1: the first oxygenation step in taxol biosynthesis. *Chemistry & Biology* 3(6), pp. 479–489.

Hellier, P., Al-Haj, L., Talibi, M., Purton, S. and Ladommatos, N. 2013. Combustion and emissions characterization of terpenes with a view to their biological production in cyanobacteria. *Fuel* 111, pp. 670–688.

van Herpen, T.W.J.M., Cankar, K., Nogueira, M., Bosch, D., Bouwmeester, H.J. and Beekwilder, J. 2010. Nicotiana benthamiana as a production platform for artemisinin precursors. *Plos One* 5(12), p. e14222.

Hohl, R.J. 1996. Monoterpenes as regulators of malignant cell proliferation. *Advances in Experimental Medicine and Biology* 401, pp. 137–146.

Horton, C.E., Huang, K.-X., Bennett, G.N. and Rudolph, F.B. 2003. Heterologous expression of the Saccharomyces cerevisiae alcohol acetyltransferase genes in Clostridium acetobutylicum and Escherichia coli for the production of isoamyl acetate. *Journal of Industrial Microbiology & Biotechnology* 30(7), pp. 427–432.

Horwitz, A.A., Walter, J.M., Schubert, M.G., Kung, S.H., Hawkins, K., Platt, D.M., Hernday, A.D., Mahatdejkul-Meadows, T., Szeto, W., Chandran, S.S. and Newman, J.D. 2015. Efficient Multiplexed Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas. *Cell systems* 1(1), pp. 88–96.

Houghton, P.J. 1988. The biological activity of valerian and related plants. *Journal of Ethnopharmacology* 22(2), pp. 121–142.

Howe, E., Holton, K., Nair, S., Schlauch, D., Sinha, R. and Quackenbush, J. 2010. MeV: MultiExperiment Viewer. In: Ochs, M. F., Casagrande, J. T., and Davuluri, R. V. eds. *Biomedical informatics for cancer research*. Boston, MA: Springer US, pp. 267–277.

Hritcu, L. and Cioanca, O. 2016. Prevalence of use of herbal medicines and complementary and alternative medicine in europe. In: Grosso, C. ed. *Herbal medicine in depression*. Cham: Springer International Publishing, pp. 135–181.

Huang, Q., Roessner, C.A., Croteau, R. and Scott, A.I. 2001. Engineering Escherichia coli for the synthesis of taxadiene, a key intermediate in the biosynthesis of taxol. *Bioorganic & Medicinal Chemistry* 9(9), pp. 2237–2242.

Hull, A., Golubkov, I., Kronberg, B., Marandzheva, T. and Stam, J. van 2006. An Alternative Fuel for Spark Ignition Engines. *International Journal of Engine Research* 7(3), pp. 203–214.

Ignea, C., Athanasakoglou, A., Ioannou, E., Georgantea, P., Trikka, F.A., Loupassaki, S., Roussis, V., Makris, A.M. and Kampranis, S.C. 2016. Carnosic acid biosynthesis elucidated by a synthetic biology platform. *Proceedings of the National Academy of Sciences of the United States of America* 113(13), pp. 3681–3686.

Ignea, C., Cvetkovic, I., Loupassaki, S., Kefalas, P., Johnson, C.B., Kampranis, S.C. and Makris, A.M. 2011. Improving yeast strains using recyclable integration cassettes, for the production of plant terpenoids. *Microbial Cell Factories* 10, p. 4.

Ignea, C., Pontini, M., Maffei, M.E., Makris, A.M. and Kampranis, S.C. 2014. Engineering monoterpene production in yeast using a synthetic dominant negative geranyl diphosphate synthase. *ACS synthetic biology [electronic resource]* 3(5), pp. 298–306.

Ignea, C., Trikka, F.A., Nikolaidis, A.K., Georgantea, P., Ioannou, E., Loupassaki, S., Kefalas, P., Kanellis, A.K., Roussis, V., Makris, A.M. and Kampranis, S.C. 2015. Efficient diterpene production in yeast by engineering Erg20p into a geranylgeranyl diphosphate synthase. *Metabolic Engineering* 27, pp. 65–75.

Ikezawa, N., Göpfert, J.C., Nguyen, D.T., Kim, S.-U., O'Maille, P.E., Spring, O. and Ro, D.-K. 2011. Lettuce costunolide synthase (CYP71BL2) and its homolog (CYP71BL1) from sunflower catalyze distinct regio- and stereoselective hydroxylations in sesquiterpene lactone metabolism. *The Journal of Biological Chemistry* 286(24), pp. 21601–21611.

Jennewein, S. and Croteau, R. 2001. Taxol: biosynthesis, molecular genetics, and biotechnological applications. *Applied Microbiology and Biotechnology* 57(1–2), pp. 13–19.

Jensen, K. and Møller, B.L. 2010. Plant NADPH-cytochrome P450 oxidoreductases. *Phytochemistry* 71(2–3), pp. 132–141.

Jiao, W., Dong, W., Li, Z., Deng, M. and Lu, R. 2009. Lathyrane diterpenes from Euphorbia lathyris as modulators of multidrug resistance and their crystal structures. *Bioorganic & Medicinal Chemistry* 17(13), pp. 4786–4792.

Johnson, H.E., Banack, S.A. and Cox, P.A. 2008. Variability in content of the anti-AIDS drug candidate prostratin in Samoan populations of Homalanthus nutans. *Journal of Natural Products* 71(12), pp. 2041–2044.

Jongedijk, E., Cankar, K., Ranzijn, J., van der Krol, S., Bouwmeester, H. and Beekwilder, J. 2015. Capturing of the monoterpene olefin limonene produced in Saccharomyces cerevisiae. *Yeast* 32(1), pp. 159–171.

Kang, M.-K., Eom, J.-H., Kim, Y., Um, Y. and Woo, H.M. 2014. Biosynthesis of pinene from glucose using metabolically-engineered Corynebacterium glutamicum. *Biotechnology Letters* 36(10), pp. 2069–2077.

Kaspera, R. and Croteau, R. 2006. Cytochrome P450 oxygenases of Taxol biosynthesis. *Phytochemistry reviews : proceedings of the Phytochemical Society of Europe* 5(2–3), pp. 433–444.

Katoh, K., Misawa, K., Kuma, K. and Miyata, T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30(14), pp. 3059–3066.

Khom, S., Baburin, I., Timin, E., Hohaus, A., Trauner, G., Kopp, B. and Hering, S. 2007. Valerenic acid potentiates and inhibits GABA(A) receptors: molecular mechanism and subunit specificity. *Neuropharmacology* 53(1), pp. 178–187.

Kim, E.-M., Eom, J.-H., Um, Y., Kim, Y. and Woo, H.M. 2015. Microbial Synthesis of Myrcene by Metabolically Engineered Escherichia coli. *Journal of Agricultural and Food Chemistry* 63(18), pp. 4606–4612.

King, A.J., Brown, G.D., Gilday, A.D., Forestier, E., Larson, T.R. and Graham, I.A. 2016. A Cytochrome P450-Mediated Intramolecular Carbon-Carbon Ring Closure in the Biosynthesis of Multidrug-Resistance-Reversing Lathyrane Diterpenoids. *Chembiochem* 17(17), pp. 1593–1597.

King, A.J., Brown, G.D., Gilday, A.D., Larson, T.R. and Graham, I.A. 2014. Production of bioactive diterpenoids in the euphorbiaceae depends on evolutionarily conserved gene clusters. *The Plant Cell* 26(8), pp. 3286–3298.

Kirby, J. and Keasling, J.D. 2009. Biosynthesis of plant isoprenoids: perspectives for microbial engineering. *Annual review of plant biology* 60, pp. 335–355.

Kirby, J., Nishimoto, M., Park, J.G., Withers, S.T., Nowroozi, F., Behrendt, D., Rutledge, E.J.G., Fortman, J.L., Johnson, H.E., Anderson, J.V. and Keasling, J.D. 2010. Cloning of casbene and neocembrene synthases from Euphorbiaceae plants and expression in Saccharomyces cerevisiae. *Phytochemistry* 71(13), pp. 1466–1473.

Kirk, J. and Dunker, K.S. 2014. Dietary counseling: the ingredient for successfully addressing the use of herbal supplements and probiotics in chronic kidney disease. *Advances in chronic kidney disease* 21(4), pp. 377–384.

Kiyota, H., Okuda, Y., Ito, M., Hirai, M.Y. and Ikeuchi, M. 2014. Engineering of cyanobacteria for the photosynthetic production of limonene from CO2. *Journal of Biotechnology* 185, pp. 1–7.

Kizer, L., Pitera, D.J., Pfleger, B.F. and Keasling, J.D. 2008. Application of functional genomics to pathway optimization for increased isoprenoid production. *Applied and Environmental Microbiology* 74(10), pp. 3229–3241.

Köllner, T.G., Gershenzon, J. and Degenhardt, J. 2009. Molecular and biochemical evolution of maize terpene synthase 10, an enzyme of indirect defense. *Phytochemistry* 70(9), pp. 1139–1145.

Kopp, S., Schweizer, W. and Altmann, K.-H. 2009. Total synthesis of valerenic acid. *Synlett : accounts and rapid communications in synthetic organic chemistry* 2009(11), pp. 1769–1772.

de Kraker, J.W., Franssen, M.C., Dalm, M.C., de Groot, A. and Bouwmeester, H.J. 2001. Biosynthesis of germacrene A carboxylic acid in chicory roots. Demonstration of a cytochrome P450 (+)-germacrene a hydroxylase and NADP+-dependent sesquiterpenoid dehydrogenase(s) involved in sesquiterpene lactone biosynthesis. *Plant Physiology* 125(4), pp. 1930–1940.

Kumar, V. 2006. Potential medicinal plants for CNS disorders: an overview. *Phytotherapy Research* 20(12), pp. 1023–1035.

Kung, Y., McAndrew, R.P., Xie, X., Liu, C.C., Pereira, J.H., Adams, P.D. and Keasling, J.D. 2014. Constructing tailored isoprenoid products by structure-guided modification of geranylgeranyl reductase. *Structure* 22(7), pp. 1028–1036.

Lange, B.M., Rujan, T., Martin, W. and Croteau, R. 2000. Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proceedings of the National Academy of Sciences of the United States of America* 97(24), pp. 13172–13177.

Langfelder, P. and Horvath, S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9, p. 559.

Lee, M.E., DeLoache, W.C., Cervantes, B. and Dueber, J.E. 2015. A highly characterized yeast toolkit for modular, multipart assembly. *ACS synthetic biology [electronic resource]* 4(9), pp. 975–986.

Leipoldt, F., Zeyhle, P., Kulik, A., Kalinowski, J., Heide, L. and Kaysser, L. 2015. Diversity of ABBA Prenyltransferases in Marine Streptomyces sp. CNQ-509: Promiscuous Enzymes for the Biosynthesis of Mixed Terpenoid Compounds. *Plos One* 10(12), p. e0143237.

Li, C., Ng, A., Xie, L., Mao, H., Qiu, C., Srinivasan, R., Yin, Z. and Hong, Y. 2016. Engineering low phorbol ester Jatropha curcas seed by intercepting casbene biosynthesis. *Plant Cell Reports* 35(1), pp. 103–114.

Li, L., Stoeckert, C.J. and Roos, D.S. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research* 13(9), pp. 2178–2189.

Liao, P., Hemmerlin, A., Bach, T.J. and Chye, M.-L. 2016. The potential of the mevalonate pathway for enhanced isoprenoid production. *Biotechnology advances* 34(5), pp. 697–713.

Lindberg, P., Park, S. and Melis, A. 2010. Engineering a platform for photosynthetic isoprene production in cyanobacteria, using Synechocystis as the model organism. *Metabolic Engineering* 12(1), pp. 70–79.

Lubertozzi, D. and Keasling, J.D. 2008. Expression of a synthetic Artemesia annua amorphadiene synthase in Aspergillus nidulans yields altered product distribution. *Journal of Industrial Microbiology & Biotechnology* 35(10), pp. 1191–1198.

Luo, D., Callari, R., Hamberger, B., Wubshet, S.G., Nielsen, M.T., Andersen-Ranberg, J., Hallström, B.M., Cozzi, F., Heider, H., Lindberg Møller, B., Staerk, D. and Hamberger, B. 2016. Oxidation and cyclization of casbene in the biosynthesis of Euphorbia factors from mature seeds

of Euphorbia lathyris L. *Proceedings of the National Academy of Sciences of the United States of America* 113(34), pp. 5082–5089.

Luo, P., Wang, Y.H., Wang, G.D., Essenberg, M. and Chen, X.Y. 2001. Molecular cloning and functional identification of (+)-delta-cadinene-8-hydroxylase, a cytochrome P450 mono-oxygenase (CYP706B1) of cotton sesquiterpene biosynthesis. *The Plant Journal: for Cell and Molecular Biology* 28(1), pp. 95–104.

Lupien, S., Karp, F., Ponnamperuma, K., Wildung, M. and Croteau, R. 1995. Cytochrome P450 limonene hydroxylases of Mentha species. *Drug metabolism and drug interactions* 12(3–4), pp. 245–260.

Maartens, G., Celum, C. and Lewin, S.R. 2014. HIV infection: epidemiology, pathogenesis, treatment, and prevention. *The Lancet* 384(9939), pp. 258–271.

Mack, J.H., Rapp, V.H., Broeckelmann, M., Lee, T.S. and Dibble, R.W. 2014. Investigation of biofuels from microorganism metabolism for use as anti-knock additives. *Fuel* 117, pp. 939–943.

Martin, D.M., Fäldt, J. and Bohlmann, J. 2004. Functional characterization of nine Norway Spruce TPS genes and evolution of gymnosperm terpene synthases of the TPS-d subfamily. *Plant Physiology* 135(4), pp. 1908–1927.

Martin, V.J.J., Yoshikuni, Y. and Keasling, J.D. 2001. The in vivo synthesis of plant sesquiterpenes byEscherichia coli. *Journal of Biochemical and Microbiological Technology and Engineering* 75(5), pp. 497–503.

Marwick, C. 2005. Researchers investigate potential use of plant as a pain killer. *BMJ (Clinical Research Ed.)* 331(7525), p. 1104.

McCaskill, D. and Croteau, R. 1997. Prospects for the bioengineering of isoprenoid biosynthesis. *Advances in Biochemical Engineering/Biotechnology* 55, pp. 107–146.

McGarvey, D.J. and Croteau, R. 1995. Terpenoid metabolism. *The Plant Cell* 7(7), pp. 1015–1026.

Meadows, A.L., Hawkins, K.M., Tsegaye, Y., Antipov, E., Kim, Y., Raetz, L., Dahl, R.H., Tai, A., Mahatdejkul-Meadows, T., Xu, L., Zhao, L., Dasika, M.S., Murarka, A., Lenihan, J., Eng, D., Leng, J.S., Liu, C.-L., Wenger, J.W., Jiang, H., Chao, L. and Tsong, A.E. 2016. Rewriting yeast central carbon metabolism for industrial isoprenoid production. *Nature* 537(7622), pp. 694–697.

Meigs, T.E., Roseman, D.S. and Simoni, R.D. 1996. Regulation of 3-hydroxy-3-methylglutaryl-coenzyme A reductase degradation by the nonsterol mevalonate metabolite farnesol in vivo. *The Journal of Biological Chemistry* 271(14), pp. 7916–7922.

Meigs, T.E. and Simoni, R.D. 1997. Farnesol as a regulator of HMG-CoA reductase degradation: characterization and role of farnesyl pyrophosphatase. *Archives of Biochemistry and Biophysics* 345(1), pp. 1–9.

Melillo, E., Setroikromo, R., Quax, W.J. and Kayser, O. 2013. Production of α-cuprenene in

Xanthophyllomyces dendrorhous: a step closer to a potent terpene biofactory. *Microbial Cell Factories* 12, p. 13.

Mercke, P., Bengtsson, M., Bouwmeester, H.J., Posthumus, M.A. and Brodelius, P.E. 2000. Molecular cloning, expression, and characterization of amorpha-4,11-diene synthase, a key enzyme of artemisinin biosynthesis in Artemisia annua L. *Archives of Biochemistry and Biophysics* 381(2), pp. 173–180.

Mercke, P., Crock, J., Croteau, R. and Brodelius, P.E. 1999. Cloning, expression, and characterization of epi-cedrol synthase, a sesquiterpene cyclase from Artemisia annua L. *Archives of Biochemistry and Biophysics* 369(2), pp. 213–222.

Meylemans, H.A., Quintana, R.L., Goldsmith, B.R. and Harvey, B.G. 2011. Solvent-free conversion of linalool to methylcyclopentadiene dimers: a route to renewable high-density fuels. *ChemSusChem* 4(4), pp. 465–469.

Meylemans, H.A., Quintana, R.L. and Harvey, B.G. 2012. Efficient conversion of pure and mixed terpene feedstocks to high density fuels. *Fuel* 97, pp. 560–568.

Mi, J., Schewe, H., Buchhaupt, M., Holtmann, D. and Schrader, J. 2016. Efficient hydroxylation of 1,8-cineole with monoterpenoid-resistant recombinant Pseudomonas putida GS1. *World journal of microbiology & biotechnology* 32(7), p. 112.

Miana, G.A., Bashir, M. and Evans, F.J. 1985. Isolation of Prostratin from Euphorbia cornigera. *Planta Medica* 51(4), pp. 353–354.

Nakano, Y., Ohtani, M., Polsri, W., Usami, T., Sambongi, K. and Demura, T. 2012. Characterization of the casbene synthase homolog from Jatropha (Jatropha curcas L.). *Plant Biotechnology* 29(2), pp. 185–189.

Narita, K., Ohnuma, S. and Nishino, T. 1999. Protein design of geranyl diphosphate synthase. Structural features that define the product specificities of prenyltransferases. *Journal of Biochemistry* 126(3), pp. 566–571.

Natarajan, P., Kanagasabapathy, D., Gunadayalan, G., Panchalingam, J., Shree, N., Sugantham, P.A., Singh, K.K. and Madasamy, P. 2010. Gene discovery from Jatropha curcas by sequencing of ESTs from normalized and full-length enriched cDNA library from developing seeds. *BMC Genomics* 11, p. 606.

Natarajan, P. and Parani, M. 2011. De novo assembly and transcriptome analysis of five major tissues of Jatropha curcas L. using GS FLX titanium platform of 454 pyrosequencing. *BMC Genomics* 12, p. 191.

Nelson, D. and Werck-Reichhart, D. 2011. A P450-centric view of plant evolution. *The Plant Journal: for Cell and Molecular Biology* 66(1), pp. 194–211.

Newman, J.D., Marshall, J., Chang, M., Nowroozi, F., Paradise, E., Pitera, D., Newman, K.L. and Keasling, J.D. 2006. High-level production of amorpha-4,11-diene in a two-phase

partitioning bioreactor of metabolically engineered Escherichia coli. *Biotechnology and Bioengineering* 95(4), pp. 684–691.

Nguyen, D.T., Göpfert, J.C., Ikezawa, N., Macnevin, G., Kathiresan, M., Conrad, J., Spring, O. and Ro, D.-K. 2010. Biochemical conservation and evolution of germacrene A oxidase in asteraceae. *The Journal of Biological Chemistry* 285(22), pp. 16588–16598.

Nomura, T., Magome, H., Hanada, A., Takeda-Kamiya, N., Mander, L.N., Kamiya, Y. and Yamaguchi, S. 2013. Functional analysis of Arabidopsis CYP714A1 and CYP714A2 reveals that they are distinct gibberellin modification enzymes. *Plant & Cell Physiology* 54(11), pp. 1837–1851.

Ohnuma, S., Nakazawa, T., Hemmi, H., Hallberg, A.M., Koyama, T., Ogura, K. and Nishino, T. 1996. Conversion from farnesyl diphosphate synthase to geranylgeranyl diphosphate synthase by random chemical mutagenesis. *The Journal of Biological Chemistry* 271(17), pp. 10087–10095.

Okamoto, S., Yu, F., Harada, H., Okajima, T., Hattan, J., Misawa, N. and Utsumi, R. 2011. A short-chain dehydrogenase involved in terpene metabolism from Zingiber zerumbet. *The FEBS Journal* 278(16), pp. 2892–2900.

Ozaki, T., Zhao, P., Shinada, T., Nishiyama, M. and Kuzuyama, T. 2014. Cyclolavandulyl skeleton biosynthesis via both condensation and cyclization catalyzed by an unprecedented member of the cis-isoprenyl diphosphate synthase superfamily. *Journal of the American Chemical Society* 136(13), pp. 4837–4840.

Özaydın, B., Burd, H., Lee, T.S. and Keasling, J.D. 2013. Carotenoid-based phenotypic screen of the yeast deletion collection reveals new genes with roles in isoprenoid production. *Metabolic Engineering* 15, pp. 174–183.

O'Connor, S.E. and Maresh, J.J. 2006. Chemistry and biology of monoterpene indole alkaloid biosynthesis. *Natural Product Reports* 23(4), pp. 532–547.

Paddon, C.J., Westfall, P.J., Pitera, D.J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M.D., Tai, A., Main, A., Eng, D., Polichuk, D.R., Teoh, K.H., Reed, D.W., Treynor, T., Lenihan, J., Fleck, M., Bajad, S., Dang, G., Dengrove, D., Diola, D. and Newman, J.D. 2013. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* 496(7446), pp. 528–532.

Payne, C.K., Mosbaugh, P.G., Forrest, J.B., Evans, R.J., Whitmore, K.E., Antoci, J.P., Perez-Marrero, R., Jacoby, K., Diokno, A.C., O'Reilly, K.J., Griebling, T.L., Vasavada, S.P., Yu, A.S., Frumkin, L.R. and ICOS RTX Study Group (Resiniferatoxin Treatment for Interstitial Cystitis) 2005. Intravesical resiniferatoxin for the treatment of interstitial cystitis: a randomized, double-blind, placebo controlled trial. *The Journal of Urology* 173(5), pp. 1590–1594.

Peralta-Yahya, P.P., Ouellet, M., Chan, R., Mukhopadhyay, A., Keasling, J.D. and Lee, T.S. 2011. Identification and microbial production of a terpene-based advanced biofuel. *Nature Communications* 2, p. 483.

Phelan, R.M., Sekurova, O.N., Keasling, J.D. and Zotchev, S.B. 2015. Engineering terpene biosynthesis in Streptomyces for production of the advanced biofuel precursor bisabolene. *ACS synthetic biology [electronic resource]* 4(4), pp. 393–399.

Phulara, S.C., Chaturvedi, P. and Gupta, P. 2016. Isoprenoid-Based Biofuels: Homologous Expression and Heterologous Expression in Prokaryotes. *Applied and Environmental Microbiology* 82(19), pp. 5730–5740.

Picaud, S., Brodelius, M. and Brodelius, P.E. 2005. Expression, purification and characterization of recombinant (E)-beta-farnesene synthase from Artemisia annua. *Phytochemistry* 66(9), pp. 961–967.

Pitera, D.J., Paddon, C.J., Newman, J.D. and Keasling, J.D. 2007. Balancing a heterologous mevalonate pathway for improved isoprenoid production in Escherichia coli. *Metabolic Engineering* 9(2), pp. 193–207.

Potter, D. and Miziorko, H.M. 1997. Identification of catalytic residues in human mevalonate kinase. *The Journal of Biological Chemistry* 272(41), pp. 25449–25454.

Potter, D., Wojnar, J.M., Narasimhan, C. and Miziorko, H.M. 1997. Identification and functional characterization of an active-site lysine in mevalonate kinase. *The Journal of Biological Chemistry* 272(9), pp. 5741–5746.

Pyle, B.W., Tran, H.T., Pickel, B., Haslam, T.M., Gao, Z., MacNevin, G., Vederas, J.C., Kim, S.-U. and Ro, D.-K. 2012. Enzymatic synthesis of valerena-4,7(11)-diene by a unique sesquiterpene synthase from the valerian plant (Valeriana officinalis). *The FEBS Journal* 279(17), pp. 3136–3146.

Ralston, L., Kwon, S.T., Schoenbeck, M., Ralston, J., Schenk, D.J., Coates, R.M. and Chappell, J. 2001. Cloning, heterologous expression, and functional characterization of 5-epi-aristolochene-1,3-dihydroxylase from tobacco (Nicotiana tabacum). *Archives of Biochemistry and Biophysics* 393(2), pp. 222–235.

Reider Apel, A., d'Espaux, L., Wehrs, M., Sachs, D., Li, R.A., Tong, G.J., Garber, M., Nnadi, O., Zhuang, W., Hillson, N.J., Keasling, J.D. and Mukhopadhyay, A. 2017. A Cas9-based toolkit to program gene expression in Saccharomyces cerevisiae. *Nucleic Acids Research* 45(1), pp. 496–508.

Reiling, K.K., Yoshikuni, Y., Martin, V.J.J., Newman, J., Bohlmann, J. and Keasling, J.D. 2004. Mono and diterpene production in Escherichia coli. *Biotechnology and Bioengineering* 87(2), pp. 200–212.

Reinsvold, R.E., Jinkerson, R.E., Radakovits, R., Posewitz, M.C. and Basu, C. 2011. The production of the sesquiterpene β-caryophyllene in a transgenic strain of the cyanobacterium Synechocystis. *Journal of Plant Physiology* 168(8), pp. 848–852.

Renault, H., Bassard, J.-E., Hamberger, B. and Werck-Reichhart, D. 2014. Cytochrome P450-

mediated metabolic engineering: current progress and future challenges. *Current Opinion in Plant Biology* 19, pp. 27–34.

Renninger, N. and McPhee, D. 2008. Fuel compositions comprising farnesane and farnesane derivatives and method of making and using same. Patent US20080098645. . Available at: https://www.google.pl/patents/US20080083158 [Accessed: 29 October 2016].

Ricigliano, V., Kumar, S., Kinison, S., Brooks, C., Nybo, S.E., Chappell, J. and Howarth, D.G. 2016. Regulation of sesquiterpenoid metabolism in recombinant and elicited Valeriana officinalis hairy roots. *Phytochemistry* 125, pp. 43–53.

Ro, D.-K., Paradise, E.M., Ouellet, M., Fisher, K.J., Newman, K.L., Ndungu, J.M., Ho, K.A., Eachus, R.A., Ham, T.S., Kirby, J., Chang, M.C.Y., Withers, S.T., Shiba, Y., Sarpong, R. and Keasling, J.D. 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* 440(7086), pp. 940–943.

Rodríguez-Concepción, M., Ahumada, I., Diez-Juez, E., Sauret-Güeto, S., Lois, L.M., Gallego, F., Carretero-Paulet, L., Campos, N. and Boronat, A. 2001. 1-Deoxy-D-xylulose 5-phosphate reductoisomerase and plastid isoprenoid biosynthesis during tomato fruit ripening. *The Plant Journal: for Cell and Molecular Biology* 27(3), pp. 213–222.

Rohdich, F., Hecht, S., Gärtner, K., Adam, P., Krieger, C., Amslinger, S., Arigoni, D., Bacher, A. and Eisenreich, W. 2002. Studies on the nonmevalonate terpene biosynthetic pathway: metabolic role of IspH (LytB) protein. *Proceedings of the National Academy of Sciences of the United States of America* 99(3), pp. 1158–1163.

Rohdich, F., Wungsintaweekul, J., Fellermeier, M., Sagner, S., Herz, S., Kis, K., Eisenreich, W., Bacher, A. and Zenk, M.H. 1999. Cytidine 5'-triphosphate-dependent biosynthesis of isoprenoids: YgbP protein of Escherichia coli catalyzes the formation of 4-diphosphocytidyl-2-C-methylerythritol. *Proceedings of the National Academy of Sciences of the United States of America* 96(21), pp. 11758–11763.

Rohmer, M., Knani, M., Simonin, P., Sutter, B. and Sahm, H. 1993. Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate. *The Biochemical Journal* 295 (Pt 2), pp. 517–524.

Roth, R.J. and Acton, N. 1989. A simple conversion of artemisinic acid into artemisinin. *Journal of Natural Products* 52(5), pp. 1183–1185.

Rude, M.A. and Schirmer, A. 2009. New microbial fuels: a biotech perspective. *Current Opinion in Microbiology* 12(3), pp. 274–281.

Sarria, S., Wong, B., García Martín, H., Keasling, J.D. and Peralta-Yahya, P. 2014. Microbial synthesis of pinene. *ACS synthetic biology [electronic resource]* 3(7), pp. 466–475.

Sato, S., Hirakawa, H., Isobe, S., Fukai, E., Watanabe, A., Kato, M., Kawashima, K., Minami, C., Muraki, A., Nakazaki, N., Takahashi, C., Nakayama, S., Kishida, Y., Kohara, M., Yamada,

M., Tsuruoka, H., Sasamoto, S., Tabata, S., Aizu, T., Toyoda, A. and Fukui, K. 2011. Sequence analysis of the genome of an oil-bearing tree, Jatropha curcas L. *DNA Research* 18(1), pp. 65–76.

Schalk, M. and 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(22), pp. 11948–11953.

Schoendorf, A., Rithner, C.D., Williams, R.M. and Croteau, R.B. 2001. Molecular cloning of a cytochrome P450 taxane 10 beta-hydroxylase cDNA from Taxus and functional expression in yeast. *Proceedings of the National Academy of Sciences of the United States of America* 98(4), pp. 1501–1506.

Schwender, J., Seemann, M., Lichtenthaler, H.K. and Rohmer, M. 1996. Biosynthesis of isoprenoids (carotenoids, sterols, prenyl side-chains of chlorophylls and plastoquinone) via a novel pyruvate/glyceraldehyde 3-phosphate non-mevalonate pathway in the green alga*Scenedesmus obliquus*. *The Biochemical journal* 316(1), pp. 73–80.

Seip, E.H. and Hecker, E. 1983. Lathyrane type diterpenoid esters from Euphorbia characias. *Phytochemistry* 22(8), pp. 1791–1795.

Shimokawa, K., Takamura, H. and Uemura, D. 2007. Concise synthesis of a highly functionalized cyclopentane segment: toward the total synthesis of kansuinine A. *Tetrahedron letters* 48(32), pp. 5623–5625.

Siller, G., Gebauer, K., Welburn, P., Katsamas, J. and Ogbourne, S.M. 2009. PEP005 (ingenol mebutate) gel, a novel agent for the treatment of actinic keratosis: results of a randomized, double-blind, vehicle-controlled, multicentre, phase IIa study. *The Australasian Journal of Dermatology* 50(1), pp. 16–22.

Singh, R., Vadlani, P.V., Harrison, M.L., Bennett, G.N. and San, K.Y. 2008. Aerobic production of isoamyl acetate by overexpression of the yeast alcohol acetyl-transferases AFT1 and AFT2 in Escherichia coli and using low-cost fermentation ingredients. *Bioprocess and biosystems engineering* 31(4), pp. 299–306.

Skeel, R.T. and Khleif, S.N. eds. 2011. *Handbook of Cancer Chemotherapy*. illustrated. Lippincott Williams & Wilkins.

Son, Y.-J., Kwon, M., Ro, D.-K. and Kim, S.-U. 2014. Enantioselective microbial synthesis of the indigenous natural product (-)-α-bisabolol by a sesquiterpene synthase from chamomile (Matricaria recutita). *The Biochemical Journal* 463(2), pp. 239–248.

Sparkes, I.A., Runions, J., Kearns, A. and Hawes, C. 2006. Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nature Protocols* 1(4), pp. 2019–2025.

Sprenger, G.A., Schörken, U., Wiegert, T., Grolle, S., de Graaf, A.A., Taylor, S.V., Begley, T.P., Bringer-Meyer, S. and Sahm, H. 1997. Identification of a thiamin-dependent synthase in Escherichia coli required for the formation of the 1-deoxy-D-xylulose 5-phosphate precursor to isoprenoids, thiamin, and pyridoxol. *Proceedings of the National Academy of Sciences of the United States of America* 94(24), pp. 12857–12862.

Srivalli, K.M.R. and Lakshmi, P.K. 2012. Overview of P-glycoprotein inhibitors: a rational outlook. *Brazilian Journal of Pharmaceutical Sciences* 48(3), pp. 353–367.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9), pp. 1312–1313.

Steele, C.L., Crock, J., Bohlmann, J. and Croteau, R. 1998. Sesquiterpene synthases from grand fir (Abies grandis). Comparison of constitutive and wound-induced activities, and cDNA isolation, characterization, and bacterial expression of delta-selinene synthase and gamma-humulene synthase. *The Journal of Biological Chemistry* 273(4), pp. 2078–2089.

Szkopińska, A., Swiezewska, E. and Karst, F. 2000. The regulation of activity of main mevalonic acid pathway enzymes: farnesyl diphosphate synthase, 3-hydroxy-3-methylglutaryl-CoA reductase, and squalene synthase in yeast Saccharomyces cerevisiae. *Biochemical and Biophysical Research Communications* 267(1), pp. 473–477.

Takahashi, S., Kuzuyama, T. and Seto, H. 1999. Purification, characterization, and cloning of a eubacterial 3-hydroxy-3-methylglutaryl coenzyme A reductase, a key enzyme involved in biosynthesis of terpenoids. *Journal of Bacteriology* 181(4), pp. 1256–1263.

Takahashi, S., Kuzuyama, T., Watanabe, H. and Seto, H. 1998. A 1-deoxy-D-xylulose 5-phosphate reductoisomerase catalyzing the formation of 2-C-methyl-D-erythritol 4-phosphate in an alternative nonmevalonate pathway for terpenoid biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America* 95(17), pp. 9879–9884.

Takahashi, S., Yeo, Y.-S., Zhao, Y., O'Maille, P.E., Greenhagen, B.T., Noel, J.P., Coates, R.M. and Chappell, J. 2007. Functional characterization of premnaspirodiene oxygenase, a cytochrome P450 catalyzing regio- and stereo-specific hydroxylations of diverse sesquiterpene substrates. *The Journal of Biological Chemistry* 282(43), pp. 31744–31754.

Takase, H., Sasaki, K., Shinmori, H., Shinohara, A., Mochizuki, C., Kobayashi, H., Ikoma, G., Saito, H., Matsuo, H., Suzuki, S. and Takata, R. 2016. Cytochrome P450 CYP71BE5 in grapevine (Vitis vinifera) catalyzes the formation of the spicy aroma compound (-)-rotundone. *Journal of Experimental Botany* 67(3), pp. 787–798.

Takemoto, H., Yagura, T. and Ito, M. 2009. Evaluation of volatile components from spikenard: valerena-4,7(11)-diene is a highly active sedative compound. *Journal of natural medicines* 63(4), pp. 380–385.

Tanino, K., Onuki, K., Asano, K., Miyashita, M., Nakamura, T., Takahashi, Y. and Kuwajima, I. 2003. Total synthesis of ingenol. *Journal of the American Chemical Society* 125(6), pp. 1498–

1500.

Tarshis, L.C., Proteau, P.J., Kellogg, B.A., Sacchettini, J.C. and Poulter, C.D. 1996. Regulation of product chain length by isoprenyl diphosphate synthases. *Proceedings of the National Academy of Sciences of the United States of America* 93(26), pp. 15018–15023.

Tarshis, L.C., Yan, M., Poulter, C.D. and Sacchettini, J.C. 1994. Crystal Structure of Recombinant Farnesyl Diphosphate Synthase at 2.6-.ANG. Resolution. *Biochemistry* 33(36), pp. 10871–10877.

Toth, M.J. and Huwyler, L. 1996. Molecular cloning and expression of the cDNAs encoding human and yeast mevalonate pyrophosphate decarboxylase. *The Journal of Biological Chemistry* 271(14), pp. 7895–7898.

Tracy, N.I., Chen, D., Crunkleton, D.W. and Price, G.L. 2009. Hydrogenated monoterpenes as diesel fuel additives. *Fuel* 88(11), pp. 2238–2240.

Trauner, G., Khom, S., Baburin, I., Benedek, B., Hering, S. and Kopp, B. 2008. Modulation of GABAA receptors by valerian extracts is related to the content of valerenic acid. *Planta Medica* 74(1), pp. 19–24.

Trikka, F.A., Nikolaidis, A., Athanasakoglou, A., Andreadelli, A., Ignea, C., Kotta, K., Argiriou, A., Kampranis, S.C. and Makris, A.M. 2015. Iterative carotenogenic screens identify combinations of yeast gene deletions that enhance sclareol production. *Microbial Cell Factories* 14(1), p. 60.

Tsuruta, H., Paddon, C.J., Eng, D., Lenihan, J.R., Horning, T., Anthony, L.C., Regentin, R., Keasling, J.D., Renninger, N.S. and Newman, J.D. 2009. High-level production of amorpha-4,11-diene, a precursor of the antimalarial agent artemisinin, in Escherichia coli. *Plos One* 4(2), p. e4489.

Unterweger, B., Bulach, D.M., Scoble, J., Midgley, D.J., Greenfield, P., Lyras, D., Johanesen, P. and Dumsday, G.J. 2016. Characterisation of CYP101J2, CYP101J3 and CYP101J4, three 1,8-cineole-hydroxylating cytochrome P450 monooxygenases from Sphingobium yanoikuyae strain B2. *Applied and Environmental Microbiology* 82(22), pp. 6507–6517.

Vasas, A. and Hohmann, J. 2014. Euphorbia diterpenes: isolation, structure, biological activity, and synthesis (2008-2012). *Chemical Reviews* 114(17), pp. 8579–8612.

Walker, K. and Croteau, R. 2001. Taxol biosynthetic genes. *Phytochemistry* 58(1), pp. 1–7.

Wang, C., Kim, J.-Y., Choi, E.-S. and Kim, S.-W. 2011. Microbial production of farnesol (FOH): Current states and beyond. *Process Biochemistry* 46(6), pp. 1221–1229.

Wang, C., Yoon, S.-H., Jang, H.-J., Chung, Y.-R., Kim, J.-Y., Choi, E.-S. and Kim, S.-W. 2011. Metabolic engineering of Escherichia coli for α-farnesene production. *Metabolic Engineering* 13(6), pp. 648–655.

Wang, H., Zou, Z., Wang, S. and Gong, M. 2013. Global analysis of transcriptome responses and

gene expression profiles to cold stress of Jatropha curcas L. *Plos One* 8(12), p. e82817.

Wang, H.-B., Wang, X.-Y., Liu, L.-P., Qin, G.-W. and Kang, T.-G. 2015. Tigliane diterpenoids from the Euphorbiaceae and Thymelaeaceae families. *Chemical Reviews* 115(9), pp. 2975–3011.

Wang, K. 2000. Isoprenyl diphosphate synthases. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* 1529(1–3), pp. 33–48.

Wang, K. and Ohnuma, S. 1999. Chain-length determination mechanism of isoprenyl diphosphate synthases and implications for molecular evolution. *Trends in Biochemical Sciences* 24(11), pp. 445–451.

Wang, Y.-B., Huang, R., Wang, H.-B., Jin, H.-Z., Lou, L.-G. and Qin, G.-W. 2006. Diterpenoids from the roots of Euphorbia fischeriana. *Journal of Natural Products* 69(6), pp. 967–970.

Wang, Y.-H. and Essenberg, M. 2010. Inhibitor and substrate activities of sesquiterpene olefins toward +-δ-cadinene-8-hydroxylase, a cytochrome P450 monooxygenase (CYP706B1). *Phytochemistry* 71(16), pp. 1825–1831.

Weaver, L.J., Sousa, M.M.L., Wang, G., Baidoo, E., Petzold, C.J. and Keasling, J.D. 2015. A kinetic-based approach to understanding heterologous mevalonate pathway function in E. coli. *Biotechnology and Bioengineering* 112(1), pp. 111–119.

Wender, P.A., Kee, J.-M. and Warrington, J.M. 2008. Practical synthesis of prostratin, DPP, and their analogs, adjuvant leads against latent HIV. *Science* 320(5876), pp. 649–652.

Wender, P.A. and McDonald, F.E. 1990. Studies on tumor promoters. 9. A second-generation synthesis of phorbol. *Journal of the American Chemical Society* 112(12), pp. 4956–4958.

Westfall, P.J., Pitera, D.J., Lenihan, J.R., Eng, D., Woolard, F.X., Regentin, R., Horning, T., Tsuruta, H., Melis, D.J., Owens, A., Fickes, S., Diola, D., Benjamin, K.R., Keasling, J.D., Leavell, M.D., McPhee, D.J., Renninger, N.S., Newman, J.D. and Paddon, C.J. 2012. Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *Proceedings of the National Academy of Sciences of the United States of America* 109(3), pp. E111-8.

Withers, S.T., Gottlieb, S.S., Lieu, B., Newman, J.D. and Keasling, J.D. 2007. Identification of isopentenol biosynthetic genes from Bacillus subtilis by a screening method based on isoprenoid precursor toxicity. *Applied and Environmental Microbiology* 73(19), pp. 6277–6283.

Wriessnegger, T., Augustin, P., Engleder, M., Leitner, E., Müller, M., Kaluzna, I., Schürmann, M., Mink, D., Zellnig, G., Schwab, H. and Pichler, H. 2014. Production of the sesquiterpenoid (+)-nootkatone by metabolic engineering of Pichia pastoris. *Metabolic Engineering* 24, pp. 18–29.

Wüst, M. and Croteau, R.B. 2002. Hydroxylation of specifically deuterated limonene enantiomers by cytochrome p450 limonene-6-hydroxylase reveals the mechanism of multiple product formation. *Biochemistry* 41(6), pp. 1820–1827.

Yang, C., Gao, X., Jiang, Y., Sun, B., Gao, F. and Yang, S. 2016. Synergy between methylerythritol phosphate pathway and mevalonate pathway for isoprene production in Escherichia coli. *Metabolic Engineering* 37, pp. 79–91.

Yang, J., Nie, Q., Ren, M., Feng, H., Jiang, X., Zheng, Y., Liu, M., Zhang, H. and Xian, M. 2013. Metabolic engineering of Escherichia coli for the biosynthesis of alpha-pinene. *Biotechnology for Biofuels* 6(1), p. 60.

Yang, Y., Dec, J.E., Dronniou, N. and Simmons, B. 2010. Characteristics of Isopentanol as a Fuel for HCCI Engines. *SAE International Journal of Fuels and Lubricants* 3(2), pp. 725–741.

Ye, J., Qu, J., Bui, H.T.N. and Chua, N.-H. 2009. Rapid analysis of Jatropha curcas gene functions by virus-induced gene silencing. *Plant Biotechnology Journal* 7(9), pp. 964–976.

Yeo, Y.-S., Nybo, S.E., Chittiboyina, A.G., Weerasooriya, A.D., Wang, Y.-H., Góngora-Castillo, E., Vaillancourt, B., Buell, C.R., DellaPenna, D., Celiz, M.D., Jones, A.D., Wurtele, E.S., Ransom, N., Dudareva, N., Shaaban, K.A., Tibrewal, N., Chandra, S., Smillie, T., Khan, I.A., Coates, R.M. and Chappell, J. 2013. Functional identification of valerena-1,10-diene synthase, a terpene synthase catalyzing a unique chemical cascade in the biosynthesis of biologically active sesquiterpenes in Valeriana officinalis. *The Journal of Biological Chemistry* 288(5), pp. 3163–3173.

Yoshikuni, Y., Ferrin, T.E. and Keasling, J.D. 2006. Designed divergent evolution of enzyme function. *Nature* 440(7087), pp. 1078–1082.

Yu, F., Okamoto, S., Harada, H., Yamasaki, K., Misawa, N. and Utsumi, R. 2011. Zingiber zerumbet CYP71BA1 catalyzes the conversion of α-humulene to 8-hydroxy-α-humulene in zerumbone biosynthesis. *Cellular and Molecular Life Sciences* 68(6), pp. 1033–1040.

Zerbe, P., Hamberger, B., Yuen, M.M.S., Chiang, A., Sandhu, H.K., Madilao, L.L., Nguyen, A., Hamberger, B., Bach, S.S. and Bohlmann, J. 2013. Gene discovery of modular diterpene metabolism in nonmodel systems. *Plant Physiology* 162(2), pp. 1073–1091.

Zhan, X., Zhang, Y.-H., Chen, D.-F. and Simonsen, H.T. 2014. Metabolic engineering of the moss Physcomitrella patens to produce the sesquiterpenoids patchoulol and α/β-santalene. *Frontiers in plant science* 5, p. 636.

Zhang, H., Liu, Q., Cao, Y., Feng, X., Zheng, Y., Zou, H., Liu, H., Yang, J. and Xian, M. 2014. Microbial production of sabinene--a new terpene-based precursor of advanced biofuel. *Microbial Cell Factories* 13, p. 20.

Zhou, K., Qiao, K., Edgar, S. and Stephanopoulos, G. 2015. Distributing a metabolic pathway among a microbial consortium enhances production of natural products. *Nature Biotechnology* 33(4), pp. 377–383.

7. Appendices

7.1. Supplementary for Chapter 2 — Development of Euphorbiaceae genetic resources for elucidation of the prostratin biosynthetic pathway

Table S2.1. List of strains and plasmids used in this study.

Table S1
Yeast strains used in this study

Strain	Parent (+ additional genetic changes)	JBEI Acc. No.	References
GTy116	MATa leu2-3,112::His3MX6_P_{GAL1}-ERG19/P_{GAL10}-ERG8 ura3-52::URA3_P_{GAL1}-mvaS(A110G)/P_{GAL10}-mvaE(CO) his3Δ1::hphMX4_P_{GAL1}-ERG12/P_{GAL10}-IDI1 trp1-289::TRP1_P_{GAL1}-crtE(X.den)/P_{GAL10}-ERG20 yprcδ15::natMX_$_{PGAL1}$-crtE(opt)/$_{PGAL10}$-crtE	JBx_063850	Ref.16
JWy1	GTy116 (pRS425::P_{GAL1}-*JcCBS1*-T_{CYC1})	JBx_040280	This study

Plasmid Name	Description	JBEI Acc. No.	References
pEAQ-CfDXS	pEAQ-CfDXS	JBx_063991	Ref.17
pEAQ-CfGGPPS	pEAQ-CfGGPPS	JBx_063992	Ref.17
pJW01	pRS425::P_{GAL1}-JcCBS1-T_{CYC1}	JBx_039606	This study
pJW02	pCAMBIA-JcCBS1	JBx_063993	This study
pJW03	pms057-JcCYP1	JBx_055074	This study
pJW04	pms057-JcCYP2	JBx_055076	This study
pJW05	pms057-JcCYP3	JBx_055078	This study
pJW06	pms057-JcCYP4	JBx_055080	This study
pJW07	pms057-JcCYP5	JBx_055082	This study
pJW08	pms057-JcCYP6	JBx_055084	This study
pJW09	pms057-JcCYP7	JBx_055086	This study
pJW10	pms057-JcCYP8	JBx_055088	This study
pJW11	pms057-JcCYP9	JBx_055090	This study
pJW12	pms057-JcCYP10	JBx_055092	This study
pJW13	pms057-JcCYP11	JBx_055094	This study
pJW14	pms057-JcCYP12	JBx_063948	This study
pJW15	pms057-JcCYP14	JBx_063949	This study
pJW16	pms057-JcCYP16	JBx_063950	This study
pJW17	pms057-JcCYP21	JBx_063951	This study
pJW18	pms057-JcCYP22	JBx_063952	This study
pJW19	pms057-JcCYP23	JBx_063953	This study
pJW20	pms057-JcCYP25	JBx_063954	This study
pJW21	pms057-JcCYP27	JBx_063932	This study
pJW22	pms057-JcCYP28	JBx_063933	This study
pJW23	pms057-JcCYP30	JBx_063934	This study
pJW24	pms057-JcCYP31	JBx_063935	This study
pJW25	pms057-JcCYP32	JBx_063936	This study
pJW26	pms057-JcCYP33	JBx_063968	This study
pJW27	pms057-JcCYP35	JBx_063969	This study
pJW28	pms057-JcCYP36	JBx_063970	This study
pJW29	pms057-JcCYP37	JBx_063971	This study
pJW30	pms057-JcCYP38	JBx_063972	This study
pJW31	pms057-JcCYP40	JBx_063973	This study
pJW32	pms057-HnCYP1	JBx_063981	This study
pJW33	pms057-HnCYP2	JBx_063982	This study
pJW34	pms057-HnCYP3	JBx_063983	This study
pJW35	pms057-HnCYP5	JBx_063984	This study
pJW36	pms057-ErCYP71.7	JBx_063985	This study
pJW37	pms057-RcCYP726A13	JBx_063986	This study
pJW38	pms057-RcCYP726A14	JBx_063995	This study
pJW39	pms057-RcCYP726A15	JBx_063987	This study
pJW40	pms057-RcCYP726A16	JBx_063988	This study
pJW41	pms057-RcCYP726A17	JBx_063989	This study
pJW42	pms057-RcCYP726A18	JBx_063990	This study

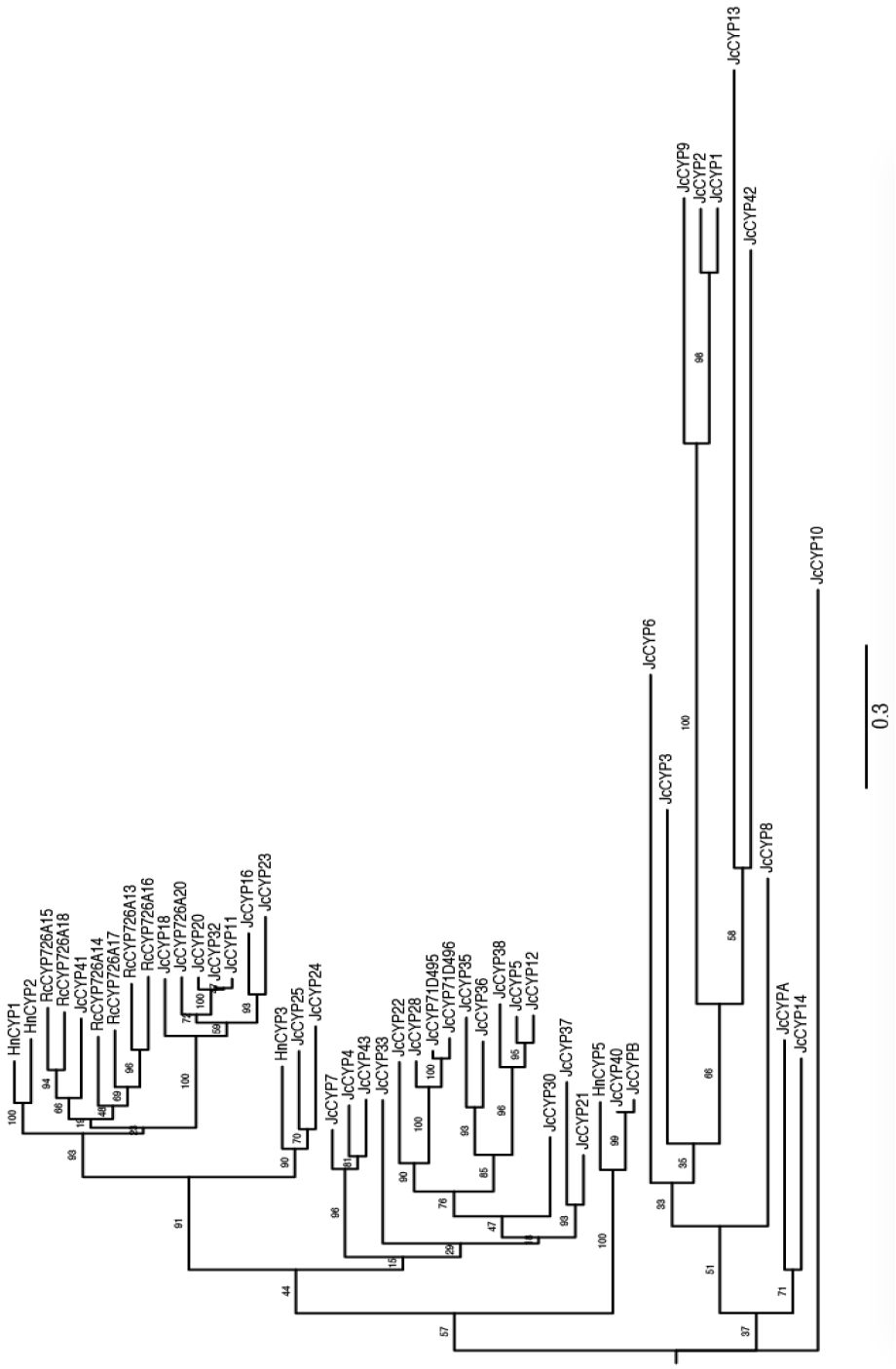
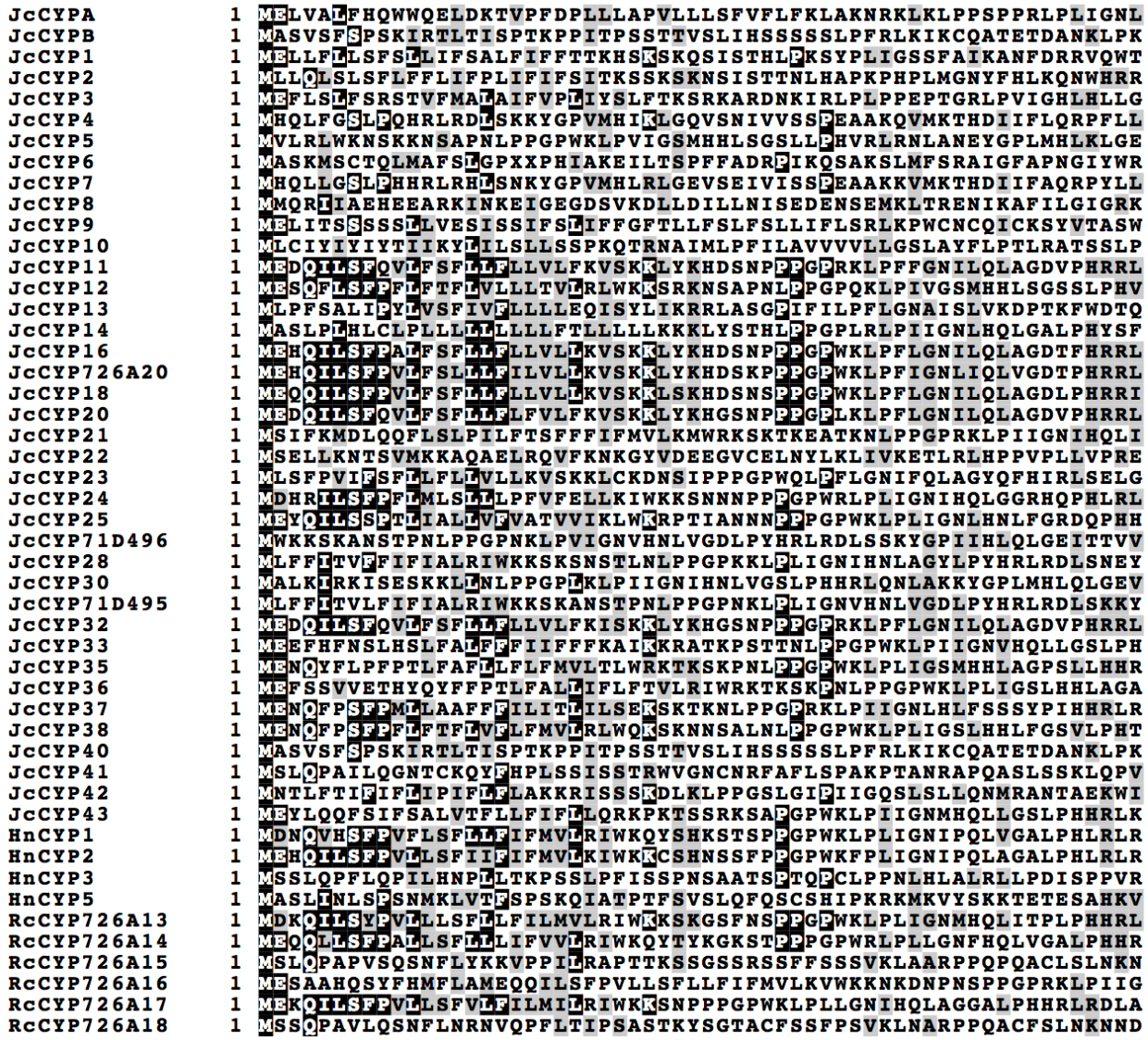


Figure S2.2. Phylogenetic tree of all P450s used in this study. *R. communis* P450s were included in the tree, along with P450 candidates from *J. curcas* and *H. nutans*. JcCYP10, (JcCYP736A12, XP_012092981.1) was used as an outgroup.The neighbor-joining tree was generated using MAFFT and RAxML. The numbers indicate the bootstrap value (%) from 100 replications. The scale bar shows the amino acid substitution ratio.

Figure S2.3. Gblocks curated alignment for all P450s used in this study. Note, *R. communis* casbene oxidases were included to generate the alignment. The multiple sequence alignment was generated using T-Coffee (http://tcoffee.crg.cat) with default parameter values. Output was formatted using Boxshade (https://embnet.vital-it.ch/software/BOX_form.html) with the output type set to postcript_portrait, 0.3 fraction of sequences similar for shading, and default parameters.



```
JcCYPA          61 HQLGPLPYRSLKKLSDKYGPLMMVHFGKVPTLVVSSAEIAQEITKNHDIAFGDRPKTAAA
JcCYPB          61 RLPPGPIKLPLIGNLHNLAGAQPHHALTELAKEYGPLMHLQLGEISAIVVSNPRIAQVIM
JcCYP1          61 SDILQTLPSATFVLHRPMGGRQIFTGNPANVQHILKTHFHLYRKGPVTRYTLFDFLGNGI
JcCYP2          61 IQWISDAVVNSPSATWVLHRPLGSGQFIITGNPANVQHILKTNFHIYEKGPIVRSTLFDF
JcCYP3          61 GSQPPHITLEKMAEKIGPIYSIKLGVHRALIVSSWEMVKECFTTNDKAFAYRPKALFLDV
JcCYP4          61 AAEILMYNFKDIAFAPYGDSWRQMRKICTLELLSTKRVRSFRPIREDEVSTFIRTISSSS
JcCYP5          61 VTNIVVSSPETAKAIMKTHDHIFAQRPFLLAANIMAYNSTDLAFAPYGDYWRQMRKICTQ
JcCYP6          61 LLRKIASSHLFSPRRILAHETLRQLECASMLRNIANEQTQNGRVYLRKHLQFASLNNIMG
JcCYP7          61 AADIILYNFKDIVFAPYGEGWRQMRKICTLQLLSTKRVRSFRVVREEETSKFIRSISGLP
JcCYP8          61 LKKVGEDFDKMMQRIIAEHEEARKINKEIGEGDSVKDLLDILLNISEDENSEMKLTRENI
JcCYP9          61 TKDFANLCDWYTHLLKKSPTGTIHVHVLNNIITANPENVEYILKTNFENYPKGKPFSALL
JcCYP10         61 SEHRENDRKLPPGPSPLPIIGNLHMLGNLPHQTLYNLAKLHGPIMSLRLGYVQTIVVSSA
JcCYP11         61 TALAKTYGPVMGIKLGQIPFLVVSSPETAKEVMKIQDPVFAERALLVAVEIVLYNRNDIV
JcCYP12         61 RLRDLAKEHGPIMHLQLGEVTNIVVSSPETAKAVMKTHDHIFAQRPFLLAANIMAYNSTD
JcCYP13         61 SALSFGHGFSVNYIIGRFIVFIRDTELSHLIFSNVRPDAFMLVGHPFGKKLFGEHNLIYM
JcCYP14         61 WQVSKKFGPVMLLQFGRVPTVIISSAETAKELIKTNDLSSCSRPRLAGTGRLSYNFLDIA
JcCYP16         61 TELAKTHGPVMSINVGQIPYVVVSSPETAKEVMKIQDPVFADHPVVLAAEVILYSPYDIF
JcCYP726A20     61 TALAKTYGPVMGVQLGQVPFLVVSSPETAKEVMKIQDPVFAERPLVLAGEIVLYNRNDIV
JcCYP18         61 TELAKKYGPVMSIKLGQHPYLVVSSPETAKEVMRTQDPIFADRPLVLAGELVLYNRNDIG
JcCYP20         61 TALAKTYGPVMGIKLGQIPFLVVSSPETAKEVMKIQDPVFAERAPLLAGEIVLYNRNDII
JcCYP21         61 GSLPHHCLRDLAKKHGGIMHLQLGEVSNIVISSPEAAKEVMKTHDIVFAQRPFLLAASII
JcCYP22         61 NSELCEINGYFIPVKSRVLINVWAIGRDPNYWKEPERFNPERFLDNSIDYKGSNFEFIPF
JcCYP23         61 QTYGPVMGIKVGQVPFLIVSSPEMAKEVLKVQDPTFVDRPVVLAAELVMYGGHDIVYAPY
JcCYP24         61 TDLARTYGPVMRLQLGQIEAVVISSAETAKQVMKTQESQFLGRPSLLAADIMLYNRTDIS
JcCYP25         61 RLRDLAGKYGAVMGFQLGQVPTVVISSAEIAKQVLKTHEFQFIDRPSLLAADIVLYNRSD
JcCYP71D496     61 ISSPELAQEVMKTHDLNFAQRPFVLAGDIVSYKCTDIALAPYGEYWRQLRKMCSLELLTA
JcCYP28         61 GPIMHLQLGEINSIVVSSPELAKEVMKTHDINFAYRPFVLAGDIVSYKCKDIAFAPYGEY
JcCYP30         61 TTIVVTSAEIAKEVMRAHDIVFSNRPSILAANIISYNATSIVFSPYGEYWRQLRKICVLE
JcCYP71D495     61 GPIMHLQLGENTTVVISSPELAQEVMKTHDVNFAQRPFVLAGDIVSYKCKDIAFAPYGEY
JcCYP32         61 TALAKTYGPVMGIKLGQIPFLVVSSPETAKEVMKIQDPVFAERALLLAGEIVLYNRNDII
JcCYP33         61 QSLQKLSGKYGPLMHLKLGEVSTVIVSSPEIAKQVLKTHDLDFAERPPNLAPKIISYDST
JcCYP35         61 VTELARKYGPIMHLQLGQVTNIFISSPEIAREVMKTHDLIFATRPSLVAVQLVTYNFTDI
JcCYP36         61 PLTHVRLRDLAKKYGPIMHLQLGEVTTIFITSPEIAKEVLKTLDIVLARRPFLQAVKLVT
JcCYP37         61 DLSKKYGAVMHLKLGQVSTIGIGSPEAAKEMMKTNDVCFADRPCYQSAEIVTYNFLDIAY
JcCYP38         61 RLRDLANEYGPIMHLQLGQVTNIVLSSPETAKAVLKTHDHIFTQRPFVLAAETMTYNFTN
JcCYP40         61 RLPPGPIKLPLIGNLHNLAGAQPHHALTELAKEYGPLMHLQLGEISAIVVSNPRIAQVIM
JcCYP41         61 VRLLTKFPASGFLAMNQSVDQFASTTTSLTKIFNKIGKPIQSSPFLVSVLLLMFMASKIQ
JcCYP42         61 EKRIQKYGPISKLSLFGKPTVFMYGQAANKFVFTSDSSTLSNSQTQSVKMILGEKCLLEL
JcCYP43         61 DLSDKYGSVMNLQLGQVSNIVISSPEAAKQVMKTHDIIFVQRPFLLAANIIMYNSKDIVF
HnCYP1          61 DLAKIYGPVMSIQLGQVPVVIISSSETAKEVLKTQDVQFADRSLILAGKMVLYDRMDIIF
HnCYP2          61 DLSKIHGPIMSIQMGQVPAVVISSPETAKQVLKTQDVQFADRPLIQAGKFVLYNQLDILY
HnCYP3          61 GNRFPTFASNQFVNQPTSTPEEDNDGNPTLPPGPWKLPLIGNIHQLLGDLPHRRLRDLAK
HnCYP5          61 LPPGPMKLPVIGNLLNLVGSEPHHALAQLAKEYGPLMHLQLGEISAVVVSNPKMAQEIMK
RcCYP726A13     61 RELAKTHGPVMSIQLGQVSAVVISSVEAAKQVLKTQGELFAERPSILASKIVLYNGMDII
RcCYP726A14     61 LTELAKIYGPVMGIQLGQISVVIISSVETAKEVLKTQGEQFADRTLVLAAKMVLYNRNDI
RcCYP726A15     61 DDSNTSASSLPPGPWKLPLLGNIHQLVGALPHHRLRDLAKAYGPVMSVKLGEVSAVVISS
RcCYP726A16     61 NMHQLAGSDLPHHPVTELSKTYGPIMSIQLGQISAIVISSVEGAKEVLKTQGELFAERPL
RcCYP726A17     61 KTYGPVMSIQLGQISAVVISSVQGAKEVLKTQGEVFAERPLIIAAKIVLYNRKDIVFGSY
RcCYP726A18     61 HSTPTSILPPGPWQLPLIGNIHQLVGHLPHSRLRDLGKIYGPVMSVQLGEVSAVVVSSVE
```

```
JcCYPA          121 DDLFFGCQNLAFCPYGEYWRQVKKVCVLELLSQKRVQYFEFVRREETANLVEKLRHASLQ
JcCYPB          121 KTHDLVFADRPEILAAKIITYGGQDIAFSKLGEYWKQMKRISLMELLGPKTVQSFASLRE
JcCYP1          121 FNADGNTWKFQRQVASHEFSTKSLRKFVETVVDTELSERLIPILSVAAANKTVVDLQDIL
JcCYP2          121 LGNGIFNTDGETWKIQRQIASHEFNTKSLRKFVQTVVDTELSQRLIPILSTAAANQTVLD
JcCYP3          121 MSYDYAMFGASPYGDYWREMRKIATLELLSVRRLELLKHVRETEVKEATEGLYQGWLKNK
JcCYP4          121 KVNLGRMVFALSNTITLRSAFGKVSERKEAFLPLVQKIVQVLEGFSVADVFPSVRFLHRI
JcCYP5          121 EILSAKRVLSFGLIREEEVSKFIRDLSSRAGSTVNFSRMFNSVTYNIIQRVAIGKLWKGE
JcCYP6          121 SVFGKRYDPAHDSKELEELRDMVREGFELLGAFNWCDYLQWLSYFYDPFRINERCLKLVP
JcCYP7          121 KVNISKMVFSLSNAITLKSAFGKVSERHDAFLPLVQKIMLVFGGFSVADFFPSVKFLHRI
JcCYP8          121 KAFILRIIAEHEEARKINKEIGEGDSVKDLLDILLNISEDENSEMKLTRENIKAFILDIF
JcCYP9          121 GDLLGKGIFNVDGDAWKFQRKMASLELGSVSIRSYAFDLITSEIKERLLPLLSSVSSENR
JcCYP10         121 NAAKLFLKTHDAVFGSRPKLRASRYMSYGTKGMAFTEYGPYWRSVRKLCTVQLLSASKIE
JcCYP11         121 FGLYGDQWRQMRKICTLELLSAKRVQSFRSVREEEVADLVKFLGSKEGSPVNLTHTLFAL
JcCYP12         121 LAFAPYGDYWRQMRKICTQEMLSNKRVQSFGLIREEEVSKLIAELSSRAGSTVNFSKMFN
JcCYP13         121 FGQDHKDLRRRIAPNFTPRALSTYSQLQQIIMLKHLKEWETMAMDNPNKPISLRLLVRDM
JcCYP14         121 FTPYGDYWRVMRKICVHELFSAKRVQSFQSIREEEVGLLIDSILKSSSSSTPVDLSEKTM
JcCYP16         121 FAPYGDHLKQMRKFCTVELLSTKRVQSFRSVREEEVADFVKFLRSKEGSSVNLTHTLFAL
JcCYP726A20     121 FGSYGDQWRQMRKFCTLELLSTKRVQSFRPVREEEVASFVKLMRTKKGTPVNLTHALFAL
JcCYP18         121 FGLYGDQWRQMRKFCALELLSTKRVQSFRSVREEEIAEFVKSLRSKEGSSVNLSHTLFAL
JcCYP20         121 FGLYGDQWRQMRKICTLELLSAKRVQSFRSVREEEVADLVKFLGSKEGSPVNLTHTLFAL
JcCYP21         121 SYNFTDIAFSPYADYWRQLRKICILELLSAKRVQSFRFIREEEVSNLITAISSSSGKAFN
JcCYP22         121 GAGRRICPGILFGIANVELPLANLLYHFDWKLPGEINPENLEMTEVFIIFLFKLLPKKSK
JcCYP23         121 GDQWRQMRKFCTLELLSTKRVQSFRSVREEEAGEFVKFLLSKEGSSVNLTHALYALSNSM
JcCYP24         121 FAPYGDYWRQMKKIAVVELLSAKRVQAYKSVMDEEVSNFINFLYSKAGSPVNLTKTFYSL
JcCYP25         121 IIFAPYGDYWRQIKKIAILELLSSKRVQSFKSVREEEVSSFFKFLYSKAGSPVNLSRTLL
JcCYP71D496     121 KRVQSFKSIREEEVFKLVESISSRSGSPINFSKMASSLTYAIISRAVCGKVSRGEEVFVP
JcCYP28         121 WRQLRKMCSLELLTAKRVQSFKSIREEEGSKLLQSISSSSGSPVNFSKMTSSLTYSIISR
JcCYP30         121 LLSAKRVQSFKSIREEEVSNIVRRISSSSDSLINLSRMLFSLTYSITSRAAFGKIRKEQE
JcCYP71D495     121 WRQLRKMCSLELLTAKRVQSFKSIREEEVSKLVESISSSSGSPINFSKMASSLTYAIISR
JcCYP32         121 FGLYGDQWRQMRKICTLELLSAKRVQSFRSVREEEVADLVKFLGSKEGSPVNLTHTLFAL
JcCYP33         121 HIVFSPYGAYWRQLRKICTMELLSPKRVQSFRFIREDEVLNLIKTISSLEGSLINISEMI
JcCYP35         121 AFAPYGDYWRQIKKICTMELLTAKRVQLFAPIRQEEVSKVITDITSNVGSTINFTNVLTS
JcCYP36         121 YNFTDVAFSPYGEYWRQLRKICTMELLTAKRVQSFGSIRQEEGSKLIRDISSNAGSPINF
JcCYP37         121 SPYEDYWRQLRKICTVELLSAKRVQSFRSIREEEVANLIRDISSSSGKPFNLSKRIFALT
JcCYP38         121 LANAPYGGYWRQIRKICTLEMLSAKRVRSFGLIREEEVSKFIRDLSSSTSAGSTVNFSRM
JcCYP40         121 KTHDLVFADRPEILAAKIITYGGQDIAFSKLGEYWKQMKRISLMELLGPKTVQSFASLRE
JcCYP41         121 NQQEEDDNSINLPPGPWRLPFIGNIHQLAGPGLPHHRLTDLAKTYGPVMGVHLGEVYAVV
JcCYP42         121 SGKDHKRVREALMSFLKPESLKQYVGKIDEEVRMHILKNWQRKQEVQVLPLMKTLTFNII
JcCYP43         121 APYGDSWRQMRKICTLELLSTKRVRSFRAIREEETSNFIRSISSLSEVNISKILLSLSNA
HnCYP1          121 GSYGDHWRQMRKICTLELLSAKRVQSFRSVREEEVENFIKHLHSKAGSPVNLTKALFALT
HnCYP2          121 APYGDHWRQMRKICTLELLSAKRVQSFRSIREEEVSNYIKFLHSKAGSPVNLTETLFSLT
HnCYP3          121 IYGPVMSIRLGEVPAVVISSVEAAKEVLRTQDVNFAERPPVLVAEIVLYNRQDIVFRSYG
HnCYP5          121 THDLIFANRPQLLASEIVTYGGKDIAFAPLGEYWKEMKRISLTELLGPRRVQSFSSIREN
RcCYP726A13     121 FGSYGDHWRQMRKICTFELLSPKRVQSFSSVRQEELSNYVRFLHSNAGSPVNLSKTLFAL
RcCYP726A14     121 VFGLYGDHWRQLRKLCTLELLSAKRVQSFKSVREELSNFVKFLHSKAGMPVNLTHTLFA
RcCYP726A15     121 VDAAKEVLRTQDVNFADRPLVLAAEIVLYNRQDIVFGSYGEQWRQMRKICTLELLSIKRV
RcCYP726A16     121 LLAAEAVLYNRMDIIFGAYGDHWRQLRKLCTLEVLSAKRIQSFSSLRQEELSHFVRFVHS
RcCYP726A17     121 GDHWRQMRKICTLELLSAKRVQSFRSVREEEVSEFVRFLQSKAGTPVNLTKTLFALTNSI
RcCYP726A18     121 AAKEVLRIQDVIFAERPPVLMAEIVLYNRHDIVFGSYGDHWRQLRKICTLELLSLKRVQS
```

```
JcCYPA          181 GSPVDLSELLVSISNNIVSRSALGTVYNNESGHSSSGDLVRGAIDLVGSFTFQDFFPSLG
JcCYPB          181 NEVEKLIQSIHLSKGKPINFTEKIFHLTNVITCKAAFGDECKDQDVVIALTKEATTIAGG
JcCYP1          181 QRFAFDNICKIAFGYDPAYLLPSLPPMDFAQAFEESIRIISDRFNCAFPIIWKIKKFFGV
JcCYP2          181 FQDILQRFAFDNICKIAFGYDPGYLLPSLPQTKFADAFEDSVRISFERFNSLFWKIKRAF
JcCYP3          181 NSNNKLEVEMIKWFWDVSLNGILKMVVGKRYVEYIKKDEDINESGHWRYALRDFFELSGK
JcCYP4          181 TGMRGKLEKLHQETDIMLENIINEHRENKRLGRSNSEGKEDDLVDVLLNIQDSDNLEFPL
JcCYP5          181 EIVIPAFKKLIEAAGGFSLSDLYPSIKLLHKISTTKFKLLRAHKETDKLFQNIIDEHRAR
JcCYP6          181 RVRKFVRGIIEEHRLDGEEKLQEDDMLAVLWEMIFRGTDTTALLTEWVMAELVLHPEMQE
JcCYP7          181 TGMRSKLEKLHQEADIMLENIINEHRENKRLGRSNSEGKEDVLVDVLLNLQNCDNLEFPL
JcCYP8          181 AAGTDTSSIATEWALAELINHPDIMRKAKEEIDFVVGKSRLVEESDITNLPYLQSIVKET
JcCYP9          181 SLDLQDVFRRFSFDSICKFSFGLDPGCLKLSLPVSEFALAFDTASKLSAERALTASPLIW
JcCYP10         181 YFAPIRKEELGFYVESLKRAAAARKVVDFSVGVGDMIQNIACRMVFGEVNNHELDLKALV
JcCYP11         181 ANSMIARNTVGHKSKNQEALLRLIDDIIESIGGVGIADIFPSLKWLPSVQRERSRIRKLH
JcCYP12         181 SVTYNIIQRVAIGKLWKGEEVVIPAIKKLIEASGGFSLSDVYPSIKLLHKISTTRFKLQR
JcCYP13         181 NLDTSQTVMVGPYLKSEARERFKFDYNLFNVGTMKLPIDLPGFAFRNARLAVDRLAQTLA
JcCYP14         181 SLTANVICRVAFGKSFQERGFNHERFQEVVREGLAMLGSFTAADFFPHVGWIVDRLTGLH
JcCYP16         181 TNSIVARTAVGHRSKNQEGLLKVIDEAVLASSGVNIADIFPSLQWLPSVKRERSRIWKTH
JcCYP726A20     181 TNSIVARNAVGHKSKNQEALLEVIDDIVVSGGGVSIVDIFPSLQWLPTAKRERSRIWKLH
JcCYP18         181 TNSIIARNTVGHKSKNQEALLKIIDDIVESLGGLSTVDIFPSLKWLPSVKRERSRIWKLH
JcCYP20         181 ANSIIARNTVGQKSKNQEALLRLIDDIIELTGSVSIADIFPSLKWLPSVQRDRSRIRKLH
JcCYP21         181 FSRKLFSLTYGIAARATFGEKCEDQEEFIPIVEEITEVAGGFSLADLFPSVKFLHSISGM
JcCYP22         181 SLNLPPGPSSLPLIGGFHHLFGAPPHHSLTNLAKKFGPIFHLQLGENSNIVISSAEMAKE
JcCYP23         181 VARSTVGHKTKNQEALLNVIDDTVSTAAGTNIADIFPSLKWLPTVKRQMSRIWKSHCQTD
JcCYP24         181 GNGIIAKTSIGKKFKKQETFLKVVDKAIRVAGGFSVGDAFPSFKLIHLITGISSTLHTAH
JcCYP25         181 SLTNGIIAKTSIGKKCKRQEEIIAVITDAIKATGGFSVADVFPSFKFLHIITGISSTIRR
JcCYP71D496     181 AVEKLVEAGRSISLADLYPSIKLFNALSVVRRRVEKIHGEVDKIIETIVMEHRERKRMVD
JcCYP28         181 AAFGKVCQGEEVFVPAVVKLTEAGRSISLADVYPSVKLFNTFSVVRRNVEKIHSEVDKIV
JcCYP30         181 AFIPLVEEIIEVGGGFSIADLFPSIKLLNRINGMKSRVERLHQEADKILENIINEHRASK
JcCYP71D495     181 AVCGKVSRGEEVFVPAVEKLVEAGRSISLADLYPSVKLFNALSVVRRRVEKIHGEVDKII
JcCYP32         181 ANSIIARNTVGHKSKNQEALLRLIDDIIESIGGVGIADIFPSLKWLPSVQRERSRIRKLH
JcCYP33         181 FSLTYGITSRAAFGKKYEDQETFIQVITEVSKIAAGFSVADLYPSIKFLEQASGLRPKLG
JcCYP35         181 LTYKILSRSTIGKILKGEEGFIRAVMDLTEEGAGFNLADFYPSIKLFRMFGSLKHKLKRI
JcCYP36         181 SKILTSSGYKIISRAAFGQVWNGEDVFLKAVNDLTEESAGFSLVDFYPSKKFLQLFTSSG
JcCYP37         181 YSITARVSFGDKCREQDAFISAAEKIMQTTGFDLADLFPSLKFLGWFSEMRTRLMNAHDE
JcCYP38         181 FSSVTYNIIQRVAIGKISKGEDTVFPAIRKLIEAFVGFNLSDAYPSIKLLHKISTKRFKL
JcCYP40         181 NEVEKLIQSIHLSKGKPINFTEKIFHLTNVITCKAAFGDECKDQDVVIALTKEATTIAGG
JcCYP41         181 VSSAETSKEVLRTQDTNFAERPLVNAAKMVLYNRNDIVFGSFGDQWRQMRKICTLELLSV
JcCYP42         181 CSLLFGIERGSRRDKLVDLFQEMIKGMWSIPINLPFTRYNSSLKASTSVRNMLKDLISEK
JcCYP43         181 ITLRSAFGKVSERQEAFLPLVQKIALMLEGFSIADIFPSVKFLHGITGMRSKLQKLHQEA
HnCYP1          181 NSIMAITSIGKKCKNQEAILSIIDDVIEVAGGFSVADVFPSFKFLHYISGEKSRLQKLHD
HnCYP2          181 NSIMLRISIGTKHKNQETILSIIEEVTEAAGGFSVADVFPSLKFLHYISGEKSRLQKLHQ
HnCYP3          181 DEWRQMKKIARLELLSAKRVHSFKSIREEQVSNFIKFIYSSTGSPVNLSKELMSLTNSII
HnCYP5          181 EVEELIESVRLSAGKPFNFTENIFRFTNVITCKAAFGDECKDQDAVIALSKQATELASGF
RcCYP726A13     181 TNSVIAKIAVGKECKNQEALLNLIEEVLVAAGGFTVADSFPSYNFLHVITGMKSNLERLH
RcCYP726A14     181 LTNNIMARTSVGKKCKNQEALLSIIDGIIDASGGFTIADVFPSVPFLHNISNMKSRLEKL
RcCYP726A15     181 QSFKSVREEELSNFIRYLHSKAGTPVNLTHHLFSLTNSIMFRISIGKKYKNQDALLRVID
RcCYP726A16     181 KAGSPINLSKVLFALTNSIIARIATGKKCKNQDALLDLIEDVIEVSGGFSIADLFPSLKF
RcCYP726A17     181 MARTSIGKKCEKQETFSSVIDGVTEVSGGFTVADVFPSLGFLHVITGMKSRLERLHRVAD
RcCYP726A18     181 FKSVREDEFSNFIKYLSSKAGTPVNLTHDLFSLTNSVMLRTSIGKKCKNQEAILRIIDSV
```

```
JcCYPA          241 LLDVLTGFTGKVKKASKELHGFLDKVIEEHLGRSQDKADDRKDIVDILLHLEKTDMLTVD
JcCYPB          241 FGIADVFPSMEFLQAITGVKGKLEKLRDELGDVFGNIIDEHKQKLMNRDGSDDVESEKED
JcCYP1          241 GSEKRLKESMSRVRDFAMKIVEEKKQELKENSALHSVDLLSRFVSSGISDETFVTDIVIS
JcCYP2          241 GIGSEKRLKEAMLEVRDFALNIVKKKKEELKQNSSLESADLLSRFLSSGHSDETFVTDIV
JcCYP3          241 FAVSDALPYLRWLDIGGVEKEMQKTTKVLDNVMTGWLEEHKKKRASGMANSEEDFMDVML
JcCYP4          241 TMEHIKAVMLDMFLGGTETSAATIEWAMAEMVKDPRVLEKAQKEVRQVFNHKENIIDETR
JcCYP5          241 KASRAKSGAKNEEEDIIDVLLQAQSEEELEYPITDDNIKAVIMDVLSGGTDTSATTVVWA
JcCYP6          241 KLCKELDGAAKDRKLTDADVANLPYLQAVVKEALRVHPPGPLLSWARLSTSDVKLTLTKC
JcCYP7          241 TMENIKAVMLDMFVAGTETSATTIEWAMSEMYLKLVIKETLRLHPPAPLLLPRECLEAVE
JcCYP8          241 LRLHPPGPLIVRQSSKDCTVNGYEIPAKTRLFVNAWALGRDPNYWENPLEFCPERFFNTE
JcCYP9          241 RIKRFFNVGSEKILKEAIRMVDELAEEMIRHRRNVGFMNNKDLLSRFMGSIKDEKYLRDI
JcCYP10         241 KEALLLAGAFNIADYIPFLGPIDLQLLSSPKQTRNAIMLPFILAVVVVLLGSLAYFLPTL
JcCYP11         241 YETDEILEDILQEHRANRQAAASRNGDQRGADNFLDVLLDLQQSGNLDVPLTDVAIKAAI
JcCYP12         241 AHKEADRVFQNIIDEHRARRASRAKSGAKNEEEDLIDVLLQAQSKEDLEFPITDDNIKAV
JcCYP13         241 GCAEQSKTRMGNNEEPTCLVDFWMQETLKEISEAKDAGKPTPPHTSNAEIGGYLFDFLFA
JcCYP14         241 ARSDRVFKEFDDFYQKIIDDHIQKGKEDPGHEDIIDVLLDLERYQTESGGIPFSQNHIKA
JcCYP16         241 RETDKILEDVLQEHRANRKAAVPKNGDQSQADNLLDVLLDLQESGNLDVPLPDAAIKGTI
JcCYP726A20     241 QNTDEILEDILQEHRAKRQATASKNWDRSEADNLLDVLLDLQQSGNLDVPLTDVAIKAAI
JcCYP18         241 CETDEILEGILEEHKANRQAAAFKNDDGSQADNLLDVLLDLQQNGNLEVPLTDVNIKAVI
JcCYP20         241 YETDEILEDILQEHRANRQAAASRKGDRRGADNLLDVLLYLQETGNLDVPLTDVAIKAAI
JcCYP21         241 RSRLIRLQKEADRVIGNIIDDHRAKNKTGKVGGEGQDDDLVDVLLRLQEHGNLEFPLTTD
JcCYP22         241 IMKTHDIIFADRPFIPSAFKSTYDGTDIAFSPYGDYWRQLRKICTTELLSVNRVQSFRSI
JcCYP23         241 EILEGILREHRAKRQTAASKNGDRAEADNLLDVLLDLQQRGDLDVPLTDINIKGAILEMF
JcCYP24         241 QEADEILEEIISEHRASKTADGDDYEADNILGVLLDIQERGNLQVPLTTDNIKAIILDMF
JcCYP25         241 IHREADTILEEIMDEHKANNESKNEPDNILDVLLDIQQRGNLEFPLTADNIKAIILEMFG
JcCYP71D496     241 TGIKSREEEDLVDVLLKFQENGDLNSSLSNDSIKAVILDMFIAGSDASSTTIEWAMSEMV
JcCYP28         241 ENIVKEHKERKRVEDIGMKSKEEEDLVDVLLKFQENGDVDSSLSDESVKAVILDMFIAGS
JcCYP30         241 ARAKPGSKGEADDLVDVLLNIQEQGDLGFALTTNNIKAVILDLFIAGSETSSTTVEWAMS
JcCYP71D495     241 ENIVIEHRERKRMAHAGINSKEEEDLVDVLLKFQENGDLDSYLSNDGIKAVILDMFIAGS
JcCYP32         241 YETDEILEDILQEHRANRQAAASRNGDQRGADNLLDVLLDLQQSGNLDVPLTDVAIKAAI
JcCYP33         241 KLHEKADGILERIVKEHRNKMNRSEEIQEDDDLVDVLLELQEHGDLEFPLSDDNIKTVIL
JcCYP35         241 HQQVDKMMQNVIDDRRATKRESGVDDEERDIVDVLLRIQEQGDLQLPLTDDNIKAVIFDM
JcCYP36         241 QKLQRVFQQVDTIMQNIIDNHRARKREAKSGDDAELEDFVDVLLKVQEQKDLELPLTDDN
JcCYP37         241 ADRIIESIINDHRANKKTTETEDIVDVLLKLQDNGNLQFPLTNTNIKAVILDLFVAGSET
JcCYP38         241 ERAHKEADKILQNIIDEHRARKASAANSEEEEDLVDILLNAQCQEDLQITDDNIKAIILD
JcCYP40         241 FGIADVFPSMEFLQAITGVKGKLEKLRDELGDVFGNIIDEHKQKLMNRDGSDDVESEKED
JcCYP41         241 KRVQSFKSVREEEMSSFIKFLSSKSGSPVNLTHHLFVLTNYIIARTSIGKKCKNQEALLR
JcCYP42         241 RMQLEEQTANSHQDLITCLLSISNQNNGEAITEKEIVDNSMLVMTAGHDTSSVVVTFLVR
JcCYP43         241 DIMLENIINEHRENKRLGRRNSEGKEDDLVDVLLNLQDHDNLELTTENMKAVMLDMFLGG
HnCYP1          241 KTDHILEDIIHEQRATNKSRDHQGEADNLLDVLLNLQENGNLQVPLTNISIKAVILVSNM
HnCYP2          241 KTDRILEDIINERRSATKSSRDDQGEADVLLDVLLDLQENGDLKVPLTNNSIKAAILVIN
HnCYP3          241 AITSIGKTFNKQEEIIGVITDAIMAAGGFSVADAFPSFKFLHLITGMSSKLHRIHRQADE
HnCYP5          241 NISDLFPSLSFLQDITGFKHQLEHVRDELSRVFGNIINEHKRKLMSKSSSGYDDSQKEDL
RcCYP726A13     241 RITDKILEDIITEHKAPRALFKRGGDEDKKEAENLLDVLLGLQEHGNLKVPLTNESVKSA
RcCYP726A14     241 HQQADDILEDIINEHRATRNRDDLEEAENLLDVLLDLQENGNLEVPLTNDSIKGAILDMF
RcCYP726A15     241 GVIEAGGGFSTADVFPSFKFLHHISGEKSSLEDLHREADYILEDIINERRASKINGDDRN
RcCYP726A16     241 IHVITGMKSRLEKLHRITDQVLEDIVNEHKATRAASKNGGGDDDKKEAKNLLDVLLDLQE
RcCYP726A17     241 QIFEDIIAEHKATRALSKNDDPKEAANLLDVLLDLQEHGNLQVPLTNDSIKAAILEMFGA
RcCYP726A18     241 VAAGGGFSVADVFPSFKLLHMISGDRSSLEALRRDTDEILDEIINEHKAGRKAGDDHDEA
```

```
JcCYPA          301 FTRESMKAVLMDMFIGGTDTTATTMDWTMAELMKNPRIMKKAQEEVRRVVGNKSKVEESD
JcCYPB          301 LVDVLLKTSKEVEGFNVLFTNNSLKAVVLMASVSFSPSKIRTLTISPTKPPITPSSTTVS
JcCYP1          301 FILAGRDTTSAALSWFFWLIFKNPEVEAEILKEIKEKSDAPIFEEVKDMVYTHASLSETM
JcCYP2          301 ISFILAGRDTTSAALTWFFWLLSRSTEVETKILKEIKEKSENPVFEEVKDMVYTHASLCE
JcCYP3          301 SLLDDSKQISNRDADTVNKATCLMEFLSLFSRSTVFMALAIFVPLIYSLFTKSRKARDNK
JcCYP4          301 LDELKYLKLVIKETLRLHPPVPLLVPRQSLDAVEIDGYKLPINTKVIINAWALGRDSRHW
JcCYP5          301 MSELLKNPDVMKRVQTEVRQVFSKKGYVDEESIGELHYLKAVVKETMRLHPTGAVLTREC
JcCYP6          301 SSHATQLQSYVDIRGGDLRLAPFGAGRRVCPGKNLGLVTVTLWVAKLVHQFKWVEDVANP
JcCYP7          301 IDGYQVPINTKVIINAWAIGRDSRYWNEAEKFYPERFLNNSIDFKGKEFEFIPFGAGRRM
JcCYP8          301 ANGNNQVDLRGLYFQLLPFGSGRRGCPGTSLALQLVQTTLAAMIQCFEWKVDGENGTVDM
JcCYP9          301 VISFLLAGRDTVASGLTSFFWLLTQNSDVEEAIRDESDRVLGLNKEITCYDQLRELHYLN
JcCYP10         301 RATSSLPSEHRENDRKLPPGPSPLPIIGNLHMLGNLPHQTLYNLAKLHGPIMSLRLGYVQ
JcCYP11         301 IDMFGAGSDTSSKTAEWAMAELMRNPEIMKKAQEELRNFFGENGKVDEAKLQELKWLYLI
JcCYP12         301 ILDVLSGGSDTSANTVVWVMSELQKNPDVMKRVQTEVRQVFSEKGYVDEESIGELHYLKA
JcCYP13         301 AQDASTSSLLWAVALLESHPDVLARVREEVSGVWSPESGKLITADQMREMKYTQAVAREV
JcCYP14         301 MLMNIFLAGVDTGAIVLIWAMAELIRDSRVMRKAQLEIRSLIGDKRNVSESDTGKLDYLK
JcCYP16         301 MEMFGAGSDTSSKTVEWAMAELMRNPEVMRKAQEELRSFFGENGEVEDAKIQELKCLKLI
JcCYP726A20     301 IDMFGAGSDTSSKTAEWAMAELMRNPEVMKKAQEELRNFFGENGKVEEAKLHELKWIKLI
JcCYP18         301 LGMFGAGSDTSSKTTEWAMAELMKNPEIMKKAQEELRSLFGESGYVDEAKLHEIKWLKLI
JcCYP20         301 IDMFGAGSDTSSKTVEWAMAELMRNPEIMKKAQEELRNFFGENGKVDEAKLQELKWLNLI
JcCYP21         301 NIKAVILDIFVAGSETSSTTVEWAMSEMLRNPRVMYKAQEEVRKLFDKKGNVEETDLQES
JcCYP22         301 REEEVSKLISSIASSAGSAINLSRMIETLMFSIISRAVFGKVCKGEEVFVPTIRKLTEAT
JcCYP23         301 GAGSDTSTKTLEWAMSELMRNPKMMKKVQQELRSFFGENGKVEEAKLQELKWLKLIIKET
JcCYP24         301 AGASDTSLTTAEWAMAEMVKHPRIMKKAQDEVRRTLNQEGNVANLLPELKYLKLVIKETL
JcCYP25         301 AASDTSSVTIEWAMSEMMKNPWTMKKAQEEVREVFNGTGDVSEASLQELQYLKLVIKETL
JcCYP71D496     301 KNPSMEKAQAEVRKVFGSKGKVDEAGLHELNYLKLVIKETLRLHPALPLLLPRQSREDCV
JcCYP28         301 DTSSTTLEWAMSEMMKNPSIMEKAQAEVRKVFGSKGKVDEAGLHELSYLKLVIKETLRLH
JcCYP30         301 ELLKNTSVMKKAQAELRQVFKNKGYVDEEGVCELNYLKLIVKETLRLHPPVPLLVPRENS
JcCYP71D495     301 DTSSTTIEWAISEMVKNPSIMEKAQAEVREVFGSKGKVDEADLHELNYLKLVIKETLRLH
JcCYP32         301 IDMFGAGRDTSSKTAEWAMAELMRNPEIMKKAQEELRNFFGENGKVDEAKLQELKWLNLI
JcCYP33         301 DMFSAGSETSSTTVEWAMSEMLKNPRILEKAQNEVRQVYKNKGTVDETSIHELKYLNSII
JcCYP35         301 FSAGSDSSAATTIWTMSELLKNPSVMEKAQAEVRQVFKKKGQVDEEGMEELHYLKAAVKE
JcCYP36         301 MKAVIFDMFSAGSDSSSTTIIWAMSELVKNPTVMEKAQAEVRRVFSKKGRVDEEGIEELH
JcCYP37         301 SSTTVEWAMSEVLKNPRILSKAQEELRRIFDKKGKVDEEGLQESHYLKLVIKETLRMHPP
JcCYP38         301 VFSGGNDPSANTVLWAMSELIRNPEVMKKAQTEVRQVFSEKGYVEEESIGELYYLKAVVK
JcCYP40         301 LVDVLLKLQGSGRFQCPVTNNSLKAVVLDLFTAGTDTSSTTVEWAMSEMMKNPRILKKAQ
JcCYP41         301 IIDDVVEAGAGFSVTDVFPSFEALHVISGDKHKFDKLHRETDKILEDIISEHKADRAVSS
JcCYP42         301 LLANDPSVYAAVLEEQEEIAKSKSKGEFLTWEDLTKMKYTWRVAQETMRIFPPIFGGFRK
JcCYP43         301 TESSSTVIEWAMSEMVKDSRVMEKAQSERQWRSMGMKSQSTPRSLYINAWAIGRDSRHWT
HnCYP1          301 -FGAGSDTSSKTTEWAMSELIRSPKALKKAQDEVRQVFGEMGKVDESRLQDLKYMKLIVK
HnCYP2          301 -MLQEMFGAGSDTSSKTTEWAMSELMRTPKALKIAQQEVRQVFAEKGKVDESRLQELKFL
HnCYP3          301 ILEDIINEHKASKPAAVSKAEADNILDVLLAVQEDGNFLFPLTTDNIKAMILEMFGAASD
HnCYP5          301 VDVLLMLQGSGRLQCPVTTSNLTSVILDLFIAGTDTSSTTVEWAMSEMMKNPRVFKKAQD
RcCYP726A13     301 ILEMLSGGSDTSAKTIEWAMSELMRSPEAMEKAQEEVRRVFGELGKIEESRLHELKYLKL
RcCYP726A14     301 GAGSDTSSKTAEWALSELMRHPEEMKKAQEEVRRIFGEDGRIDEARFQELKFLNLVIKET
RcCYP726A15     301 QADNLLDVLLDLQENGNLEIALTNDSIKAAILEMFGAGSDTSSKTAEWALSELMRHPEEM
RcCYP726A16     301 DGSLLQVPLTDDSIKAAILEMLGGGSDTSAKTTEWAMSEMMRYPETMKKAQEEVRQAFGN
RcCYP726A17     301 GSDTSSKTTEWAMSELMRNPTEMRKAQEEVRRVFGETGKVDETRLHELKFLKLVVKETXR
RcCYP726A18     301 ENLLDVLLDLQENGDLEVPLTNDSIKATILDMFGAGSDTSSKTAEWALSELMRHPEIMKK
```

```
JcCYPA          361 LDHMVYLKCVVKETLRHHVSGMIPRQTTSDVKLEGYDISANTRVLINAWGIQRDPRLWEQ
JcCYPB          361 LIHSSSSSLPFRLKIKCQATETDANKLPKRLPPGPIKLPLIGNLHNLAGAQPHHALTELA
JcCYP1          361 RLYPPVPVDSKMAMEDNVLPDGTLVKKGTRVTYHPFAMGRLEALWGKDWADFKPERWLAR
JcCYP2          361 SMRLYPPVPSDSKFAVADDVLPDGTPVRKGSRVTYHPYAMGRLEMLWGSDWADFKPERWL
JcCYP3          361 IRLPLPPEPTGRLPVIGHLHLLGGSQPPHITLEKMAEKIGPIYSIKLGVHRALIVSSWEM
JcCYP4          361 NEAEKFYPERFQNNSIDFKGNDFQFIPFGAGRRMCPGVGYGMALVELALANLLYHFDWKL
JcCYP5          361 REDCVINGYDIPYKSRIIINAWALGRDPDYWPEAERFNPDRFLNSSIDYKGKHFEFLPFG
JcCYP6          361 VDLSEILKLSCEMKYPLSAMALQRNN--------------------------------
JcCYP7          361 CPGMAYGMAVVELAVANLLYHFDWKLPSGIEPHQLDMSESSGVTARRKNELHLMPIPYNP
JcCYP8          361 EEGPGITLPRANPLICFPVTRLNPFPSV-------------------------------
JcCYP9          361 AAIYESLRLFPPVQFDSKFAQEDDILPDGTFVSKGTRVTYHQYAMGRMDRVWGPDCLEFK
JcCYP10         361 TIVVSSANAAKLFLKTHDAVFGSRPKLRASRYMSYGTKGMAFTEYGPYWRSVRKLCTVQL
JcCYP11         361 NKETLRLHPAVAVVPRVCRERTKVSGYDVYPGTRVFINAWAIGRDPKVWSEAEKFKPERF
JcCYP12         361 VVKETMRLRPPGTILTRECMADCVINGYDIPHKSRVIINAWALGRDPDYWPEAERFNPDR
JcCYP13         361 LRYRAPATLVPHVAMKDFPLTESYTIPKGTIVFPSVYESSFQGFTEADRFDPDRFSEERQ
JcCYP14         361 LVLKETMRLHPPVPLLVPREAISQFSINGYEVYPKTQIRVNVWAIGRDPKIWKNPEEFSP
JcCYP16         361 IKETLRLHPPGAVIPRLCRERTKVAGYDIYPNTKIFVNTWAIGRDPEIWSEAEKFNPDRF
JcCYP726A20     361 IKETLRLHPAVAVIPRVCREKTKVYGYDVEPGTRVFINVWSIGRDPKVWSEAERFKPERF
JcCYP18         361 INETLRLHPAVTLIPRLCREKTKVSGYDVYPNTRVFINTWAIGRDPTIWSEPEKFVPERF
JcCYP20         361 NKETLRLHPAAAVVPRVCRERTKVSGYDVYPGTRVFINAWAIGRDPKVWSEAEKFKPERF
JcCYP21         361 KYLKLVIKETLRLHPPAPLLLPRESTEKCEINGYDIPEKSKVIVNAWAIGRDPNHWTEAE
JcCYP22         361 TGFNLVDLYPSNKLLQRMSIGLPIIKRLHSEVDKIIQDVVDEHRARKQAGKIVAEGEEED
JcCYP23         361 LRLHPPIAVIPRLCRERTKVCGYDVYPNTRVFVNVWAMGRDPKIWNEAEKFNPERFIDSS
JcCYP24         361 RLHPPVALIPRECDGRCELNGYDVNPKTKILVNAWAIGRDHNLWNDPERFDPERFLDNSS
JcCYP25         361 RLHPPLTLIPRECNQKCQINEYDIYPKTRVLVNAWAIGRDPNWWTDPERFDPERFRCGSV
JcCYP71D496     361 IEGYNIATKSTVIVNAWAIARDPKYWDGAERFYPERFINSSIDFKGTNFEFTPFGAGRRR
JcCYP28         361 PPVPLLLPRENKENVVIEGYDIPAKSKVVVNAWAIARDPKYWDEAERFYPERFINSSIDF
JcCYP30         361 ELCEINGYFIPVKSRVLINVWAIGRDPNYWKEPERFNPERFLDNSIDYKGSNFEFIPFGA
JcCYP71D495     361 PAVPLLLPRQSREDCVIEGYNIATKSTVIVNAWAIARDPKYWDEAERFYPERFINSSIDF
JcCYP32         361 NKETLMRLHPAVAVVPRICRERTKVSGYDVYPGTRVFINAWAIGRDPKVWSEAEKFKPER
JcCYP33         361 KETLRLHPPLPLIPRESRARVEIIGYDIPIKTKVLVNAWAIGRDPKNWTEPENFCPERFL
JcCYP35         361 TLRLHAPGPLVVPRECTENCVIAGYDIPAKSRINVNSWALGRDPEYWTEPERYSPERFLD
JcCYP36         361 YLKAVAKETLRLHAPGPLLLPRECAENCVINDFDIPAKSRIAVNFWAIGRDPQYWTEPER
JcCYP37         361 GPLLLPRECRESCQLNGYNIPAKTKVIVNAWALARDPNYWTEPDTFYPERFLDSTVDFKG
JcCYP38         361 ETLRLHPPGTFLARECMEDCVINGYDISVKSIFVINTWALGRHPDYWPEAERFNPDRFLN
JcCYP40         361 EEVREAFKGKKIICEEDVKQLKYLPLVIKETLRLHPPAPLLLPRESREACEIDGYEIPMR
JcCYP41         361 KKSDGEVENLLDVLLDLQENGNLQFPLTNDAIKGAILDTFGAGSDTSSKTAEWTLSELIR
JcCYP42         361 VVKDIEYDGYLIPKGWQIFWVTSMTQMDDSIFQKPQKFDPARFENPSSIPPYCYVPFGGG
JcCYP43         361 EAEKFCPERFQNNSIDIKGNDFQFIPFGAGRRMCPGIAYGMAVVELVLANLLFHFDWKLP
HnCYP1          360 ETLRLHPAVALIPRECREKTKVEGYDIHQKTKVLVNTYAIGRDPKVWSEAEKFMPERFEE
HnCYP2          360 KLVVKETLRLHPAVSLIPRECREKTRVDGYDIYPKTRVIVNAYAIGRDPNVWSEAETFWP
HnCYP3          361 TSSVTIEWAMSEMMKNPRVMIKAQEEVRRVFSEKGNVDEDGLHELKYLKLIIKETLRLHP
HnCYP5          361 EVREAFKGKKTIKETDVQKLSYLPLVLKETLRLHPPAPLLLPRESNKSCEIDGYQIPIKT
RcCYP726A13     361 VIKETLRLHPALALIPRECMKRTKIDGYDISPKTKALVNVWAIGRDPSVWNEPEKFFPER
RcCYP726A14     361 LRLHPPVALIPRECREKTKVNGYDIYPKTRTLINVWSMGRDPSVWTEAEKFYPERFLDGT
RcCYP726A15     361 EKAQTEVRQVFGKDGNLDETRLHELKFLKLVIKETLRLHPPVALIPRECRQRTKVNGYDI
RcCYP726A16     361 AGKIDEARIHELKYLRAVFKETLRLHPPLAMIPRECRQKTKINGYDIYPKTKTLINVYAI
RcCYP726A17     361 LHPAIALIPRECRERTKVDGYDIKPTARVLVNVWAIGRDPNVWSEPERFHPERFVNSSVD
RcCYP726A18     361 AQEEVRGVFGDSGEVDETRLHELKYLKLVIKETLRLHPAIPLIPRECRERTKINGYDVYP
```

```
JcCYPA        421 AEDFIPERFVDNPADFKGQHKEYIPFGSGRRLCPGISYALKEVEYVLANLLFLFDWKLPE
JcCYPB        421 KEYGPLMHLQLGEISAIVVSNPRIAQVIMKTHDLVFADRPEILAAKIITYGGQDIAFSKL
JcCYP1        421 DGEHKKWNFVGRDAYTYPVFQAGPRICLGKEMAFLQMKRVASAILSRFKVVPAAEDGAEP
JcCYP2        421 ERDAVNEKWNFVGKDPYTYPVFQAGPRICLGKEMAFLQMKRVVAGVLNKFKVVPAAEDGF
JcCYP3        421 VKECFTTNDKAFAYRPKALFLDVMSYDYAMFGASPYGDYWREMRKIATLELLSVRRLELL
JcCYP4        421 PNGLEPHLLDMSDSFGASARRKHELHLIPIPYNSSPSPVK--------------------
JcCYP5        421 AGRRICPGILFGISNVQFPLARLLYHFDWKLPNGMRPEDLDMNEKYGIAVRRANDLQLIP
JcCYP6            ------------------------------------------------------------
JcCYP7        421 FPSQVC----------------------------------------------------------
JcCYP8            ------------------------------------------------------------
JcCYP9        421 PERWLKNGVFVPENPFKYPVFQAGFRVCLGKEMALVEMTNVALAVIRAFNVRVVDPEQAP
JcCYP10       421 LSASKIEYFAPIRKEELGFYVESLKRAAAARKVVDFSVGVGDMIQNIACRMVFGEVNNHE
JcCYP11       421 IDSAIDYKGTNFELIPFGAGKRICPGMTLGMANLEIFLANLLYHFDWKFPKGVTAENLDM
JcCYP12       421 FLNSSIDFKGNHFEFLPFGAGRRVCPGILFGISNVQFPLARLLYHFDWKLPNGMRPEDLD
JcCYP13       421 EDQLFKRNFLAFGAGAHQCVGQRYALNHLVLFIAMFSTLLDFKRHKTDGCDDIVYNPTIC
JcCYP14       421 ERFINSSIDFRGQNYEFLPFGGGRRICPGMTMGMTLVELALANLLFCFDWKLPYNMKEED
JcCYP16       421 IDSSIDYKGNNFELIPFGAGRRICPGITLASANMELFLANLLYHFDWKFPQGITAENLDM
JcCYP726A20   421 IDSAIDYRGLNFELIPFGAGKRICPGMTLGMANLEIFLANLLYHFDWKFPKGVTAENLDM
JcCYP18       421 IDSSIDYRGNHFEYTPFGAGRRICPGMAFGMVNLEIFLANLLYHFDWKLPKGITSENLDM
JcCYP20       421 IDSAIDYRGTNFELIPFGAGKRICPGMTLGMANLEIFLANLLYHFDWKFPKGVTAENLDM
JcCYP21       421 TFYPERFLDSSIDYKGNNFEFIPFGAGRRMCPGILFGIANVELPLAQLLYHFDWELPSGL
JcCYP22       421 LVDVLLNLLEKGDLDFSLSTENIKAVILDMFIAGSDTSSTPIEWAMSEMMKNPEVMEKAQ
JcCYP23       421 IDYRGNNFELIPFGAGKRICPGITLAIVHVETVLANLLYHFDWKFPEGVTAENFDMNETF
JcCYP24       421 DFRGTDFKFIPFGAGKRICPGITMAITIIEVLLAQLLYHFDWKLPDGAKPESLDMSDTFG
JcCYP25       421 DFKGTDFEFIPFGAGKRMCPGITMAMANIELILAQLLYHFNWELPGKAKPETLDMSESFG
JcCYP71D496   421 CPGMLFGLASVELPLAQLLYHFDWKLPGGQKPEDLDMSDNPGGTATRRHALYLTATPNLP
JcCYP28       421 KGANFEFIPFGAGRRMCPGMIFGLASVELPLAQLLYHFDWKLPGGQKPEDLDMSEDLGGT
JcCYP30       421 GRRICPGILFGIANVELPLANLLYHFDWKLPGEINPENLEMTEASGIAVRRKNDLNLIPI
JcCYP71D495   421 KGTNFEFIPFGAGRRMCPGMLFGLASVELPLAQLLYHFDWKLPGGQKPEDLDMSDDLDGT
JcCYP32       421 FIDSAIDYRGTNFELIPFGAGKRICPGMTLGMANLEIFLANLLYHFDWKFPKGVTAENLD
JcCYP33       421 DNAISYKGTDFEFIPFGSGRRICPGISFALPNIELPLAQLLYHFDWKLGNGMKNEDLDMT
JcCYP35       421 GSIDHKGNNFTYLPFGSGRRLCPGILFGLTNIEFILANLLYYFDWKLPNGLQPEDLDMTE
JcCYP36       421 FNPERFLDCPVDYNYKGTNYTYLPFGSGRRICPGMSFAIANMEFTLAQMLFYFDWKLPNG
JcCYP37       421 NNYEFIPFGAGRRICPGISFATPNMELPLANFLYHFDWQLPGGMKLENLDMRDGDFGSTI
JcCYP38       421 CSIDYKGNNFEFLPFGAGRRMCPGILFGISNVQFPLARLLYHFDWKLPNGMQPEDLDMNE
JcCYP40       421 SKVIVNAYAIGRDPEAWPEPDNFKPERFINSSVDFKGMNFEFIPFGAGRRICPGIAFGLA
JcCYP41       421 NPEAMRKAQAEIRRVFDETGYVDEDKFEELKYLKLVVKETLRLHPAVPLIPRECRGKTKI
JcCYP42       421 PRICPGYEFAKIETLVTIHYLVTQFTWKLSADNFFRRDPMPVPTKGLPIQITPKHQIL--
JcCYP43       421 NGLEPHLLDMSESFGVSARRKNELNLIPIPYNPSHSQEIS--------------------
HnCYP1        420 SSIDYKGTNFELIPFGAGKRICPGMTLGVTNLELFLANLLYHFDWKLPDGVDTLDMAEAF
HnCYP2        420 ERFEESSIDYKGTNFELIPFGAGKRICPGLTLSGTHLELFLANLLYHFDWKLPDGVDTVD
HnCYP3        421 PLALIPRECREKCQVSGYDIHPKTKVLINVWAIGRDPNSWTEPERFYPERFVESSIDFKG
HnCYP5        421 KIIVNAWAIGRDPETWEDADKFIPDRFMDNSVDFKGMHFEFIPFGAGRRICPGIAFGLAN
RcCYP726A13   421 FVDSSIDFRGNNFELLPFGSGKRICPGMTLGLATVELFLSYLLYYFDWKLVGGVPLDMTE
RcCYP726A14   421 IDYRGTNFELIPFGAGKRICPGMTLGIVNLELFLAHLLYHFDWKLVDGVAPDTLDMSEGF
RcCYP726A15   421 DPKTKVLVNVWAISRDPNIWTEAEKFYPERFLHSSIDYKGNHCEFAPFGSGKRICPGMNL
RcCYP726A16   421 GRDPNVWSEPEKFYPERHLDSPIDFRGSNFELIPFGAGKRICPGMTLAITTVELFLAHLL
RcCYP726A17   421 FKGTDFELLPFGAGKRICPGILVGITNLELVLAHLLYHFDWKFVDGVTSDSFDMREGFGG
RcCYP726A18   421 KTKVLVNIWAISRDPNIWSEADKFKPERFLNSSLDYKGNYLEFAPFGSGKRVCPGMTLGI
```

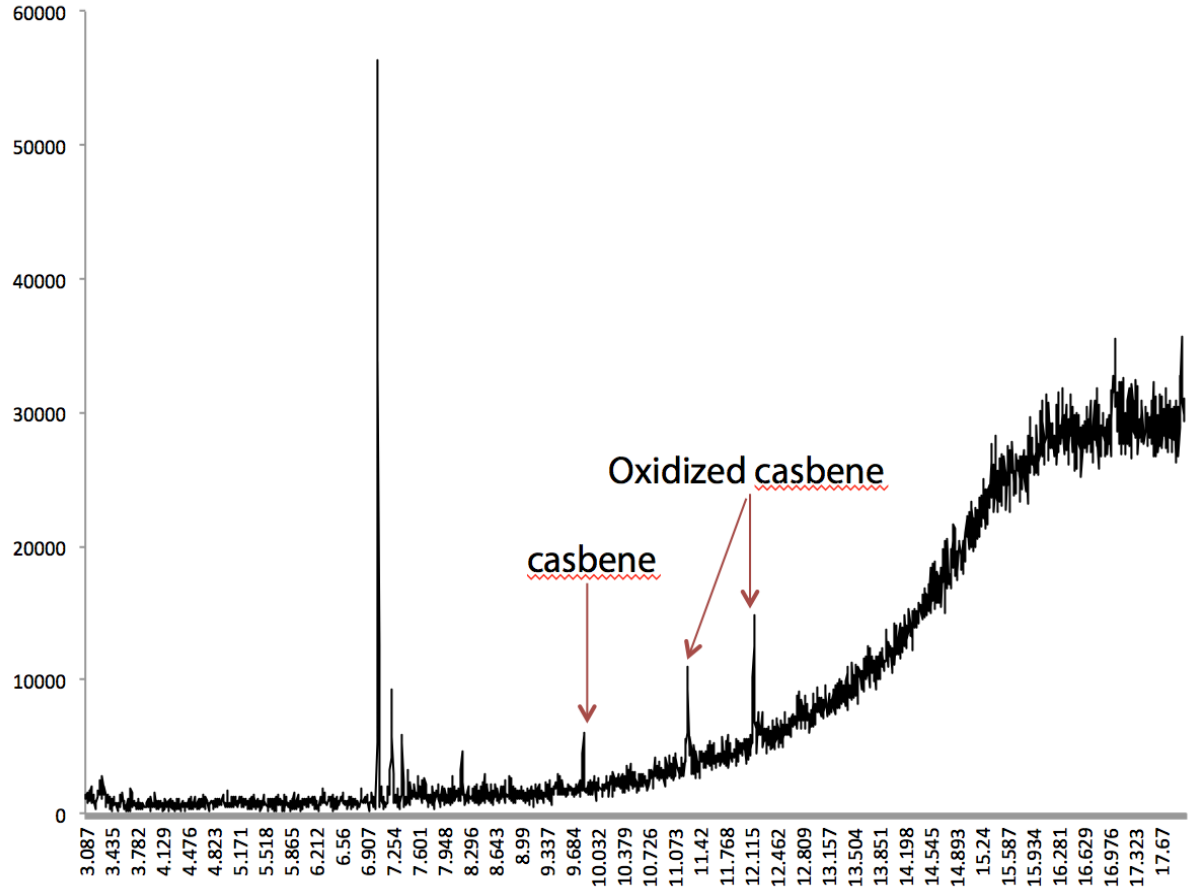


Figure S2.4. Total ion chromatogram (TIC) of GCMS profiles of heterologous expression of *R. communis* P450 positive control in *Nicotiana benthamiana.*

Figure S2.5. Sequences of genes used in this study.

>JcCBS1

MASTKSETEARPLAYFPPTVWGDRLASLTFNQPAFELLSKQVELLNEKIKKEMLNVSTS
DLAEKIILIDSLCRLGVSYHFEEEIQENLTRIFNTQPNFLNEKDYDLFTVAVIFRVFRQHGF
KISSDVFNKFKDSDGKFKESLLNDIKGILSLFEATHVSMPNEPILDEALAFTKAFLESSAV
KSFPNFAKHISSALEQPVHKGIPRLEARKYIDLYEVDESRNETVLELAKLDFNRVQLLHQ
EELSQFSKWWKSLNISAEVPYARNRMAEIFFWAVSMYFEPQYAKARMIVSKVVLLISLI
DDTIDAYATIDEIHRVADAIERWDMRLVDQLPNYMKVIYRLIINTFDEFEKDLEAEGKSY
SVKYGREAYQELVRGYYLEAIWKADGKVPSFDEYIYNGGVTTGLPLVATVSFMGVKEI
KGTKAFQWLKTYPKLNQAGGEFIRLVNDVMSHETEQDRGHVASCIDCYMKQYGVSKE
EAVEEIQKMATNEWKKLNEQLIVRSTEVVPVNLLMRIVNLVRLTDVSYKYGDGYTDSS
QLKEYVKGLFIEPIAT*

>JcCYPA

MELVALFHQWWQELDKTVPFDPLLLAPVLLLSFVFLFKLAKNRKLKLPPSPPRLPLIGNL
HQLGPLPYRSLKKLSDKYGPLMMVHFGKVPTLVVSSAEIAQEITKNHDIAFGDRPKTAA
ADDLFFGCQNLAFCPYGEYWRQVKKVCVLELLSQKRVQYFEFVRREETANLVEKLRHA
SLQGSPVDLSELLVSISNNIVSRSALGTVYNNESGHSSSGDLVRGAIDLVGSFTFQDFFPS
LGLLDVLTGFTGKVKKASKELHGFLDKVIEEHLGRSQDKADDRKDIVDILLHLEKTDML
TVDFTRESMKAVLMDMFIGGTDTTATTMDWTMAELMKNPRIMKKAQEEVRRVVGNK
SKVEESDLDHMVYLKCVVKETLRHHVSGMIPRQTTSDVKLEGYDISANTRVLINAWGIQ
RDPRLWEQAEDFIPERFVDNPADFKGQHKEYIPFGSGRRLCPGISYALKEVEYVLANLLF
LFDWKLPEGQGPEDLDMDEVFYLVIRKKIPLMVVPSLHN*

>JcCYPB

MASVSFSPSKIRTLTISPTKPPITPSSTTVSLIHSSSSSLPFRLKIKCQATETDANKLPK

RLPPGPIKLPLIGNLHNLAGAQPHHALTELAKEYGPLMHLQLGEISAIVVSNPRIAQVIMK
THDLVFADRPEILAAKIITYGGQDIAFSKLGEYWKQMKRISLMELLGPKTVQSFASLREN
EVEKLIQSIHLSKGKPINFTEKIFHLTNVITCKAAFGDECKDQDVVIALTKEATTIAGGFGI
ADVFPSMEFLQAITGVKGKLEKLRDELGDVFGNIIDEHKQKLMNRDGSDDVESEKEDLV
DVLLKTSKEVEGFNVLFTNNSLKAVVLMASVSFSPSKIRTLTISPTKPPITPSSTTVSLIHSS
SSSLPFRLKIKCQATETDANKLPKRLPPGPIKLPLIGNLHNLAGAQPHHALTELAKEYGPL
MHLQLGEISAIVVSNPRIAQVIMKTHDLVFADRPEILAAKIITYGGQDIAFSKLGEYWKQ
MKRISLMELLGPKTVQSFASLRENEVEKLIQSIHLSKGKPINFTEKIFHLTNVITCKAAFG
DECKDQDVVIALTKEATTIAGGFGIADVFPSMEFLQAITGVKGKLEKLRDELGDVFGNII
DEHKQKLMNRDGSDDVESEKEDLVDVLLKTSKEVEGFNVLFTNNSLKAVVLDLFTAGT
DTSSTTVEWAMSEMMKNPRILKKAQEEVREAFKGKKIICEEDVKQLKYLPLVIKETLRL
HPPAPLLLPRESREACEIDGYEIPMRSKVIVNAYAIGRDPEAWPEPDNFKPERFINSSVDF

KGMNFEFIPFGAGRRICPGIAFGLANIELPLARILYHFDWKLPEGITSENLDMTEAFGATV
GRKNQLYLIPIPYTSKAESSHHSEAMKLVN*

>JcCYP1

MELLFLLSFSLLIFSALFIFFTTKHSKSKQSISTHLPKSYPLIGSSFAIKANFDRRVQWT

SDILQTLPSATFVLHRPMGGRQIFTGNPANVQHILKTHFHLYRKGPVTRYTLFDFLGNGI
FNADGNTWKFQRQVASHEFSTKSLRKFVETVVDTELSERLIPILSVAAANKTVVDLQDIL
QRFAFDNICKIAFGYDPAYLLPSLPPMDFAQAFEESIRIISDRFNCAFPIIWKIKKFFGVGSE
KRLKESMSRVRDFAMKIVEEKKQELKENSALHSVDLLSRFVSSGISDETFVTDIVISFILA
GRDTTSAALSWFFWLIFKNPEVEAEILKEIKEKSDAPIFEEVKDMVYTHASLSETMRLYP
PVPVDSKMAMEDNVLPDGTLVKKGTRVTYHPFAMGRLEALWGKDWADFKPERWLAR
DGEHKKWNFVGRDAYTYPVFQAGPRICLGKEMAFLQMKRVASAILSRFKVVPAAEDG
AEPAFIAYLTTKMNGGFPVRFEERSVQ*

>JcCYP2

MLLQLSLSFLFFLIFPLIFIFSITKSSKSKNSISTTNLHAPKPHPLMGNYFHLKQNWHRR

IQWISDAVVNSPSATWVLHRPLGSGQFIITGNPANVQHILKTNFHIYEKGPIVRSTLFDF

LGNGIFNTDGETWKIQRQIASHEFNTKSLRKFVQTVVDTELSQRLIPILSTAAANQTVLDF
QDILQRFAFDNICKIAFGYDPGYLLPSLPQTKFADAFEDSVRISFERFNSLFWKIKRAFGIG
SEKRLKEAMLEVRDFALNIVKKKKEELKQNSSLESADLLSRFLSSGHSDETFVTDIVISFI
LAGRDTTSAALTWFFWLLSRSTEVETKILKEIKEKSENPVFEEVKDMVYTHASLCESMR
LYPPVPSDSKFAVADDVLPDGTPVRKGSRVTYHPYAMGRLEMLWGSDWADFKPERWL
ERDAVNEKWNFVGKDPYTYPVFQAGPRICLGKEMAFLQMKRVVAGVLNKFKVVPAAE
DGFEPEFIAYLTSKMKGGFPVRIVERS*

>JcCYP3

MEFLSLFSRSTVFMALAIFVPLIYSLFTKSRKARDNKIRLPLPPEPTGRLPVIGHLHLLG

GSQPPHITLEKMAEKIGPIYSIKLGVHRALIVSSWEMVKECFTTNDKAFAYRPKALFLDV
MSYDYAMFGASPYGDYWREMRKIATLELLSVRRLELLKHVRETEVKEATEGLYQGWL
KNKNSNNKLEVEMIKWFWDVSLNGILKMVVGKRYVEYIKKDEDINESGHWRYALRDF
FELSGKFAVSDALPYLRWLDIGGVEKEMQKTTKVLDNVMTGWLEEHKKKRASGMANS
EEDFMDVMLSLLDDSKQISNRDADTVNKATCLMEFLSLFSRSTVFMALAIFVPLIYSLFT
KSRKARDNKIRLPLPPEPTGRLPVIGHLHLLGGSQPPHITLEKMAEKIGPIYSIKLGVHRAL
IVSSWEMVKECFTTNDKAFAYRPKALFLDVMSYDYAMFGASPYGDYWREMRKIATLE
LLSVRRLELLKHVRETEVKEATEGLYQGWLKNKNSNNKLEVEMIKWFWDVSLNGILK
MVVGKRYVEYIKKDEDINESGHWRYALRDFFELSGKFAVSDALPYLRWLDIGGVEKEM
QKTTKVLDNVMTGWLEEHKKKRASGMANSEEDFMDVMLSLLDDSKQISNRDADTVN

KATCLALILAASDTTKTTLTWTLSLLLNNPEVLKKAKDELDMQVGKERQVKDSDIKHLT
YLQAIIKESLRLYPAAPLSVPRVCIEDCVVGGYRIPAGTRLFVHISKIQRDPRVWENPLEF
QPERFLTTHKDIDVKGQNYELIPFSTGRRICPGASFALQVLNLSLATLLHAFDIETPLGKP
VDMTEGHGITNLKATPLDVLLTPRLPAHLY*

>JcCYP4

MHQLFGSLPQHRLRDLSKKYGPVMHIKLGQVSNIVVSSPEAAKQVMKTHDIIFLQRPFL
LAAEILMYNFKDIAFAPYGDSWRQMRKICTLELLSTKRVRSFRPIREDEVSTFIRTISSSSK
VNLGRMVFALSNTITLRSAFGKVSERKEAFLPLVQKIVQVLEGFSVADVFPSVRFLHRIT
GMRGKLEKLHQETDIMLENIINEHRENKRLGRSNSEGKEDDLVDVLLNIQDSDNLEFPLT
MEHIKAVMLDMFLGGTETSAATIEWAMAEMVKDPRVLEKAQKEVRQVFNHKENIIDET
RLDELKYLKLVIKETLRLHPPVPLLVPRQSLDAVEIDGYKLPINTKVIINAWALGRDSRH
WNEAEKFYPERFQNNSIDFKGNDFQFIPFGAGRRMCPGVGYGMALVELALANLLYHFD
WKLPNGLEPHLLDMSDSFGASARRKHELHLIPIPYNSSPSPVK*

>JcCYP5

MVLRLWKNSKKNSAPNLPPGPWKLPVIGSMHHLSGSLLPHVRLRNLANEYGPLMHLKL
GEVTNIVVSSPETAKAIMKTHDHIFAQRPFLLAANIMAYNSTDLAFAPYGDYWRQMRKI
CTQEILSAKRVLSFGLIREEEVSKFIRDLSSRAGSTVNFSRMFNSVTYNIIQRVAIGKLWK
GEEIVIPAFKKLIEAAGGFSLSDLYPSIKLLHKISTTKFKLLRAHKETDKLFQNIIDEHRAR
KASRAKSGAKNEEEDIIDVLLQAQSEEELEYPITDDNIKAVIMDVLSGGTDTSATTVVWA
MSELLKNPDVMKRVQTEVRQVFSKKGYVDEESIGELHYLKAVVKETMRLHPTGAVLTR
ECREDCVINGYDIPYKSRIIINAWALGRDPDYWPEAERFNPDRFLNSSIDYKGKHFEFLPF
GAGRRICPGILFGISNVQFPLARLLYHFDWKLPNGMRPEDLDMNEKYGIAVRRANDLQL
IPIPCFPPQPQVK*

>JcCYP6

MASKMSCTQLMAFSLGPXXPHIAKEILTSPFFADRPIKQSAKSLMFSRAIGFAPNGIYWR
LLRKIASSHLFSPRRILAHETLRQLECASMLRNIANEQTQNGRVYLRKHLQFASLNNIMG
SVFGKRYDPAHDSKELEELRDMVREGFELLGAFNWCDYLQWLSYFYDPFRINERCLKL
VPRVRKFVRGIIEEHRLDGEEKLQEDDMLAVLWEMIFRGTDTTALLTEWVMAELVLHP
EMQEKLCKELDGAAKDRKLTDADVANLPYLQAVVKEALRVHPPGPLLSWARLSTSDV
KLTLTKCSSHATQLQSYVDIRGGDLRLAPFGAGRRVCPGKNLGLVTVTLWVAKLVHQF
KWVEDVANPVDLSEILKLSCEMKYPLSAMALQRNN*

>JcCYP7

MHQLLGSLPHHRLRHLSNKYGPVMHLRLGEVSEIVISSPEAAKKVMKTHDIIFAQRPYLL
AADIILYNFKDIVFAPYGEGWRQMRKICTLQLLSTKRVRSFRVVREEETSKFIRSISGLPK
VNISKMVFSLSNAITLKSAFGKVSERHDAFLPLVQKIMLVFGGFSVADFFPSVKFLHRITG

MRSKLEKLHQEADIMLENIINEHRENKRLGRSNSEGKEDVLVDVLLNLQNCDNLEFPLT
MENIKAVMLDMFVAGTETSATTIEWAMSEMYLKLVIKETLRLHPPAPLLLPRECLEAVE
IDGYQVPINTKVIINAWAIGRDSRYWNEAEKFYPERFLNNSIDFKGKEFEFIPFGAGRRM
CPGMAYGMAVVELAVANLLYHFDWKLPSGIEPHQLDMSESSGVTARRKNELHLMPIPY
NPFPSQVC*

>JcCYP8

MMQRIIAEHEEARKINKEIGEGDSVKDLLDILLNISEDENSEMKLTRENIKAFILGIGRK

LKKVGEDFDKMMQRIIAEHEEARKINKEIGEGDSVKDLLDILLNISEDENSEMKLTRENI
KAFILRIIAEHEEARKINKEIGEGDSVKDLLDILLNISEDENSEMKLTRENIKAFILDIF

AAGTDTSSIATEWALAELINHPDIMRKAKEEIDFVVGKSRLVEESDITNLPYLQSIVKET

LRLHPPGPLIVRQSSKDCTVNGYEIPAKTRLFVNAWALGRDPNYWENPLEFCPERFFNTE
ANGNNQVDLRGLYFQLLPFGSGRRGCPGTSLALQLVQTTLAAMIQCFEWKVDGENGTV
DMEEGPGITLPRANPLICFPVTRLNPFPSV*

>JcCYP9

MELITSSSSSLLVESISSIFSLIFFGFTLLFSLFSLLIFLSRLKPWCNCQICKSYVTASW

TKDFANLCDWYTHLLKKSPTGTIHVHVLNNIITANPENVEYILKTNFENYPKGKPFSALL
GDLLGKGIFNVDGDAWKFQRKMASLELGSVSIRSYAFDLITSEIKERLLPLLSSVSSENRS
LDLQDVFRRFSFDSICKFSFGLDPGCLKLSLPVSEFALAFDTASKLSAERALTASPLIWRIK
RFFNVGSEKILKEAIRMVDELAEEMIRHRRNVGFMNNKDLLSRFMGSIKDEKYLRDIVIS
FLLAGRDTVASGLTSFFWLLTQNSDVEEAIRDESDRVLGLNKEITCYDQLRELHYLNAAI
YESLRLFPPVQFDSKFAQEDDILPDGTFVSKGTRVTYHQYAMGRMDRVWGPDCLEFKP
ERWLKNGVFVPENPFKYPVFQAGFRVCLGKEMALVEMTNVALAVIRAFNVRVVDPEQ
APRFSPGLTATVRGGLPVVIQERESSSSSKSS*

>JcCYP10

MLCIYIYIYTIIKYLILSLLSSPKQTRNAIMLPFILAVVVVLLGSLAYFLPTLRATSSLP

SEHRENDRKLPPGPSPLPIIGNLHMLGNLPHQTLYNLAKLHGPIMSLRLGYVQTIVVSSA
NAAKLFLKTHDAVFGSRPKLRASRYMSYGTKGMAFTEYGPYWRSVRKLCTVQLLSAS
KIEYFAPIRKEELGFYVESLKRAAAARKVVDFSVGVGDMIQNIACRMVFGEVNNHELDL
KALVKEALLLAGAFNIADYIPFLGPIDLQLLSSPKQTRNAIMLPFILAVVVVLLGSLAYFL
PTLRATSSLPSEHRENDRKLPPGPSPLPIIGNLHMLGNLPHQTLYNLAKLHGPIMSLRLGY
VQTIVVSSANAAKLFLKTHDAVFGSRPKLRASRYMSYGTKGMAFTEYGPYWRSVRKLC
TVQLLSASKIEYFAPIRKEELGFYVESLKRAAAARKVVDFSVGVGDMIQNIACRMVFGE
VNNHELDLKALVKEALLLAGAFNIADYIPFLGPIDLQMLPFILAVVVVLLGSLAYFLPTL
RATSSLPSEHRENDRKLPPGPSPLPIIGNLHMLGNLPHQTLYNLAKLHGPIMSLRLGYVQ

TIVVSSANAAKLFLKTHDAVFGSRPKLRASRYMSYGTKGMAFTEYGPYWRSVRKLCTV
QLLSASKIEYFAPIRKEELGFYVESLKRAAAARKVVDFSVGVGDMIQNIACRMVFGEVN
NHELDLKALVKEALLLAGAFNIADYIPFLGPIDLQILWITP*

>JcCYP11

MEDQILSFQVLFSFLLFLLVLFKVSKKLYKHDSNPPPGPRKLPFFGNILQLAGDVPHRRLT
ALAKTYGPVMGIKLGQIPFLVVSSPETAKEVMKIQDPVFAERALLVAVEIVLYNRNDIVF
GLYGDQWRQMRKICTLELLSAKRVQSFRSVREEEVADLVKFLGSKEGSPVNLTHTLFAL
ANSMIARNTVGHKSKNQEALLRLIDDIIESIGGVGIADIFPSLKWLPSVQRERSRIRKLHYE
TDEILEDILQEHRANRQAAASRNGDQRGADNFLDVLLDLQQSGNLDVPLTDVAIKAAII
DMFGAGSDTSSKTAEWAMAELMRNPEIMKKAQEELRNFFGENGKVDEAKLQELKWLY
LINKETLRLHPAVAVVPRVCRERTKVSGYDVYPGTRVFINAWAIGRDPKVWSEAEKFKP
ERFIDSAIDYKGTNFELIPFGAGKRICPGMTLGMANLEIFLANLLYHFDWKFPKGVTAEN
LDMNEAFGAAVKRKVDLELVPIPYRP*

>JcCYP12

MESQFLSFPFLFTFLVLLLTVLRLWKKSRKNSAPNLPPGPQKLPIVGSMHHLSGSSLPHV
RLRDLAKEHGPIMHLQLGEVTNIVVSSPETAKAVMKTHDHIFAQRPFLLAANIMAYNST
DLAFAPYGDYWRQMRKICTQEMLSNKRVQSFGLIREEEVSKLIAELSSRAGSTVNFSKM
FNSVTYNIIQRVAIGKLWKGEEVVIPAIKKLIEASGGFSLSDVYPSIKLLHKISTTRFKLQR
AHKEADRVFQNIIDEHRARRASRAKSGAKNEEEDLIDVLLQAQSKEDLEFPITDDNIKAV
ILDVLSGGSDTSANTVVWVMSELQKNPDVMKRVQTEVRQVFSEKGYVDEESIGELHYL
KAVVKETMRLRPPGTILTRECMADCVINGYDIPHKSRVIINAWALGRDPDYWPEAERFN
PDRFLNSSIDFKGNHFEFLPFGAGRRVCPGILFGISNVQFPLARLLYHFDWKLPNGMRPE
DLDMNAKYGIAVTRVNDLQLIPIPYLPAKVK*

>JcCYP13

MLPFSALIPYLVSFIVFLLLLEQISYLIKRRLASGPIFILPFLGNAISLVKDPTKFWDTQ

SALSFGHGFSVNYIIGRFIVFIRDTELSHLIFSNVRPDAFMLVGHPFGKKLFGEHNLIYM

FGQDHKDLRRRIAPNFTPRALSTYSQLQQIIMLKHLKEWETMAMDNPNKPISLRLLVRD
MNLDTSQTVMVGPYLKSEARERFKFDYNLFNVGTMKLPIDLPGFAFRNARLAVDRLAQ
TLAGCAEQSKTRMGNNEEPTCLVDFWMQETLKEISEAKDAGKPTPPHTSNAEIGGYLFD
FLFAAQDASTSSLLWAVALLESHPDVLARVREEVSGVWSPESGKLITADQMREMKYTQ
AVAREVLRYRAPATLVPHVAMKDFPLTESYTIPKGTIVFPSVYESSFQGFTEADRFDPDR
FSEERQEDQLFKRNFLAFGAGAHQCVGQRYALNHLVLFIAMFSTLLDFKRHKTDGCDDI
VYNPTICPKDGCTVFLSRRYKRFPNLSTE*

>JcCYP14

MASLPLHLCLPLLLLLLLLLFTLLLLLKKKLYSTHLPPGPLRLPIIGNLHQLGALPHYSF

WQVSKKFGPVMLLQFGRVPTVIISSAETAKELIKTNDLSSCSRPRLAGTGRLSYNFLDIAF
TPYGDYWRVMRKICVHELFSAKRVQSFQSIREEEVGLLIDSILKSSSSSTPVDLSEKTMSL
TANVICRVAFGKSFQERGFNHERFQEVVREGLAMLGSFTAADFFPHVGWIVDRLTGLH
ARSDRVFKEFDDFYQKIIDDHIQKGKEDPGHEDIIDVLLDLERYQTESGGIPFSQNHIKAM
LMNIFLAGVDTGAIVLIWAMAELIRDSRVMRKAQLEIRSLIGDKRNVSESDTGKLDYLK
LVLKETMRLHPPVPLLVPREAISQFSINGYEVYPKTQIRVNVWAIGRDPKIWKNPEEFSPE
RFINSSIDFRGQNYEFLPFGGGRRICPGMTMGMTLVELALANLLFCFDWKLPYNMKEED
INMEEEFGLTAEKKEALLLVPIKYHLE*

>JcCYP16

MEHQILSFPALFSFLLFLLVLLKVSKKLYKHDSNPPPGPWKLPFLGNILQLAGDTFHRRL
TELAKTHGPVMSINVGQIPYVVVSSPETAKEVMKIQDPVFADHPVVLAAEVILYSPYDIF
FAPYGDHLKQMRKFCTVELLSTKRVQSFRSVREEEVADFVKFLRSKEGSSVNLTHTLFA
LTNSIVARTAVGHRSKNQEGLLKVIDEAVLASSGVNIADIFPSLQWLPSVKRERSRIWKT
HRETDKILEDVLQEHRANRKAAVPKNGDQSQADNLLDVLLDLQESGNLDVPLPDAAIK
GTIMEMFGAGSDTSSKTVEWAMAELMRNPEVMRKAQEELRSFFGENGEVEDAKIQELK
CLKLIIKETLRLHPPGAVIPRLCRERTKVAGYDIYPNTKIFVNTWAIGRDPEIWSEAEKFN
PDRFIDSSIDYKGNNFELIPFGAGRRICPGITLASANMELFLANLLYHFDWKFPQGITAEN
LDMNECFGGAVKRKVDLELIPIPFRT*

>JcCYP726A20

MEHQILSFPVLFSLLLFILVLLKVSKKLYKHDSKPPPGPWKLPFIGNLIQLVGDTPHRRL

TALAKTYGPVMGVQLGQVPFLVVSSPETAKEVMKIQDPVFAERPLVLAGEIVLYNRNDI
VFGSYGDQWRQMRKFCTLELLSTKRVQSFRPVREEEVASFVKLMRTKKGTPVNLTHAL
FALTNSIVARNAVGHKSKNQEALLEVIDDIVVSGGGVSIVDIFPSLQWLPTAKRERSRIW
KLHQNTDEILEDILQEHRAKRQATASKNWDRSEADNLLDVLLDLQQSGNLDVPLTDVAI
KAAIIDMFGAGSDTSSKTAEWAMAELMRNPEVMKKAQEELRNFFGENGKVEEAKLHE
LKWIKLIIKETLRLHPAVAVIPRVCREKTKVYGYDVEPGTRVFINVWSIGRDPKVWSEAE
RFKPERFIDSAIDYRGLNFELIPFGAGKRICPGMTLGMANLEIFLANLLYHFDWKFPKGV
TAENLDMNEAFGGAVKRKVDLELIPIPFRP*

>JcCYP18

MEQQILSFPVLFSFLLFLLVLLKVSKKLSKHDSNSPPGPWKLPFLGNILQLAGDLPHRRIT
ELAKKYGPVMSIKLGQHPYLVVSSPETAKEVMRTQDPIFADRPLVLAGELVLYNRNDIG
FGLYGDQWRQMRKFALELLSTKRVQSFRSVREEEIAEFVKSLRSKEGSSVNLSHTLFALT
NSIIARNTVGHKSKNQEALLKIIDDIVESLGGLSTVDIFPSLKWLPSVKRERSRIWKLHCET
DEILEGILEEHKANRQAAAFKNDDGSQADNLLDVLLDLQQNGNLEVPLTDVNIKAVILG

MFGAGSDTSSKTTEWAMAELMKNPEIMKKAQEELRSLFGESGYVDEAKLHEIKWLKLII
NETLRLHPAVTLIPRLCREKTKVSGYDVYPNTRVFINTWAIGRDPTIWSEPEKFVPERFID
SSIDYRGNHFEYTPFGAGRRICPGMAFGMVNLEIFLANLLYHFDWKLPKGITSENLDMT
ENFGGVIKRKQDLELIPAPFRP*

>JcCYP20

MEDQILSFQVLFSFLLFLFVLFKVSKKLYKHGSNPPPGPLKLPFLGNILQLAGDVPHRRLT
ALAKTYGPVMGIKLGQIPFLVVSSPETAKEVMKIQDPVFAERAPLLAGEIVLYNRNDIIFG
LYGDQWRQMRKICTLELLSAKRVQSFRSVREEEVADLVKFLGSKEGSPVNLTHTLFALA
NSIIARNTVGQKSKNQEALLRLIDDIIELTGSVSIADIFPSLKWLPSVQRDRSRIRKLHYET
DEILEDILQEHRANRQAAASRKGDRRGADNLLDVLLYLQETGNLDVPLTDVAIKAAIID
MFGAGSDTSSKTVEWAMAELMRNPEIMKKAQEELRNFFGENGKVDEAKLQELKWLNL
INKETLRLHPAAAVVPRVCRERTKVSGYDVYPGTRVFINAWAIGRDPKVWSEAEKFKPE
RFIDSAIDYRGTNFELIPFGAGKRICPGMTLGMANLEIFLANLLYHFDWKFPKGVTAENL
DMNEAFGAAVKRKVDLELVPIPFRP*

>JcCYP21

MSIFKMDLQQFLSLPILFTSFFFIFMVLKMWRKSKTKEATKNLPPGPRKLPIIGNIHQLI

GSLPHHCLRDLAKKHGGIMHLQLGEVSNIVISSPEAAKEVMKTHDIVFAQRPFLLAASIIS
YNFTDIAFSPYADYWRQLRKICILELLSAKRVQSFRFIREEEVSNLITAISSSSGKAFN

FSRKLFSLTYGIAARATFGEKCEDQEEFIPIVEEITEVAGGFSLADLFPSVKFLHSISGM

RSRLIRLQKEADRVIGNIIDDHRAKNKTGKVGGEGQDDDLVDVLLRLQEHGNLEFPLTT
DNIKAVILDIFVAGSETSSTTVEWAMSEMLRNPRVMYKAQEEVRKLFDKKGNVEETDL
QESKYLKLVIKETLRLHPPAPLLLPRESTEKCEINGYDIPEKSKVIVNAWAIGRDPNHWTE
AETFYPERFLDSSIDYKGNNFEFIPFGAGRRMCPGILFGIANVELPLAQLLYHFDWELPSG
LKAETLDMIESFGATVRRKNDLHLIPNPYTPSSAS*

>JcCYP22

MSELLKNTSVMKKAQAELRQVFKNKGYVDEEGVCELNYLKLIVKETLRLHPPVPLLVP
RENSELCEINGYFIPVKSRVLINVWAIGRDPNYWKEPERFNPERFLDNSIDYKGSNFEFIPF
GAGRRICPGILFGIANVELPLANLLYHFDWKLPGEINPENLEMTEVFIIFLFKLLPKKSKSL
NLPPGPSSLPLIGGFHHLFGAPPHHSLTNLAKKFGPIFHLQLGENSNIVISSAEMAKEIMKT
HDIIFADRPFIPSAFKSTYDGTDIAFSPYGDYWRQLRKICTTELLSVNRVQSFRSIREEEVS
KLISSIASSAGSAINLSRMIETLMFSIISRAVFGKVCKGEEVFVPTIRKLTEATTGFNLVDL
YPSNKLLQRMSIGLPIIKRLHSEVDKIIQDVVDEHRARKQAGKIVAEGEEEDLVDVLLNL
LEKGDLDFSLSTENIKAVILDMFIAGSDTSSTPIEWAMSEMMKNPEVMEKAQAEVRLVF
EAKGDVDEASLNELNYLKLVIKETLRLHPPVPLLVPRECREHCIVNGYDIPEKTRVIVNA

WAIGRNPEYWREPEKFNPERFLDNSIDYKGAHFEFIPFGAGRRMCPGISFGMANVELPLA
HLLYHFDWKLPSGVNPENLDMTESLSVTVKRSNALYVIPIPCCHSRVM*

>JcCYP23

MLSFPVIFSFLLFLLVLLKVSKKLCKDNSIPPPGPWQLPFLGNIFQLAGYQFHIRLSELG

QTYGPVMGIKVGQVPFLIVSSPEMAKEVLKVQDPTFVDRPVVLAAELVMYGGHDIVYA
PYGDQWRQMRKFCTLELLSTKRVQSFRSVREEEAGEFVKFLLSKEGSSVNLTHALYALS
NSMVARSTVGHKTKNQEALLNVIDDTVSTAAGTNIADIFPSLKWLPTVKRQMSRIWKSH
CQTDEILEGILREHRAKRQTAASKNGDRAEADNLLDVLLDLQQRGDLDVPLTDINIKGAI
LEMFGAGSDTSTKTLEWAMSELMRNPKMMKKVQQELRSFFGENGKVEEAKLQELKWL
KLIIKETLRLHPPIAVIPRLCRERTKVCGYDVYPNTRVFVNVWAMGRDPKIWNEAEKFN
PERFIDSSIDYRGNNFELIPFGAGKRICPGITLAIVHVETVLANLLYHFDWKFPEGVTAEN
FDMNETFAGIIRRKVDLELIPVAFRP*

>JcCYP24

MDHRILSFPFLMLSLLLPFVFELLKIWKKSNNNPPPGPWRLPLIGNIHQLGGRHQPHLRLT
DLARTYGPVMRLQLGQIEAVVISSAETAKQVMKTQESQFLGRPSLLAADIMLYNRTDIS
FAPYGDYWRQMKKIAVVELLSAKRVQAYKSVMDEEVSNFINFLYSKAGSPVNLTKTFY
SLGNGIIAKTSIGKKFKKQETFLKVVDKAIRVAGGFSVGDAFPSFKLIHLITGISSTLHTAH
QEADEILEEIISEHRASKTADGDDYEADNILGVLLDIQERGNLQVPLTTDNIKAIILDMFA
GASDTSLTTAEWAMAEMVKHPRIMKKAQDEVRRTLNQEGNVANLLPELKYLKLVIKET
LRLHPPVALIPRECDGRCELNGYDVNPKTKILVNAWAIGRDHNLWNDPERFDPERFLDN
SSDFRGTDFKFIPFGAGKRICPGITMAITIIEVLLAQLLYHFDWKLPDGAKPESLDMSDTF
GLVVKRRIDLNLIPIP*

>JcCYP25

MEYQILSSPTLIALLVFVATVVIKLWKRPTIANNNPPPGPWKLPLIGNLHNLFGRDQPHH
RLRDLAGKYGAVMGFQLGQVPTVVISSAEIAKQVLKTHEFQFIDRPSLLAADIVLYNRS
DIIFAPYGDYWRQIKKIAILELLSSKRVQSFKSVREEEVSSFFKFLYSKAGSPVNLSRTLLS
LTNGIIAKTSIGKKCKRQEEIIAVITDAIKATGGFSVADVFPSFKFLHIITGISSTIRR

IHREADTILEEIMDEHKANNESKNEPDNILDVLLDIQQRGNLEFPLTADNIKAIILEMFG

AASDTSSVTIEWAMSEMMKNPWTMKKAQEEVREVFNGTGDVSEASLQELQYLKLVIK
ETLRLHPPLTLIPRECNQKCQINEYDIYPKTRVLVNAWAIGRDPNWWTDPERFDPERFRC
GSVDFKGTDFEFIPFGAGKRMCPGITMAMANIELILAQLLYHFNWELPGKAKPETLDMS
ESFGLAVKRKVELNLIPTAFNP*

>JcCYP71D496

MWKKSKANSTPNLPPGPNKLPVIGNVHNLVGDLPYHRLRDLSSKYGPIIHLQLGEITTVV
ISSPELAQEVMKTHDLNFAQRPFVLAGDIVSYKCTDIALAPYGEYWRQLRKMCSLELLT
AKRVQSFKSIREEEVFKLVESISSRSGSPINFSKMASSLTYAIISRAVCGKVSRGEEVFVPA
VEKLVEAGRSISLADLYPSIKLFNALSVVRRRVEKIHGEVDKIIETIVMEHRERKRMVDT
GIKSREEEDLVDVLLKFQENGDLNSSLSNDSIKAVILDMFIAGSDASSTTIEWAMSEMVK
NPSMEKAQAEVRKVFGSKGKVDEAGLHELNYLKLVIKETLRLHPALPLLLPRQSREDCV
IEGYNIATKSTVIVNAWAIARDPKYWDGAERFYPERFINSSIDFKGTNFEFTPFGAGRRRC
PGMLFGLASVELPLAQLLYHFDWKLPGGQKPEDLDMSDNPGGTATRRHALYLTATPNL
PSAVGKI*

>JcCYP28

MLFFITVFFIFIALRIWKKSKSNSTLNLPPGPKKLPLIGNIHNLAGYLPYHRLRDLSNEY

GPIMHLQLGEINSIVVSSPELAKEVMKTHDINFAYRPFVLAGDIVSYKCKDIAFAPYGEY
WRQLRKMCSLELLTAKRVQSFKSIREEEGSKLLQSISSSSGSPVNFSKMTSSLTYSIISR

AAFGKVCQGEEVFVPAVVKLTEAGRSISLADVYPSVKLFNTFSVVRRNVEKIHSEVDKI
VENIVKEHKERKRVEDIGMKSKEEEDLVDVLLKFQENGDVDSSLSDESVKAVILDMFIA
GSDTSSTTLEWAMSEMMKNPSIMEKAQAEVRKVFGSKGKVDEAGLHELSYLKLVIKET
LRLHPPVPLLLPRENKENVVIEGYDIPAKSKVVVNAWAIARDPKYWDEAERFYPERFINS
SIDFKGANFEFIPFGAGRRMCPGMIFGLASVELPLAQLLYHFDWKLPGGQKPEDLDMSE
DLGGTATRSNALYLIPTPYIPPTVGKISR*

>JcCYP30

MALKIRKISESKKLLNLPPGPLKLPIIGNIHNLVGSLPHHRLQNLAKKYGPLMHLQLGEV
TTIVVTSAEIAKEVMRAHDIVFSNRPSILAANIISYNATSIVFSPYGEYWRQLRKICVLELL
SAKRVQSFKSIREEEVSNIVRRISSSSDSLINLSRMLFSLTYSITSRAAFGKIRKEQE

AFIPLVEEIIEVGGGFSIADLFPSIKLLNRINGMKSRVERLHQEADKILENIINEHRASK

ARAKPGSKGEADDLVDVLLNIQEQGDLGFALTTNNIKAVILDLFIAGSETSSTTVEWAMS
ELLKNTSVMKKAQAELRQVFKNKGYVDEEGVCELNYLKLIVKETLRLHPPVPLLVPRE
NSELCEINGYFIPVKSRVLINVWAIGRDPNYWKEPERFNPERFLDNSIDYKGSNFEFIPFG
AGRRICPGILFGIANVELPLANLLYHFDWKLPGEINPENLEMTEASGIAVRRKNDLNLIPI
TFPSVLVA*

>JcCYP71D495

MLFFITVLFIFIALRIWKKSKANSTPNLPPGPNKLPLIGNVHNLVGDLPYHRLRDLSKKYG
PIMHLQLGENTTVVISSPELAQEVMKTHDVNFAQRPFVLAGDIVSYKCKDIAFAPYGEY
WRQLRKMCSLELLTAKRVQSFKSIREEEVSKLVESISSSSGSPINFSKMASSLTYAIISRAV
CGKVSRGEEVFVPAVEKLVEAGRSISLADLYPSVKLFNALSVVRRRVEKIHGEVDKIIENI

VIEHRERKRMAHAGINSKEEEDLVDVLLKFQENGDLDSYLSNDGIKAVILDMFIAGSDTS
STTIEWAISEMVKNPSIMEKAQAEVREVFGSKGKVDEADLHELNYLKLVIKETLRLHPA
VPLLLPRQSREDCVIEGYNIATKSTVIVNAWAIARDPKYWDEAERFYPERFINSSIDFKGT
NFEFIPFGAGRRMCPGMLFGLASVELPLAQLLYHFDWKLPGGQKPEDLDMSDDLDGTA
TRRHALYLTATPYLPSAVGKISR*

>JcCYP32

MEDQILSFQVLFSFLLFLLVLFKISKKLYKHGSNPPPGPRKLPFLGNILQLAGDVPHRRL

TALAKTYGPVMGIKLGQIPFLVVSSPETAKEVMKIQDPVFAERALLLAGEIVLYNRNDIIF
GLYGDQWRQMRKICTLELLSAKRVQSFRSVREEEVADLVKFLGSKEGSPVNLTHTLFAL
ANSIIARNTVGHKSKNQEALLRLIDDIIESIGGVGIADIFPSLKWLPSVQRERSRIRKLHYE
TDEILEDILQEHRANRQAAASRNGDQRGADNLLDVLLDLQQSGNLDVPLTDVAIKAAII
DMFGAGRDTSSKTAEWAMAELMRNPEIMKKAQEELRNFFGENGKVDEAKLQELKWL
NLINKETLMRLHPAVAVVPRICRERTKVSGYDVYPGTRVFINAWAIGRDPKVWSEAEKF
KPERFIDSAIDYRGTNFELIPFGAGKRICPGMTLGMANLEIFLANLLYHFDWKFPKGVTA
ENLDMNEVFGGAVKRKVDLELVPIPYRP*

>JcCYP33

MEEFHFNSLHSLFALFFFIIFFFKAIKKRATKPSTTNLPPGPWKLPIIGNVHQLLGSLPH

QSLQKLSGKYGPLMHLKLGEVSTVIVSSPEIAKQVLKTHDLDFAERPPNLAPKIISYDSTH
IVFSPYGAYWRQLRKICTMELLSPKRVQSFRFIREDEVLNLIKTISSLEGSLINISEMI

FSLTYGITSRAAFGKKYEDQETFIQVITEVSKIAAGFSVADLYPSIKFLEQASGLRPKLG

KLHEKADGILERIVKEHRNKMNRSEEIQEDDDLVDVLLELQEHGDLEFPLSDDNIKTVIL
DMFSAGSETSSTTVEWAMSEMLKNPRILEKAQNEVRQVYKNKGTVDETSIHELKYLNSI
IKETLRLHPPLPLIPRESRARVEIIGYDIPIKTKVLVNAWAIGRDPKNWTEPENFCPERFLD
NAISYKGTDFEFIPFGSGRRICPGISFALPNIELPLAQLLYHFDWKLGNGMKNEDLDMTE
GYGLTIRRKQDLFLVPMYPVNSEINRSSN*

>JcCYP35

MENQYFLPFPTLFAFLLFLFMVLTLWRKTKSKPNLPPGPWKLPLIGSMHHLAGPSLLHH
RVTELARKYGPIMHLQLGQVTNIFISSPEIAREVMKTHDLIFATRPSLVAVQLVTYNFTDI
AFAPYGDYWRQIKKICTMELLTAKRVQLFAPIRQEEVSKVITDITSNVGSTINFTNVLTSL
TYKILSRSTIGKILKGEEGFIRAVMDLTEEGAGFNLADFYPSIKLFRMFGSLKHKLKRIHQ
QVDKMMQNVIDDRRATKRESGVDDEERDIVDVLLRIQEQGDLQLPLTDDNIKAVIFDMF
SAGSDSSAATTIWTMSELLKNPSVMEKAQAEVRQVFKKKGQVDEEGMEELHYLKAAV
KETLRLHAPGPLVVPRECTENCVIAGYDIPAKSRINVNSWALGRDPEYWTEPERYSPERF

LDGSIDHKGNNFTYLPFGSGRRLCPGILFGLTNIEFILANLLYYFDWKLPNGLQPEDLDM
TEVFGSAARRKNDLLVIPFQYLPPAVN*

>JcCYP36

MEFSSVVETHYQYFFPTLFALLIFLFTVLRIWRKTKSKPNLPPGPWKLPLIGSLHHLAGAP
LTHVRLRDLAKKYGPIMHLQLGEVTTIFITSPEIAKEVLKTLDIVLARRPFLQAVKLVTYN
FTDVAFSPYGEYWRQLRKICTMELLTAKRVQSFGSIRQEEGSKLIRDISSNAGSPINFSKIL
TSSGYKIISRAAFGQVWNGEDVFLKAVNDLTEESAGFSLVDFYPSKKFLQLFTSSGQKLQ
RVFQQVDTIMQNIIDNHRARKREAKSGDDAELEDFVDVLLKVQEQKDLELPLTDDNMK
AVIFDMFSAGSDSSSTTIIWAMSELVKNPTVMEKAQAEVRRVFSKKGRVDEEGIEELHY
LKAVAKETLRLHAPGPLLLPRECAENCVINDFDIPAKSRIAVNFWAIGRDPQYWTEPERF
NPERFLDCPVDYNYKGTNYTYLPFGSGRRICPGMSFAIANMEFTLAQMLFYFDWKLPN
GLPGESLDMTEKFGVTVRRESDLILIPFPYIPASVN*

>JcCYP37

MENQFPSFPMLLAAFFFILITLILSEKSKTKNLPPGPRKLPIIGNLHLFSSSYPIHHRLR

DLSKKYGAVMHLKLGQVSTIGIGSPEAAKEMMKTNDVCFADRPCYQSAEIVTYNFLDIA
YSPYEDYWRQLRKICTVELLSAKRVQSFRSIREEEVANLIRDISSSSGKPFNLSKRIFALTY
SITARVSFGDKCREQDAFISAAEKIMQTTGFDLADLFPSLKFLGWFSEMRTRLMNAHDE
ADRIIESIINDHRANKKTTETEDIVDVLLKLQDNGNLQFPLTNTNIKAVILDLFVAGSETSS
TTVEWAMSEVLKNPRILSKAQEELRRIFDKKGKVDEEGLQESHYLKLVIKETLRMHPPG
PLLLPRECRESCQLNGYNIPAKTKVIVNAWALARDPNYWTEPDTFYPERFLDSTVDFKG
NNYEFIPFGAGRRICPGISFATPNMELPLANFLYHFDWQLPGGMKLENLDMRDGDFGSTI
TRRNDLVLIPVPYHPPTMEA*

>JcCYP38

MENQFPSFPFLFTFLVFLFMVLRLWQKSKNNSALNLPPGPWKLPLIGSLHHLFGSVLPHT
RLRDLANEYGPIMHLQLGQVTNIVLSSPETAKAVLKTHDHIFTQRPFVLAAETMTYNFT
NLANAPYGGYWRQIRKICTLEMLSAKRVRSFGLIREEEVSKFIRDLSSSTSAGSTVNFSR
MFSSVTYNIIQRVAIGKISKGEDTVFPAIRKLIEAFVGFNLSDAYPSIKLLHKISTKRFKLER
AHKEADKILQNIIDEHRARKASAANSEEEEDLVDILLNAQCQEDLQITDDNIKAIILDVFS
GGNDPSANTVLWAMSELIRNPEVMKKAQTEVRQVFSEKGYVEEESIGELYYLKAVVKE
TLRLHPPGTFLARECMEDCVINGYDISVKSIFVINTWALGRHPDYWPEAERFNPDRFLNC
SIDYKGNNFEFLPFGAGRRMCPGILFGISNVQFPLARLLYHFDWKLPNGMQPEDLDMNE
KSGPCSDKAK*

>JcCYP40

MASVSFSPSKIRTLTISPTKPPITPSSTTVSLIHSSSSSLPFRLKIKCQATETDANKLPK

RLPPGPIKLPLIGNLHNLAGAQPHHALTELAKEYGPLMHLQLGEISAIVVSNPRIAQVIMK
THDLVFADRPEILAAKIITYGGQDIAFSKLGEYWKQMKRISLMELLGPKTVQSFASLREN
EVEKLIQSIHLSKGKPINFTEKIFHLTNVITCKAAFGDECKDQDVVIALTKEATTIAGGFGI
ADVFPSMEFLQAITGVKGKLEKLRDELGDVFGNIIDEHKQKLMNRDGSDDVESEKEDLV
DVLLKLQGSGRFQCPVTNNSLKAVVLDLFTAGTDTSSTTVEWAMSEMMKNPRILKKAQ
EEVREAFKGKKIICEEDVKQLKYLPLVIKETLRLHPPAPLLLPRESREACEIDGYEIPMRSK
VIVNAYAIGRDPEAWPEPDNFKPERFINSSVDFKGMNFEFIPFGAGRRICPGIAFGLANIEL
PLARILYHFDWKLPEGITSENLDMTEAFGATVGRKNQLYLIPIPYTSKAESSHHSEAMKL
VN*

>JcCYP41

MSLQPAILQGNTCKQYFHPLSSISSTRWVGNCNRFAFLSPAKPTANRAPQASLSSKLQPV
VRLLTKFPASGFLAMNQSVDQFASTTTSLTKIFNKIGKPIQSSPFLVSVLLLMFMASKIQN
QQEEDDNSINLPPGPWRLPFIGNIHQLAGPGLPHHRLTDLAKTYGPVMGVHLGEVYAVV
VSSAETSKEVLRTQDTNFAERPLVNAAKMVLYNRNDIVFGSFGDQWRQMRKICTLELL
SVKRVQSFKSVREEEMSSFIKFLSSKSGSPVNLTHHLFVLTNYIIARTSIGKKCKNQEALL
RIIDDVVEAGAGFSVTDVFPSFEALHVISGDKHKFDKLHRETDKILEDIISEHKADRAVSS
KKSDGEVENLLDVLLDLQENGNLQFPLTNDAIKGAILDTFGAGSDTSSKTAEWTLSELIR
NPEAMRKAQAEIRRVFDETGYVDEDKFEELKYLKLVVKETLRLHPAVPLIPRECRGKTK
INGYDIFPKTKVLVNVWAISRDPAIWPEPEKFNPERFIDNPIDYKSINCELTPFGAGKRICP
GMTLGITNLELFLANLLYHFDWKLPDGKMPEDLDMSESFGGAIKRKTDLKLIPVLARPL
TPRNANSGNTFTTTDADSPASMCPHLKAL

>JcCYP42

MNTLFTIFIFLIPIFLFLAKKRISSSKDLKLPPGSLGIPIIGQSLSLLQNMRANTAEKWIEKRI
QKYGPISKLSLFGKPTVFMYGQAANKFVFTSDSSTLSNSQTQSVKMILGEKCLLELSGKD
HKRVREALMSFLKPESLKQYVGKIDEEVRMHILKNWQRKQEVQVLPLMKTLTFNIICSL
LFGIERGSRRDKLVDLFQEMIKGMWSIPINLPFTRYNSSLKASTSVRNMLKDLISEKRMQ
LEEQTANSHQDLITCLLSISNQNNGEAITEKEIVDNSMLVMTAGHDTSSVVVTFLVRLLA
NDPSVYAAVLEEQEEIAKSKSKGEFLTWEDLTKMKYTWRVAQETMRIFPPIFGGFRKVV
KDIEYDGYLIPKGWQIFWVTSMTQMDDSIFQKPQKFDPARFENPSSIPPYCYVPFGGGPRI
CPGYEFAKIETLVTIHYLVTQFTWKLSADNFFRRDPMPVPTKGLPIQITPKHQIL*

>JcCYP43

MEYLQQFSIFSALVTFLLFIFLLQRKPKTSSRKSAPGPWKLPIIGNMHQLLGSLPHHRLKD
LSDKYGSVMNLQLGQVSNIVISSPEAAKQVMKTHDIIFVQRPFLLAANIIMYNSKDIVFA
PYGDSWRQMRKICTLELLSTKRVRSFRAIREEETSNFIRSISSLSEVNISKILLSLSNAITLR
SAFGKVSERQEAFLPLVQKIALMLEGFSIADIFPSVKFLHGITGMRSKLQKLHQEADIMLE
NIINEHRENKRLGRRNSEGKEDDLVDVLLNLQDHDNLELTTENMKAVMLDMFLGGTES
SSTVIEWAMSEMVKDSRVMEKAQEERQWRSMGMKSQSTPRSLYINAWAIGRDSRHWT

EAEKFCPERFQNNSIDIKGNDFQFIPFGAGRRMCPGIAYGMAVVELVLANLLFHFDWKL
PNGLEPHLLDMSESFGVSARRKNELNLIPIPYNPSHSQEIS*

>HnCYP1

MDNQVHSFPVFLSFLLFIFMVLRIWKQYSHKSTSPPGPWKLPLIGNIPQLVGALPHLRLR
DLAKIYGPVMSIQLGQVPVVIISSSETAKEVLKTQDVQFADRSLILAGKMVLYDRMDIIF
GSYGDHWRQMRKICTLELLSAKRVQSFRSVREEEVENFIKHLHSKAGSPVNLTKALFAL
TNSIMAITSIGKKCKNQEAILSIIDDVIEVAGGFSVADVFPSFKFLHYISGEKSRLQKLHDK
TDHILEDIIHEQRATNKSRDHQGEADNLLDVLLNLQENGNLQVPLTNISIKAVILVSNMF
GAGSDTSSKTTEWAMSELIRSPKALKKAQDEVRQVFGEMGKVDESRLQDLKYMKLIVK
ETLRLHPAVALIPRECREKTKVEGYDIHQKTKVLVNTYAIGRDPKVWSEAEKFMPERFE
ESSIDYKGTNFELIPFGAGKRICPGMTLGVTNLELFLANLLYHFDWKLPDGVDTLDMAE
AFGAALKRKTDLTLIPIQFTPLATD*

>HnCYP2

MEHQILSFPVLLSFIIFIFMVLKIWKKCSHNSSFPPGPWKFPLIGNIPQLAGALPHLRLR

DLSKIHGPIMSIQMGQVPAVVISSPETAKQVLKTQDVQFADRPLIQAGKFVLYNQLDILY
APYGDHWRQMRKICTLELLSAKRVQSFRSIREEEVSNYIKFLHSKAGSPVNLTETLFSLT
NSIMLRISIGTKHKNQETILSIIEEVTEAAGGFSVADVFPSLKFLHYISGEKSRLQKLHQKT
DRILEDIINERRSATKSSRDDQGEADVLLDVLLDLQENGDLKVPLTNNSIKAAILVINML
QEMFGAGSDTSSKTTEWAMSELMRTPKALKIAQQEVRQVFAEKGKVDESRLQELKFLK
LVVKETLRLHPAVSLIPRECREKTRVDGYDIYPKTRVIVNAYAIGRDPNVWSEAETFWPE
RFEESSIDYKGTNFELIPFGAGKRICPGLTLSGTHLELFLANLLYHFDWKLPDGVDTVDM
TEAFGGALKRKTDLILIPVPFSP*

>HnCYP3

MSSLQPFLQPILHNPLLTKPSSLPFISSPNSAATSPTQPCLPPNLHLALRLLPDISPPVR

GNRFPTFASNQFVNQPTSTPEEDNDGNPTLPPGPWKLPLIGNIHQLLGDLPHRRLRDLAK
IYGPVMSIRLGEVPAVVISSVEAAKEVLRTQDVNFAERPPVLVAEIVLYNRQDIVFRSYG
DEWRQMKKIARLELLSAKRVHSFKSIREEQVSNFIKFIYSSTGSPVNLSKELMSLTNSIIAI
TSIGKTFNKQEEIIGVITDAIMAAGGFSVADAFPSFKFLHLITGMSSKLHRIHRQADEILED
IINEHKASKPAAVSKAEADNILDVLLAVQEDGNFLFPLTTDNIKAMILEMFGAASDTSSV
TIEWAMSEMMKNPRVMIKAQEEVRRVFSEKGNVDEDGLHELKYLKLIIKETLRLHPPLA
LIPRECREKCQVSGYDIHPKTKVLINVWAIGRDPNSWTEPERFYPERFVESSIDFKGNDYE
FLPFGAGKRICPGITMAMANIQLTIAQMLYHFDWKLPDGAESESIDMTESIGLAIKRKLH
LNLIPIPYP*

>HnCYP5

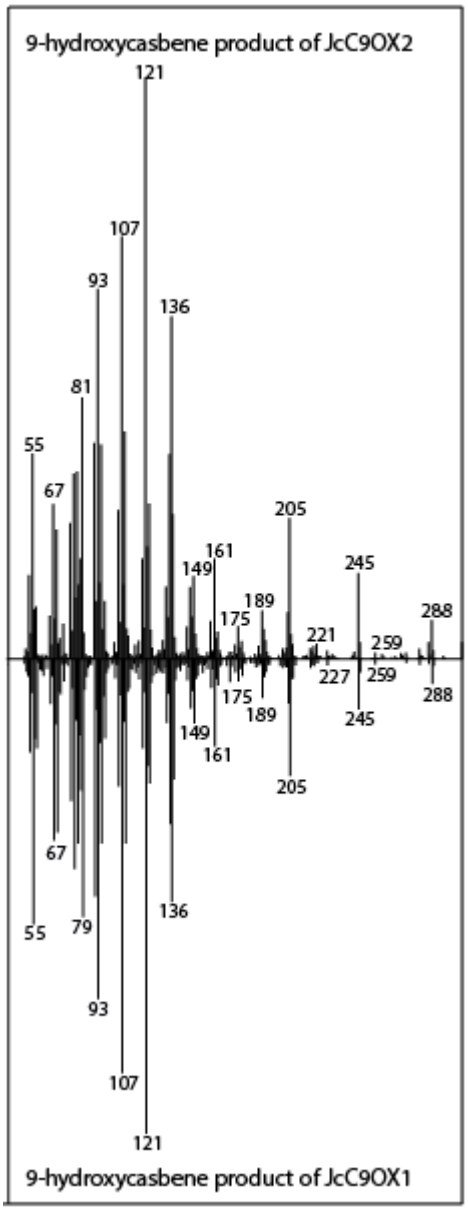
MASLINLSPSNMKLVTFSPSKQIATPTFSVSLQFQSCSHIPKRKMKVYSKKTETESAHKVL
PPGPMKLPVIGNLLNLVGSEPHHALAQLAKEYGPLMHLQLGEISAVVVSNPKMAQEIM
KTHDLIFANRPQLLASEIVTYGGKDIAFAPLGEYWKEMKRISLTELLGPRRVQSFSSIREN
EVEELIESVRLSAGKPFNFTENIFRFTNVITCKAAFGDECKDQDAVIALSKQATELASGFN
ISDLFPSLSFLQDITGFKHQLEHVRDELSRVFGNIINEHKRKLMSKSSSGYDDSQKEDLVD
VLLMLQGSGRLQCPVTTSNLTSVILDLFIAGTDTSSTTVEWAMSEMMKNPRVFKKAQD
EVREAFKGKKTIKETDVQKLSYLPLVLKETLRLHPPAPLLLPRESNKSCEIDGYQIPIKTKI
IVNAWAIGRDPETWEDADKFIPDRFMDNSVDFKGMHFEFIPFGAGRRICPGIAFGLANM
ELPLARLLYHFDWKLPDGITAENFDMTESFGATVGRKNNLYLIPTLYKPQHEHHYSPQP
TEPLHLLN*

7.2. Supplementary for Chapter 3 — High-titer production of lathyrane diterpenoids from sugar by engineered Saccharomyces cerevisiae

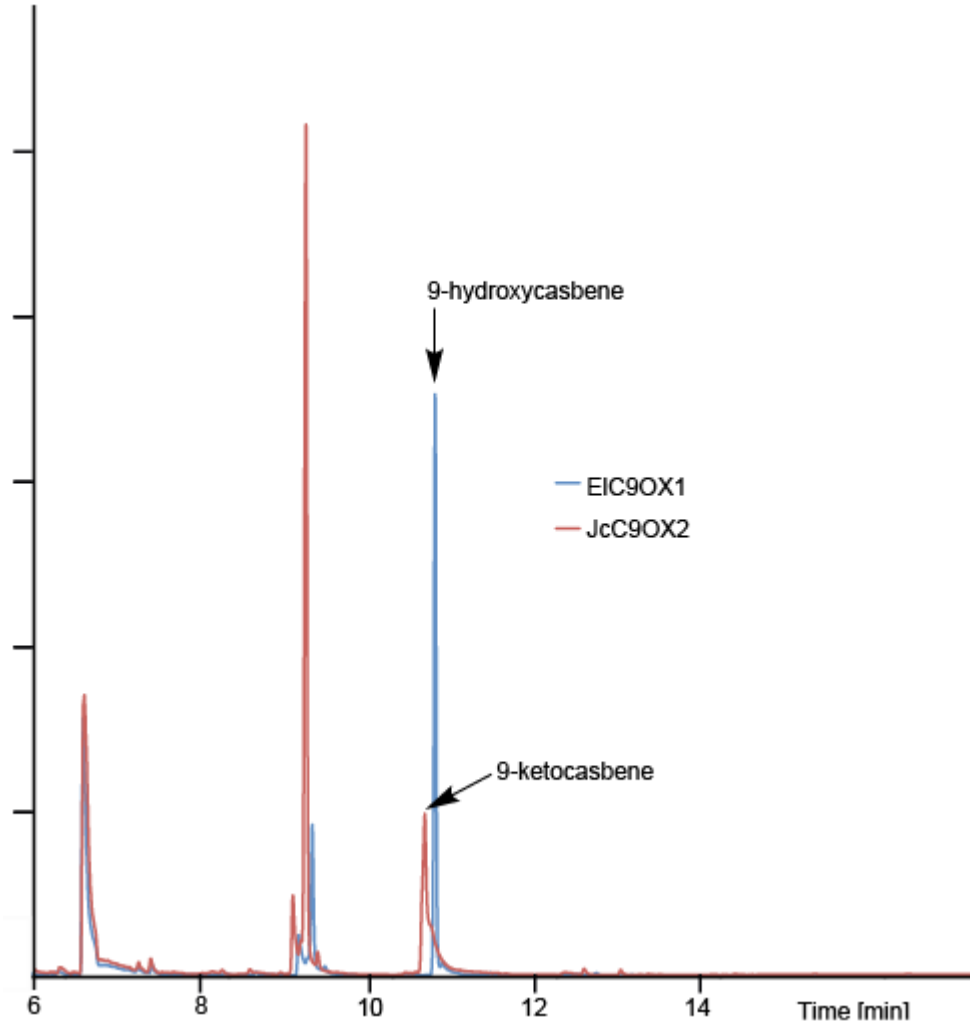
Supplementary Table 1| Yeast strains

Strain	Parent (+ additional genetic changes)	Compound	References
GTY116	*MATa leu2-3,112::His3MX6_P_{GAL1}-ERG19/P_{GAL10}-ERG8 ura3-52::URA3_P_{GAL1}-mvaS(A110G)/P_{GAL10}-mvaE(CO) his3Δ1::hphMX4_P_{GAL1}-ERG12/P_{GAL10}-IDI1 trp1-289::TRP1_P_{GAL1}-crtE(X.den)/P_{GAL10}-ERG20 yprcδ15::natMX_$_{PGAL1}$-crtE(opt)/$_{PGAL10}$-crtE*	GGPP	Reider et al.,2016
JWY501	GTY116 (*ura3-52* prototrophy removed for use of Cas9 system)	GGPP	This work
JWY502	JWY501 (ARS1014a::P_{GAL1}-CBS)	Casbene	This work
JWY503	JWY501 (ARS1014a::P_{GAL1}-CBSopt)	Casbene	This work
JWY504	JWY501 (ARS1014a::P_{GAL1}-CBS-GFP)	Casbene	This work
JWY505	JWY501 (ARS1014a::P_{GAL1}-MBP-CBS)	Casbene	This work
JWY506	JWY501 (ARS1014a::P_{GAL1}-CBS-ERG20F96C)	Casbene	This work
JWY507	JWY501 (ARS1014a::P_{GAL1}-MBP-CBS-ERG20F96C)	Casbene	This work
JWY508	JWY501 (ARS1014a::P_{GAL1}-MBP-CBS-ERG20F96C-GFP)	Casbene	This work
JWY522	JWY519 (*HIS3B::P_{GAL1}-MBP-JcADH1*)	Jolkinol	This work
JWY523	JWY519 (*HIS3B::P_{GAL1}-MBP-EIADH1*)	Jolkinol	This work
JWY524	JWY519 (*HIS3B::P_{GAL1}-JcADH1-GFP*)	Jolkinol	This work
JWY525	JWY519 (*HIS3B::P_{GAL1}-EIADH1-GFP*)	Jolkinol	This work

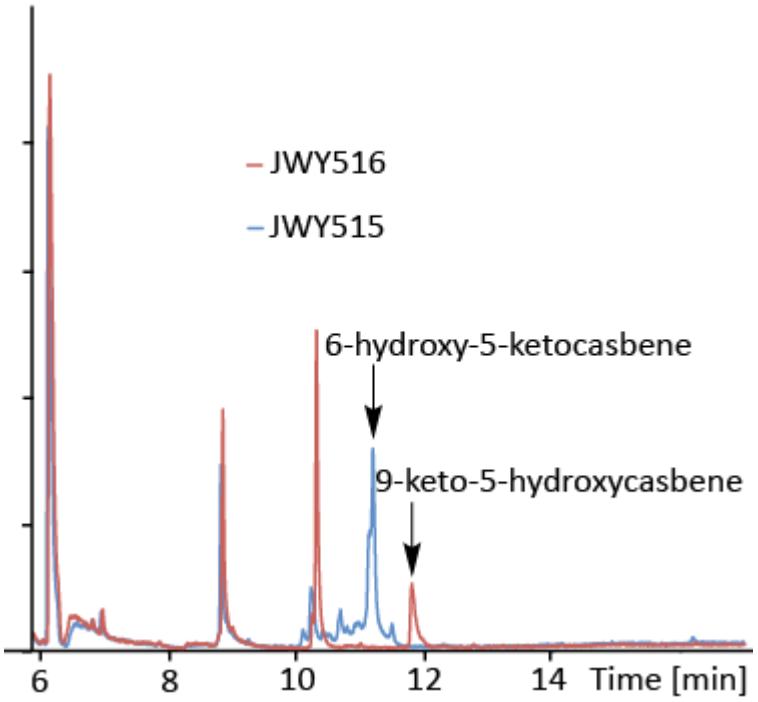
Supplemental Table 3.1. Genotype of yeast host strains used in this study

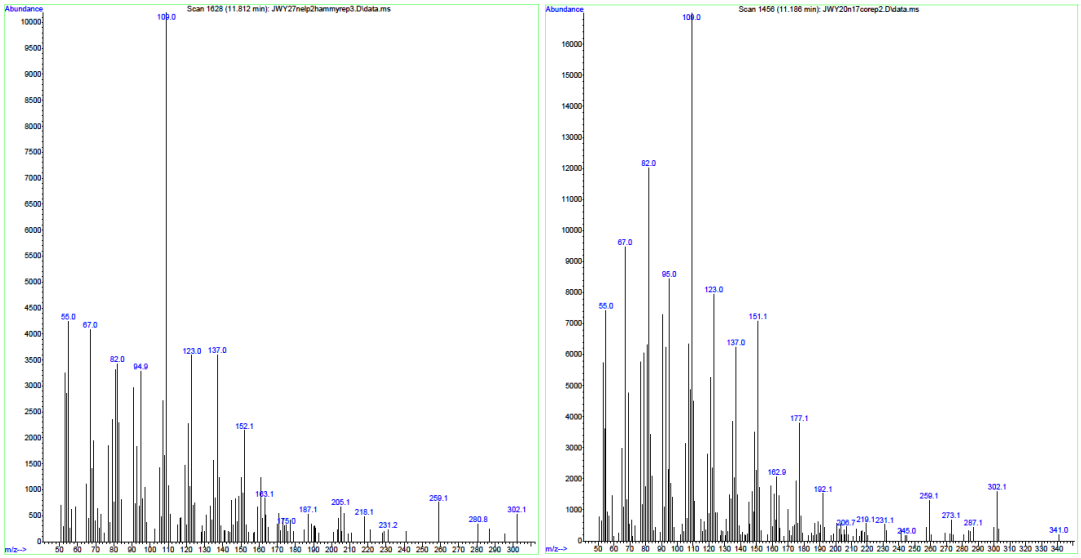


Supplemental Fig. 3.1. GC-EIMS spectra of JcC9OX1 9-hydroxycasbene product. Expression of JcC9OX1 in casbene producing strain JWY509 results in production of 9-hydroxycasbene. Spectra show the fragmentation of the product is identical to that of JcC9OX2, a CYP that produces 9-hydroxycasbene.

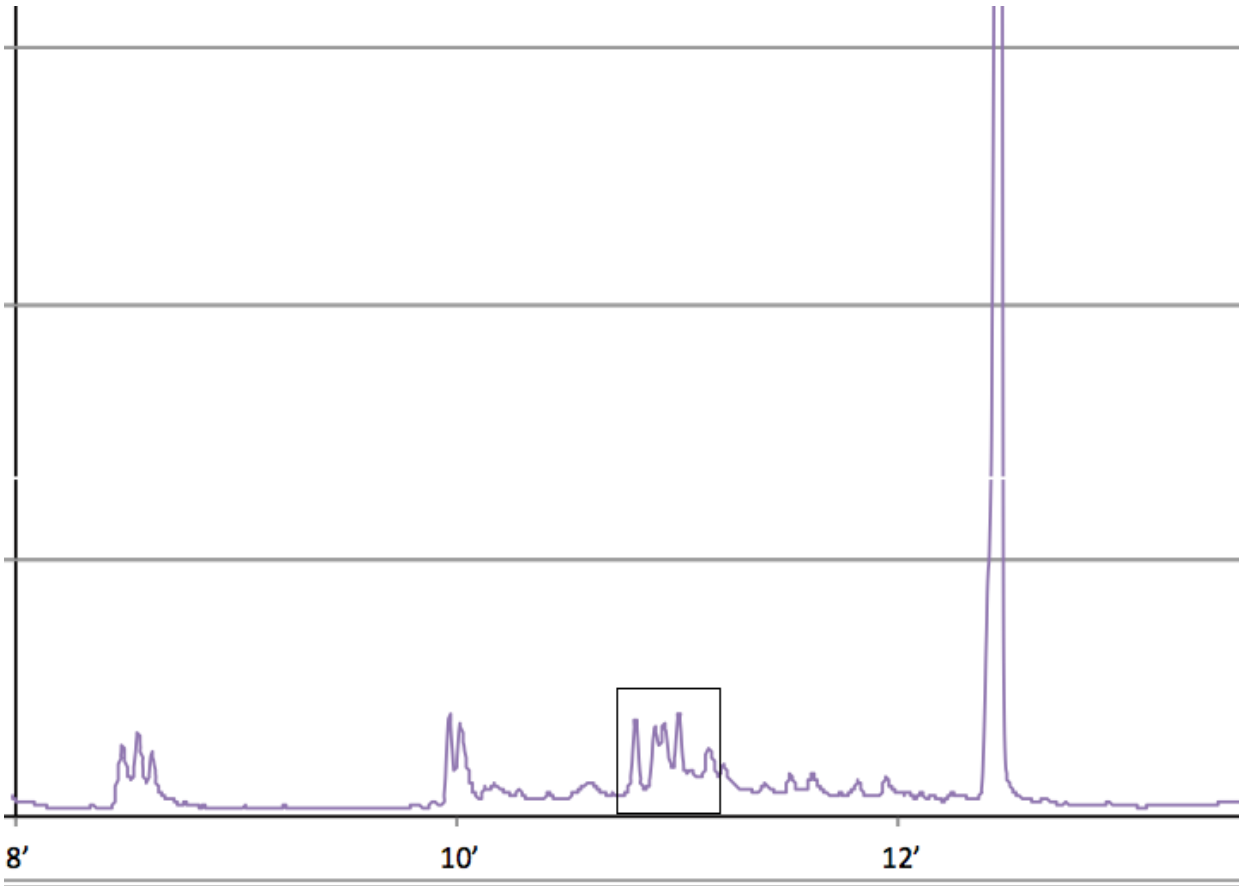


Supplemental Fig. 3.2. GC-EIMS spectra product profiles of select CYPs. Extracts of yeast cultures expressing JcCBS1 and ElC9OX1 or JcC9OX2 show different oxidized product profiles. Chromatogram comparisons of ElC9OX1 produces primarily 9-hydroxycasbene, while JcC9OX2 produces a relatively equal mix of 9-ketocasbene and 9-hydroxycasbene. In total, ElC9OX1 produces more C-9 oxidized casbane product.

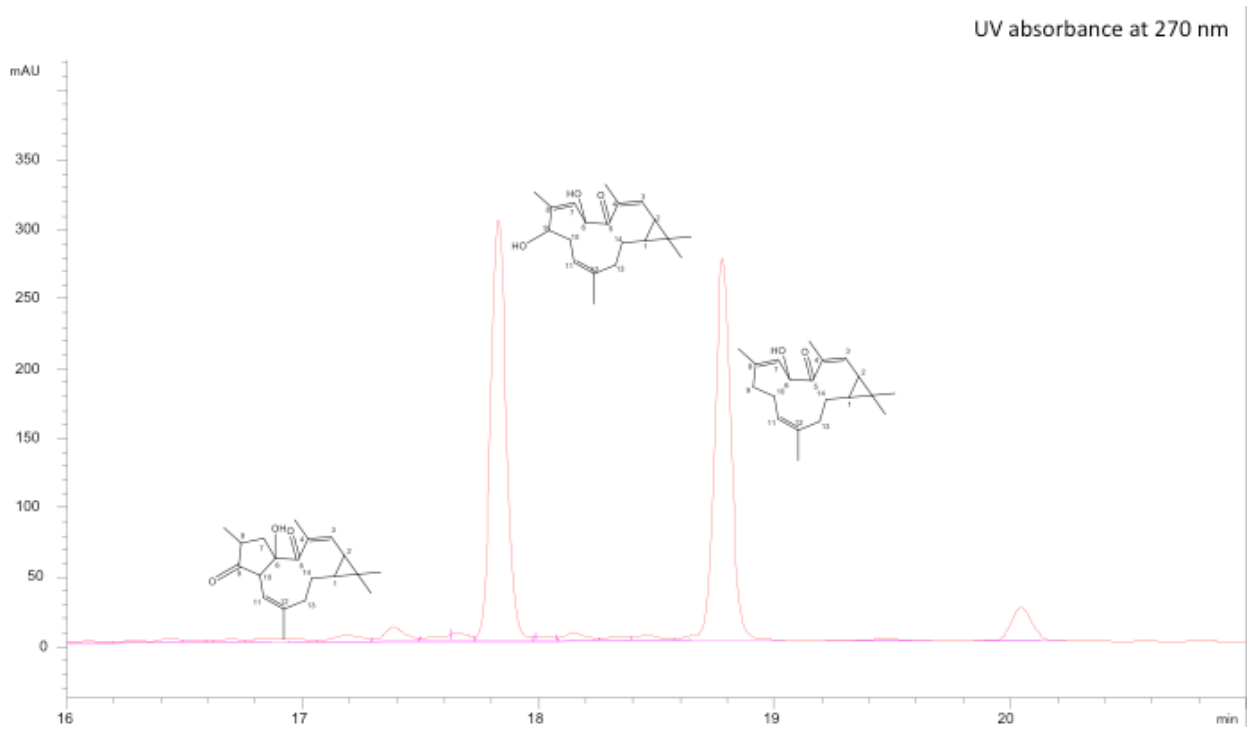




Supplemental Fig. 3.3. Chromatogram and GC-EIMS spectra of coexpression of ElC9OX1 + ElC5OX1 in comparison to 6-hydroxy-5-ketocasbene product of JcC5OX2. (a) Chromatogram of products of ElC9OX1 and ElC5OX1 coexpression show new peak eluting around 11.9 min. As a reference, the chromatogram of JcC5OX2 expression in JWY509 was overlaid to show difference in elution times of oxidized products. (b) JcC5OX2 expression in JWY509 produces 6-hydroxy-5-ketocasbene (left), while ElC9OX1 and ElC5OX1 coexpression produces a new, unknown doubly oxidized casbane product (right). PRevious reports suggest this is 9-keto-5-hydroxycasbene, a purported dead end product.

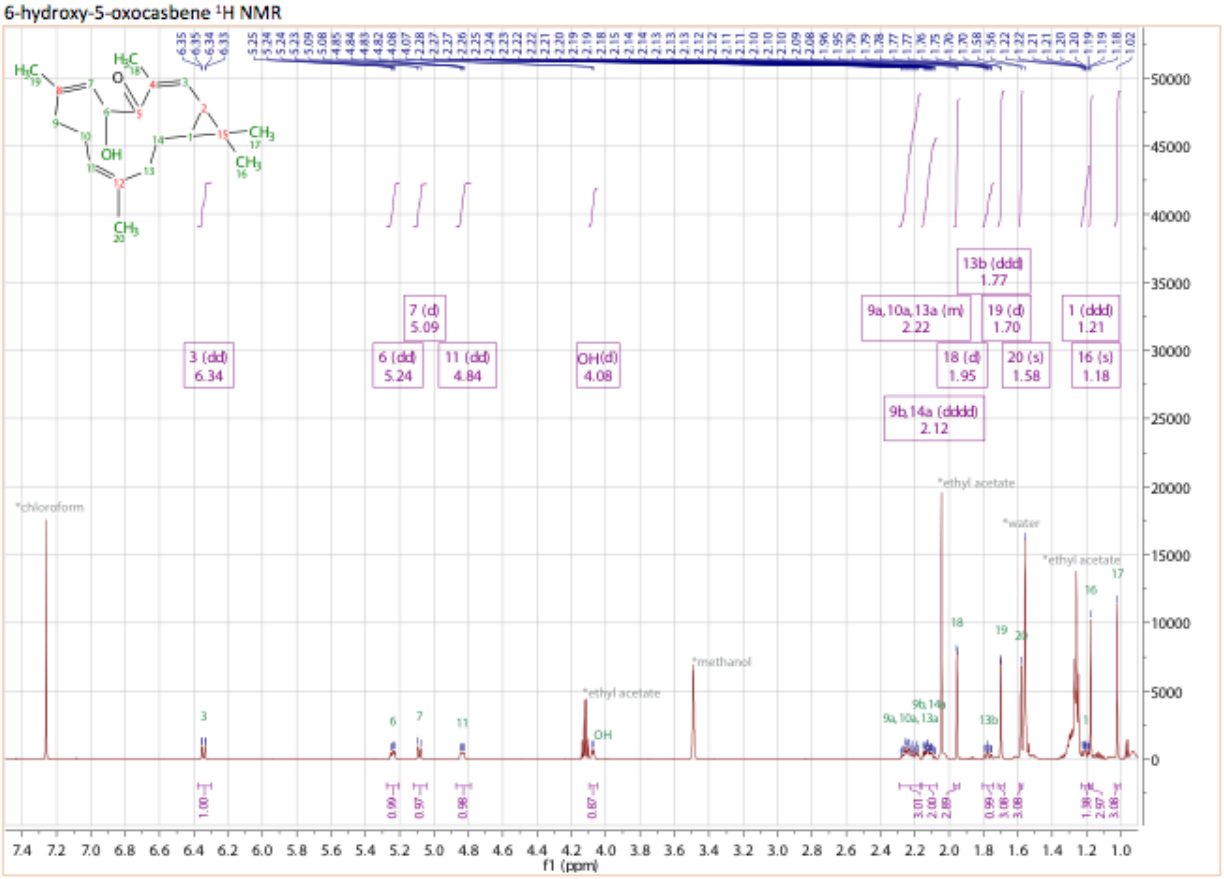


Supplemental Fig. 3.4. Chromatogram and GC-EIMS spectra of jolkinol. Initially, we attempted to quantify the products via GC-EIMS. However, we noticed the appearance of multiple peaks (peaks designated by box) from highly active C5OX CYPs that eluted after the C-9 oxidized casbene and contributed to the significant decrease of C-9 oxidized product in these strains. All five of these unknown peaks had parent ions with m/z 316, corresponding to jolkinol and 6,9-dihydroxy-5-ketocasbene, but we surmised that these were degradation products of these compounds that formed during the GC-EIMS analysis.

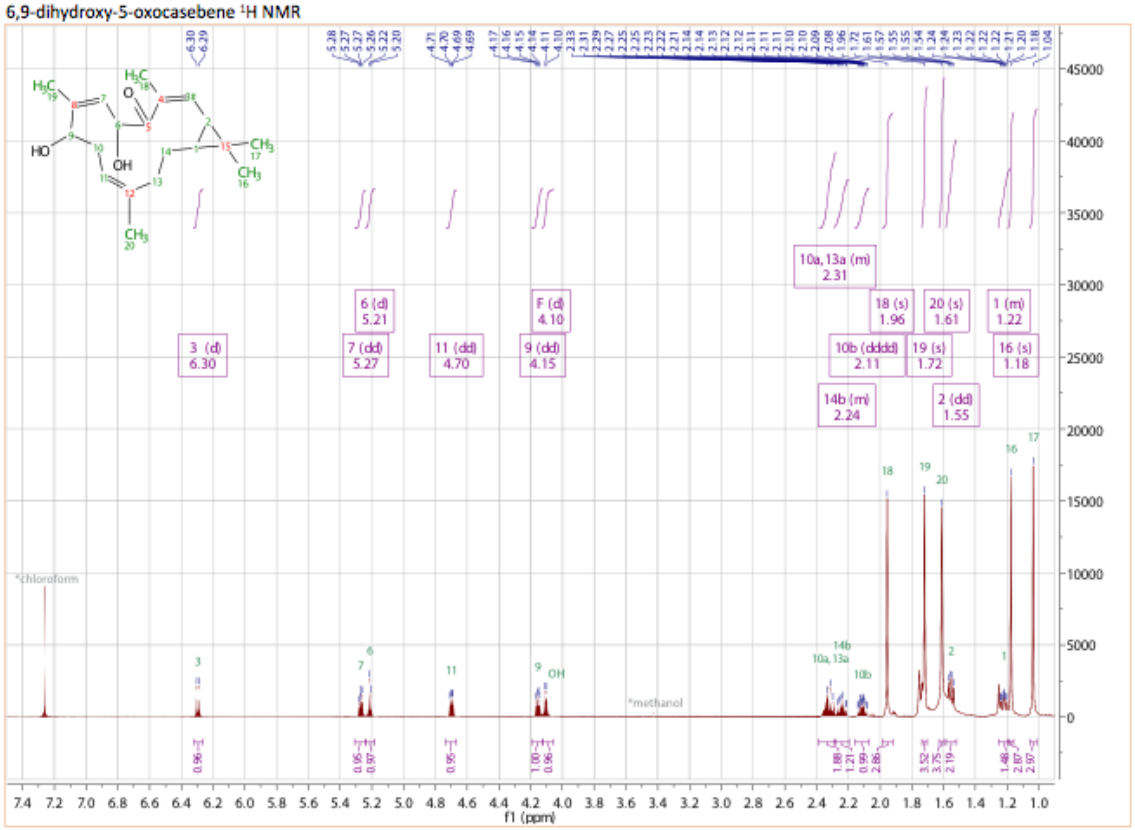


Supplemental Fig. 5: HPLC chromatograms of 6-hydroxy-5-ketocasbene, 6-hydroxy-5,9-diketocasbene, and jolkinol.

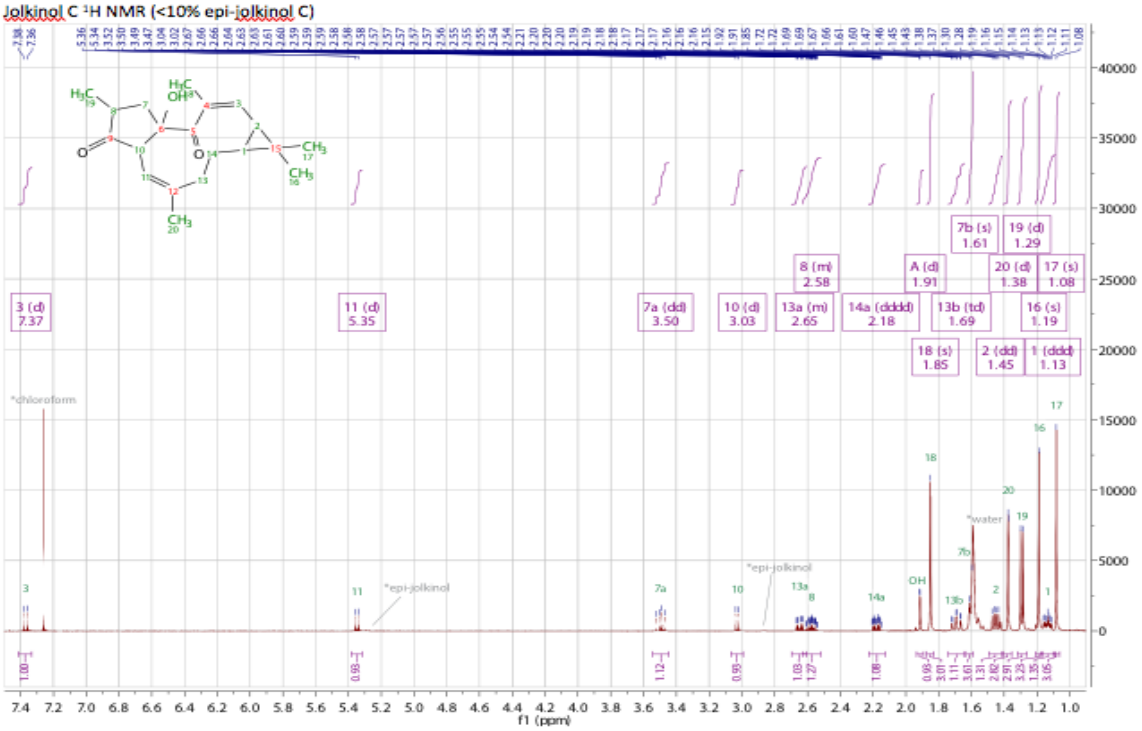
6-hydroxy-5-oxocasbene ^{1}H NMR



6,9-dihydroxy-5-oxocasebene ^{1}H NMR



Jolkinol C ^{1}H NMR (<10% epi-jolkinol C)



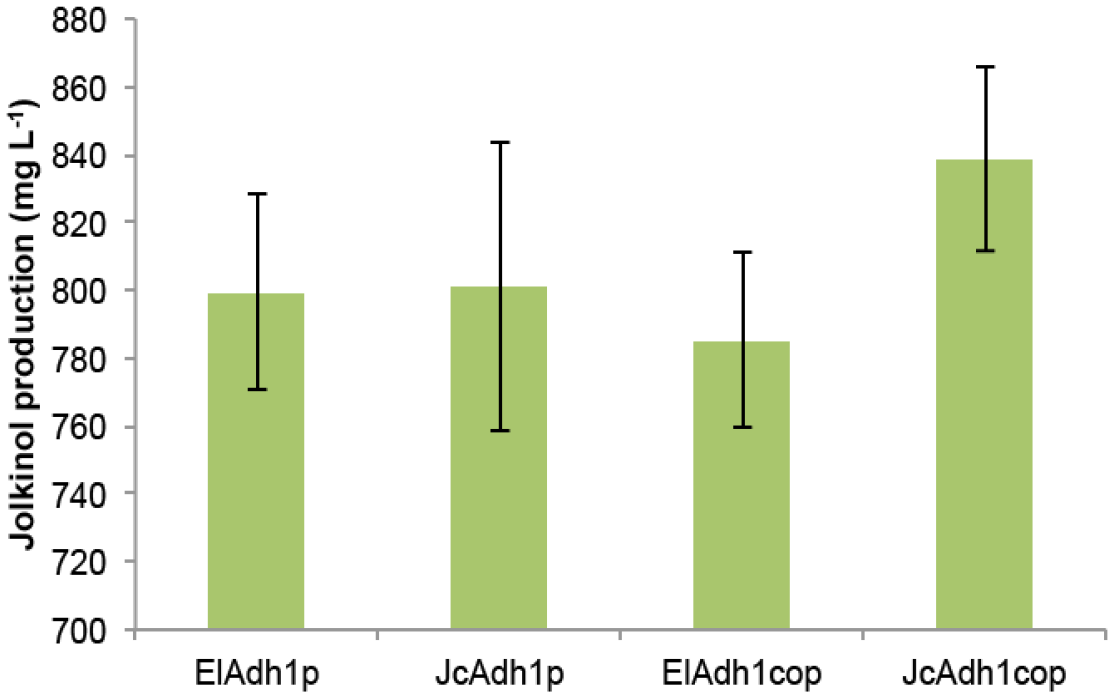
Supplemental Fig. 3.6. ^{1}H NMRs of 6-hydroxy,5-ketocasbene, 6-hydroxy-5,9-diketocasbene, and jolkinol. NMR spectra were obtained on Bruker DRX 500 MHz and AV 600 MHz spectrometers at the UC Berkeley College of Chemistry NMR facility, funded in part by NSF grant CHE-013086. Strain JWY519 was grown in YPG media under two distinct conditions: In one case, a 500 mL culture was shaken at 200 rpm in a 2L shake flask for 3 days; This condition yielded 6-hydroxy-5-oxocasbene and 6,9-dihydroxy-5-oxocasbene. In another case, 300 mL of culture was split into 5 24-well plates (2.5mL per well) and shaken at 800 rpm; This condition yielded 6,9-dihydroxy-5-oxocasbene and jolkinol. Both cultures were extracted with 2 volumes of ethyl acetate with 1% methanol, washed with brine, dried under reduced pressure and subjected to silica flash chromatography (RediSep Rf Gold 24g cartridge) on a Teledyne Isco Combiflash Rf with a hexane:ethyl acetate gradient ranging from 0% to 40% ethyl acetate. 5-hydroxy-6-oxocasbene eluted between 5% and 6% ethyl acetate, 5,9-dihydroxy-6-oxocasbene eluted between 17% and 19% ethyl acetate and Jolkinol eluted between 14% and 17% ethyl acetate. Each fraction was subjected to preparative reverse-phase HPLC on a Algilent Zorbax PrepHT C18 column (150 mm 21 mm) with a water-methanol gradient (50% methanol to 100% methanol over 10 min, flow rate: 20 mL/min). The major reverse-phase peak for each normal-phase fraction was isolated and identified by NMR spectrometry. Pure preps matching literature spectra were combined, weighed and dissolved in a known quantity of methanol for use as HPLC standards.

6-hydroxy-5-oxocasbene: 1.0 mg isolated yield, colorless oil. λ_{max}=274 nm. ^{1}H NMR (600 MHz, Chloroform-d) δ 6.34 (dd, J = 10.7, 1.6 Hz, 1H, 3), 5.24 (dd, J = 9.3, 5.8 Hz, 1H, 6), 5.09 (d, J = 9.2 Hz, 1H, 7), 4.84 (dd, J = 9.4, 4.3 Hz, 1H, 11), 4.08 (d, J = 6.5 Hz, 1H, 22), 2.29 – 2.17 (m, 3H, 9a, 10a, 13a), 2.12 (dddd, J = 20.1, 10.3, 7.8, 3.6 Hz, 2H, 9b, 14a), 1.95 (d, J = 1.3 Hz, 3H, 18), 1.77 (ddd, J = 12.7, 9.5, 2.9 Hz, 1H, 13b), 1.70 (d, J = 1.4 Hz, 3H, 19), 1.58 (s, 3H, 20), 1.21 (ddd, J = 12.5, 8.3, 2.3 Hz, 1H, 1), 1.18 (s, 3H, 16), 1.02 (s, 3H, 17), 0.85 (d, J = 2.0 Hz, 4H, 14b).

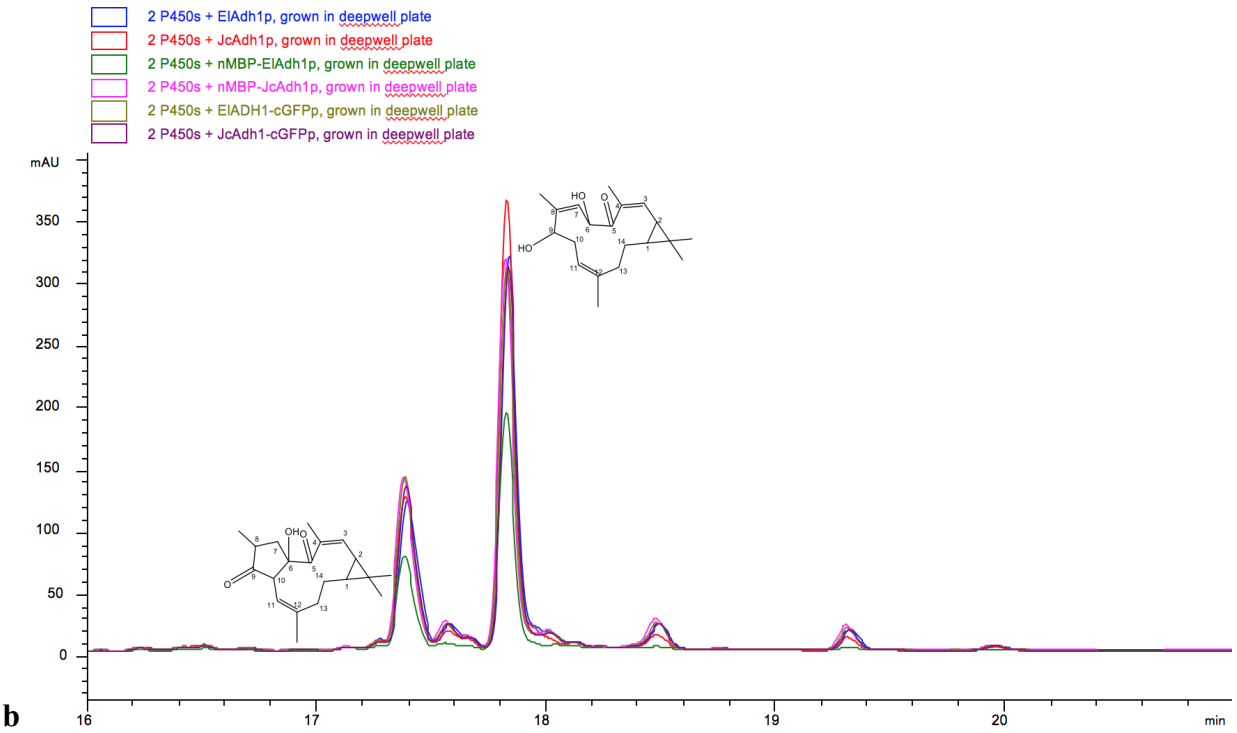
6,9-dihydroxy-5-oxocasbene: 4.1 mg isolated yield, colorless oil. λ_{max}=274 nm. ^{1}H NMR (600 MHz, Chloroform-d) δ 6.30 (d, J = 10.5 Hz, 1H, 3#), 5.27 (dd, J = 9.4, 5.8 Hz, 1H, 7), 5.21 (d, J = 9.4 Hz, 1H, 6), 4.70 (dd, J = 9.5, 3.1 Hz, 1H, 11), 4.15 (dd, J = 9.5, 6.3 Hz, 1H, 9), 4.10 (d, J = 6.5 Hz, 1H), 2.38 – 2.28 (m, 2H, 10a, 13a), 2.29 – 2.20 (m, 1H, 14b), 2.11 (dddd, J = 13.3, 10.6, 8.4, 2.3 Hz, 1H, 10b), 1.96 (s, 3H, 18), 1.72 (s, 4H, 19), 1.61 (s, 4H, 20), 1.55 (dd, J = 10.6, 8.3 Hz, 1H, 2), 1.25 – 1.19 (m, 1H, 1), 1.18 (s, 3H, 16), 1.04 (s, 3H, 17), 0.80 (q, J = 12.5 Hz, 1H, 14a).

Jolkinol D: 10.0 mg yield, yellowish viscous oil. λ_{max}=284 nm. ^{1}H NMR (500 MHz, Chloroform-d) δ = 7.37 (d, J=11.9, 1H, 3), 5.35 (d, J=9.7, 1H, 11), 3.50 (dd, J=14.8, 11.0, 1H, 7a), 3.03 (d, J=9.6, 1H, 10), 2.70 – 2.62 (m, 1H, 13a), 2.61 – 2.53 (m, 1H, 8), 2.18 (dddd, J=14.6, 5.6, 3.5, 1.8, 1H, 14a), 1.91 (d, J=1.7, 1H, 1), 1.85 (s, 3H, 18), 1.69 (td, J=13.2, 2.0, 1H, 13b), 1.61 (s, 1H, 7b), 1.45 (dd, J=11.9, 8.3, 1H, 2), 1.38 (d, J=1.4, 3H, 20), 1.29 (d, J=7.6, 3H, 19), 1.19 (s, 3H, 16), 1.13 (ddd, J=12.0, 8.3, 3.8, 1H, 1), 1.08 (s, 3H, 17).

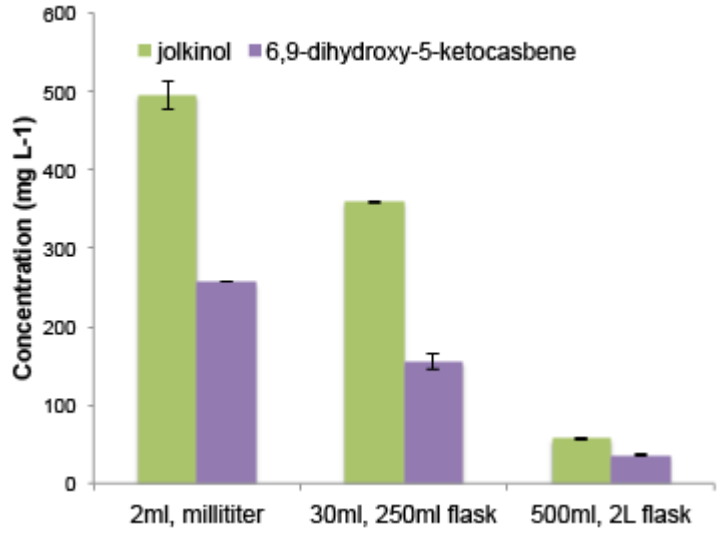
a


Jolkinol production (mg L-1)
880
860
840
820
800
780
760
740
720
700
ElAdh1p
JcAdh1p
ElAdh1cop
JcAdh1cop

b


2 P450s + ElAdh1p, grown in deepwell plate
2 P450s + JcAdh1p, grown in deepwell plate
2 P450s + nMBP-ElAdh1p, grown in deepwell plate
2 P450s + nMBP-JcAdh1p, grown in deepwell plate
2 P450s + ElADH1-cGFPp, grown in deepwell plate
2 P450s + JcAdh1-cGFPp, grown in deepwell plate
mAU
350
300
250
200
150
100
50
0
16
17
18
19
20
min

c



Supplemental Fig. 3.7. ADH codon optimization, tagging strategies, and effects of increased culture size on jolkinol production. (a,b) The JcADH1 and ElADH1 proteins were codon optimized and/or tagged with both MBP and GFP protein tags, individually, and integrated into JWY519. All of these modifications did not improve nor diminish jolkinol titer, thus we continued to use non-codon optimized JcADH1 for additional studies. **(c)** Culture size affects production of jolkinol. To test production in glass and in larger volumes, we grew our jolkinol strain in 30ml in 250ml baffled flasks and 500ml in 2L baffled flasks, alongside the 2ml YPgal cultures grown in millititer plates. Increased culture size decreases jolkinol production, likely due to decreased oxygen availability for CYP function. All compounds were measured by HPLC-UV. Bars represent mean values ±1 s.d. of three replicate cultures. All compounds were quantified using authentic standards purified from high-producing strains.

Supplemental Fig. 3.8. Sequences of synthetic genes and JcCPR1 used in this study.

CBSopt

ATGGCTAGTACGAAGTCTGAAACCGAGGCTAGGCCATTGGCATATTTTCCGCCGACA
GTGTGGGGTGACCGTCTTGCCTCCCTAACCTTCAATCAGCCCGCATTCGAGCTATTA
AGTAAGCAGGTGGAACTGCTGAATGAAAAAATCAAAAAGGAGATGTTGAATGTATC
TACGTCAGACCTAGCCGAAAAGATTATCCTAATAGACTCCCTTTGCAGACTTGGTGT
CTCTTACCACTTTGAAGAAGAGATTCAGGAAAACCTAACAAGAATATTCAATACCCA
ACCAAACTTTCTGAATGAGAAGGACTATGACTTATTTACGGTTGCGGTAATTTTTCG
TGTATTTCGTCAACATGGCTTCAAGATTAGTTCTGATGTCTTCAACAAATTCAAAGAT
AGCGATGGGAAGTTCAAAGAGTCTTTGCTAAATGATATAAAAGGTATTCTTTCTCTA
TTCGAAGCTACGCATGTGTCTATGCCGAACGAACCGATTCTAGATGAAGCCCTTGCG
TTTACGAAGGCATTCCTAGAGTCAAGTGCCGTTAAGTCTTTTCCAAATTTCGCGAAA
CACATAAGTAGTGCCCTAGAACAGCCAGTGCATAAAGGCATACCGCGTCTTGAAGC
TAGAAAATATATAGATCTGTACGAAGTGGACGAATCCAGGAATGAAACGGTTCTTG
AACTGGCGAAGCTGGACTTTAATAGGGTACAATTATTACATCAGGAGGAGTTGTCAC
AATTTAGCAAGTGGTGGAAGAGCCTAAACATCTCCGCAGAGGTTCCCTATGCAAGG
AACAGGATGGCGGAAATTTTCTTCTGGGCGGTAAGCATGTATTTCGAACCTCAATAT
GCAAAGGCGCGTATGATAGTATCTAAGGTAGTTCTTTTAATTTCACTAATCGACGAC
ACAATCGACGCATATGCGACGATTGACGAAATCCACCGTGTAGCAGACGCCATAGA
GCGTTGGGACATGAGATTGGTCGATCAACTACCCAACTACATGAAGGTGATTTACAG
GCTTATTATCAATACGTTCGACGAATTTGAAAAAGATCTTGAGGCAGAGGGAAAGA
GCTATTCAGTGAAGTATGGCAGAGAGGCTTACCAAGAGTTGGTCAGAGGGTACTAC
CTTGAAGCTATTTGGAAAGCTGATGGTAAAGTGCCTTCCTTTGATGAATATATTTAC
AATGGAGGGGTAACTACGGGTTTGCCTCTGGTTGCAACGGTGTCTTTCATGGGCGTT
AAAGAGATAAAAGGAACTAAGGCGTTTCAGTGGCTTAAAACTTATCCAAAACTGAA
CCAAGCAGGTGGTGAGTTTATTAGGTTGGTAAATGACGTAATGAGCCATGAAACTG
AACAAGATAGAGGTCACGTCGCGTCATGCATAGACTGCTACATGAAACAGTACGGT
GTCTCCAAGGAAGAGGCCGTCGAGGAGATCCAGAAAATGGCGACTAATGAATGGAA
GAAACTAAACGAACAATTGATTGTCCGTAGCACTGAGGTTGTTCCGGTGAACCTTTT
GATGAGGATCGTAAATCTGGTCCGTTTAACGGACGTATCATACAAGTATGGCGATGG
ATACACTGACTCTTCCCAACTTAAAGAGTATGTTAAAGGATTATTTATAGAACCCAT
CGCCACTTGA

JcCYP71D495

ATGCTGTTCTTTATTACGGTGCTTTTTATCTTTATCGCACTTAGGATATGGAAAAAAA
GTAAGGCAAACTCTACACCAAACCTTCCCCCTGGTCCCAATAAATTGCCTCTGATAG
GAAATGTCCACAATTTAGTTGGTGATTTGCCCTACCATAGATTGAGGGATCTATCAA
AGAAGTACGGTCCAATTATGCACCTTCAGTTAGGAGAAAACACAACTGTGGTTATTT
CATCACCGGAACTTGCCCAAGAGGTTATGAAGACGCACGACGTGAACTTCGCACAA

AGGCCTTTTGTATTAGCAGGTGACATAGTGAGCTATAAATGTAAAGATATCGCGTTT
GCTCCCTATGGTGAGTACTGGCGTCAACTTAGAAAGATGTGCTCCCTTGAGCTGCTT
ACCGCGAAGCGTGTACAGTCATTCAAATCCATCAGGGAAGAAGAAGTCAGTAAATT
GGTAGAGTCTATAAGCAGCTCATCCGGTTCCCCCATAAATTTTTCTAAAATGGCCTC
CTCCTTAACATATGCGATTATATCTAGAGCAGTTTGTGGGAAGGTATCAAGAGGCGA
AGAGGTTTTTGTACCAGCTGTAGAAAAGTTAGTGGAGGCTGGGAGAAGTATCTCACT
AGCAGATTTATACCCAAGCGTTAAACTGTTCAACGCGTTGAGTGTTGTGAGGAGACG
TGTCGAAAAAATTCACGGGGAAGTAGATAAAATCATCGAGAATATCGTTATCGAAC
ATAGGGAGAGAAAGCGTATGGCGCATGCGGGTATTAATTCCAAGGAAGAGGAAGAT
TTAGTAGACGTATTACTGAAGTTTCAAGAAAACGGAGATTTGGATTCATACTTGTCA
AATGATGGTATCAAAGCAGTCATCTTAGATATGTTCATTGCAGGTAGTGACACTAGT
TCAACTACTATAGAGTGGGCGATAAGTGAAATGGTTAAGAATCCGTCAATCATGGA
GAAGGCTCAGGCTGAAGTTAGAGAAGTATTCGGATCCAAGGGTAAAGTAGATGAGG
CCGATCTGCACGAGCTTAACTACCTGAAACTAGTCATTAAAGAAACGCTTAGGCTAC
ATCCCGCCGTGCCTTTACTATTACCCCGTCAAAGTAGAGAGGATTGTGTTATCGAAG
GATATAATATCGCTACCAAGTCTACGGTTATTGTAAATGCCTGGGCTATAGCTAGGG
ATCCTAAATATTGGGACGAGGCCGAACGTTTCTATCCAGAAAGGTTTATAAACAGCT
CAATTGACTTTAAGGGAACCAATTTCGAGTTTATACCGTTTGGGGCAGGGCGTAGGA
TGTGCCCTGGTATGCTATTTGGGCTGGCCAGCGTAGAGTTGCCTCTTGCACAGTTGCT
GTATCACTTCGATTGGAAATTGCCGGGTGGACAGAAGCCTGAGGATCTGGACATGA
GCGACGATTTAGACGGCACCGCAACCAGGAGGCATGCACTATACCTGACAGCTACA
CCTTATCTTCCTTCTGCGGTCGGCAAAATTAGTCGTTAG

JcCYP71D496

ATGTGGAAAAAGTCTAAAGCGAATTCAACCCCTAATTTGCCACCCGGACCAAACAA
GCTACCTGTTATCGGGAATGTACATAATTTGGTAGGGGATTTGCCATACCACAGGCT
AAGGGACTTAAGTAGCAAGTATGGGCCGATCATTCACCTTCAGCTTGGGGAAATCA
CTACAGTGGTCATCTCCTCTCCGGAGCTGGCTCAGGAGGTGATGAAGACGCACGATC
TTAATTTCGCCCAACGTCCTTTTGTCCTGGCTGGCGATATTGTAAGTTACAAGTGTAC
TGACATAGCTTTAGCACCCTACGGAGAATACTGGAGACAACTTAGGAAGATGTGCT
CCCTAGAACTACTGACGGCCAAACGTGTGCAATCATTTAAGAGTATAAGGGAAGAG
GAGGTTTTCAAGCTGGTGGAGAGCATCTCCTCCAGGAGTGGCTCTCCTATAAACTTT
AGTAAAATGGCTAGTTCTTTAACGTATGCTATAATATCAAGGGCGGTCTGCGGCAAG
GTATCCAGAGGTGAGGAGGTCTTTGTCCCCGCGGTAGAAAAGTTGGTCGAGGCGGG
TAGAAGTATAAGCTTAGCCGACCTTTATCCTAGTATTAAGCTTTTCAATGCGCTTTCT
GTTGTTAGACGTCGTGTCGAGAAGATTCACGGAGAAGTTGACAAAATAATCGAGAC
CATCGTAATGGAACATCGTGAGAGGAAACGTATGGTAGATACCGGCATAAAAAGCC
GTGAGGAGGAGGACTTAGTGGACGTTCTACTAAAATTTCAAGAAAATGGTGACTTA
AACAGTAGCTTAAGTAATGACTCAATTAAAGCAGTAATCTTAGACATGTTTATTGCG
GGCTCCGATGCCAGTTCTACTACAATCGAGTGGGCTATGTCTGAAATGGTTAAAAAT

CCGAGTATGGAAAAGGCCCAGGCTGAGGTAAGGAAAGTTTTTGGTTCTAAGGGCAA
AGTTGATGAGGCGGGGTTACACGAACTGAATTACTTGAAGCTGGTAATTAAAGAAA
CACTAAGACTGCATCCTGCCTTACCGTTATTGTTACCAAGGCAGAGCCGTGAAGATT
GCGTCATCGAAGGCTACAACATCGCCACGAAATCCACGGTTATCGTCAATGCTTGGG
CTATTGCTAGGGATCCTAAATACTGGGATGGAGCCGAAAGGTTTTATCCAGAGAGAT
TCATTAACTCCAGCATAGACTTCAAAGGAACAAATTTCGAGTTCACTCCGTTCGGCG
CAGGGAGGAGAAGATGCCCTGGAATGTTGTTCGGGCTTGCCAGTGTAGAACTTCCGT
TAGCTCAGTTACTGTATCATTTTGACTGGAAATTACCAGGCGGACAAAAACCGGAGG
ATCTTGACATGTCCGACAATCCTGGGGGCACGGCAACCAGGAGACATGCCCTTTACC
TTACGGCTACCCCCAACCTGCCCAGTGCTGTCGGAAAAATATAA

ElCYP71D445

ATGGAACTGGAATTTCGTTCACCTAGCAGCCCGTCAGAGTGGGCCATCACTTCTACC
ATCACACTACTGTTTCTAATTCTGCTTAGGAAAATTCTAAAACCAAAAACGCCCACA
CCCAACCTGCCACCGGGCCCAAAAAAGCTTCCCCTTATAGGAAACATTCATCAGCTA
ATCGGCGGAATCCCCCACCAGAAGATGCGTGACTTAAGTCAAATCCACGGGCCGAT
TATGCATCTTAAATTGGGTGAGCTTGAAAACGTAATTATTTCCTCCAAGGAGGCTGC
TGAAAAAATTCTGAAGACCCATGATGTTCTATTCGCGCAAAGACCGCAAATGATTGT
AGCAAAGAGCGTTACATACGACTTCCACGACATTACATTTTCACCCTACGGCGACTA
TTGGAGGCAGCTTCGTAAAATAACTATGATTGAATTATTAGCCGCTAAGAGGGTCCT
ATCCTTCAGAGCGATCAGGGAGGAAGAGACAACAAAATTAGTGGAGTTGATTAGGG
GCTTCCAGAGCGGCGAAAGTATAAACTTCACGCGTATGATTGATTCAACGACATACG
GCATCACTAGCAGAGCGGCCTGTGGCAAGATATGGGAAGGTGAAAACCTATTCATC
TCAAGCTTGGAAAAAATCATGTTCGAGGTTGGTAGCGGGATATCATTTGCCGACGCT
TACCCATCTGTTAAACTTTTGAAGGTCTTCTCAGGTATCAGGATTAGAGTGGACCGT
TTGCAAAAGAACATCGACAAAATCTTCGAATCTATCATAGAAGAACACAGGGAAGA
GAGAAAGGGGAGGAAGAAAGGCGAGGATGATCTTGATCTAGTGGACGTCTTGTTGA
ATTTACAAGAGAGTGGCACATTAGAAATCCCGCTAAGCGACGTTACTATAAAAGCT
GTAATCATGGATATGTTTGTAGCGGGGGTAGACACCAGCGCAGCCACTACGGAGTG
GCTGATGAGCGAGTTGATCAAAAACCCCGAAGTTATGAAAAAGGCTCAAGCAGAGA
TCAGAGAAAAATTCAAGGGCAAAGCTAGCATTGATGAAGCTGATTTGCAAGACCTG
CACTATCTAAAGCTAGTGATCAAGGAAACATTCCGTTTGCATCCTAGCGTACCGCTT
CTTGTCCCACGTGAATGTCGTGAATCCTGTGTGATTGAGGGATATGATATACCTGTT
AAAACTAAGATCATGGTGAATGCTTGGGCGATGGGTCGTGATACAAAATACTGGGG
AGAAGACGCCGAGAAATTCAAGCCTGAGAGGTTCATCGACTCCCCTATCGATTTCAA
AGGCCATAACTTTGAGTACCTACCGTTCGGGTCAGGTCGTAGGTCTTGTCCTGGAAT
GGCCTTTGGGGTAGCCAACGTCGAGATTGCAGTCGCAAAATTATTGTATCATTTTGA
CTGGAGACTGGGCGACGGAATGGTTCCCGAAAACTTAGACATGACGGAAAAAATCG
GCGGAACAACAAGGAGACTATCAGAGTTGTACATCATACCGACTCCGTACGTACCG
CAAAATTCAGCGTAA

JcCYP726A35 chimera

ATGGACTTACAGCTACAAATACCCTCTTACCCTATAATCTTTTCTTTTTTTATCTTCAT
TTTCATGCTGATTAAAATCTGGAAAAAACAGACACAAACCAGTATTTTCCCTCCCGG
GCCTTGGAGGCTTCCTTTCATTGGCAACATTCACCAGTTAGCGGGGCCAGGGCTGCC
GCACCACCGTTTGACAGATTTAGCAAAAACGTATGGTCCGGTCATGGGAGTACATCT
TGGCGAAGTTTATGCCGTCGTAGTTTCATCTGCCGAGACATCTAAGGAAGTTCTAAG
GACTCAAGACACCAATTTCGCGGAACGTCCATTGGTAAATGCGGCTAAAATGGTACT
TTATAACCGTAATGATATTGTATTTGGCTCTTTTGGAGATCAGTGGCGTCAAATGCGT
AAGATATGCACTCTGGAGTTGCTATCAGTGAAAAGGGTTCAATCCTTTAAGTCCGTC
CGTGAAGAGGAGATGAGCTCCTTCATTAAGTTTTTGTCTAGCAAGAGTGGATCACCC
GTGAATTTGACCCATCATCTATTTGTTCTGACCAATTACATAATCGCCAGGACAAGC
ATAGGTAAAAAATGTAAGAATCAGGAAGCGCTACTTAGAATTATCGACGATGTGGT
AGAAGCTGGCGCAGGATTCTCAGTCACCGACGTTTTTCCATCCTTTGAGGCCCTTCA
TGTGATCTCAGGGGATAAGCACAAATTTGATAAGCTGCACCGTGAGACTGACAAGA
TACTGGAAGACATAATATCTGAGCACAAAGCAGACAGAGCAGTGAGCAGCAAGAA
GTCAGATGGGGAGGTAGAGAATTTGCTAGATGTTTTACTAGACCTGCAGGAGAATG
GTAACCTGCAGTTTCCATTGACGAACGACGCAATTAAAGGTGCGATTTTAGACACGT
TTGGAGCCGGTTCAGACACCTCCTCAAAGACGGCCGAATGGACGCTTTCAGAGCTTA
TCCGTAACCCGGAAGCAATGAGGAAAGCTCAAGCAGAAATCAGAAGGGTATTTGAT
GAGACTGGCTATGTGGATGAGGATAAGTTCGAAGAACTTAAATACTTGAAACTTGT
AGTCAAGGAAACTCTAAGGCTGCATCCCGCCGTCCCCTTGATTCCAAGAGAGTGCA
GGGGCAAGACCAAGATCAACGGGTACGACATTTTCCCGAAGACGAAAGTTTTAGTT
AATGTTTGGGCTATATCTCGTGATCCTGCTATTTGGCCGGAACCAGAGAAATTCAAC
CCAGAGAGGTTCATCGATAATCCGATAGACTACAAATCTATAAATTGCGAGTTAACC
CCATTTGGGGCCGGAAAGAGGATTTGTCCCGGCATGACCTTAGGTATAACCAACCTA
GAACTGTTTTTAGCTAACCTATTGTATCACTTCGATTGGAAACTACCCGATGGTAAG
ATGCCGGAAGACTTGGATATGTCAGAAAGCTTTGGAGGAGCGATTAAAAGGAAGAC
TGACCTTAAGCTTATACCAGTCTTAGCACGTCCACTAACCCCTCGTAATGCCAATAG
CGGTAATACATTCACAACAACTGACGCAGACTCTCCTGCGAGCATGTGTCCACATTT
GAAGGCGCTTTAA

JcCYP726A20

ATGGAGCATCAAATTCTATCCTTTCCCGTCTTGTTCTCATTGCTATTATTCATCCTGGT
ACTACTGAAAGTATCCAAAAAGCTTTACAAGCACGATTCTAAGCCTCCACCGGGTCC
GTGGAAATTACCTTTCATAGGCAATTTAATACAGCTTGTGGGTGACACCCCGCACAG
GAGGCTGACCGCGCTAGCGAAGACTTACGGCCCCGTCATGGGCGTTCAGCTTGGAC
AGGTCCCTTTCCTGGTAGTTTCCAGCCCGGAGACTGCGAAAGAAGTTATGAAGATAC
AAGACCCTGTCTTTGCGGAGAGGCCATTGGTTCTTGCTGGAGAAATAGTTTTATACA
ACCGTAACGATATTGTATTTGGCAGTTATGGGGATCAATGGAGGCAAATGAGAAAA
TTCTGTACTTTGGAATTACTTTCTACGAAGAGGGTTCAAAGCTTCAGACCTGTAAGG

GAAGAGGAAGTTGCTTCATTTGTGAAGCTAATGCGTACGAAAAAAGGAACCCCCGT
GAACCTTACACACGCACTGTTTGCACTTACAAATTCAATCGTAGCTAGAAACGCGGT
GGGTCACAAGTCCAAGAACCAAGAGGCCCTGTTAGAGGTCATTGACGACATCGTGG
TGAGCGGAGGGGGGGTCTCAATAGTTGACATTTTCCCATCCCTACAGTGGTTGCCGA
CTGCAAAGAGAGAGAGGTCAAGGATCTGGAAGCTACACCAAAATACCGACGAGATT
CTTGAAGATATACTACAAGAGCACCGTGCCAAGCGTCAGGCAACTGCCTCCAAGAA
CTGGGATAGGTCTGAGGCCGATAATCTACTAGACGTGCTTTTGGATTTACAACAAAG
CGGAAATCTAGATGTCCCCCTGACTGATGTTGCCATAAAGGCCGCAATTATCGATAT
GTTCGGGGCTGGGAGCGATACTAGCTCAAAGACAGCGGAGTGGGCCATGGCCGAGC
TGATGAGGAACCCTGAAGTCATGAAGAAGGCTCAAGAGGAGTTACGTAATTTCTTT
GGTGAAAACGGAAAGGTAGAAGAGGCTAAGTTGCATGAACTTAAATGGATTAAATT
GATAATTAAGGAGACCTTGCGTCTGCACCCGGCGGTTGCTGTAATTCCTAGGGTGTG
CAGGGAGAAAACTAAAGTCTACGGGTACGACGTAGAGCCAGGGACAAGAGTTTTCA
TAAATGTTTGGAGTATAGGTAGGGACCCAAAGGTGTGGAGTGAGGCTGAAAGATTT
AAGCCTGAACGTTTTATCGATTCAGCTATAGATTACAGGGGTCTTAATTTCGAGCTT
ATTCCCTTTGGGGCGGGCAAACGTATTTGCCCGGGGATGACTCTGGGTATGGCTAAT
CTGGAAATATTCCTGGCGAATTTATTGTACCACTTCGATTGGAAGTTTCCCAAGGGC
GTAACAGCTGAGAATTTAGATATGAACGAGGCCTTCGGCGGGGCTGTGAAGCGTAA
GGTCGATCTAGAATTGATTCCTATACCCTTTAGGCCATAA

ElCYP726A27

ATGGACTTGCAATTACAAATCCCATCCTACCCAATCATCTTCTCATTCTTCATTTTCA
TCTTCATGTTGATCAAGATCTGGAAGAAGCAAACCCAAACCTCTATTTTTCCACCAG
GTCCATTCAAGTTTCCAATCGTTGGTAATATTCCACAATTGGCTACTGGTGGTACTTT
GCCACATCATAGATTGAGAGATTTGGCTAAAATCTACGGTCCAATCATGACCATTCA
ATTGGGTCAAGTTAAGTCCGTTGTTATCTCTTCACCAGAAACCGCTAAAGAAGTCTT
GAAAACCCAAGATATTCAATTCGCCGATAGACCTTTGTTGTTGGCTGGTGAAATGGT
CTTGTACAACAGAAAGGATATCTTGTACGGTACTTACGGTGATCAATGGAGACAAAT
GAGAAAGATCTGCACCTTGGAATTATTGTCCGCCAAGAGAATTCAATCCTTCAAGTC
CGTTAGAGAAAAAGAAGTCGAATCCTTCATCAAGACCTTGAGATCTAAATCCGGTAT
CCCAGTTAATTTGACCAACGCTGTTTTCGAATTGACCAACACCATTATGATGATTAC
CACCATTGGTCAAAAGTGCAAGAATCAAGAAGCCGTTATGTCCGTTATCGATAGAGT
TTCTGAAGCTGCTGCTGGTTTTTCTGTTGCTGATGTTTTTCCATCCTTGAAGTTCTTGC
ATTACTTGTCTGGTGAAAAGACCAAGTTGCAAAAGTTGCACAAAGAAACCGACCAA
ATCTTGGAAGAAATCATCTCTGAACATAAGGCTAACGCTAAAGTTGGTGCTCAAGCT
GATAATTTGTTGGATGTTTTGTTGGACTTACAAAAGAACGGTAACTTACAAGTCCCA
TTGACTAACGATAACATTAAGGCTGCTACTTTGGAAATGTTTGGTGCTGGTTCTGAT
ACTTCTTCTAAGACTACTGATTGGGCTATGGCTCAAATGATGAGAAAACCTACTACT
ATGAAGAAGGCCCAAGAAGAAGTTAGAAGAGTTTTTGGTGAAAACGGTAAGGTTGA
AGAATCCAGAATCCAAGAATTGAAGTACTTGAAGTTGGTCGTCAAAGAAACTTTGA

GATTGCATCCAGCTGTTGCCTTGATTCCTAGAGAATGTAGAGAAAAGACTAAGATCG
ATGGTTTCGACATCTACCCAAAGACAAAGATTTTGGTTAACCCATGGGCTATTGGTA
GAGATCCAAAAGTTTGGAATGAACCAGAATCTTTCAACCCAGAAAGATTCCAAGAC
TCCCCAATTGATTACAAGGGTACAAATTTTGAATTGATCCCATTCGGTGCCGGTAAG
AGAATTTGTCCAGGTATGACTTTGGGTATCACCAACTTAGAATTATTCTTGGCCAAC
TTGTTGTACCACTTCGATTGGAAGTTTCCAGATGGTATTACCTCCGAAAACTTGGAT
ATGACTGAAGCTATTGGTGGTGCCATTAAGAGAAAGTTGGACTTGGAATTGATCTCT
ATCCCATACACCTCTTCATAA

JcADH1

ATGGCAAGCTCTTCTAGCCCGGCGCCTACCGCAAAACGTCTAGAGGGCAAGGTAGC
CCTAATAACGGGGGGGGCGAGCGGGATCGGCGAATGCACGGCGCGTCTGTTCGCCA
GACATGGTGCCAAAGTGATTATCGCAGACGTACAATCAGAATTGGGTAGAAGTGTG
GCAGAGAAAATTGGTTCAGAAACAGGGCAGCCTGTGACCTATGTTGATTGTAATGT
GACAGTGGAGTCTGATGTGGAAAATGCAGTGAACACCGCCGTTTCTTTACACGGGA
AATTGGATATCATGTTCAATAACGCGGGGATTGCAGGCAACAACCATGATAAGATC
CTTAGCACTGAGCGTGAGGATTTTATGCGTGTACTTGACATTAACATATATGGCGGA
GTCTTGGGGGCCAAACATGCAGCTAGAGTGATGATTCCGGAAAAGAAAGGGTGCAT
ACTTTTTACTGCATCAGTTTCATCTGCTTTGTATGGAGGACCGTACGCCTACACGGCA
TCTAAACATGCCGTTGTTGGGCTTGCCAAGAATTTAGCCATAGAGCTTGGTCAACAC
GGCATACGTGTCAATTGCATTAGTCCGGGAGCTGTCCAGACTGGCCTTGCTAAACAA
TTGGGACTATCAGAACAGCAGGTGCAAGAGTGGTCTAGTGCCCTGGCTAACTTAAA
AGTGGTGAAACTAGAAGTCAACGATATAGCTGAAGCGGCGTTATACTTGGCATCTG
ATGATAGTAAGTTTGTAAGTGGATTAAATCTGTTGGTTGACGGAGCTGCCAGTCTGC
CGACCACGACGAGGGCATACTAG

ElADH1

ATGAATGGTTGTTGTTCTCAGGACCCCACGTCCAAAAGATTAGAAGGAAAAGTCGC
GGTAATCACGGGAGGGGCGAGTGGAATCGGCGCGTGTACTGTTAAGCTTTTCGTGA
AGCACGGGGCTAAGGTTGTTATCGCCGATGTCCAAGACGAGTTGGGACATTCCTTGT
GTAAAGAAATAGGCAGCGAAGACGTAGTCACTTATGTACATTGCGACGTATCTTCCG
ATAGCGATGTTAAAAATGTGGTAGACTCTGCGGTAAGTAAATACGGGAAGCTTGAT
ATTATGTTCTCAAACGCCGGCGTTAGCGGTGGTTTAGACCCTCGTATATTAGCGACA
GAAAATGACGAATTTAAGAAGGTATTTGAGGTTAATGTTTTCGGTGGGTTTTTAGCC
GCCAAACACGCTGCTAGGGTCATGATACCTGAAAAGAAAGGTTGTATATTATTCACG
TCCTCCAACAGCGCTGCAATCGCCATTCCTGGACCCCATTCATACGTTGTGTCCAAG
CACGCGCTGAATGGTTTAATGAAGAACCTAAGTGCAGAGTTAGGACAGCATGGGAT
AAGGGTGAATTGCGTCTCACCATTCGGCGTAGTAACGCCTATGATGGCTACGGCATT
TGGCATGAAGGATGCTGATCCTGAGGTCGTCAAGGCAACCATAGAAGGTCTGTTAG
CTAGTGCCGCTAACCTTAAAGAAGTAACACTAGGCGCTGAGGATATAGCTAACGCC

GCCTTGTACCTGGCAAGCGACGAGGCAAAGTATGTTAGCGGCCTTAACCTGGTGGTA
GATGGTGGTTACTCCGTTACCAACCCATCCTTCACGGCAACCCTACAAAAAGCATTC
GCAGTTGCTCATGTGTAA

JcCPR1

ATGAGTTCGGATTTGGTTAGGTATGTTGAGTCTGTCCTCGGGGTCTCGCTTGGTGGTT
CTGTGACTGATTCGCTTCTTCTGATTGTCACGACGTCGGTTGCGGTGATAGTTGGGCT
GTTTGTGTTTTTGTTGAAGAAATCGTCGGATCGAAGCAAAGAGGTGAAACCTGTGGT
CGTTCCAAAGTCACTTACGGTGAAGAAAGAGGAGGATGACGCTGAGGCTCTTGCTG
GTAAAACTAGAGTTATTATCTTTTATGGGACTCAGACTGGAACAGCTGAGGGCTTTG
CTAAGTCCTTATCTGAGGAGATCAAGGCAAGATACGAGAAAGCAGCTGTTAAAGTT
GTTGACCTGGATGATTATGCTGCGGATGATGAAGAATATGAAGAAAAATTGAAGAA
GGAGACTTTGTCATTTTTCATGGTTGCCACTTATGGAGATGGAGAGCCAACTGATAA
TGCTGCAAGATTTTATAAGTGGTTGACTGAGGAAAATGAAAGGGGAGTCTGGCTTC
AACAGCTCACTTTTGGAGTTTTTGGCTTGGGTAACCGTCAATATGAACATTTTAATA
AGATAGCAAAAGTGCTCGATGAACAACTTAGCGAACAAGGTGCAAAACGCCTCATT
CCTGTCGGTCTTGGTGATGATGATCAATGCATTGAAGATGATTTTTCTGCTTGGAAA
GAATTATTATGGCCTGAGTTAGACCAGATACTCAGAGATGAAGATGATGTGAATACT
CCTTCTACTCCATATACAGCTGCTATTCTGGAATATCGAGTGGTGATTCATGATGCTT
CTATGACATCTTTTGATGATAAATCATCACACTTGGCAAATGGGAATACTGTTTTCG
ATATTCACCATCCATGCAGGGCTAATGTTGCTGTTCAAAAAGAGCTCCACAAACCAG
AGTCTGACAGGTCTTGCATACATCTGGAGTTTGACGTAGCAGGGACTGGTATTACGT
ATGAAACTGGTGACCATGTGGGTGTTTATTCTGAGAATTTTGACGAAACTGTTGAAG
AAGCAGCAAAATTGTTGGGTCAACCATTAGATTTGCTATTTTCTGTTTACACTGACA
ATGAAGATGGCACACCCCTTGGAAGTTCATTGCCCCCTGCGTTCCCAGGGCCATGCA
CACTGCGAACTGCCTTGGCACGCTATGCAGACCTCTTGAACTCACCTCGAAAGGCGG
CTTTGATTGCTTTGGCTGCCCATGCCAGTGATCCTAGTGAGGCAGAGAGACTCAGAT
TTTTAGCATCACCACAAGGGAAGGATGAGTATGCTCAATGGATTGTTGCTAGCCAGA
GAAGTCTTCTTGAGGTAATGGCTGAGTTCCCTTCTGCAAAACCTCCCCTTGGTGTATT
TTTTGCAGCAGTAGCTCCTCGTCTACAGCCTCGCTACTATTCGATCTCATCCTCCCCC
AGGTTTGCTCCCAGTAGAGTGCATGTGACCTGTGCTTTAGTTTATGGTCCAACACCG
ACTGGTAGAATCCACAAAGGGGTGTGCTCAACCTGGATGAAGAATGCAGTTCCTCT
GGAGAGAAGCAGTGACTGTAGCTGGGCTCCCATTTTCATTCGGACATCTAATTTCAA
GTTACCATCTGATCCATCAGTTCCAATTATCATGGTGGGACCTGGTACTGGATTGGC
ACCTTTTAGAGGATTTCTACAGGAAAGAATGGCCTTGAAGCAGGAGGGCGCTCAAC
TTGGTCCTGCTCTGCTCTTCTTTGGATGTAGAAATCGTCGAATGGATTTCATATATGA
GGATGAACTCAATAATTTCGTAGAACAAGGTGTGATATCTGAGTTGATTGTTGCATT
CTCAAGAGAGGGGCCACAAAAGGAGTATGTTCAACATAAGATGGTGCAGAAAGCA
GCACAAATATGGGCCATAATTTCTCAGGGCGGATACCTTTATGTCTGTGGTGATGCA
AAGGGTATGGCCAGAGACGTCCACCGTACTTTGCATAATATTGTTCAGGAGCAGGG

AAATTTGGACGCATCGAAGACAGAATCAATGGTTAAGAAACTCCAAATGGATGGGC
GATATCTCAGAGATGTCTGGTGA

7.3. Supplementary for Chapter 4 — De novo synthesis of the sedative valerenic acid in Saccharomyces cerevisiae

Figure S4.1. GCMS profiles of transiently expressed valerenic acid pathway enzymes in *Nicotiana benthamiana.* Plasmids expressing *Arabidopsis thaliana* HMGR (AtHMGR, P14891), *Arabidopsis thaliana* FPPS (AtFPPS, NM_124151.3), *Valeriana officinalis* VDS, and select *V. officinalis* P450s were transformed into *Agrobacteria tumefaciens* and coinfiltrated into young *N. benthamiana* leaves. Valerena-4,7(11)-diene (Peak 2, retention time: 9.47 min), denoted by an asterisk, was produced at variable levels, but no detectable valerenic acid was formed. Dodecane (Peak 1, retention time: 4.63 min, 50mg/L) was used as an internal standard. For GC analysis, a different program was used; an initial temperature of 100°C was held for 1 min, followed by ramping to 250°C at a rate of 15°C/min to 250°C, followed by ramping to 300°C at a rate of 30°C/min, and then held at 300°C for 3 min.

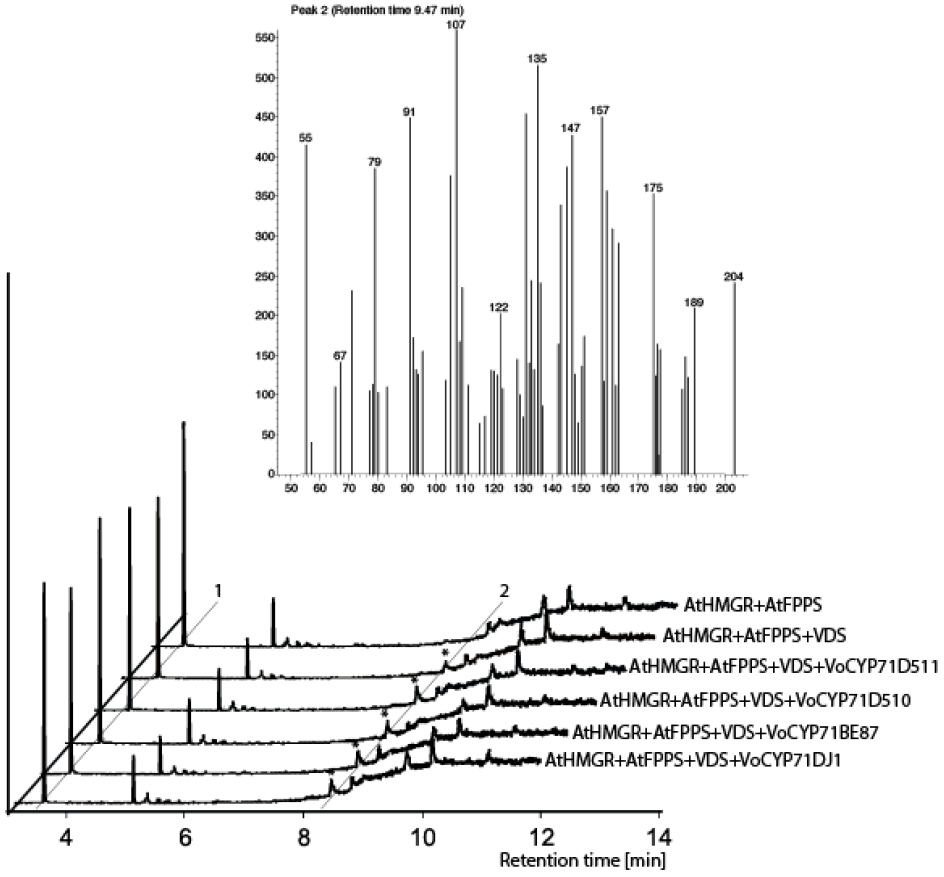


Figure S4.2. GFP tagging of VDS indicates protein insolubility. VDS was fused to either GFP or with two additional protein tags, MBP on the N-terminal and ERG20 on the C-terminal (3-tag VDS). The 3-tag VDS showed improved cytosolic expression.

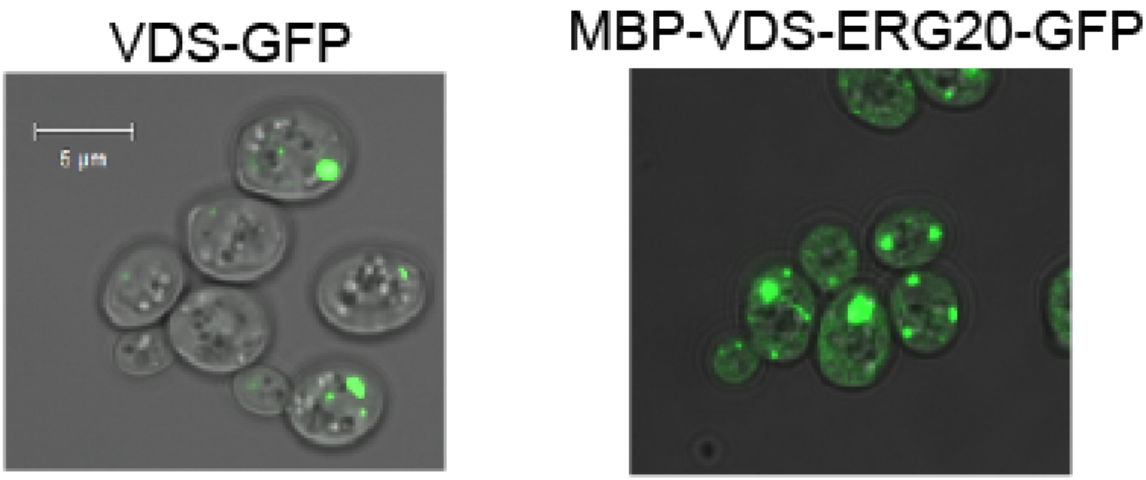
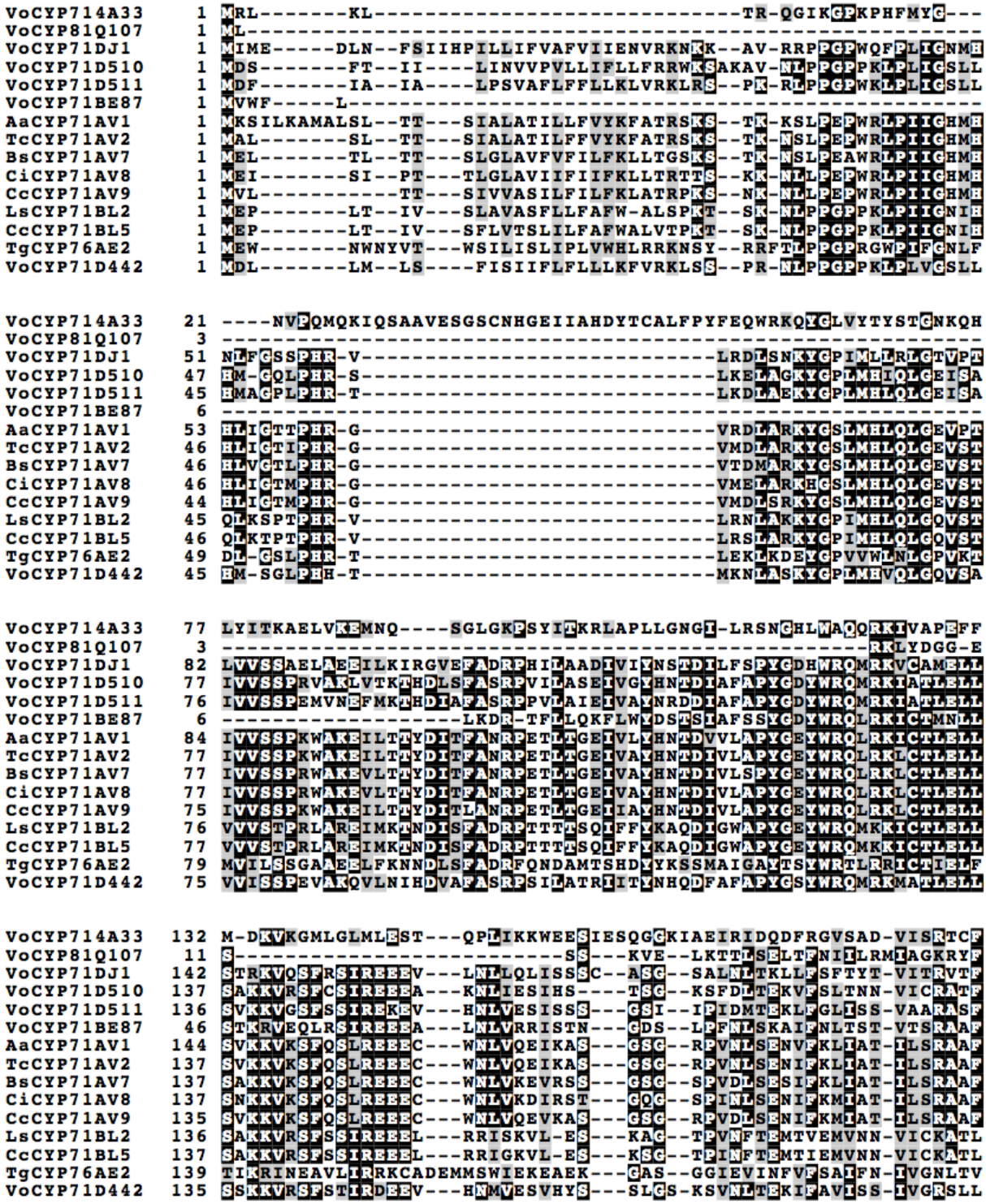
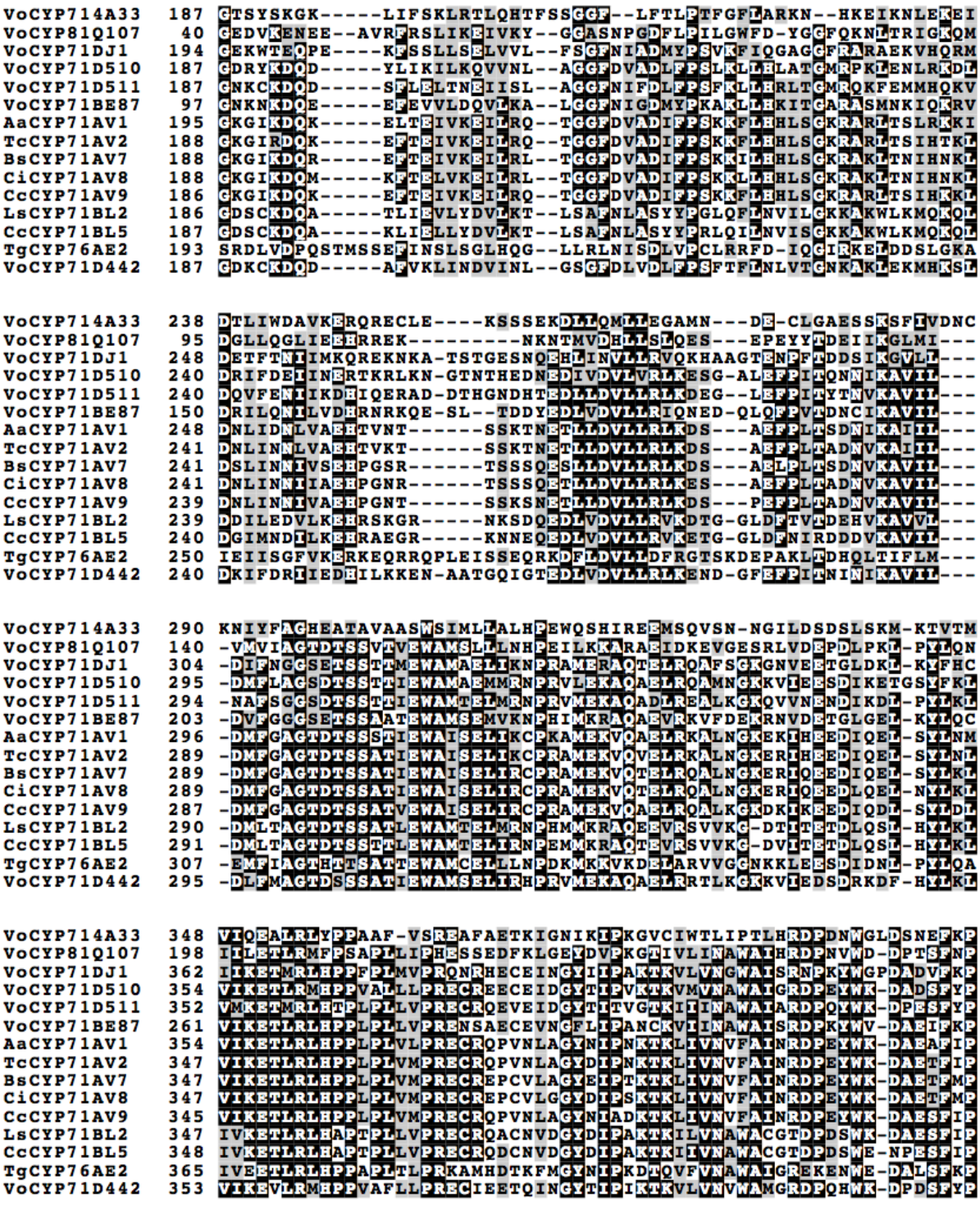


Figure S4.3. Gblocks curated alignment for V. officinalis P450s and related Asteraceae P450s.




VoCYP714A33 187 GTSYSKGK-----LIFSKLRTLQHTFSSGGF--LFTLPTFGFLARKN--HKEIKNLEKEI
VoCYP81Q107 40 GEDVKENEE--AVRFRSLIKEIVKY--GGASNPGDFLPILGWFD-YGGFQKNLTRIGKQM
VoCYP71DJ1 194 GEKWTEQPE----KFSSLLSELVVL--FSGFNIADMYPSVKFIQGAGGFRARAEKVHQRM
VoCYP71D510 187 GDRYKDQD-----YLIKILKQVVNL--AGGFDVADLFPSLKLLHLATGMRPKLENLRKDL
VoCYP71D511 187 GNKCKDQD-----SFLELTNEIISL--AGGFNIFDLFPSFKLLHRLTGMRQKFEMMHQKV
VoCYP71BE87 97 GNKNKDQE-----EFEVVLDQVLKA--LGGFNIGDMYPKAKLLHKITGARASMNKIQKRV
AaCYP71AV1 195 GKGIKDQK-----ELTEIVKEILRQ--TGGFDVADIFPSKKFLHHLSGKRARLTSLRKKI
TcCYP71AV2 188 GKGIRDQK-----EFTEIVKEILRQ--TGGFDVADIFPSKKFLHHLSGKRARLTSIHTKL
BsCYP71AV7 188 GKGIKDQR-----EFTEIVKEILRL--TGGFDVADIFPSKKILHHLSGKRAKLTNIHNKL
CiCYP71AV8 188 GKGIKDQM-----KFTELVKEILRL--TGGFDVADIFPSKKLLHHLSGKRAKLTNIHNKL
CcCYP71AV9 186 GKGIKDQK-----EFTEIVKEILRQ--TGGFDVADIFPSKKFLHHLSGKRARLTSIHKKL
LsCYP71BL2 186 GDSCKDQA-----TLIELVLYDVLKT--LSAFNLASYYPGLQFLNVILGKKAKWLKMQKQL
CcCYP71BL5 187 GDSCKDQA-----KLIELLYDVLKT--LSAFNLASYYPRLQILNVISGKKAKWLKMQKQL
TgCYP76AE2 193 SRDLVDPQSTMSSEFINSLSGLHQG--LLRLNISDLVPCLRRFD-IQGIRKELDDSLGKA
VoCYP71D442 187 GDKCKDQD-----AFVKLINDVINL--GSGFDLVDLFPSFTFLNLVTGNKAKLEKMHKSL
VoCYP714A33 238 DTLIWDAVKERQRECLE----KSSSEKDLLQMLIEGAMN---DE-CLGAESSKSFIVDNC
VoCYP81Q107 95 DGLLQGLIEEHRREK---------NKNTMVDHLLSLQES---EPEYYTDEIIKGLMI---
VoCYP71DJ1 248 DETFTNIIMKQREKNKA-TSTGESNQEHLINVLLRVQKHAAGTENPFTDDSIKGVLL---
VoCYP71D510 240 DRIFDEIINERTKRLKN-GTNTHEDNEDIVDVLVRLKESG-ALEFPITQNNIKAVIL---
VoCYP71D511 240 DQVFENIIKDHIQERAD-DTHGNDHTEDLLDVLLRLKDEG--LEFPITYTNVKAVIL---
VoCYP71BE87 150 DRILQNILVDHRNRKQE-SL--TDDYEDLVDVLLRIQNED-QLQFPVTDNCIKAVIL---
AaCYP71AV1 248 DNLIDNLVAEHTVNT------SSKTNETLLDVLLRLKDS---AEFPLTSDNIKAIIL---
TcCYP71AV2 241 DNLINNLVAEHTVKT------SSKTNETLLDVLLRLKDS---AEFPLTADNVKAIIL---
BsCYP71AV7 241 DSLINNIVSEHPGSR------TSSSQESLLDVLLRLKDS---AELPLTSDNVKAVIL---
CiCYP71AV8 241 DNLINNIIAEHPGNR------TSSSQETLLDVLLRLKES---AEFPLTADNVKAVIL---
CcCYP71AV9 239 DNLINNIVAEHPGNT------SSKSNETLLDVLLRLKDS---PEFPLTADNVKAVIL---
LsCYP71BL2 239 DDILEDVLKEHRSKGR-----NKSDQEDLVDVLLRVKDTG-GLDFTVTDEHVKAVVL---
CcCYP71BL5 240 DGIMNDILKEHRAEGR-----KNNEQEDLVDVLLRVKETG-GLDFNIRDDDVKAVIL---
TgCYP76AE2 250 IEIISGFVKERKEQRRQPLEISSEQRKDFLDVLLDFRGTSKDEPAKLTDHQLTIFLM---
VoCYP71D442 240 DKIFDRIIEDHILKKEN-AATGQIGTEDLVDVLLRLKEND-GFEFPITNINIKAVIL---
VoCYP714A33 290 KNIYFAGHEATAVAASWSIMLLALHPEWQSHIREEMSQVSN-NGILDSDSLSKM-KTVTM
VoCYP81Q107 140 -VMVIAGTDTSSVTVEWAMSLLLNHPEILKKARAEIDKEVGESRLVDEPDLPKL-PYLQN
VoCYP71DJ1 304 -DIFNGGSETSSTTMEWAMAELIKNPRAMERAQTELRQAFSGKGNVEETGLDKL-KYFHC
VoCYP71D510 295 -DMFLAGSDTSSTTIEWAMAEMMRNPRVLEKAQAELRQAMNGKKVIEESDIKETGSYFKL
VoCYP71D511 294 -NAFSGGSDTSSTTIEWAMTELMRNPRVMEKAQADLREALKGKQVVNENDIKDL-PYLKL
VoCYP71BE87 203 -DVFGGGSETSSAATEWAMSEMVKNPHIMKRAQAEVRKVFDEKRNVDETGLGEL-KYLQC
AaCYP71AV1 296 -DMFGAGTDTSSSTIEWAISELIKCPKAMEKVQAELRKALNGKEKIHEEDIQEL-SYLNM
TcCYP71AV2 289 -DMFGAGTDTSSATIEWAISELIKCPRAMEKVQVELRKALNGKERIHEEDIQEL-SYLNL
BsCYP71AV7 289 -DMFGAGTDTSSATIEWAISELIRCPRAMEKVQTELRQALNGKERIQEEDIQEL-SYLKL
CiCYP71AV8 289 -DMFGAGTDTSSATIEWAISELIRCPRAMEKVQTELRQALNGKERIQEEDLQEL-NYLKL
CcCYP71AV9 287 -DMFGAGTDTSSATVEWAISELIRCPRAMEKVQAELRQALKGKDKIKEEDIQDL-SYLDL
LsCYP71BL2 290 -DMLTAGTDTSSATLEWAMTELMRNPHMMKRAQEEVRSVVKG-DTITETDLQSL-HYLKL
CcCYP71BL5 291 -DMLTAGTDTSSTTLEWAMTELIRNPEMMKRAQTEVRSVVKG-DVITETDLQSL-HYLKL
TgCYP76AE2 307 -EMFIAGTHTTSATTEWAMCELLLNPDKMKKVKDELARVVGGNKKLEESDIDNL-PYLQA
VoCYP71D442 295 -DLFMAGTDSSSATIEWAMSELIRHPRVMEKAQAELRRTLKGKKVIEDSDRKDF-HYLKL
VoCYP714A33 348 VIQEALRLYPPAAF-VSREAFAETKIGNIKIPKGVCIWTLIPTLHRDPDNWGLDSNEFKP
VoCYP81Q107 198 IILETLRMFPSAPLLIPHESSEDFKLGEYDVPKGTIVLINAWAIHRDPNVWD-DPTSFNP
VoCYP71DJ1 362 IIKETMRLHPPFPLMVPRQNRHECEINGYIIPAKTKVLVNGWAISRNPKYWGPDADVFKP
VoCYP71D510 354 VIKETLRMHPPVALLLPRECREECEIDGYTIPVKTKVMVNAWAIGRDPEYWK-DADSFYP
VoCYP71D511 352 VMKETMRLHTPLPLLVPRECRQEVEIDGYTITVGTKIIINAWAIARDPQYWK-DPESFYP
VoCYP71BE87 261 VIKETLRLHPPLPLLVPRENSAECEVNGFLIPANCKVIINAWAISRDPKYWV-DAEIFKP
AaCYP71AV1 354 VIKETLRLHPPLPLVLPRECRQPVNLAGYNIPNKTKLIVNVFAINRDPEYWK-DAEAFIP
TcCYP71AV2 347 VIKETLRLHPPLPLVMPRECRQPVNLAGYDIPNKTKLIVNVFAINRDPEYWK-DAETFIP
BsCYP71AV7 347 VIKETLRLHPPLPLVMPRECREPCVLAGYEIPTKTKLIVNVFAINRDPEYWK-DAETFMP
CiCYP71AV8 347 VIKETLRLHPPLPLVMPRECREPCVLGGYDIPSKTKLIVNVFAINRDPEYWK-DAETFMP
CcCYP71AV9 345 VIKETLRLHPPLPLVMPRECRQPVNLAGYNIADKTKLIVNVFAINRDPEYWK-DAESFIP
LsCYP71BL2 347 IVKETLRLHAPTPLLVPRECRQACNVDGYDIPAKTKILVNAWACGTDPDSWK-DAESFIP
CcCYP71BL5 348 IVKETLRLHAPTPLLVPRECRQDCNVDGYDIPAKTKIIVNAWACGTDPDSWE-NPESFIP
TgCYP76AE2 365 IVEETLRLHPPAPLTLPRKAMHDTKFMGYNIPKDTQVFVNAWAIGREKENWE-DALSFKP
VoCYP71D442 353 VIKEVLRMHPPVAFLLPRECIEETQINGYTIPIKTKVLVNVWAMGRDPQHWK-DPDSFYP

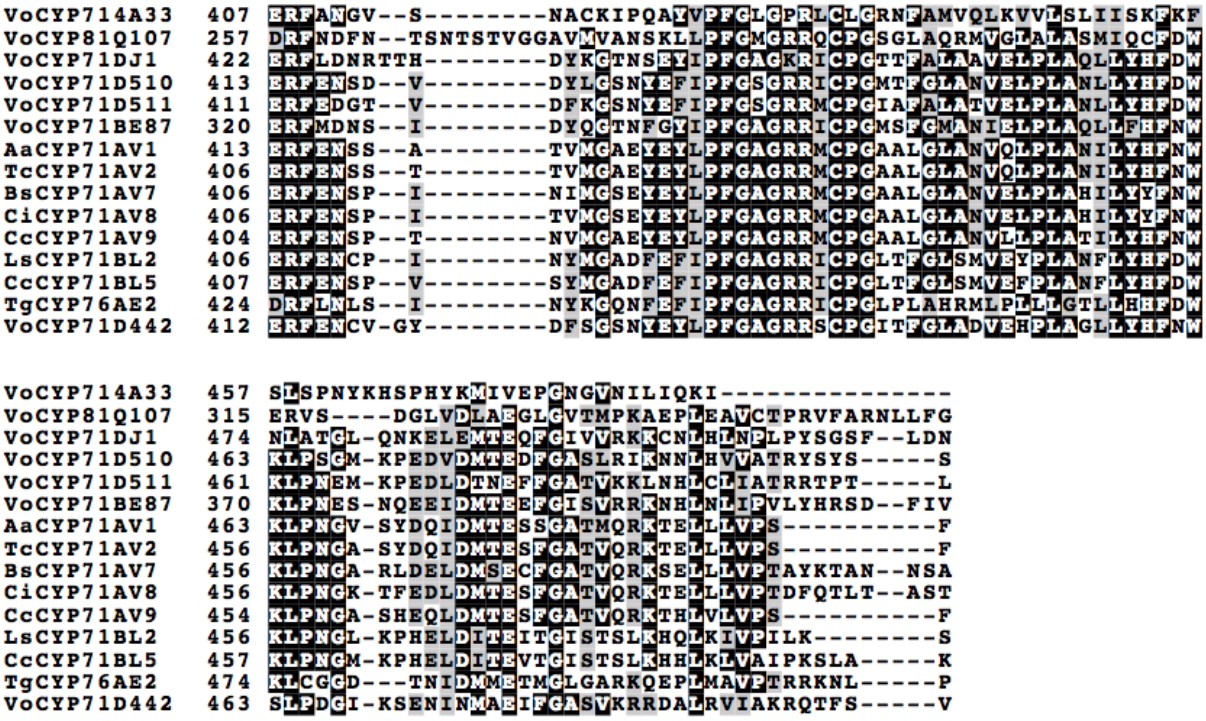

VoCYP714A33 407 ERFANGV--S--------NACKIPQALVPFGLGPRLCLGRNFAMVQLKVVLSLIISKFKF
VoCYP81Q107 257 DRFNDFN--TSNTSTVGGAVMVANSKLLPFGMGRRQCPGSGLAQRMVGLALASMIQCFDW
VoCYP71DJ1 422 ERFLDNRTTH--------DYKGTNSEYIPFGAGKRICPGTTFALAAVELPLAQLLYHFDW
VoCYP71D510 413 ERFENSD--V--------DYLGSNYEFIPFGSGRRICPGMTFGLANVELPLANLLYHFDW
VoCYP71D511 411 ERFEDGT--V--------DFKGSNYEFIPFGSGRRMCPGIAFALATVELPLANLLYHFDW
VoCYP71BE87 320 ERFMDNS--I--------DYQGTNFGYIPFGAGRRICPGMSFGMANIELPLAQLLFHFNW
AaCYP71AV1 413 ERFENSS--A--------TVMGAEYEYLPFGAGRRMCPGAALGLANVQLPLANILYHFNW
TcCYP71AV2 406 ERFENSS--T--------TVMGAEYEYLPFGAGRRMCPGAALGLANVQLPLANILYHFNW
BsCYP71AV7 406 ERFENSP--I--------NIMGSEYEYLPFGAGRRMCPGAALGLANVELPLAHILYYFNW
CiCYP71AV8 406 ERFENSP--I--------TVMGSEYEYLPFGAGRRMCPGAALGLANVELPLAHILYYFNW
CcCYP71AV9 404 ERFENSP--T--------NVMGAEYEYLPFGAGRRMCPGAALGLANVLLPLATILYHFNW
LsCYP71BL2 406 ERFENCP--I--------NYMGADFEFIPFGAGRRICPGLTFGLSMVEYPLANFLYHFDW
CcCYP71BL5 407 ERFENSP--V--------SYMGADFEFIPFGAGRRICPGLTFGLSMVEFPLANFLYHFDW
TgCYP76AE2 424 DRFLNLS--I--------NYKGQNFEFIPFGAGRRICPGLPLAHRMLPLLLGTLLHHFDW
VoCYP71D442 412 ERFENCV-GY--------DFSGSNYEYLPFGAGRRSCPGITFGLADVEHPLAGLLYHFNW
VoCYP714A33 457 SLSPNYKHSPHYKMIVEPGNGVNILIQKI----------
VoCYP81Q107 315 ERVS----DGLVDLAEGLGVTMPKAEPLEAVCTPRVFARNLLFG
VoCYP71DJ1 474 NLATGL-QNKELEMTEQFGIVVRKKCNLHLNPLPYSGSF--LDN
VoCYP71D510 463 KLPSGM-KPEDVDMTEDFGASLRIKNNLHVVATRYSYS-----S
VoCYP71D511 461 KLPNEM-KPEDLDTNEFFGATVKKLNHLCLIATRRTPT-----L
VoCYP71BE87 370 KLPNES-NQEEIDMTEEFGISVRRKNHLNLIPVLYHRSD--FIV
AaCYP71AV1 463 KLPNGV-SYDQIDMTESSGATMQRKTELLLVPS----------F
TcCYP71AV2 456 KLPNGA-SYDQIDMTESFGATVQRKTELLLVPS----------F
BsCYP71AV7 456 KLPNGA-RLDELDMSECFGATVQRKSELLLVPTAYKTAN--NSA
CiCYP71AV8 456 KLPNGK-TFEDLDMTESFGATVQRKTELLLVPTDFQTLT--AST
CcCYP71AV9 454 KLPNGA-SHEQLDMTESFGATVQRKTHLVLVPS----------F
LsCYP71BL2 456 KLPNGL-KPHELDITEITGISTSLKHQLKIVPILK--------S
CcCYP71BL5 457 KLPNGM-KPHELDITEVTGISTSLKHHLKLVAIPKSLA-----K
TgCYP76AE2 474 KLCGGD---TNIDMMETMGLGARKQEPLMAVPTRRKNL-----P
VoCYP71D442 463 SLPDGI-KSENINMAEIFGASVKRRDALRVIAKRQTFS-----V

Figure S4.4. Heatmap of *Valeriana officinalis* dehydrogenase expression. AaADH1 and AaALDH1 were BLASTed against the *V. officinalis* transcriptome. Expression profiles for hits were analyzed in Excel. Select dehydrogenases are highlighted in yellow; VoADH1 = voa_locus_40753_iso_1_len_634_ver_2, VoALDH1 = voa_locus_940_iso_1_len_1560_ver_2. VoADH1 and VoALDH1 were selected on basis of homology for use in this work.

V. officinalis protein isoform	Homology	Leaf	Stem	Root	Callus
voa_locus_10192_iso_3_len_424_ver_2	Alcohol dehydrogenase	2.38	2.29	3.78	4.18
voa_locus_12186_iso_2_len_408_ver_2	Alcohol dehydrogenase	7.78	9.91	10.02	9.84
voa_locus_1256_iso_3_len_1472_ver_2	Alcohol dehydrogenase	6.71	6.45	7.94	6.74
voa_locus_13569_iso_1_len_711_ver_2	Alcohol dehydrogenase	0.00	0.94	4.10	0.00
voa_locus_17676_iso_1_len_293_ver_2	Alcohol dehydrogenase	6.41	6.43	5.77	6.29
voa_locus_2011_iso_4_len_1194_ver_2	Alcohol dehydrogenase	6.26	5.57	4.70	5.69
voa_locus_2246_iso_2_len_436_ver_2	Alcohol dehydrogenase	8.06	9.87	9.99	9.83
voa_locus_27773_iso_3_len_1072_ver_2	Alcohol dehydrogenase	5.76	6.45	7.98	9.17
voa_locus_36439_iso_1_len_1092_ver_2	Alcohol dehydrogenase	2.01	-0.47	2.25	0.00
voa_locus_40753_iso_1_len_634_ver_2	Alcohol dehydrogenase	3.08	1.03	0.00	2.22
voa_locus_43224_iso_1_len_360_ver_2	Alcohol dehydrogenase	2.71	3.55	3.29	2.92
voa_locus_46353_iso_1_len_959_ver_2	Alcohol dehydrogenase	0.00	0.49	4.23	2.21
voa_locus_51835_iso_1_len_297_ver_2	Alcohol dehydrogenase	0.00	0.00	4.08	0.00
voa_locus_52514_iso_1_len_595_ver_2	Alcohol dehydrogenase	2.24	0.00	2.27	1.07
voa_locus_57703_iso_1_len_328_ver_2	Alcohol dehydrogenase	0.00	0.00	1.62	2.65
voa_locus_5914_iso_5_len_1585_ver_2	Alcohol dehydrogenase	6.42	6.49	6.48	5.97
voa_locus_70032_iso_1_len_457_ver_2	Alcohol dehydrogenase	0.00	0.00	1.42	1.65
voa_locus_76277_iso_1_len_315_ver_2	Alcohol dehydrogenase	0.00	3.74	0.00	2.07
voa_locus_8119_iso_7_len_2405_ver_2	Alcohol dehydrogenase	5.43	5.75	5.43	5.51
voa_locus_8280_iso_5_len_854_ver_2	Alcohol dehydrogenase	6.56	7.77	8.55	7.26
voa_locus_86433_iso_1_len_315_ver_2	Alcohol dehydrogenase	0.00	0.00	3.23	0.00
voa_locus_87007_iso_1_len_474_ver_2	Alcohol dehydrogenase	0.00	0.00	3.29	0.00
voa_locus_940_iso_1_len_1560_ver_2	Aldehyde dehydrogenase	2.82	4.51	4.47	3.68
voa_locus_10288_iso_1_len_1603_ver_2	Aldehyde dehydrogenase	6.11	6.38	5.83	6.50
voa_locus_1044_iso_2_len_447_ver_2	Aldehyde dehydrogenase	6.90	7.83	7.54	7.64
voa_locus_13329_iso_1_len_500_ver_2	Aldehyde dehydrogenase	4.36	4.25	3.30	0.00
voa_locus_14000_iso_1_len_854_ver_2	Aldehyde dehydrogenase	4.76	4.84	3.78	1.33
voa_locus_15015_iso_1_len_387_ver_2	Aldehyde dehydrogenase	4.99	4.70	3.58	1.20
voa_locus_15837_iso_2_len_1736_ver_2	Aldehyde dehydrogenase	3.72	4.61	3.76	4.22
voa_locus_1633_iso_2_len_561_ver_2	Aldehyde dehydrogenase	6.42	6.51	6.46	5.82
voa_locus_28555_iso_1_len_1686_ver_2	Aldehyde dehydrogenase	4.10	4.21	5.26	4.53
voa_locus_40187_iso_1_len_447_ver_2	Aldehyde dehydrogenase	2.68	3.06	3.90	3.54
voa_locus_43657_iso_1_len_1034_ver_2	Aldehyde dehydrogenase	2.75	2.75	2.97	3.08
voa_locus_46346_iso_1_len_470_ver_2	Aldehyde dehydrogenase	4.33	4.68	4.91	5.34
voa_locus_7635_iso_4_len_1785_ver_2	Aldehyde dehydrogenase	4.49	5.22	5.25	5.43
voa_locus_2461_iso_6_len_3831_ver_2	Aldehyde dehydrogenase fami	6.28	6.43	6.67	6.23
voa_locus_87178_iso_1_len_326_ver_2	Aldehyde dehydrogenase MIS	0.00	0.00	0.00	0.00

Figure S4.5. VoCYP71DJ1 produces oxidized valerenadiene. GC-MS chromatograms of yeast extracts of JWy608 and JWy614. The addition of VoCYP71DJ1 produces a new peak (Peak 1), likely valerenic alcohol, with a mass of 220m/z (RT = 7.8 min).

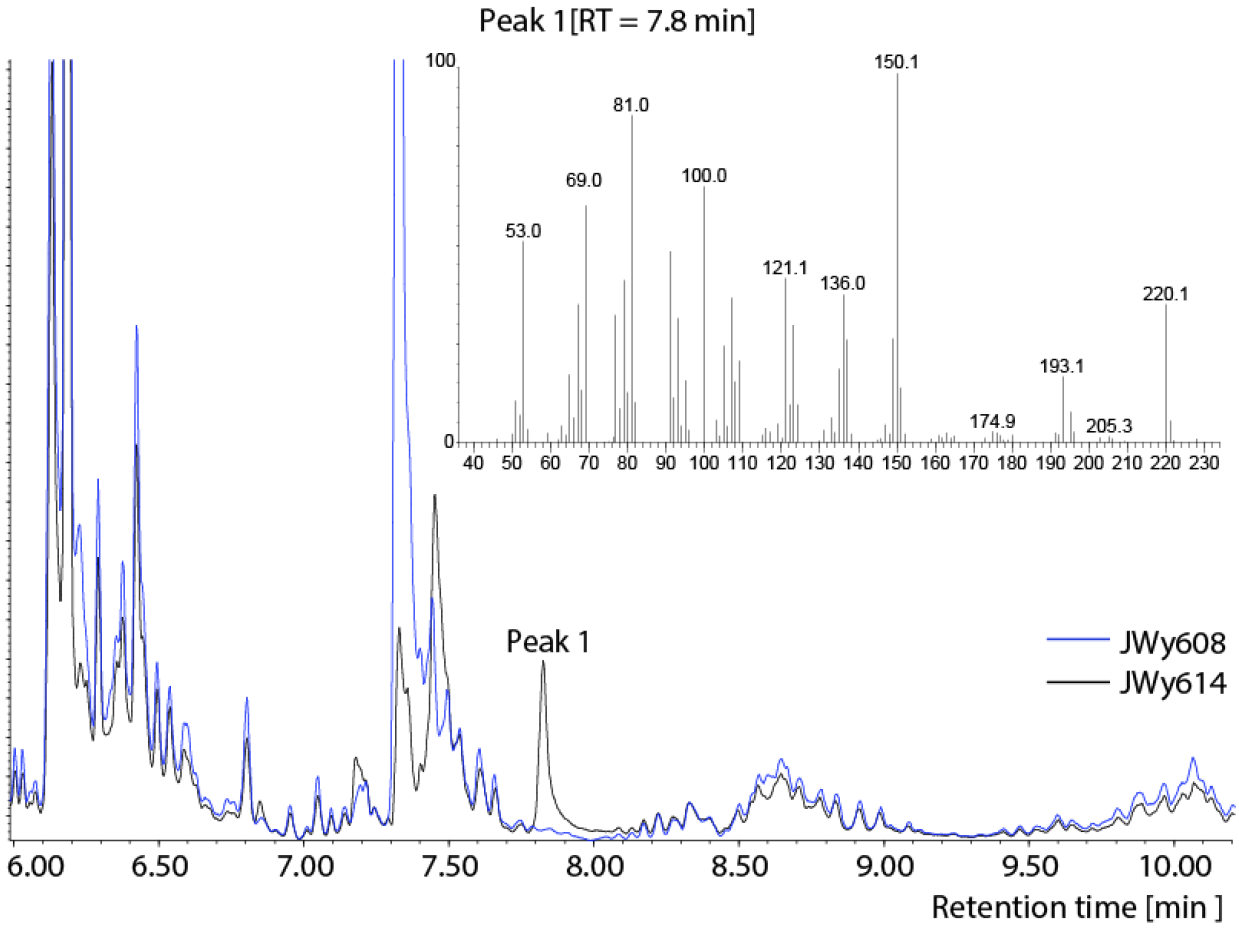


Figure S4.6. Expression of *V. officinalis* P450 candidates in VoCYP71DJ1 expressing strain, JWy615. Valerenic acid levels were determined [arbitrary units (AU] and normalized to internal standard trans-caryophyllene standard. Evaluation of the data shows expression of P450 candidates with VoCYP71DJ1 did not increase valerenic acid production. Data represent the averages of three replicate cultures; error bars show s.d.

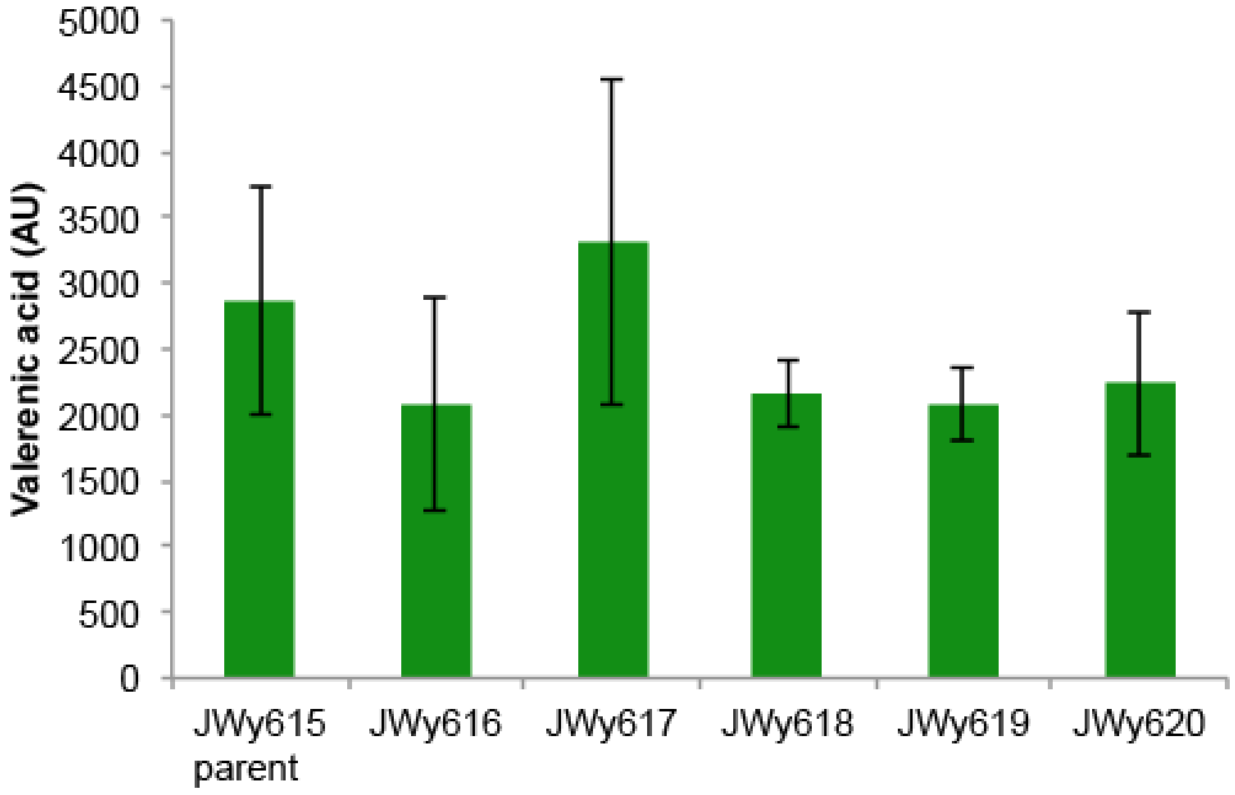
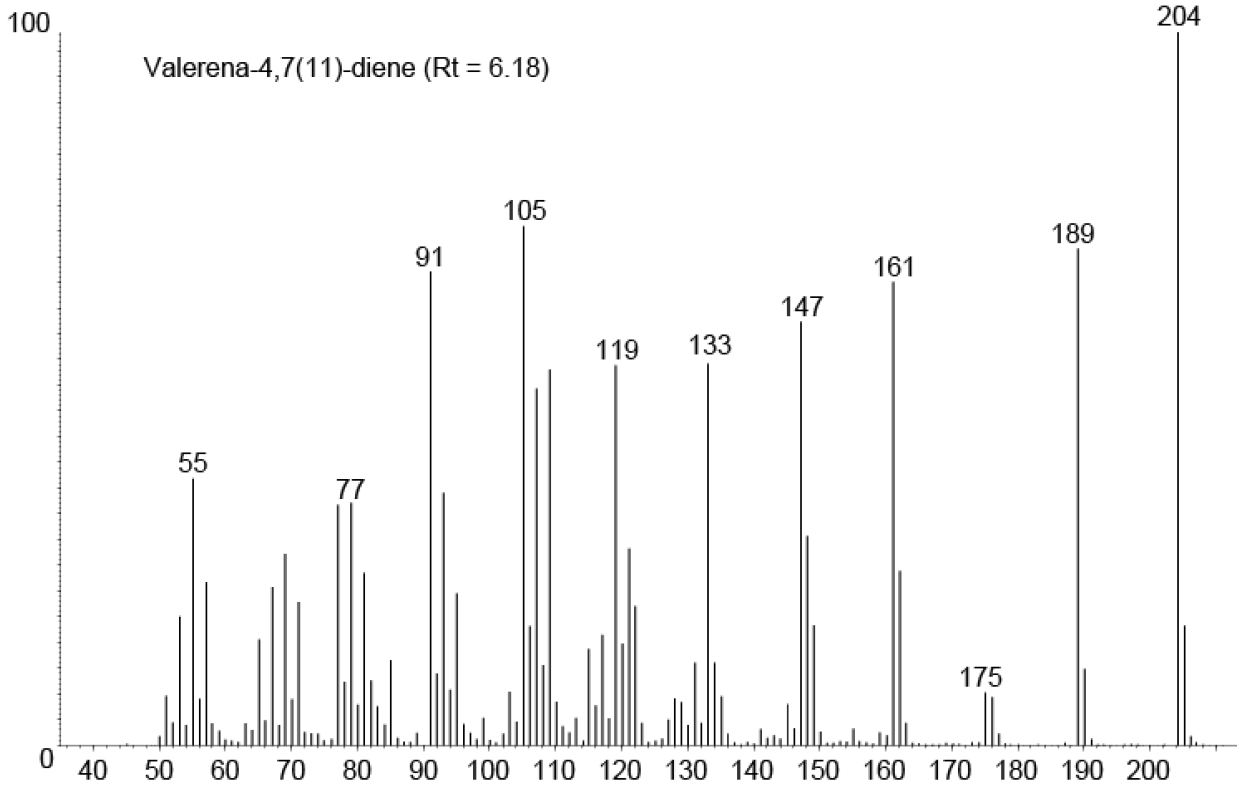


Figure S4.7. GCMS spectra of (A) valerena-4,7(11)-diene from yeast culture and (B) underivatized valerenic acid standard. Underivatized valerenic acid was difficult to detect at low concentrations, and exhibited broad peaks when analyzed on GC.



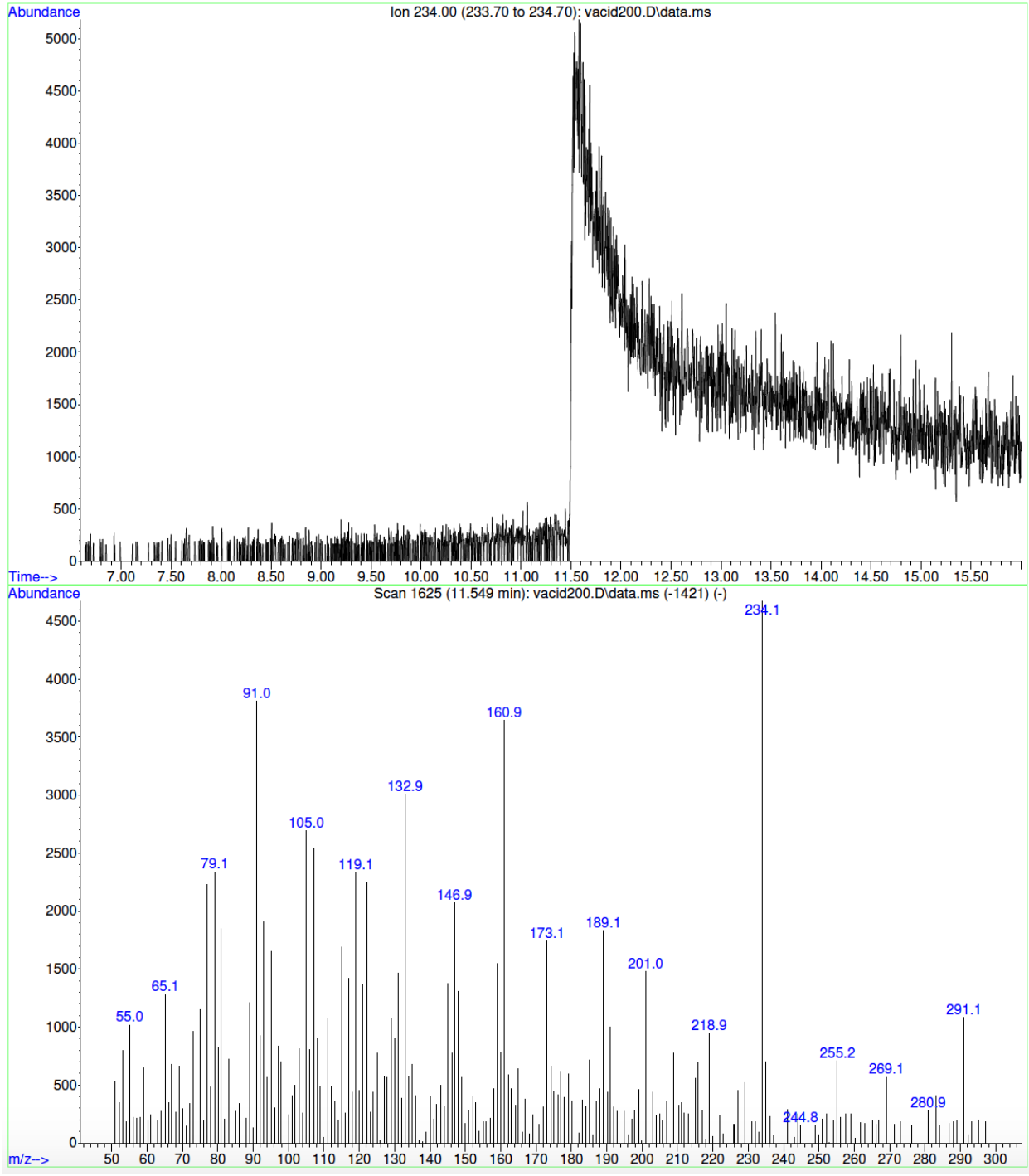

Abundance
Ion 234.00 (233.70 to 234.70): vacid200.D\data.ms
Time-->
Abundance
Scan 1625 (11.549 min): vacid200.D\data.ms (-1421) (-)
234.1
91.0
160.9
132.9
105.0
79.1
119.1
146.9
173.1
189.1
201.0
65.1
55.0
218.9
291.1
255.2
269.1
280.9
244.8
m/z-->

Figure S4.8. Sequences of synthetic genes used in this study.

>VDS

ATGTCTGAGAGCTGTCTTTCATTTTCAAGTCCACCTCCCACAAAGAAAAATATTCAA
GAGCCCGTAAGGCCGAATGCAAAGTTTCACAAAAGTGTCTGGGGCAACCACTTTTTA
AAGTACGCCTCTAATCCAGAACAAATCGACTACGACGCTGACGAACAACATGAACA
ATTGAAGGAGGAGCTGAGGAAAAAGCTGGTGGTGAATGTAACTAACGAAAGGGTG
GAGGAACAACTAAAGCTTATAGACGCTATCCAGCGTTTGGGCGTAGCATATCATTTT
CAACGTGAGATCGATGCAGTCCTGAATAATCTGTTACTATTCAGGTCCAACAAGGAC
TCTGATGATATTTATATGGTGTCTCTGAGATTCAGACTTTTAAGACAACAAGGGCAC
AATGTTTCCTGCTCAGTATTTGAGAAATTTAAGAATATAGATGGCAGATTCAAGGAC
TCCCTAAGAGATGATGTAAGAGGGTTGCTGTCACTATATGAGGCAACCCATATGAG
GGTTCACAAAGAGGATATCTTAGAAGAAGCGTTAGAATTTACCATCTATGAACTGG
AGCAAGTTGTTAAGTTGTCCTCTAACGACACCTTACTTGCTTCCGAGGTGATTCATGC
GCTGAATATGCCTATCCGTAAGGGGTTGACGAGGATTGAAGCGAGGCACTTTATCA
GTGTTTATCAACATGACAAAAGCCATGATGAGACCCTGCTTAAATTTTCTAAAATTG
ACTTTAACATGCTGCAAAAGTTGCATCAACGTGAGTTGGCAGACTTAACTATCTGGT
GGGAAAAATTGAACGTGGCTGAAAAGATGCCATACGCGAGAGATAGATTCGTGGAA
TGTTATTTTTGGGGGCTTGGTGTATACTTTGAGCCTCAATATTCCAGAGCTCGTAAAA
TGTTTGTAAAGGTGATAAATTTAACTTCCCTTATTGATGACACGTATGATTCATATGG
TACATTCGATGAATTGGATTTATTCACAGATGCGGTTAAGAGGTGGAACGTAAATGA
AACTGACAAATTGCCGGAGTATATGCGTCCATTGTTTATGGAGCTTCTTAATGTTTAT
AACGCGATGGAGGAGGAGCTAAAGGAGGAGGGGGTTAGCTACAGAGTCGAGTACG
CAAAACAATCTATGATACAGATAGTGACTGCATACAACGATGAGGCTATCTGGTAC
CACAACGGTTATGTCCCTACGTTTGACGAATATCTTAAAGTTGCCCTGATCAGCTGC
GGCTACATGTTATTGTCCACTATAAGCTTTGTCGGCATGGGGGTCACCACAGTCACG
AAACCCGCCTTTGACTGGGTAACAAATAACCCACTAATCCTAATTGCCTCATGTACA
ATTAACAGACTGGCGGACGACAAGGTGGGGCACGAGCTAGAGCAGGAGAGAGGGC
ATGTTGCTTCAGGGGTGGAGTGCTATATGAAACATAATAATGCAACGAAACAGGAA
GTGGTAATCGAATTTAACAAGAGAATTAGTAACGCTTGGAAAGACATCAATCAAGA
GTGCTTGCACCCACTACCAGTACCGCTTCACTTGGTAGTCCGTCCCTTATATCTAGCG
TGCTTCATGAACGTATTTTATAAAGATGAAGATTGGTATACCCATAGCAACACGCAG
ATGAAAGAGTGCATCAATTCTTTGTTAGTTGAAAGCGTCCCTTATTAA

>VoCYP71DJ1

ATGATTATGGAGGATCTTAACTTCTCAATTATACACCCAATACTACTAATTTTCGTGG
CATTCGTCATTATTGAGAACGTGAGAAAAAACAAAAAGGCGGTACGTAGACCGCCC
GGCCCGTGGCAATTCCCGCTAATTGGAAATATGCACAATTTATTTGGTTCATCCCCG
CACAGAGTACTGCGTGATCTAAGTAATAAATACGGGCCCATCATGTTGCTGCGTTTA
GGGACAGTACCTACACTTGTGGTATCCTCTGCTGAGTTAGCTGAAGAGATACTGAAA

ATCAGAGGTGTTGAGTTTGCGGATAGGCCACACATTCTAGCAGCTGACATAGTCATC
TATAATAGCACCGACATACTGTTTTCACCGTATGGCGATCATTGGCGTCAAATGAGG
AAGGTCTGTGCAATGGAGCTTCTTAGTACTAGAAAAGTACAGTCCTTCCGTTCTATC
AGGGAAGAGGAGGTACTAAATTTACTTCAACTAATATCTTCATCTTGTGCAAGCGGG
TCCGCGTTGAACTTGACTAAACTACTATTTTCCTTTACTTACACAGTTATCACACGTG
TGACGTTTGGTGAGAAGTGGACCGAACAACCCGAAAAATTCAGCTCCCTTTTAAGTG
AACTAGTCGTGCTTTTTTCTGGCTTTAACATAGCTGATATGTACCCGAGCGTCAAGTT
TATTCAAGGCGCGGGAGGGTTCAGGGCGAGAGCAGAGAAGGTCCACCAAAGGATG
GATGAGACGTTCACCAACATAATCATGAAACAACGTGAAAAAAACAAGGCAACAA
GTACCGGAGAGTCCAACCAGGAACACCTGATTAACGTTCTATTACGTGTTCAGAAAC
ATGCAGCTGGAACCGAGAATCCATTCACTGATGACAGCATCAAAGGCGTTCTACTA
GATATTTTTAATGGTGGCAGTGAAACTAGCTCCACTACGATGGAGTGGGCGATGGCT
GAGTTGATTAAAAACCCAAGGGCAATGGAACGTGCGCAAACAGAGTTAAGACAAGC
TTTCAGCGGGAAAGGAAATGTAGAGGAAACAGGACTTGATAAGCTGAAGTATTTTC
ACTGTATTATAAAAGAGACAATGAGATTGCACCCTCCGTTTCCACTGATGGTCCCAC
GTCAGAATAGGCATGAATGCGAGATAAATGGGTACATCATTCCGGCCAAGACTAAA
GTATTAGTGAACGGATGGGCTATCAGTAGAAACCCCAAATACTGGGGACCCGACGC
CGACGTCTTTAAGCCGGAGAGATTCCTGGATAACAGGACTACCCATGACTACAAAG
GGACCAACTCTGAATATATACCCTTCGGCGCTGGAAAGAGGATCTGTCCAGGGACT
ACATTCGCTTTGGCGGCTGTAGAGCTACCCCTGGCGCAGCTTTTATACCATTTCGACT
GGAATTTAGCAACTGGCCTACAGAATAAAGAGCTGGAAATGACCGAACAATTCGGC
ATTGTGGTAAGGAAGAAATGCAACTTACACCTAAACCCCCTACCTTACAGCGGTTCC
TTCCTAGATAACTAA

>VoCYP71D510

ATGGACTCATTTACCATAATCCTTATCAACGTTGTACCAGTTTTACTGATTTTCCTAC
TGTTTCGTAGATGGAAAAGTGCTAAAGCTGTCAATCTTCCGCCAGGTCCCCCCAAGC
TACCATTGATAGGCAGTCTTCTGCACATGGGTCAGTTGCCGCATAGAAGCTTGAAGG
AGCTGGCGGGAAAGTACGGACCTCTAATGCACATTCAGTTAGGAGAAATTAGCGCA
ATTGTAGTGTCATCTCCAAGAGTTGCTAAACTAGTGACAAAAACGCACGACCTAAGC
TTTGCGTCAAGACCAGTTATCTTGGCAAGTGAGATAGTTGGATATCATAACACTGAC
ATCGCATTTGCCCCATATGGTGACTACTGGCGTCAAATGCGTAAAATAGCCACGTTA
GAGTTATTGTCTGCGAAAAAAGTGAGGTCCTTCTGTTCTATCAGAGAGGAGGAGGC
AAAAAATTTGATTGAATCTATACATTCTACCAGTGGGAAGTCATTCGATCTAACCGA
AAAGGTTTTCTCACTTACAAACAACGTAATCTGCAGAGCGACCTTTGGCGATAGGTA
TAAGGACCAAGATTATTTGATAAAGATCCTTAAACAAGTAGTGAATTTAGCTGGAG
GGTTTGACGTAGCTGACCTATTCCCCAGCTTGAAGCTACTTCACTTAGCGACAGGAA
TGAGGCCAAAACTTGAGAATCTAAGGAAGGATCTAGATAGAATATTTGATGAAATT
ATCAATGAGCGTACCAAGAGGTTGAAAAACGGCACAAATACGCATGAAGACAACG
AAGATATCGTTGATGTTCTAGTTAGGCTGAAGGAATCCGGGGCGCTAGAGTTCCCTA

TCACCCAAAACAACATTAAGGCGGTAATACTAGATATGTTCTTAGCCGGATCCGACA
CGTCCTCAACAACTATCGAATGGGCTATGGCAGAAATGATGCGTAATCCAAGAGTTT
TGGAAAAGGCACAGGCGGAGCTTCGTCAGGCTATGAACGGTAAAAAGGTAATAGAA
GAATCAGATATTAAGGAAACAGGGTCATACTTCAAGTTAGTAATTAAAGAGACATT
GAGAATGCATCCGCCTGTCGCGCTTCTTCTACCAAGGGAATGTCGTGAGGAGTGCGA
AATCGATGGCTACACCATCCCAGTGAAGACAAAGGTAATGGTCAACGCTTGGGCGA
TTGGGAGGGATCCTGAGTATTGGAAAGACGCTGACAGCTTTTATCCCGAGCGTTTCG
AGAACTCTGATGTAGACTATTTGGGGAGTAACTACGAATTTATCCCCTTTGGAAGTG
GGCGTAGGATATGTCCTGGCATGACCTTTGGACTGGCGAACGTCGAGTTGCCACTAG
CGAATTTGCTTTACCATTTCGATTGGAAACTGCCCTCAGGTATGAAACCGGAAGATG
TGGACATGACCGAGGACTTTGGGGCTTCTCTTAGAATAAAAAACAACCTTCACGTTG
TGGCGACGAGGTACAGCTACTCCTCTTAG

>VoCYP71D511

ATGGACTTCATTGCCATAGCTCTTCCCTCAGTGGCGTTCCTGTTCTTTTTGTTGAAAT
TAGTCAGGAAATTGCGTTCCCCGAAAAGATTACCACCTGGGCCTTGGAAACTGCCAT
TGATCGGGTCCCTTTTGCATATGGCCGGACCATTACCACACAGAACATTGAAAGATC
TAGCTGAGAAGTACGGCCCTTTAATGCATTTACAATTAGGAGAGATCTCTGCAATTG
TGGTCTCTAGCCCGGAGATGGTAAATGAATTTATGAAAACACACGATATCGCTTTTG
CTAGTAGGCCGCCCGTTTTAGCGATCGAGATTGTTGCTTACAATAGAGATGATATAG
CGTTCGCACCATATGGTGATTACTGGAGACAAATGCGTAAGATCGCTACATTAGAAC
TACTTAGTGTAAAGAAGGTGGGTAGCTTCTCTTCTATCCGTGAGAAAGAAGTGCACA
ATCTTGTTGAGTCAATCTCCAGCTCAGGATCAATCATCCCGATAGACATGACAGAAA
AGTTATTTGGTCTTATTAGCAGCGTCGCAGCTAGGGCATCATTCGGTAACAAATGTA
AAGACCAAGACAGTTTCCTAGAGTTAACAAATGAGATTATCTCTTTAGCAGGTGGAT
TTAACATTTTCGACCTTTTTCCGTCTTTTAAACTTCTGCACAGGCTGACAGGAATGAG
GCAGAAATTTGAGATGATGCATCAAAAGGTTGACCAAGTTTTTGAGAACATTATTAA
AGATCATATACAAGAAAGAGCCGATGATACCCACGGGAACGACCACACGGAAGAT
CTATTAGACGTTCTGCTACGTCTAAAGGACGAAGGCCTTGAGTTTCCTATAACTTAC
ACCAACGTCAAAGCGGTTATACTTAATGCTTTCTCCGGAGGCTCCGACACGTCCAGC
ACCACTATAGAGTGGGCGATGACGGAGTTGATGAGGAATCCTCGTGTAATGGAAAA
AGCGCAAGCAGACTTAAGAGAAGCACTGAAGGGAAAACAGGTAGTCAACGAGAAT
GATATTAAAGACTTACCGTATTTGAAGTTGGTTATGAAAGAAACCATGCGTCTTCAT
ACACCTCTACCACTACTGGTGCCTAGAGAGTGCCGTCAGGAAGTAGAAATTGACGG
GTACACAATAACAGTAGGTACTAAGATCATTATAAATGCTTGGGCTATTGCGAGGG
ACCCCCAGTATTGGAAAGATCCGGAATCTTTTTACCCGGAGAGGTTCGAGGACGGTA
CGGTGGACTTTAAAGGTTCTAACTACGAGTTTATCCCGTTCGGCTCTGGAAGAAGGA
TGTGCCCAGGCATTGCTTTTGCACTAGCTACAGTCGAGTTACCACTGGCCAATTTACT
ATATCATTTCGATTGGAAGCTACCAAACGAAATGAAACCCGAGGACTTGGACACTA

ACGAATTTTTCGGCGCAACAGTCAAAAAACTTAACCACCTGTGTCTTATCGCCACTA
GGAGGACTCCCACTCTATAA

>VoCYP81Q107

ATGCTCCGAAAATTGTACGATGGTGGTGAGTCATCATCTAAGGTTGAATTGAAAACC
ACACTCTCTGAGCTAACATTTAACATCATATTGAGGATGATTGCCGGAAAAAGATAT
TTTGGGGAAGATGTGAAAGAGAACGAGGAGGCTGTTCGGTTTAGAAGTTTGATAAA
GGAGATTGTTAAGTACGGCGGAGCATCGAATCCCGGAGATTTCTTGCCGATTCTTGG
GTGGTTTGATTATGGCGGGTTTCAGAAGAATCTGACGAGGATCGGTAAGCAGATGG
ACGGTTTGTTGCAGGGATTAATCGAAGAGCATAGACGTGAGAAGAATAAAAATACG
ATGGTGGATCATCTTCTTTCCCTTCAAGAATCAGAACCGGAATATTACACCGATGAA
ATCATCAAAGGCCTAATGATTGTGATGGTAACCGCCGGAACAGACACATCGTCAGT
GACAGTCGAATGGGCAATGTCGCTATTACTCAACCACCCCGAAATACTAAAGAAAG
CACGAGCCGAAATCGATAAAGAAGTCGGAGAAAGCCGTTTAGTAGACGAACCGGAT
CTCCCAAAACTACCTTATCTCCAAAACATAATTCTCGAAACACTCAGAATGTTCCCA
TCGGCACCACTACTAATCCCGCACGAGTCGTCAGAGGATTTCAAGCTTGGAGAATAC
GATGTACCAAAAGGAACAATCGTTTTAATCAATGCATGGGCTATACATCGGGATCCG
AACGTATGGGATGATCCGACGAGTTTTAATCCCGATAGATTCAATGATTTTAATACT
AGTAATACTAGTACTGTCGGTGGAGCGGTAATGGTGGCGAATAGTAAGTTGTTACCT
TTTGGGATGGGACGGAGGCAGTGCCCAGGATCGGGTTTGGCTCAACGGATGGTCGG
GTTAGCGTTGGCTTCGATGATACAGTGTTTTGATTGGGAAAGAGTTAGTGACGGGTT
AGTCGATTTAGCCGAAGGACTCGGAGTTACAATGCCAAAAGCTGAGCCGCTCGAGG
CTGTGTGTACACCGCGGGTGTTCGCGCGGAATCTTCTATTTGGATGA

>VoCYP714A33

ATGAGATTGAAGCTAACAAGACAAGGCATAAAAGGTCCAAAGCCTCATTTTATGTA
TGGAAATGTTCCTCAGATGCAGAAAATCCAATCTGCCGCTGTAGAATCTGGTAGCTG
CAACCATGGCGAAATCATCGCGCACGATTACACTTGTGCCCTCTTCCCATATTTCGA
ACAATGGAGAAAACAATACGGTTTAGTATACACATACTCAACTGGGAACAAACAGC
ATTTGTATATAACCAAAGCCGAATTAGTCAAGGAGATGAATCAGTCTGGATTAGGG
AAGCCTTCTTATATTACCAAGAGACTAGCTCCTCTGCTTGGCAATGGCATTTTAAGA
TCAAATGGCCATCTCTGGGCCCAACAGAGAAAAATTGTTGCCCCTGAATTCTTCATG
GATAAAGTCAAGGGTATGCTGGGTCTGATGTTGGAATCAACACAGCCACTGATTAA
GAAATGGGAAGAATCAATTGAAAGCCAAGGTGGAAAAATAGCTGAGATTAGAATTG
ATCAAGATTTCAGGGGCGTCTCTGCTGACGTCATCTCAAGAACTTGCTTTGGAACTT
CTTACTCCAAAGGCAAACTCATTTTCTCCAAGCTTAGAACTCTTCAACACACCTTCTC
TTCTGGAGGTTTCCTTTTTACTCTTCCTACATTCGGATTTCTTGCGAGAAAGAACCAC
AAGGAGATCAAGAATTTGGAGAAAGAGATAGACACGTTAATCTGGGATGCAGTTAA
AGAACGGCAAAGAGAGTGTTTAGAGAAATCATCTTCAGAGAAGGATCTTTTACAGA
TGTTATTAGAAGGAGCCATGAATGATGAATGTCTAGGAGCAGAATCGTCAAAATCA

TTCATCGTTGATAATTGCAAAAACATTTATTTCGCCGGCCATGAAGCAACCGCCGTC
GCAGCTTCATGGTCCATAATGTTGCTTGCTTTGCATCCAGAATGGCAGTCTCACATTC
GAGAAGAAATGTCTCAAGTTTCCAATAATGGAATCCTAGATTCAGATTCCCTATCCA
AAATGAAAACTGTAACGATGGTGATCCAAGAAGCGTTGCGATTATATCCACCAGCA
GCATTTGTGTCGAGAGAAGCATTTGCAGAAACAAAAATAGGAAATATTAAGATTCC
AAAAGGGGTATGCATATGGACATTAATACCCACGCTGCATCGTGATCCTGATAATTG
GGGATTAGATTCTAATGAATTTAAACCAGAGAGATTCGCTAATGGAGTATCAAATGC
TTGCAAGATTCCACAAGCTTACGTTCCATTTGGGCTTGGTCCTAGGCTGTGTTTAGGG
CGAAATTTCGCTATGGTTCAGCTGAAAGTTGTTTTATCTCTAATCATATCCAAGTTCA
AATTCTCCTTGTCGCCGAATTACAAGCACTCGCCTCATTATAAAATGATCGTCGAAC
CTGGGAATGGAGTCAATATCTTGATTCAAAAGATTTAA

>VoCYP71BE87

ATGGTATGGTTTTTACTGAAAGACAGGACTTTCCTATTGCAGAAATTTTTGTGGTATG
ATAGTACTTCTATCGCCTTCAGCTCTTACGGGGACTACTGGCGTCAGCTAAGGAAGA
TTTGTACGATGAACCTATTGAGTACAAAGCGTGTGGAACAGCTGAGAAGCATTAGG
GAGGAGGAAGCTTTGAACCTAGTTAGGAGAATATCCACAAACGGCGATTCTTTGCC
ATTTAATCTAAGCAAGGCTATTTTTAATTTGACCAGCACGGTCACGAGTCGTGCGGC
GTTCGGGAATAAAAACAAAGACCAAGAAGAATTCGAGGTAGTTCTGGACCAAGTTC
TGAAGGCGCTGGGCGGTTTTAACATAGGCGACATGTACCCGAAAGCTAAATTACTTC
ATAAGATAACCGGGGCGCGTGCCTCCATGAACAAGATACAGAAGAGAGTGGATAGA
ATTTTACAAAACATCTTAGTCGACCACCGTAATAGGAAGCAGGAATCTTTAACAGAT
GATTACGAGGATCTAGTGGACGTACTACTTAGAATCCAAAATGAAGACCAACTTCA
ATTCCCAGTAACTGATAACTGTATAAAAGCAGTCATCTTAGACGTGTTCGGTGGTGG
GAGCGAAACAAGTAGCGCTGCTACTGAGTGGGCAATGAGTGAAATGGTGAAGAATC
CCCACATCATGAAAAGGGCTCAAGCTGAAGTTCGTAAAGTTTTCGATGAGAAAAGA
AATGTAGATGAGACCGGCTTGGGGGAGCTTAAATATCTGCAGTGTGTAATAAAAGA
GACCCTGAGGCTGCATCCACCGTTGCCATTACTGGTTCCCAGAGAGAATAGTGCCGA
GTGTGAAGTTAACGGATTTCTGATTCCAGCGAACTGTAAAGTTATCATAAATGCGTG
GGCGATCTCCCGTGATCCTAAGTACTGGGTGGATGCAGAGATTTTCAAACCCGAAAG
ATTTATGGATAATAGTATTGATTACCAGGGGACGAACTTTGGATACATACCATTCGG
CGCTGGTAGGAGGATTTGTCCGGGAATGTCCTTCGGTATGGCGAACATCGAACTTCC
CCTGGCCCAGCTATTATTCCACTTCAACTGGAAGCTACCTAACGAGTCAAATCAGGA
GGAGATCGACATGACAGAAGAGTTCGGAATCTCAGTAAGGCGTAAAAACCATCTAA
ACTTGATCCCCGTCCTGTACCATAGATCAGACTTTATAGTGTAA

>VoALDH1

ATGGCCGGGGATAGCAACGGATCTCTAGATAGTTTTGTCAAGATCCCTGACATTAAA
TTTACTAAATTATTTATAAACGGTGAATTTATTGACTCAATTAGCGGATCAACATTTG
AAACTATTGACCCCCGTACAGGAGAAGTCATTACTAGAGTAGCTGAAGGCAGAAAA

GAGGATATTGATTTAGCGGTCAAGGCAGCAAGGAATGCATTCGACCATGGACCGTG
GCCTAGGTTCAGCGGAAGCGCGCGTGGTAAGATTATGATGAAATTTGCCAATCTGGT
AGACGAGAATGCTGAAGAACTTGCGATCTTAGATACCATCGATGGGGGCAAGTTAT
TTGGGATAGGTAAGGGACACGATATACCACAAGCTGCAGAATGTCTTAGATATTAT
GCAGGAGCCGCCGATAAAATCCACGGAGAAACACTAAAGATGTCATCTGAATTTCA
GGCCTACACACTGAAGGAACCCGTAGGGGTTGTCGGTCATATCATACCTTGGAACTT
TCCTTCTCAAATGTTCCTTATGAAGGTAGGGCCGGCTTTGGCAGCAGGATGTACTAT
GGTGGTTAAACCTGCCGAGCAGACTCCGCTTAGCGCTCTATTCTACGCGCATCTAGC
TAAGTTGGCAGGGGTTCCGGATGGGGTCATAAACGTCGTAACAGGTTTCGGTGGTAC
CGCCGGAAGTGCCATTTCTAGTCACATGGACATTGACATGGTTAGCTTTACAGGCTC
AACTGAGGTAGGCAGGCTAGTTATGCAAGCCGCAGCTCTGAGTAACCTGAAGCCAG
TATCACTTGAGTTGGGCGGTAAAAGCCCACTAATGATTTTCGACGACGCGGATGTGG
ACAAAGCGGTTGACTTAGCTCTGCTAGGATCTTTGTATAATAAAGGGGAAATATGCG
TAGCGGGAACCAGGATCTTCGTGCAGGAGGGTATTTACGACAAATTTCTGGAGAAA
CTGGCAGTGGGGATAAAGACGTGGGTCGTAGGGGACCCCTTCCATCCTAGCACTAG
GCAAGGTCCGCAAGTCGATAAGAAACAATATGAAAAGGTTTTATCTTACATAGAGC
ATGGGAAGACCGAGGGCGCAACCCTGTTTGCTGGAGGAAATCCATGTGGAAAGAAG
GGGTACTTCATTGAACCCACCATTTTCACGGACGTAAAAGACCACATGAAAATCGCC
AAAGAAGAAATCTTTGGTCCGGTGATGAGTGTATTCAAGTTTAAGACCGTTGAGGA
GGGAATTGAAAGAGCAAACGCAACGAAGTACGGCCTTGCGGCCGGGATCGTAACCA
ACAATCTGAATATCGCCAACACTGTTAGTAGATCCATAAGGGCGGGCGTCATATGG
ATTAACTGTTATTTTGCGTTTGATAGAGACTGTCCATACGGAGGGTATAAACAGAGC
GGTTTCGGCAGAGACCTTGGAATGGATGCGCTTCATAAATATCTGCATGTGAAGGCT
GTCGCAACCCCAATATACAACTCCCCGTGGCTATAA

>VoADH1

ATGACCAAGAGTTCCGGTGAGGTAATCAGTTGCAAAGCCGCCGTTATATACAAATCT
GGAGAGCCTGCAAAAGTTGAAGAGATTAGAGTGGACCCACCGAAATCATCAGAGGT
AAGAATAAAGATGTTGTATGCATCTCTATGCCATACGGATATTCTATGCTGTAACGG
ACTACCCGTCCCTTTATTTCCGCGTATACCTGGCCATGAGGGAGTTGGCGTAGTCGA
ATCAGCGGGTGAGGACGTGAAAGATGTCAAGGAGGGTGACATCGTGATGCCACTGT
ATCTTGGAGAATGTGGGGAGTGCCTTAACTGTTCATCCGGGAAAACAAATTTATGTC
ACAAGTATCCACTTGACTTTAGCGGTGTTCTACCCAGCGATGGGACCAGTCGTATGT
CCGTCGCTAAGTCAGGCGAAAAAATTTTCCATCATTTCAGTTGTTCCACATGGTCCG
AGTACGTCGTGATAGAGTCTAGTTATGTCGTAAAAGTCGACTCTCGTCTGCCCCTAC
CACACGCCTCCTTTCTGGCGTGCGGGTTTACCACCGGATACGGTGCTGCATGGAAAG
AAGCTGATATCCCCAAGGGAAGTACGGTAGCAGTGTTAGGGCTGGGCGCGGTGGGT
TTAGGTGTAGTTGCTGGCGCGCGTTCACAAGGTGCCAGTCGTATAATTGGAGTTGAC
ATCAACGACAAAAAAAAAGCCAAAGCTGAGATATTTGGAGTTACTGAGTTTCTTAA
TCCGAAACAGTTGGGTAAAAGTGCTTCCGAAAGCATCAAAGACGTTACTGGCGGCC

TGGGCGTGGACTATTGCTTTGAGTGTACCGGAGTACCGGCTTTATTAAATGAGGCTG
TAGACGCTTCAAAGATTGGATTAGGGACTATCGTCATGATCGGAGCCGGAATGGAA
ACCTCTGGTGTTATCAATTACATTCCTCTTCTATGTGGAAGAAAGCTGATCGGCTCTA
TCTATGGTGGAGTTCGTATAAGGTCCGACCTACCCTTGATTATCGAAAAATGCATAA
ATAAAGAAATTCCACTTAATGAGTTACAAACTCACGAAGTGTCTTTGGAGGGCATAA
ACGATGCATTCGGTATGCTGAAGCAGCCAGACTGTGTGAAAATAGTAATCAAGTTTG
AACAGAAGTGA

>VoCPR1

ATGGACACGAACAGCGATCTATTGAGGTCCATCGAGAGTTATCTGGGCGTAAGTATC
TCTAGCAACACACTTGTCCTTATTGCGACGACTTCAGTTGCTATTGTCGTCGGGCTTT
TGGTCTTTGTTTGGAAGAAAAGCAGCGGTGGGTCCAAAGAATTTAAGCCGGTTGTCT
TGCCTAAATCACCCACAGTCGAAGACGACGAGGACGAGGCCGAGGCCCCTCCGGGG
AAGACAAAACTAAGCATTTTCTTTGGTACTCAAACTGGTACTGCGGAGGGATTCGCT
AAGGCGCTTGCCGACGAAATTAAAGCAAAGTATGAGAAGGCTATTGTCAAGGTGAT
CGATTTAGACGACTATGCAGCCGATGATGACGTATACGAGGAGAAATTAAAAAAAG
AGACCTTGGCCTTCTTTATGGTTGCAACATATGGTGACGGGGAACCCACCGACAACG
CTGCAAGATTTTACAAGTGGTTCTCTGAGGGTCAAGACAGAGGTACTTGGCTTCAGC
AATTAACTTATGGGGTCTTTGCACTTGGCAACAGACAATATGAACATTTCAACAAGA
TTGGGAAAGTAATAGATGATCAGTTGGTTGAGCAGGGGGCCAAGAGGCTAGTGCCA
ATTGGCCTTGGAGACGATGATCAGTGCATCGAGGACGATTTTGCCGCCTGGAAGGA
GCTACTTTGGCCTGAATTAGACCAGCTGTTAAGAGACGAGGATGACACAAGCACCG
TAGCCACACCTTACACAGCGGTAATTCCGGAATATAGAGTTGTAATACATGATCCCG
ATACAACGACGTCCGATGATATGAACCTTCACGTCCCAAACGGTAATGTTAGTTTTG
ACATTCACCACCCTTGTAGAGCCAATGTTGCAGTCCAGAGAGAGCTTCATAAGCCGG
AGAGCGACCGTTCTTGCATACATTTAGAGTTCGACATAAGCGGTACCGGCATCACCT
ACGAGACAGGGGATCACGTAGGAGTGTTTGCCGAGAATTGCGAAGAGACAGTTGAG
GAGGCGGCGAAGCTGTTGGGACAACCGCTAGACATGTTGTTTTCCATCCATACAGAT
AAGGATGATGGCTCCTCACAGGGCAGCTCTTTACCACCGCCCTTCCCAGGGCCGTGC
ACCCTTCGTACTGCCCTGGCAAGGTATGCGGACGTGCTGAACCCTCCTAGGAAAGCT
GCATTAGTGGCACTAGCAGCCCATGCGACTGAACCGGCAGAAGCAGAGAGGTTGAA
GTTCCTGTCTAGCCCTCAGGGTAAAGATGAATACTCACAGTGGGTGTTAGGCTCCCA
AAGAAGTTTGCTAGAGGTCATGGCTGAATTCCCTAGTGCGAAACCTCCCATTGGAGT
CTTCTTCGCTGCTGTCGCGCCACGTCTACCGCCCCGTTACTATTCTATCTCATCCTCC
CCTAGGTACGCCGGTGACCGTGTACACGTGACTTGTGCCCTGGTTTATGGACCAAGT
CCCACCGGAAGGATCCATAAGGGGGTATGCAGTACCTGGATGAAGAACGCGGTACC
TTTGGGCAAATCTGACGACTGTTCATGGGCGCCCATTTTCATCAGGACGAGCAATTT
CAAATTACCCGCCGATCCAAGCGTCCCGATAATAATGGTCGGACCAGGGACTGGAT
TAGCGCCGTTCCGTGGATTCCTGCAAGAACGTCTTTCTCTGAAGGAGGAGGGTGCCC
AGTTGGGTCCCGCCCTACTGTTTTTCGGTTGCAGGAACAGGAGGATGGATTTTATAT

ATGAAGAGGAACTGAACAATTTTGTCGACGAAGGCGTGATTTCAGAATTGATTGTTG
CATTTTCTCGTGAGGGCCCCACCAAAGAATACGTCCAACACAAAATCATAGAGAAG
GCTGCGGATATTTGGAGTCTTATAAGCGAGGGAGCTTATTTATATGTGTGTGGTGAT
GCGAAGGGTATGGCTAGAGACGTCCACCGTACCTTACACACAGTAGTTCAAGAACA
GGAGAAAGTTGATAGTACTAAGGCTGAAGCTATTGTGAAGAAGTTACAAATGGATG
GACGTTATCTTAGGGACGTATGGTAA