Total biosynthesis of the cyclic AMP booster forskolin from *Coleus forskohlii*

Irini Pateraki[1,2,$*], Johan Andersen-Ranberg[1,2,†,$], Niels Bjerg Jensen[3], Sileshi Gizachew Wubshet[4,†], Allison Maree Heskes[1,2], Victor Forman[1], Björn Hallström[5], Britta Hamberger[1,2,†], Mohammed Saddik Motawia[1,2], Carl Erik Olsen[1,2], Dan Staerk[4], Jørgen Hansen[3], Birger Lindberg Møller[1,2] and Björn Hamberger[1,2,†]

[1]Plant Biochemistry Laboratory, Department of Plant and Environmental Sciences, University of Copenhagen, Denmark;

[2]Center for Synthetic Biology "bioSYNergy", Copenhagen, Denmark;

[3]Evolva A/S, Copenhagen, Denmark;

[4]Department of Drug Design and Pharmacology, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark;

[5]Science for Life Laboratory, KTH - Royal Institute of Technology, Stockholm, Sweden;

[†]current addresses:

JAR, Department of Plant & Microbial Biology, University of California, Berkeley, 111 Koshland Hall, Berkeley, CA, 94720;

SGW, Nofima AS, Osloveien 1, N-1430 Ås, Norway

BjH, BrH; Department of Biochemistry & Molecular Biology, Michigan State University, Molecular Plant Sciences Building, 1066 Bogue Street, East Lansing, MI 48824, USA

[$]These authors have contributed equally to the article

*Correspondence to: Irini Pateraki, eipa@plen.ku.dk

Abstract (150)

Forskolin is a unique structurally complex labdane type diterpenoid used in the treatment of glaucoma and heart failure based on its activity as a cyclic AMP booster. Commercial production of forskolin relies exclusively on extraction from its only known natural source, the plant *Coleus forskohlii*, in which forskolin accumulates in the root cork. Here we report the discovery of five cytochrome P450s and two acetyltransferases which catalyze a cascade of reactions converting the forskolin precursor 13*R*-manoyl oxide into forskolin and a diverse array of additional labdane-type diterpenoids. A minimal set of three P450s in combination with a single acetyl transferase was identified that catalyzes the conversion of 13*R*-manoyl oxide into forskolin as demonstrated by transient expression in *Nicotiana benthamiana*. The entire pathway for forskolin production from glucose encompassing expression of nine genes was stably integrated into *Saccharomyces cerevisiae* and afforded forskolin titers of 40 mg/L.

Introduction

Plants synthesize an impressive diversity of specialized metabolites enabling them to communicate and adapt to environmental challenges (***Mithöfer et al., 2012; Woldemariam et al., 2011***). Throughout history, humans have benefited from the medicinal properties of many of these phytochemicals (***Hardy et al., 2012***). Specialized plant metabolites and direct derivatives thereof still constitute more than a third of approved pharmaceuticals (***Cragg et al., 2013; David et al., 2015***). With over 50,000 known structures according to the "Dictionary of natural products"

(http://dnp.chemnetbase.com/), terpenoids are the largest class of plant specialized metabolites and constitute a vast repository of bio-active natural products including many structurally complex compounds (***Pateraki et al., 2015***). Examples of widely used plant derived terpenoid pharmaceuticals are the anti-cancer drug paclitaxel (taxol®) (***Liu et al., 2010***), the therapeutic ingenol mebutate (picato®) that is used for treatment of actinic keratosis (***King et al., 2016; Luo et al., 2016***) and artemisinin which is the most efficient treatment against malaria caused by *Plasmodium* parasites (***Graham et al., 2010; Paddon et al., 2014***). Traditional chemical synthesis of plant derived diterpenoid pharmaceuticals remains economically challenging, despite recent examples of elegant strategies mimicking natural routes (***Appendino 2014; Kawamura et al., 2016; Yuan et al., 2016***). Extraction from plant biomass and semi-synthesis from biotechnologically produced intermediates have been approached as alternative strategies (***Graham et al., 2010; Paddon et al., 2013; Roberts 2007***). In contrast to recent examples demonstrating complete pathway reconstruction and production of opiate alkaloids in yeast (***Galanie et al., 2015; Nakagawa et al., 2016***), engineered total biosynthesis of terpenoid therapeutics—including paclitaxel and ingenol esters—has not yet been achieved. Challenges on the way to achieving this goal include the identification of pathway enzymes in native systems, particularly for those belonging to multi-enzyme families catalyzing the biosynthesis of specialized metabolites in plants, engineering of poorly understood multi-step enzymatic pathways and difficulties encountered in heterologous expression of key enzymes catalyzing monooxygenations critical for diterpenoid biosynthesis (***Pateraki et al., 2015; Renault et al., 2014***).

The diterpenoid forskolin is the active hypotensive principle accumulating in the root cork of *Coleus forskohlii* (***Pateraki et al., 2014***), a perennial shrub of the Lamiaceae family, indigenous to India and Southeast Asia with numerous reported applications in traditional medicine (***Alasbahi et***

al., 2010b; Kavitha et al., 2010). The pharmaceutical properties of forskolin are based on its ability to directly activate adenylate cyclase enzyme resulting in elevated levels of the second messenger cyclic adenosine monophosphate (cAMP) (***Doseyici et al., 2014; Seamon et al., 1981***). Approved applications of forskolin range from alleviation of glaucoma (Ocufors™Eye drop solutions, Sabinsa, India), treatment of hypertension and heart failure (Colforsin daropate hydrochloride, a water-soluble derivative of forskolin, Nippon Kayaku, Japan) to lipolysis and body weight control (***Godard et al., 2005; Kikura et al., 2004; Toya et al., 1998; Wagh et al., 2012; Yoneyama et al., 2002***). Therapeutic opportunities were also suggested in animal tests, where forskolin induced pigmentation of the skin, increasing protection against UV-associated carcinogenesis (***D'Orazio et al., 2006***). The complex chemical structure of forskolin with a decalin core characteristic of labdane type diterpenoids, a tetrahydropyran ring, five oxidized positions and eight chiral centers (***Figure 1A***) represents a challenge for classical organic chemical synthesis, although a key intermediate for stereoselective total synthesis has been reported (***Ye et al., 2009***). Hence, commercially available forskolin is extracted from *C. forskohlii* roots and purified from a mixture of over 60 structurally related abietane and epoxylabdane diterpenoids with a forskolin content varying from 0.013 % to 0.728 % of root dry weight (***Alasbahi et al., 2010a; Asada et al., 2012; Srivastava et al., 2017***). As the demand for forskolin grows, reliable and sustainable commercial production from *C. forskholii* will become unachievable due to low yields, susceptibility of this species to diseases, changing climatic conditions and the resource intensive extraction and purification procedure required to obtain pharmaceutical grade forskolin (***Mora-Pale et al., 2014***). Elucidation of the biosynthetic pathway to forskolin and subsequent engineering of the pathway into microbial hosts offers a cleaner and more stable alternative production system that will be better able to address future needs.

Recently, we reported specific accumulation of forskolin and its diterpene scaffold 13*R*-manoyl oxide in the root cork cells of *C. forskohlii*. A pair of diterpene synthases (*Cf*TPS2 and *Cf*TPS3), exclusively present in the root cork, was found to catalyze cyclization of the C_{20} diterpenoid precursor geranylgeranyl diphosphate (GGPP) into 13*R*-manoyl oxide, the diterpene scaffold of forskolin (***Pateraki et al., 2014***). As a proof of concept, biosynthesis of 13*R*-manoyl oxide in enantiomerically pure form but in low yields was achieved by expressing *Cf*TPS2 and *Cf*TPS3 in *Saccharomyces cerevisiae, E. coli* and *Synechocystis* sp. (***Andersen-Ranberg et al., 2016; Englund et al., 2015; Nielsen et al., 2014***). Taking into account the functionalization steps needed for conversion of 13*R*-manoyl oxide to forskolin (***Figure 1A),*** involvement of enzymes from the families of cytochrome P450s (CYPs) and acetyltransferases would be predicted.

Here we report an integrated biochemical and functional genomics approach, including metabolomics, single cell-type transcriptome studies and a synthetic biology modular approach involving transient combinatorial expression of candidate genes in *Nicotiana benthamiana* to identify the panel of enzymes catalyzing functionalization of the *C. forskohlii* diterpene backbones and more specifically the biosynthesis of forskolin. Pathway intermediates were identified using GC- or HPLC-HRMS-SPE-NMR. To demonstrate the downstream application of the present work regarding biotechnological production of forskolin, the entire forskolin biosynthetic pathway was reconstituted in engineered *Saccharomyces cerevisiae* for fermentation-based production of forskolin from glucose. Forskolin is the first example of a pharmaceutical diterpenoid produced entirely in yeast at titers relevant for industrial consideration. The outlined combinatorial biochemistry approach paves the way for development of yeast-engineered platforms for biosynthesis of other known or new-to-nature diterpenoids

Results

Discovery of multifunctional cytochromes P450 in *Coleus forskohlii* producing a multitude of 13*R*-manoyl oxide-derived diterpenoids and identification of a biosynthetic pathway for forskolin.

The conversion of 13*R*-manoyl oxide to forskolin requires six regio- and stereospecific monooxygenations and a single regiospecific acetylation (***Figure 1A***). Considering the strict localization of forskolin in the root cork cells of *C. forskohlii* and the almost exclusive expression of the pair of diterpene synthases forming 13*R*-manoyl oxide within the same tissue (**Pateraki et al., 2014**), the root cork was selected for deep RNA-Seq transcriptome analysis. The generated transcriptome contained 263,652 assembled putative cDNAs. The transcriptome was queried for transcripts encoding cytochrome P450s belonging to the CYP71 clan, based on their established role in monooxygenation reactions in the biosynthesis of specialized metabolites (***Nelson 2013; Werck-Reichhart et al., 2000***). Their relative levels in the root cork transcriptome were also taken into consideration. Within the CYP71 clan, focus was also placed on P450 subfamilies that showed extensive, recent expansions in the cork transcriptome (***Nelson et al., 2011; Werck-Reichhart et al., 2000***). Cytochrome P450 members of the CYP76AH subfamily, part of the CYP71 clan, have recently been shown to catalyze monooxygenation of abietane-type diterpenoids like miltiradiene and dehydroabietadiene in Lamiaceae species, closely related to *C. forskohlii* (***Božić et al., 2015; Ignea et al., 2016a; Zi et al., 2013***). Members of this P450 subfamily were therefore of high interest as enzymes putatively involved in diterpenoid biosynthesis in *C. forskohlii*. Based on these

considerations, a total of 29 cytochrome P450 candidates (***Figure 1-source data 1***) were selected and cloned in full length from cDNA synthesized from root cork total RNA. Among these CYPs, seven members were assigned to the CYP76AH subfamily by the "P450 Nomenclature committee" (***Nelson 2009***), rendering this CYP subfamily the highest represented in the transcriptome (***Figure 1-source data 1***)**.** Five were full length sequences (*Cf*CYP76AH8, *Cf*CYP76AH9, *Cf*CYP76AH10, *Cf*CYP76AH11, and *Cf*CYP76AH11), while two (*Cf*CYP76AH15 and *Cf*CYP76AH16) were represented by partial cDNAs. For the latter, 5'RACE experiments afforded the full-length cDNAs. Similarly to the previously identified diTPSs, *Cf*TPS2 and *Cf*TPS3 (***Pateraki et al., 2014***), gene expression studies showed that the identified members of the CYP76AH family were highly or exclusively expressed in the root cork cells (***Figure 1B***).

We have recently reported an *Agrobacterium* mediated modular transient expression system in *Nicotiana benthamiana* enabling biosynthesis of labdane-type diterpenes in quantities permitting purification and structural elucidation (***Andersen-Ranberg et al., 2016***). Utilizing this system, each of the candidate P450 genes were heterologously expressed in combination with genes necessary for the production of high amounts of 13*R*-manoyl oxide (***Andersen-Ranberg et al., 2016; Pateraki et al., 2014***). Of the CYPs tested, six efficiently converted 13*R*-manoyl oxide into oxygenated derivatives (***Figure 2A, 3 and Figure 3-supplement 1 and 2***). *Cf*CYP76AH15, *Cf*CYP76AH8 and *Cf*CYP76AH17 catalyzed formation of 11-oxo-13*R*-manoyl oxide (**2**) as the main product (***Figure 2 and 3***). The structure of forskolin harbors a keto-group at the C-11 position and this also applies for the majority of 13*R*-manoyl oxide-derived diterpenoids found in *C. forskohlii* (***Asada et al., 2012; Zhang et al., 2009***). Of the three CYP76AHs tested in this experiment, *Cf*CYP76AH15 showed the highest efficiency and specificity for the conversion of 13*R*-manoyl oxide to **2** with no concomitant formation of multi-oxygenated products (***Figure 2***). Compound **5d** produced by

158 *Cf*CYP76AH8 as well as by *Cf*CYP76AH17 was identified as 1,11-dihydroxy-13*R*-manoyl oxide
159 (***Figures 2B, 4 and Table 1 and 2)***. This specific oxygenation pattern is also found in forskolin.
160 Minor amounts of several di- and trihydroxylated 13*R*-manoyl oxide-derived compounds were also
161 produced by *Cf*CYP76AH8 and *Cf*CYP76AH17 (***Figure 2A and Figure 3-Figure Supplement 1***).
162 Although we managed to identify the chemical structures of a number of 13*R*-manoyl oxide
163 derivatives (***Figure 2B***), it was not possible to do so the procedure for all the compounds shown in
164 ***Figure 3-Figure Supplement 1***. The main obstacle was the high complexity of the diterpenoids
165 produced in *N. benthamiana* leaves expressing the *Cf*CYP76AHs as well as the small amounts
166 present of most of these diterpenoids. Additional limiting factors were the instability of several
167 these compounds and the limiting plant material available. Production of higher amounts of the
168 compounds in microbial hosts was not pursued, because the terpenoid profiles observed following
169 expression of the enzymes in plants and yeast cells were not identical.
170 Two additional CYPs of the CYP76AH subfamily catalyzed oxygenation of 13*R*-manoyl oxide at
171 different positions without substantial formation of C-11 keto derivatives. *Cf*CYP76AH16 yielded
172 predominantly **3a** which was identified by NMR as 9-hydroxy-13*R*-manoyl oxide (or coleorol) and
173 *Cf*CYP76AH11 produced a range of monooxygenated derivatives including traces of **10b** which was
174 identified by NMR as 9-deoxydeacetylforskolin (***Figure 2***). The positions of the carbonyl and
175 hydroxyl-groups in **3a** and **10b** were consistent with those in forskolin. Thus the individual
176 activities of the CYPs catalyzing formation of these compounds can be considered complementary
177 in forskolin biosynthesis: **10b** carries the carbonyl-function at C-11 and the three hydroxyl groups
178 observed in forskolin at positions C-1, C-6 and C-7, but lacks hydroxylation at the C-9 position,
179 which is observed in **3a**. The only enzyme outside the CYP76AH subfamily that displayed activity

towards 13*R*-manoyl oxide was CYP71D381, which resulted in oxidized derivatives at positions not compatible with forskolin (***Figure 3-Figure Supplement 2***).

To probe the role of the different CYP76AH enzymes in forskolin biosynthesis, they were co-expressed by combining one of the three P450s catalyzing formation of **2** with the functionally distinct *Cf*CYP76AH11 and *Cf*CYP76AH16, first in pairs, then in all possible permutations as triplets (***Figure 5 and Figure 5-Figure Supplement 1*).** The combination of *Cf*CYP76AH15 and *Cf*CYP76AH11 afforded production of 6,7-dihydroxy-11-keto-manoyl oxide (**7h**) as part of a complex mixture. Formation of **7h** demonstrated combined introduction of the carbonyl group at C-11 together with two hydroxyl groups at positions C-6 and C-7, again consistent with the oxygenation pattern of forskolin (***Figure 2 and Figure 5-Figure Supplement 1***).

When the CYP76AH enzymes were assayed in triplet combinations, the product profiles were further shifted towards multi-oxygenated 13*R*-manoyl oxide derivatives. The formation of minor amounts of deacetylforskolin (**13b**) and several compounds with identical mass to charge ratio *(m/z)* but different retention times were detected using different enzyme combinations (***Figure 5, Table 3).*** The triplet combination *Cf*CYP76AH15, *Cf*CYP76AH11 and *Cf*CYP76AH16 led to the highest amounts of **13b.** Thus, this combination of multifunctional P450s appeared to constitute the optimal biosynthetic pathway for specific formation of **13b** from **1** using the engineering approaches of synthetic biology (***Figure 5***).

Monooxygenase activity of the CYP76AH subfamily towards other terpene scaffolds

In addition to 13*R*-manoyl oxide-derived diterpenoids, the root cork of *C. forskohlii* contains numerous abietane diterpenoids derived from miltiradiene (***Alasbahi et al., 2010a***). Recently a

pair of diterpene synthases (*Cf*TPS1 and *Cf*TPS3) mainly expressed in the root cork of *C. forskohlii* was demonstrated to produce miltiradiene (***Pateraki et al., 2014***). It has been shown previously that members of the CYP76AH subfamily in Lamiaceae are able to oxygenate miltiradiene or derivatives thereof. CYP76AH1 from *Salvia miltiorrhiza* (***Guo et al., 2013***), CYP76AH4 as well as *Ro*FS1 and *Ro*FS2 from *Rosmarinus officinalis* (***Bozic et al., 2015; Zi et al., 2013***) and CYP76AH24 from *S. pomifera* catalyze synthesis of ferruginol, the precursor of carnosic acid and tanshinones (***Guo et al., 2013; Ignea et al., 2016a***), from miltiradiene or dehydroabietadiene. Additionally, CYP76AH3 from *S. miltiorrhiza* has been shown to accept ferruginol as a substrate to produce sugiol, 11-hydroxy-ferruginol and 11-hydroxy-sugiol (***Guo et al., 2015***). These miltiradiene-accepting CYP76AHs show high sequence homology, ranging from 60% to 85% at the amino acid level, with those identified in *C. forskohlii*. Therefore, it was tempting to study the ability of the *Cf*CYP76AHs to metabolize miltiradiene. The *Cf*CYP76AHs were co-expressed individually in *N. benthamiana* leaves producing miltiradiene (***Andersen-Ranberg et al., 2016***), and the product profile monitored by unbiased LC-MS analysis. *Cf*CYP76AH15 was shown to convert miltiradiene to ferruginol (***Figure 6***). Ferruginol was identified in extracts of *C. forskohlii* root cork (***Figure 6***), so it is possible that *Cf*CYP76AH15 is also involved in the biosynthesis of ferruginol *in planta*. In a parallel series of experiments, CYP76AHs from rosemary and salvia known to accept miltiradiene as substrate were tested for their ability to use 13*R*-manoyl oxide as a substrate. Transient expression of *Ro*CYP76AH4 (***Zi et al., 2013***), *Ro*FS1 and *Sf*FS (***Božić et al., 2015***) in *N. benthamiana* leaves able to synthesize 13*R*-manoyl oxide (***Pateraki et al. 2014***) demonstrated that *Ro*CYP76AH4 efficiently converted 13*R*-manoyl oxide to 11-oxo-13*R*-manoyl, while *Ro*FS1 and *Sf*FS, in addition to 11-oxo-13*R*-manoyl produced 11-hydroxy manoyl oxide (***Figure 7***).

Establishing regiospecific acetylation as the final step of forskolin biosynthesis

To complete the biosynthetic route to forskolin, specific acetylation of the C-7 hydroxyl group of deacetylforskolin (**13b**) is required. The root cork transcriptome was mined for acyltransferases (ACTs) from clade III of the BAHD family earlier reported to predominantly use acetyl CoA as acetyl donor for acetylation of hydroxyl groups (***D'Auria 2006***). Ten ACTs (***Figure 1-source data 1***) were identified, cloned and tested functionally by *Agrobacterium*-mediated transient expression in *N. benthamiana* leaves, engineered to produce deacetylforskolin by co-expression of the enzymes *Cf*DXS, *Cf*GGPPS, *Cf*TPS2, *Cf*TPS3, *Cf*CYP76AH15, *Cf*CYP76AH11 and *Cf*CYP76AH16. Two ACT candidates, *Cf*ACT1-6 and *Cf*ACT1-8, were found to catalyze acetylation of **13b** (***Figure 8***). Expression of *Cf*ACT1-6 resulted in formation of a broad range of acetylated products of which forskolin constituted a minor fraction. In contrast, *Cf*ACT1-8 exhibited high activity and specificity, with efficient conversion of **13b** to forskolin and absence of detectable acetylated side products. This establishes the entire and highly specific biosynthetic route to forskolin from its precursor, GGPP (***Figure 9***).

Engineering of the entire pathway of forskolin in *Saccharomyces cerevisiae*

Expression and engineering of plant biosynthetic pathways in microbial organisms provides a method for sustainable production of high value compounds like the structurally complex bioactive diterpenoids (***Guo et al., 2013; Ignea et al., 2016a; Jia et al., 2016***). With the genes encoding the entire biosynthetic pathway of forskolin identified, we proceeded to establish stable forskolin production in *S. cerevisiae*, an excellent host organism for biosustainable and scalable

production of numerous bio-active natural products (***Brochado et al., 2010; Brown et al., 2015; Galanie et al., 2015; Hansen et al., 2009; Ignea et al., 2016a; Jeandet et al., 2012***). For expression in yeast, all *C. forskohlii* genes were codon-optimized and stably integrated in neutral loci in the yeast genome. Genomic integration was chosen versus expression via episomal plasmids as the former strategy favors the simultaneous expression of a large number of genes as well as effective selection marker recycling (***Jensen et al., 2014***). Additionally, a sequence encoding a NADPH-dependent cytochrome P450 oxidoreductase (*Cf*POR), required to support the P450 activity, was identified from the *C. forskohlii* root cork transcriptome and cloned for co-integration in the yeast genome. The isolated *Cf*POR was the only one present in the root cork transcriptome. Genomic integration enabled stable, simultaneous expression of a total of nine heterologous genes in the microbial host.

*Cf*POR, *Cf*CYP76AH15, *Cf*CYP76AH11, *Cf*CYP76AH16 and *Cf*ACT1-8 were co-expressed in the yeast strain EFSC4498, previously engineered to produce 350 mg/L 13*R*-manoyl oxide (***Andersen-Ranberg et al., 2016***). Transformed yeast strains verified for the integration of all forskolin biosynthetic cDNAs into their genome, were further analyzed and production titers of forskolin and pathway intermediates were monitored. The highest forskolin producing strain, EVST21543, demonstrated genetic stability through several rounds of cultivation and was grown in a 5 L fermenter using minimal medium under glucose limited conditions. During fermentation, accumulation of forskolin (**16c**), 13*R*-manoyl oxide (**1**), 9-hydroxy-13*R*-manoyl oxide (**3a**), and biomass formation (***Figure 10***) were monitored over time. Forskolin levels reached more than 40 mg/L of yeast culture. Simultaneous accumulation of high titers of 13*R*-manoyl oxide (**1**) and 9-hydroxy-13*R*-manoyl oxide (**3a**) at levels of 200 and 500 mg/L of yeast culture, respectively, indicated that the conversion of 13*R*-manoyl oxide (**1**) to forskolin was far from complete. These

intermediates did not accumulate in the *C. forskohlii* root cork. A comparison of the total intermediate profiles of *C. forskohlii* root cork versus the fermenter grown EVST21543 yeast strain is shown in (***Figure 10-Figure Supplement 1*****).**

Discussion

The terpenoid biosynthetic pathways active in the root of *C. forskohlii* produce an array of 13*R*-manoyl oxide and miltiradiene-derived diterpenoids, including forskolin. Forskolin is one of the most complex and highly oxygenated diterpenoids reported in *C. forskohlii*. In the current study, the genes encoding the entire biosynthetic pathway for forskolin were identified. Availability of the transcriptome from the root cork cells of *C. forskohlii,* where forskolin biosynthesis takes place, the option to achieve rapid functional characterization of the gene candidates *in planta* by transient expression in *N. benthamiana* and high sensitivity techniques for structural characterization made the pathway elucidation possible. With the genes encoding the entire forskolin biosynthetic pathway in hand, we achieved *de novo* production of forskolin in engineered yeast was achieved.

Initially, a number of genes encoding CYP76AH subfamily members, expressed mainly in the root cork of *C. forskohlii*, was cloned and transiently expressed in *N. benthamiana* leaves being able to produce 13*R*-manoyl oxide. The products profiles obtained with these enzymes revealed that *Cf*CYP76AHs have discrete roles in forskolin biosynthesis. Efficient monooxygenation at C-11 is catalyzed mainly by *Cf*CYP76AH15 (but also by *Cf*CYP76AH8, *Cf*CYP76AH17 and *Cf*CYP76AH11).

Monooxygenation at position C-9 is catalyzed exclusively by *Cf*CYP76AH16 and monooxygenation at C-1, C-6 and C-7 mainly by *Cf*CYP76AH11. Monooxygenation at C-1 was also observed using CfCYP76AH8 and CfCYP76AH17 (**4c**, ***Figure 3-Figure Supplement 1***). Collectively this set of data displays the multifunctional roles of these enzymes. Together they could account for all the oxygenated positions in forskolin. Co-expression of *Cf*CYP76AH15, *Cf*CYP76AH11 and *Cf*CYP76AH16 resulted in specific and efficient formation of the final intermediate, deacetylforskolin.

Despite the complementary multifunctionality of the CYP76AH enzymes, partial functional redundancy is possible, as demonstrated by the varied oxygenation patterns observed in experiments with single enzymes. Overlapping functionalities may contribute to a coordinated action for efficient conversion of **1** to **13b.** Moreover, certain *C. forskohlii* CYP76AH enzymes seem able to accept oxygenated forms of **1** as substrates which results in the observed shifts in profile towards higher decorated products when they are co-expressed in *N. benthamiana* (***Table 3***)**.**

Although co-expression of CYP76AH11 and CYP76AH16 with either one of the three CYP76AH8, CYP76AH17 or CYP76AH15 in the engineered system resulted in biosynthesis of deacetylforskolin (**13b**) (***Figure 5***), the precise sequence of *in planta* 13*R*-manoyl oxide oxygenations cannot be deduced from the experimental results partly because all identified *Cf*CYP76AHs accept **1** as substrate. The co-expression of the CYP76AH encoding genes in the root cork of *C. forskohlii* and their partial functional redundancy or complementarity may *in vivo* constitute the basis for the chemical diversity of labdane terpenoids present in the root cork of *C. forskohlii. In planta*, the forskolin biosynthetic pathway would thus appear to be entangled within a metabolic grid offering simultaneous production of a multitude of other diterpenoids.

311 Recently CYP76AH enzymes accepting miltiradiene as substrate were reported from other
312 Lamiaceae species (***Božić et al., 2015; Guo et al., 2015; Ignea et al., 2016a; Zi et al., 2013***), a non-
313 epoxylabdane, abietane diterpenoid, which is also present in *C. forskohlii* roots. This prompted us
314 to examine whether the promiscuous and multifunctional *Cf*CYP76AH could accept miltiradiene as
315 substrate, and vice versa, e.g. if different Lamiaceae CYP76AHs can catalyze oxygenations of 13*R*-
316 manoyl oxide (***Figure 6 and 7)***. *Cf*CYP76AH15 was found to have very similar catalytic activities
317 compared to the rosemary CYP76AH4, an enzyme with a suggested role in biosynthesis of
318 ferruginol from miltiradiene (***Zi et al., 2013***). Efficient formation of ferruginol as well as 11-oxo-
319 13*R*-manoyl oxide, by both enzymes indicates that they may represent orthologues. Two additional
320 ferruginol synthases of the CYP76AH subfamily, one from *Salvia fruticosa* (*Sf*FS) and one from
321 *Rosemary officinalis* (*Ro*FS), were found to catalyze the conversion of 13*R*-manoyl oxide to 11-oxo-
322 13*R*-manoyl oxide and 11-hydroxy-manoyl oxide when expressed in *N. benthamiana* leaves (***Figure***
323 ***6 and 7)***. These findings are not reflected in the phylogenetic analysis of the known CYP76AHs. All
324 *C. forskohlii* CYP76AHs able to produce 11-oxo-13*R*-manoyl oxide are clustered together, while
325 CYP76AHs from *Salvia* spp. and *R. officinalis* that can catalyze the formation of the same
326 compound, as well as those CYPs able to accept miltiradiene as substrate, form a different cluster
327 when analyzed with currently known CYP76AHs. Thus, it is likely that the ability of CYP76AHs to
328 catalyze 11-oxo-13*R*-manoyl oxide has evolved convergently in these plants (***Figure 11***).

329 These data highlight the functional versatility of the CYP76 family. The enzymes can exhibit broad
330 substrate specificity which may advance metabolic evolution as they provide metabolic plasticity
331 and flexibility affording synthesis of new diterpenoids. This facilitates the expansion of the number
332 of possible diterpenoids produced in nature and potentially serves to diversify and augment the
333 phytochemical defense of plants. The promiscuity of the CYP76AHs also provides potentials for

their use in combinatorial approaches for synthesizing a range of diterpenoids with pharmaceutical relevance. The exclusive presence of the CYP76AH subfamily in Lamiaceae species may reflect gene duplications in a Lamiaceae ancestral species and expansion and neofunctionalization after speciation (***Figure 11***).

The identification of *Cf*ACT1-8 as an ACT catalyzing regiospecific acetylation of deacetylforskolin (**13b**) to afford forskolin completed the entire biosynthetic pathway for forskolin from its precursor, GGPP. Interestingly, the only currently identified acyltransferases in diterpenoid biosynthesis are those involved in biosynthesis of paclitaxel. Those acetyltransferases show substantial regioselective promiscuity (***Ondari et al., 2008; Walker et al., 2000***) and belong to Clade V (***D'Auria 2006; Tuominen et al., 2011***), whereas the majority of the ACTs identified in the root cork transcriptome of *C. forskohlii*, including ACT1-6 and ACT1-8, belong to Clade III (***Figure 12***).

With all forskolin biosynthetic pathway genes identified, we moved to the generation of a stable forskolin producing *S. cerevisiae* strain. To engineer a stable microbial production platform, we proceeded to integrate the minimum required set of functional parts into the *S. cerevisiae* genome. Specific *de novo* production of the highly functionalized diterpenoid from glucose at titers above 40 mg/L was achieved through a pathway consisting of a total of 10 enzymatic steps catalyzed by 7 heterologously co-expressed enzymes. This high forskolin titer, achieved with no optimization steps, highlights the potential to develop a microbial manufacturing platform for efficient and stereospecific production of forskolin with further fine-tuning of the biosynthetic pathway. Currently it is not possible to accurately estimate the forskolin titers necessary for industrially profitable production, as the exact commercial applications, market size and price as

well as the production cost including downstream processing cannot be determined at present. Given the knowledge gained in our present study and experiences with other compounds (***Paddon et al., 2014***) we find it realistic to aim for yields ranging from a single to double digits of gram per liter of yeast culture. To achieve higher forskolin yields, it is important upfront to ascertain a proper flux towards GGPP through the mevalonate pathway (***Kampranis et al., 2012***). Specifically for forskolin pathway, it seems clear that there is a limitation in flux through one or several of the P450s involved as we encounter accumulation of the intermediates, 300 mg/L and 500 mg/L of compounds **1** and **3a** respectively, compared to a forskolin production of 40 mg/L. Only minute amounts of deacetylforskolin. Accumulation of **1** and **3a** intermediates was not observed *in planta* (***Figure 10-Figure Supplement 1***). This signifies that the expression levels of the heterologously expressed genes are not properly balanced or that some of the CYPs have lower specific activity when inserted into a yeast membrane. In the native plant host, the pathway exhibits optimized carbon flux. P450s are notorious difficult to express in high amounts in yeast and recognized as exhibiting rather low K_{cat} values (***Jung et al., 2011; Renault et al., 2014***)). Hence, to increase forskolin production in the yeast system, efforts should obviously be focused on optimizing *Cf*CYP76AHs expression and enzyme kinetics, specificity and catalytic efficiency as well as pathway scaffolding to facilitate formation of a metabolon improving pathway flux and efficiency and minimizing accumulation of pathway intermediates (***Laursen et al., 2016***). The high forskolin titers already obtained though in the engineered yeast strain highlights the potential to develop a microbial manufacturing platform for efficient and stereospecific production of forskolin and other labdane terpenoids by fine tuning the biosynthetic pathways.

A yeast based production platform constitutes a sustainable alternative to traditional crop-based production but the gains always need to be compared to yield improvements obtained by classical

or molecular breeding of the traditional host plant (***Graham et al., 2010***). The model-example from the literature and industry towards production of a pharmaceutically relevant terpenoid in *S. cerevisiae* is the sesquiterpenoid artemisinic acid, a pathway intermediate to the anti-malarial compound artemisinin (***Paddon et al., 2013***). However, this prominent model example is dependent on a costly organic chemical synthesis component to chemically convert artemisinic acid to artemisinin (***Peplow 2016***). Recent approaches of engineered *de novo* production of structurally complex diterpenoids, triterpenoids or alkaloids in microbial systems have also been limited to proof-of-concept studies, expression of partial pathways and sub-milligram yields, highlighting the challenges in synthetic biology to offer an economically realistic and sustainable alternative to isolation of the desired medicinal compounds from medicinal plants bred to produce elevated levels. The constraints to achieve high yields are connected to expression of multiple P450s and reconstruction of pathways with multiple functionally divergent steps (***Brown et al., 2015; Li et al., 2016; Zhou et al., 2015***). Strategies addressing these issues are the development of synthetic microbial consortia of *S. cerevisiae* and *E. coli*, optimization of P450s for functional expression in *E. coli*, optimization of interactions between the P450s and their reductase partner and N-terminal modifications (***Biggs et al., 2016; Laursen et al., 2016; Vazquez-Albacete et al., 2016; Zhou et al., 2015***).

Our current study demonstrates that mining for additional members of the CYP76AH family has the potential to facilitate the assembly of further optimized panels of mixed-species P450s for the biosynthesis of bioactive diterpenoids. This shows the great promise that combinatorial assembly including CYPs outside the CYP76AH subfamily may offer the opportunity to design production systems for diterpenoids that are currently inaccessible due to their exclusive presence in rare or

red-listed plants and to further expand the chemical diversity of diterpenoids to production of diterpenoids currently not known from nature.

Materials and Methods

Materials

All chemicals including forskolin were acquired from Sigma-Aldrich. An authentic standard of 13*R*-manoyl oxide was prepared as previously described (*Nielsen et al., 2014*). CYP76AH4 (***Zi et al., 2013***) was cloned from rosemary plants acquired at a local market in Copenhagen, Denmark. *Ro*FS1 and *Sf*FS (***Božić et al., 2015***) were kindly provided by Dr. Angelos Kanellis (University of Thessaloniki, Greece).

Chemical synthesis of deacetylforskolin

The deacetylation of forskolin has been achieved previously with the use of methanolic potassium carbonate which can provide 7-desacetylforskolin in 65 % yield (***Kosley et al., 1989***). Here we carried out the deacetylation of forskolin using a solution of methanolic ammonia solution (2M) to afford 7-desacetylforskolin quantitatively. The 1H NMR data of the deacetylated forskolin were in agreement with the reported data in literature.

Transcriptome sequencing

Coleus forskohlii root cork total RNA was extracted as previously described (***Pateraki et al., 2014***). RNA was prepared for sequencing using the Illumina TruSeq sample preparation kit v2 using poly-A selection (Illumina San Diego, USA). The fragments were clustered on cBot and sequenced with paired ends (2 × 100bp) on a HiSeq 2500 (Illumina San Diego, USA) according to the manufacturer's instructions. A total 106.2 million read-pairs were generated. Adaptor sequences were removed from raw reads and reads were trimmed at the ends to phred score 20, using the fastq-mcf tool from ea-utils (https://code.google.com/p/ea-utils/). Processed reads were assembled using Trinity (r2013-02-16) resulting in a total of 263,652 assembled putative transcripts. Transcript abundance estimation was performed using RSEM and the scripts provided with Trinity. Likewise, the putative coding sequences were predicted using the TransDecoder scripts from Trinity. The high throughput RNA sequences reported here have been submitted to the short read archive (SRA) at the NCBI [accession no SAMN06013363]

Identification and cloning of genes involved in forskolin biosynthesis

Mining of the *C. forskohlii* transcriptome database was performed as previously described (***Zerbe et al., 2013***) using tBLASTx software and known cytochrome P450 (CYP) or acetyl transferase (ACT) sequences as query. The identified contigs were amplified from single stranded cDNA generated from root cork total RNA using the “SuperScript III First-Strand Synthesis System for RT-PCR” (Invitrogen) and oligo-dT primer. Cloning of the putative CYP and ACT cDNAs was achieved after PCR amplification using gene specific primers (***Figure 1-source data 3***) that were designed based on the *in silico* sequences of the identified CYP and ACT contigs (***Figure 1-source data 1***). PCR products were cloned into the pJET1.2 vector and verified by sequencing. For the identified non-

443 full-length cDNAs, (*Cf*CYP76AH15, *Cf*CYP76AH16 and *Cf*ACT1-8), full-length transcripts were
444 obtained using 5'RACE techniques.
445
446 Phylogenetic analysis of CYP76AH and ACT candidate enzymes
447 Amino acid sequences of the CYP76AHs of *C. forskohlii* and of currently available CYP76AHs from
448 other plant species were used to construct a phylogenetic tree to infer the evolutionary history of
449 these enzymes. Sequences were retrieved from the current work, GenBank database
450 (http://www.ncbi.nlm.nih.gov/) and original articles. Peptide alignments were performed using
451 the MUSCLE program included in the MEGA6 software. Phylogenetic analyses were performed by
452 the Maximum Likelihood method based on the Dayhoff matrix-based model with "uniform rates"
453 (***Schwarz et al., 1979***) and using all sites with "Nearest-Neighbor-Interchange" heuristic method
454 conducted using MEGA6 software (***Tamura et al., 2013***). The tree is drawn to scale, with branch
455 lengths measured in the number of substitutions per site. Bootstrap values shown in % were
456 inferred from 1000 replicates. Branches supported by bootstrap values higher than 75% are
457 shown.
458 For the phylogenetic analyses of *Cf*ACT candidates, the identified sequences from *C. forskohlii*
459 were analyzed together with BAHD family acyltransferase representatives from all clades (***D'Auria***
460 ***2006***). The analysis includes only functionally characterized members. The analysis was inferred by
461 using the Maximum Likelihood method based on the Dayhoff matrix-based model with "uniform
462 rates" and using all sites (***Schwarz et al., 1979***). Initial trees for the heuristic search were obtained
463 automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances
464 estimated using a JTT model, and then selecting the topology with superior log likelihood value.

The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Evolutionary analyses were conducted using MEGA6 software (***Tamura et al., 2013***).

RNA extraction and Quantitative Real-Time PCR

Total RNAs from *C. forskohlii* tissues were extracted as previously described (***Pateraki et al., 2014***) and digested by DNase I on-column. The integrity of the RNA samples was evaluated using the RNA-nano assay using an Agilent 2100 Bioanalyser (Agilent Technologies). First-strand cDNAs were synthesized from 0.5 µg of total RNA from an oligo-dT primer, using the "SuperScript III First-Strand Synthesis System for RT-PCR" (Invitrogen). The resulting cDNAs were diluted 10-fold. For the qRT-PCR reactions, gene specific primers were used (***Figure 1-source data 3***) with Maxima SYBR Green/Fluorescein qPCR Master Mix (Fermentas) on a Rotor-Gene Q cycler (Qiagen) using the following cycling parameters: 95 °C for 7 min, 35 cycles of 95 °C for 15 s, 60 °C for 30 s and 72 °C for 30 s followed by a melting curve cycle from 60 °C to 90 °C. Eukaryotic initiation factor 4A (TIF4a) and Elongation Factor 1A (EF1a) were used as reference genes because they showed the lowest variation across different tissues (***Pateraki et al., 2014***). No statistically significant differences were observed between the results obtained using the 2 different reference genes and the results presented were normalized based on TIF4a. Relative transcript abundance was calculated as the mean of three biological replications obtained using three different *C. forskohlii* plants, while the reactions were performed in three technical replicates. Amplification efficiency was calculated with the "Real Time PCR Miner" (http://www.miner.ewindup.info/Version2). Efficiency-corrected ΔCT values were used to quantify relative differences in target gene transcript accumulation. Primer specificity was assessed by agarose gel analysis and sequencing of amplicons from representative reactions, as well as from melting curve analysis of every reaction.

Functional characterization of *Coleus forskohlii* cytochrome P450 enzymes (*Cf*CYPs) by transient expression in *N. benthamiana*

Functional characterization of the selected candidate genes was obtained using transient expression in *N. benthamiana* which offers optimal native plant protein translation and processing, convenient rapid and optional combinatorial co-expression of multiple genes from independent vectors, native subcellular location of diTPS and CYPs as well as the availability of an endogenous native pathway providing GGPP.

For the functional characterization of *C. forskohlii* selected CYPs and testing of their ability to hydroxylate 13*R*-manoyl oxide (**1**), genes encoding candidate *Cf*CYPs were transiently co-expressed in *N. benthamiana* leaves together with *C. forskohlii* enzymes boosting the formation of **1**, namely 1-deoxy-D-xylulose 5-phosphate synthase (*Cf*DXS), geranylgeranyl diphosphate synthase (*Cf*GGPPS), *Cf*TPS2 and *Cf*TPS3 (***Pateraki et al., 2014***). Transient expression in *N. benthamiana* was performed as previously described (***Spanner et al., 2014***). *Cf*CYP cDNAs selected for functional testing were subcloned into pCAMBIA130035Su by USER cloning (***Nour-Eldin et al., 2010***). Vectors carrying the selected cDNAs were then transformed into the agrobacterium strain AGL-1-GC3850 (***Spanner et al., 2014***). For the agro-infiltration, the OD of the agrobacterium cultures of all transformed agrobacteria strains was normalized to OD_{600} = 1. Different combinations prepared from equal volumes of each culture of transformed agrobacterium strains expressing individual genes encoding *Cf*CYPs or 13*R*-manoyl oxide biosynthetic enzymes were infiltrated into the leaves of 4-6 weeks old *N. benthamiana* plants. Controls encompassing expression of the *Cf*CYP encoding genes without the 13*R*-manoyl oxide biosynthetic genes were used to assess the possibility of *Cf*CYPs cross-reactivity with tobacco endogenous diterpenoids.

Metabolites from transgenic *N. benthamiana* leaves were extracted by hexane and 85% MetOH and were analyzed by GC-MS and LC-MS-qTOF, respectively. LC-MS-qTOF analysis was performed on the system that was comprised of an Agilent G1312B SL binary pump, Agilent G1367C SL WP autosampler, Agilent G1316B column oven, Agilent G1315C Starlight DAD detector and Bruker microTOF II Mass Spectrometer using Electron Spray Ionization (ESI). Samples were separated on a Synergi 2.5 mm Fusion-RP C18 column (50 × 3.2 mm i.d., Phenomenex Inc., Torrance, CA, USA) at a flow rate of 0.2 mL/min with column temperature held at 25 °C. The mobile phase consisted of water with 0.1% formic acid (v/v; solvent A) and 80% acetonitrile with 0.1% formic acid (v/v; solvent B). The gradient program was 0 min, 60% B; 25 min, 98% B; 31 min, 98% B; 32 min, 60% B; 42min, 60% B. Mass spectra were acquired in positive ion mode using a drying temperature of 200 °C, a nebulizer pressure of 3.0 bar, and a drying gas flow of 7 L/min (***Luo et al., 2016***).

GC-MS analysis was performed on a Shimadzu GCMS-QP2010 Ultra using a 3 Agilent HP-5MS column (20 m × 0.180 mm i.d., 0.18 µm film thickness). Injection volume and temperature was set to 1 µL and 250 °C. GC program: 60 °C for 1 min, ramp at rate 30 °C min-1 to 180 °C, ramp at rate 10 °C min-1 to 290 °C, ramp at rate 30 °C min-1 to 320 °C and hold for 2 min. H_2 was used as carrier gas. Transfer line temperature was set to 280 °C and electron impact (EI) was used as ionization method in the mass spectrometer (MS) with the ion source temperature and voltage set to 300 °C and 70 eV. MS spectra were recorded from 50 m/z to 400 m/z.

<u>Large scale biosynthesis of oxygenated 13*R*-manoyl oxide compounds for NMR analysis</u>

11-oxo-13*R*-manoyl oxide (**2**), 2-hydroxy-13*R*-manoyl oxide (**3b**), 19-hydroxy-13*R*-manoyl oxide (**3c**), 1,11-dihydroxy-13*R*-manoyl oxide (**5d**), 1,9-deoxydeacetylforskolin (**7h**) and 9-deoxydeacetylforskolin (**10b**) were produced using the biosynthetic scheme described above by

large-scale expression of the relevant gene combinations in *N. benthamiana (**Andersen-Ranberg et al., 2016**)*. *CYP76AH8* was expressed to obtain biosynthesis of compound (**2**), *CYP71D381* to obtain (**3b**) and (**3c**), whereas combined expression of *CfCYP76AH8* and Cf*CYP76AH11* afforded (**5d)**, (**7h)** and (**10b)**. For the large-scale experiments, agroinfiltration was performed by vacuum infiltration. For biosynthesis of each compound in amounts sufficient for NMR analysis, 100-200 g of fresh weight *N. benthamiana* leaf material were harvested 7 d after infiltration. Leaf material chopped into small pieces was extracted using 0.5 L of *n*-hexane. The solvent was thrice evaporated and recovered by rotor evaporation for repeated extraction of the same leaf material. Concentrated extracts from *N. benthamiana* leaves biosynthesizing target compounds were subjected to solid phase extraction (SPE), using silica gel with *n*-hexane and ethyl acetate mixtures (100:0, 99:1, 98:2, 94:6, 92:8, 88:12 (v/v)) in steps of 100 mL. Fractions containing diterpenoids were identified by GC-MS. Diterpenoid-containing fractions were combined and the solvent removed by rotor evaporation and the samples resuspended in 1 mL *n*-hexane affording a crude fraction for further purification.

Isolation of oxygenated 13*R*-manoyl oxide-derived diterpenoids from *N. benthamiana* extracts

Final isolation of individual diterpenoids was achieved using an Agilent 7890B GC installed with an Agilent 5977A MSD, GERSTEL Preparative Fraction Collector (PFC) AT 6890/7890 and a GERSTEL CIS 4C Bundle injection port. Separation was carried out using a RESTEK Rtx-5 column (30 m $\times$ 0.53 mm i.d. $\times$ 1 µm d_f) with H_2 as carrier gas. At the column outlet, a splitter was mounted with a split vent ratio of 1:100 to the MS and the PFC, respectively. Sufficient amounts of oxidized 13*R*-manoyl oxide-derived diterpenoids for NMR analysis (0.5-1 mg) were obtained by 100 repeated injections of 5 µL extract aliquots. Injection port was set in solvent vent mode with a H_2 gas flow of 100

mL/min until 0.17 min, combined with a sample injection speed of 1.5 mL/min. Purge flow was set to 3 mL/min from 0.17 min to 2.17 min. Injection temperature was held at 40 °C for 0.1 min followed by ramping at 12 °C/sec until 320 °C, which was held for 2 min. Column flow was set to 7.5 mL, which was held constant throughout the GC program. The GC program was set to hold at 60 °C for 1 min, ramp 30 °C/min to 220 °C, ramp 2 °C/min to 250 °C and a final ramp of 30 °C/min to 220 °C, which was held for 2 min. Temperature of the transfer line from GC to PFC and the PFC itself was set to 250 °C. The PFC was customized to collect the peaks of **2**, **3b**, **3c**, **5d** and **7h**, and **10b** by their retention time identified by the MS. The MS for monitoring the PFC purification was set in scan mode from *m*/*z* 35 to *m*/*z* 500, with a threshold of 150 ion counts. Solvent cut-off was set to 4 min and the temperature of the MS source and the MS quadropole - to 300 °C and 150 °C, respectively.

Large scale biosynthesis in *S. cereviseae* and isolation of 9-hydroxy-13*R*-manoyl oxide (**3a**) for NMR analysis

For biosynthesis of **3a** in amounts sufficient for NMR analysis, the *S. cerevisiae* strain EFSC4494 carrying chromosomally integrated *CYP76AH16* was inoculated into a pre-culture of 5 mL selective media (SC-Ura) and grown for 16 h at 30 °C and 400 rpm. A 1 mL aliquot of the pre-culture was used for inoculation of 100 mL non-selective media (SC) and grown in a 500 mL Erlenmeyer shake-flask for 120 h at 30 °C with horizontal shaking at 180 rpm. Following addition of 100 mL UV-grade 99.9% ethanol and maintenance of the sample at 60 °C for 20 min, 200 mL of *n*-hexane was added and the sample shaken for 2 h at room temperature. The hexane phase was concentrated by rotor evaporation and subjected to column chromatography (dual layer Florisil/Na_2SO_4 6mL PP SPE TUBE, Supelco Analytical) with a gradient composed of *n*-hexane and 1-15% ethyl acetate.

Metabolic engineering of *S. cerevisiae* for the production of forskolin

All genes selected for functional expression in *S. cerevisiae* were codon optimized for efficient expression in *S. cerevisiae,* and purchased as DNA STRINGS (Geneart, LifeTechnologies). Genomic integration was chosen over expression via episomal plasmids to favor simultaneous expression of a number of genes as well as to enable the use of selection marker recycling (***Jensen et al., 2014***). The 13*R*-manoyl oxide producing *S. cerevisiae* strain EFSC4498 (***Andersen-Ranberg et al., 2016***) was used to test all the selected gene combinations for their ability of affording synthesis of forskolin and intermediate products. All genes were cloned into yeast genome integration plasmids by the USER technique (***Nour-Eldin et al., 2010***) targeting incorporation into site XI-2 (***Mikkelsen et al., 2012***). Transformants were verified by PCR on genomic DNA for correct insertion of heterologous genes and grown and tested in 96 deep-well plates (***Andersen-Ranberg et al., 2016***). The yeast strain found to produce the highest amount of forskolin and which exhibited stability through several cultivation rounds (EVST21543) was selected for cultivation for 140 h in a 5 L fermenter using minimal medium and glucose limited conditions. Forskolin production was monitored using withdrawn culture aliquots. Forskolin was extracted from the mixture of yeast cells and culture broth using 85% methanol and incubation for 20 min at 75 °C and the extract centrifuged (10000 g for 5 min) to precipitate yeast debris. The supernatant obtained was used after filtration for LC-MS analysis and forskolin quantification.

Forskolin quantification synthesized from yeast strain EVST21543 growing in the fermenter and comparison to *C. forskohlii* root extract

For forskolin quantification, aliquots of the yeast samples (with broth) collected at specific time points (***Figure 3***) were combined with methanol to give a concentration of 85% methanol, incubated at 75 °C for 20 min, filtered and then analyzed by LC-MS. Quantification was based on a standard calibration curve of forskolin purchased from Sigma-Aldrich. An Ultimate 3000 UHPLC^{+} Focused system (Dionex Corporation, Sunnyvale, CA, USA) coupled to a Bruker Compact ESI-QTOF-MS (Bruker Daltonik, Bremen, Germany) was used to quantify forskolin. Samples were separated on a Kinetex XB-C18 column (100 × 2.1 mm i.d., 1.7 µm particle size, 100 Å pore size; Phenomenex Inc., Torrance, CA, USA) maintained at 40°C with a flow rate of 0.3 mL/min and mobile phase consisting of 0.05% (v/v) formic acid in water (solvent A) and 0.05% (v/v) formic acid in acetonitrile (solvent B). The gradient LC method used for quantification was as follows: solvent B was held at 20% for 30 s, then ramped to 100% over 8.5 min, held at 100% for 2 min, decreased to 20% over 30 s and held for 3.5 min to give an overall run time of 15 min. The ESI source parameters were as follows: capillary voltage, 4500 V; nebulizer pressure 1.2 bar; dry gas flow, 8 l/min; dry gas temperature, 250°C. The QTOF-MS was operated in MS only mode with collision cell energy of 7 eV and collision cell RF of 500 Vpp. Ions were monitored in the positive mode over a range of 50-1300 *m*/*z* and spectra collected at a rate of 2 Hz. For comparison of yeast profiles to *C. forskohlii* root extract, roots were grinded and then extracted with 85% methanol, incubated at 75 °C for 30 min, filtered and analyzed by LC-MS. Analysis was performed as described for forskolin quantification but with the following gradient method: 20 % B for 1 min, increased to 100 % B over 22 min and then returned to 20% B in 0.5 min and held for 4 min.

Quantification of 13*R*-manoyl oxide (**1**) and 9-hydroxy-13*R*-manoyl oxide (**3a**) synthesized from yeast strain EVST21543 growing in the fermenter

Yeast samples were collected at specific time points (***Figure 10***) and samples kept at −20 °C in glass vials. For diterpenoid extraction, 500 µL of *n*-hexane was added to 500 µL yeast broth, shaken for 1 h at room temperature and separated into two phases by centrifugation at 2500 rpm and stored overnight at 4°C. The hexane phase was then diluted 10 times and run on a SCION 436 GC-FID (Bruker). Sample (1 µL) was injected in splitless mode at 280 °C. The GC-program was as follows: 60 °C for 1 min, ramp at 20 °C/min to 160 °C, ramp at 10 °C/min from 160 to 240 °C, ramp at 20 °C/min from 240 to 320 °C, hold at 320 °C for 2 min. H_2 was used as carrier gas with a linear flow of 50 mL/min. The FID was set at 300 °C, with N_2 flow of 25 mL/min, H_2 at 30 mL/min and air at 300 mL/min. Data sampling rate was 10 Hz. Compound **1** and **3a** was identified by comparing the retention time with an authentic standard and quantification was based on FID peak area and a standard curve of **1**.

NMR analysis

All NMR experiments were recorded at 300 K in $CDCl_3$ using a Bruker Avance III 600 MHz NMR spectrometer (^{1}H operating frequency 600.13 MHz) equipped with a Bruker SampleJet sample changer and a cryogenically cooled gradient inverse triple-resonance 1.7-mm TCI probe-head (Bruker Biospin, Rheinstetten, Germany). The experiments were acquired in automation (temperature equilibration to 300 K, optimization of lock parameters, gradient shimming, and setting of receiver gain). Both one-dimensional ^{1}H and ^{13}C spectra were acquired with 30°-pulses and 64k data points. The ^{1}H spectra were recorded with 3.66 s inter-pulse intervals and the FID was multiplied with an exponential function corresponding to line-broadening of 0.3 Hz prior to Fourier transform. An acquisition time of 0.9 s were used for the ^{13}C experiments with an additional relaxation delay of 2.0 s. Protons were decoupled during acquisition using waltz16

composite pulse sequence. Backward linear prediction was used to correct the first complex data points of ^{13}C FIDs before zero-filling to 128k data points and application of exponential window function with a line-broadening factor of 1.0 Hz. Two-dimensional homo- and heteronuclear experiments were acquired with 2048 data points in the direct dimension and 128 (DQF-COSY and HMBC) or 256 (multiplicity edited HSQC and phase sensitive NOESY) data points in the indirect dimension; with spectral widths optimized from the corresponding ^{1}H spectra. The HMBC and HSQC experiments were optimized for $^{n}J_{H,C}$ = 10 Hz and $^{1}J_{H,C}$ = 145 Hz, respectively. Acquisition and processing of NMR data were performed using Topspin ver. 3.0 (Bruker Biospin GmbH), and IconNMR ver. 4.2 (Bruker Biospin GmbH) was used for controlling automated sample change and acquisition.

Acknowledgements: We thank Dr. Sotirios Kampranis for providing ferruginol and 11β-hydroxy-13*R*-manoyl oxide reference compounds and Dr. Angelos Kanellis for providing RoFS1 and SfFS cDNAs. This work was supported by the VILLUM research center "Plant Plasticity" (B.L.M.), the Center for Synthetic Biology "bioSYNergy" supported by the UCPH Excellence Program for Interdisciplinary Research (B.L.M.), the Novo Nordisk Foundation (Bj.H.), an ERC Advanced Grant to BLM (ERC-2012-ADG_20120314), individual Marie Skłodowska-Curie fellowships to I.P. and A.M.H. Bj.H is in part supported by the DOE Great Lakes Bioenergy Research Center (DOE Office of Science BER DE-FC02-07ER64494) and gratefully acknowledges the Strategic Partnership Grant (15-SPG-Full-3101), MSU Foundation, startup funding from the Department of Molecular Biology and Biochemistry, Michigan State University and support from Michigan State University AgBioResearch (MICL02454).

Competing interests: J.A-R., N.B.J., I.P., B.L.M. and Bj.H. have filed international patent applications (PCT/DK2015/050020) covering 'Biosynthesis of forskolin and related compounds'. N.B.J. and J.H. are employees of Evolva SA

675 **References**

676 Alasbahi R, Melzig M. 2010a. Plectranthus barbatus: a review of phytochemistry, ethnobotanical uses and
677 pharmacology - Part 1. *Planta Medica* **76:** 653-661.
678 Alasbahi R, Melzig M. 2010b. Plectranthus barbatus: a review of phytochemistry, ethnobotanical uses and
679 pharmacology - part 2. *Planta Medica* **76:** 753-765.
680 Andersen-Ranberg J, Kongstad KT, Nielsen MT, Jensen NB, Pateraki I, Bach SS, Hamberger B, Zerbe P, Staerk
681 D, Bohlmann J, Møller BL, Hamberger B. 2016. Expanding the landscape of diterpene structural diversity
682 through stereochemically controlled combinatorial biosynthesis. *Angewandte Chemie International Edition*
683 **55:** 2142-2146.
684 Appendino G. 2014. Omnia praeclara rara. The quest for ingenol heats up. *Angewandte Chemie*
685 *International Edition* **53:** 927-929.
686 Asada Y, Li W, Terada T, Kuang X, Li Q, Yoshikawa T, Hamaguchi S, Namekata I, Tanaka H, Koike K. 2012.
687 Labdane-type diterpenoids from hairy root cultures of *Coleus forskohlii*, possible intermediates in the
688 biosynthesis of forskolin. *Phytochemistry* **79:** 141-146.
689 Biggs BW, Lim CG, Sagliani K, Shankar S, Stephanopoulos G, De Mey M, Ajikumar PK. 2016. Overcoming
690 heterologous protein interdependency to optimize P450-mediated Taxol precursor synthesis in Escherichia
691 coli. *Proceedings of the National Academy of Sciences of the United States of America* **113:** 3209-3214.
692 Bozic D, Papaefthimiou D, Bruckner K, Vos RC, Tsoleridis CA, Katsarou D, Papanikolaou A, Pateraki I,
693 Chatzopoulou FM, Dimitriadou E. 2015. Towards elucidating carnosic acid biosynthesis in Lamiaceae:
694 functional characterization of the three first steps of the pathway in Salvia fruticosa and Rosmarinus
695 officinalis. *PLoS ONE* **10:**
696 Božić D, Papaefthimiou D, Brückner K, de Vos RCH, Tsoleridis CA, Katsarou D, Papanikolaou A, Pateraki I,
697 Chatzopoulou FM, Dimitriadou E, Kostas S, Manzano D, Scheler U, Ferrer A, Tissier A, Makris AM, Kampranis
698 SC, Kanellis AK. 2015. Towards elucidating carnosic acid biosynthesis in Lamiaceae: Functional
699 characterization of the three first steps of the pathway in *Salvia fruticosa* and *Rosmarinus officinalis*. *PLoS*
700 *ONE* **10:** e0124106.
701 Brochado AR, Matos C, Møller BL, Hansen J, Mortensen UH, Patil KR. 2010. Improved vanillin production in
702 baker's yeast through in silico design. *Microbial Cell Factories* **9:** 1-15.
703 Brown S, Clastre M, Courdavault V, O'Connor SE. 2015. De novo production of the plant-derived alkaloid
704 strictosidine in yeast. *Proceedings of the National Academy of Sciences of the United States of America* **112:**
705 3205-3210.
706 Cragg GM, Newman DJ. 2013. Natural products: A continuing source of novel drug leads. *Biochimica et*
707 *Biophysica Acta (BBA) - General Subjects* **1830:** 3670-3695.
708 D'Orazio JA, Nobuhisa T, Cui R, Arya M, Spry M, Wakamatsu K, Igras V, Kunisada T, Granter SR, Nishimura
709 EK, Ito S, Fisher DE. 2006. Topical drug rescue strategy and skin protection based on the role of Mc1r in UV-
710 induced tanning. *Nature* **443:** 340-344.
711 D'Auria JC. 2006. Acyltransferases in plants: a good time to be BAHD. *Current Opinion in Plant Biology* **9:**
712 331-340.
713 David B, Wolfender J-L, Dias D. 2015. The pharmaceutical industry and natural products: historical status
714 and new trends. *Phytochemistry Reviews* **14:** 299-315.
715 Doseyici S, Mehmetoglu I, Toker A, Yerlikaya F, Erbay E. 2014. The effects of forskolin and rolipram on
716 cAMP, cGMP and free fatty acid levels in diet induced obesity. *Biotechnic & Histochemistry* **89:** 388-392.
717 Englund E, Andersen-Ranberg J, Miao R, Hamberger B, Lindberg P. 2015. Metabolic engineering of
718 Synechocystis sp. PCC 6803 for production of the plant diterpenoid manoyl oxide. *ACS Synthetic Biology* **4:**
719 Gabetta B, Zini G, Danieli B. 1989. Minor diterpenoids of *Coleus forskohlii*. *Phytochemistry* **28:** 859-862.
720 Galanie S, Thodey K, Trenchard IJ, Filsinger Interrante M, Smolke CD. 2015. Complete biosynthesis of
721 opioids in yeast. *Science* **349:** 1095-1100.

Godard MP, Johnson BA, Richmond SR. 2005. Body composition and hormonal adaptations associated with forskolin consumption in overweight and obese men. *Obes Res* **13:** 1335-1343.
Graham IA, Besser K, Blumer S, Branigan CA, Czechowski T, Elias L, Guterman I, Harvey D, Isaac PG, Khan AM, Larson TR, Li Y, Pawson T, Penfield T, Rae AM, Rathbone DA, Reid S, Ross J, Smallwood MF, Segura V, Townsend T, Vyas D, Winzer T, Bowles D. 2010. The Genetic Map of *Artemisia annua* L. Identifies Loci Affecting Yield of the Antimalarial Drug Artemisinin. *Science* **327:** 328-331.
Guo J, Ma X, Cai Y, Ma Y, Zhan Z, Zhou YJ, Liu W, Guan M, Yang J, Cui G, Kang L, Yang L, Shen Y, Tang J, Lin H, Ma X, Jin B, Liu Z, Peters RJ, Zhao ZK, Huang L. 2015. Cytochrome P450 promiscuity leads to a bifurcating biosynthetic pathway for tanshinones. *New Phytologist* **210:** 525–534.
Guo J, Zhou YJ, Hillwig ML, Shen Y, Yang L, Wang Y, Zhang X, Liu W, Peters RJ, Chen X, Zhao ZK, Huang L. 2013. CYP76AH1 catalyzes turnover of miltiradiene in tanshinones biosynthesis and enables heterologous production of ferruginol in yeasts. *Proceedings of the National Academy of Sciences of the United States of America* **110:** 12108-12113.
Hansen EH, Møller BL, Kock GR, Bünner CM, Kristensen C, Jensen OR, Okkels FT, Olsen CE, Motawia MS, Hansen J. 2009. De Novo Biosynthesis of Vanillin in Fission Yeast (Schizosaccharomyces pombe) and Baker's Yeast (Saccharomyces cerevisiae). *Applied and Environmental Microbiology* **75:** 2765-2774.
Hardy K, Buckley S, Collins M, Estalrrich A, Brothwell D, Copeland L, García-Tabernero A, García-Vargas S, de la Rasilla M, Lalueza-Fox C, Huguet R, Bastir M, Santamaría D, Madella M, Wilson J, Cortés Á, Rosas A. 2012. Neanderthal medics? Evidence for food, cooking, and medicinal plants entrapped in dental calculus. *Naturwissenschaften* **99:** 617-626.
Ignea C, Athanasakoglou A, Ioannou E, Georgantea P, Trikka FA, Loupassaki S, Roussis V, Makris AM, Kampranis SC. 2016a. Carnosic acid biosynthesis elucidated by a synthetic biology platform. *Proceedings of the National Academy of Sciences of the United States of America* **113:** 3681-3686.
Ignea C, Ioannou E, Georgantea P, Trikka FA, Athanasakoglou A, Loupassaki S, Roussis V, Makris AM, Kampranis SC. 2016b. Production of the forskolin precursor 11β-hydroxy-manoyl oxide in yeast using surrogate enzymatic activities. *Microbial Cell Factories* **15:** 1-11.
Jeandet P, Delaunois B, Aziz A, Donnez D, Vasserot Y, Cordelier S, Courot E. 2012. Metabolic engineering of yeast and plants for the production of the biologically active hydroxystilbene, resveratrol. *Journal of Biomedical Biotechnololy* **2012:**
Jensen NB, Strucko T, Kildegaard KR, David F, Maury J, Mortensen UH, Forster J, Nielsen J, Borodina I. 2014. EasyClone: method for iterative chromosomal integration of multiple genes Saccharomyces cerevisiae. *FEMS Yeast Res* **14:** 238-248.
Jia M, Potter KC, Peters RJ. 2016. Extreme promiscuity of a bacterial and a plant diterpene synthase enables combinatorial biosynthesis. *Metabolic Engineering* **37:** 24-34.
Jung ST, Lauchli R, Arnold FH. 2011. Cytochrome P450: taming a wild type enzyme. *Current Opinion in Biotechnology* **22:** 809-817.
Kampranis SC, Makris AM. 2012. Developing a yeast cell factory for the production of terpenoids. *Computational and Structural Biotechnology Journal* **3:** e201210006.
Kavitha C, Rajamani K, Vadivel E. 2010. Coleus forskohlii – A comprehensive review on morphology, phytochemistry and pharmacological aspects. *Journal of Medicinal Plants Research* **4:** 278-285.
Kawamura S, Chu H, Felding J, Baran PS. 2016. Nineteen-step total synthesis of (+)-phorbol. *Nature* **532:** 90-93.
Kikura M, Morita K, Sato S. 2004. Pharmacokinetics and a simulation model of colforsin daropate, new forskolin derivative inotropic vasodilator, in patients undergoing coronary artery bypass grafting. *Pharmacological Research* **49:** 275-281.
King AJ, Brown GD, Gilday AD, Forestier E, Larson TR, Graham IA. 2016. A Cytochrome P450-Mediated Intramolecular Carbon–Carbon Ring Closure in the Biosynthesis of Multidrug-Resistance-Reversing Lathyrane Diterpenoids. *ChemBioChem* **17:** 1593-1597.
Kosley RW, Cherill RJ. 1989. Regioselective acylations of 7-desacetylforskolin. *The Journal of Organic Chemistry* **54:** 2972-2975.

Laursen T, Borch J, Knudsen C, Bavishi K, Torta F, Martens H, Silvestro D, Hatzakis N, Wenk M, Dafforn T, Olsen C, Motawia M, Hamberger B, Møller BL, Bassard J-E. 2016. Characterization of a dynamic metabolon producing the defense compound dhurrin in sorghum. *Science* **in press:**
Li Y, Smolke CD. 2016. Engineering biosynthesis of the anticancer alkaloid noscapine in yeast. *Nat Commun* **7:** 12137.
Liu T, Khosla C. 2010. A Balancing Act for Taxol Precursor Pathways in *E. coli*. *Science* **330:** 44-45.
Luo D, Callari R, Hamberger B, Wubshet SG, Nielsen MT, Andersen-Ranberg J, Hallström BM, Cozzi F, Heider H, Lindberg Møller B, Staerk D, 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:** E5082-E5089.
Mikkelsen MD, Buron LD, Salomonsen B, Olsen CE, Hansen BG, Mortensen UH, Halkier BA. 2012. Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. *Metabolic Engineering* **14:** 104-111.
Mithöfer A, Boland W. 2012. Plant Defense Against Herbivores: Chemical Aspects. *Annual Review of Plant Biology* **63:** 431-450.
Mora-Pale M, Sanchez-Rodriguez SP, Linhardt RJ, Dordick JS, Koffas MAG. 2014. Biochemical strategies for enhancing the in vivo production of natural products with pharmaceutical potential. *Current Opinion in Biotechnology* **25:** 86-94.
Nakagawa A, Matsumura E, Koyanagi T, Katayama T, Kawano N, Yoshimatsu K, Yamamoto K, Kumagai H, Sato F, Minami H. 2016. Total biosynthesis of opiates by stepwise fermentation using engineered Escherichia coli. *Nat Commun* **7:10390:**
Nelson D, Werck-Reichhart D. 2011. A P450-centric view of plant evolution. *The Plant Journal* **66:** 194-211.
Nelson DR. 2009. The Cytochrome P450 Homepage. *Human Genomics* **4:** 59-65.
Nelson DR. 2013. A world of cytochrome P450s. *Philosophical Transactions of the Royal Society B: Biological Sciences* **368:**
Nielsen MT, Ranberg JA, Christensen U, Christensen HB, Harrison SJ, Olsen CE, Hamberger B, Møller BL, Nørholm MHH. 2014. Microbial Synthesis of the Forskolin Precursor Manoyl Oxide in an Enantiomerically Pure Form. *Applied and Environmental Microbiology* **80:** 7258-7265.
Nour-Eldin H, Geu-Flores F, Halkier B (2010). USER Cloning and USER Fusion: The Ideal Cloning Techniques for Small and Big Laboratories. Plant Secondary Metabolism Engineering. AG Fett-Neto, Humana Press. **643:** 185-200.
Nour-Eldin HH, Hansen BG, Nørholm MHH, Jensen JK, Halkier BA. 2006. Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. *Nucleic Acids Research* **34:** e122.
Ondari ME, Walker KD. 2008. The Taxol Pathway 10-O-Acetyltransferase Shows Regioselective Promiscuity with the Oxetane Hydroxyl of 4-Deacetyltaxanes. *Journal of the American Chemical Society* **130:** 17187-17194.
Paddon CJ, Keasling JD. 2014. Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nature Reviews in Microbiology* **12:** 355-367.
Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, McPhee D, Leavell MD, Tai A, Main A, Eng D. 2013. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **496:**
Pateraki I, Andersen-Ranberg J, Hamberger B, Heskes AM, Martens HJ, Zerbe P, Bach SS, Møller BL, Bohlmann J, Hamberger B. 2014. Manoyl oxide (13*R*), the biosynthetic brecursor of forskolin, is synthesized in specialized root cork cells in *Coleus forskohlii*. *Plant Physiology* **164:** 1222-1236.
Pateraki I, Heskes AM, Hamberger B (2015). Cytochromes P450 for Terpene Functionalisation and Metabolic Engineering. Biotechnology of Isoprenoids. J Schrader, J Bohlmann. Cham, Springer International Publishing, 10.1007/10_2014_301**:** 107-139.
Peplow M. 2016. Synthetic biology's first malaria drug meets market resistance. *Nature* **530:** 389-390.
Renault H, Bassard J-E, Hamberger B, Werck-Reichhart D. 2014. Cytochrome P450-mediated metabolic engineering: current progress and future challenges. *Current Opinion in Plant Biology* **19:** 27-34.

Roberts SC. 2007. Production and engineering of terpenoids in plant cell culture. *Nature Chemical Biology* **3:** 387-395.
Schwarz R, Dayhoff M. 1979. Matrices for detecting distant relationships. *Atlas of protein sequences*, National Biomedical Research Foundation.
Seamon KB, Padgett W, Daly JW. 1981. Forskolin: unique diterpene activator of adenylate cyclase in membranes and in intact cells. *Proceedings of the National Academy of Sciences of the United States of America* **78:** 3363-3367.
Spanner SB, Bassard J-É, Andersen-Ranberg J, Møldrup ME, Simonsen HT, Hamberger B. 2014. High throughput testing of terpenoid biosynthesis candidate genes using transient expression in *Nicotiana benthamiana*. *Methods in Molecular Biology*, Humana Press.
Srivastava S, Misra A, Mishra P, Shukla P, Kumar M, Sundaresan V, Negi KS, Agrawal PK, Singh Rawat AK. 2017. Molecular and chemotypic variability of forskolin in Coleus forskohlii Briq., a high value industrial crop collected from Western Himalayas (India). *RSC Advances* **7:** 8843-8851.
Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* **30:** 2725-2729.
Toya Y, Schwencke C, Ishikawa Y. 1998. Forskolin Derivatives with Increased Selectivity for Cardiac Adenylyl Cyclase. *Journal of Molecular and Cellular Cardiology* **30:** 97-108.
Tuominen L, Johnson V, Tsai C-J. 2011. Differential phylogenetic expansions in BAHD acyltransferases across five angiosperm taxa and evidence of divergent expression among Populus paralogues. *BMC Genomics* **12:** 236.
Vazquez-Albacete D, Cavaleiro AM, Christensen U, Seppälä S, Møller BL, Nørholm MHH. 2016. An expression tag toolbox for microbial production of membrane bound plant cytochromes P450. *Biotechnology and Bioengineering*, 10.1002/bit.26203in press.
Wagh VD, Patil PN, Surana SJ, Wagh KV. 2012. Forskolin: Upcoming antiglaucoma molecule. *Journal of Postgraduate Medicine* **58:** 199-202.
Walker K, Croteau R. 2000. Molecular cloning of a 10-deacetylbaccatin III-10-O-acetyl transferase cDNA from Taxus and functional expression in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **97:** 583-587.
Werck-Reichhart D, Feyereisen R. 2000. Cytochromes P450: a success story. *Genome Biology* **1:** reviews 3003.3001.
Woldemariam MG, Baldwin IT, Galis I. 2011. Transcriptional regulation of plant inducible defenses against herbivores: a mini-review. *Journal of Plant Interactions* **6:** 113-119.
Ye H, Deng G, Liu J, Qiu FG. 2009. Expedient Construction of the Ziegler Intermediate Useful for the Synthesis of Forskolin via Consecutive Rearrangements. *Organic Letters* **11:** 5442-5444.
Yoneyama M, Sugiyama A, Satoh Y, Takahara A, Nakamura Y, Hashimoto K. 2002. Cardiovascular and Adenylate Cyclase Stimulating Effects of Colforsin Daropate, a Water-Soluble Forskolin Derivative, Compared With Those of Isoproterenol, Dopamine and Dobutamine. *Circulation Journal* **66:** 1150-1154.
Yuan C, Jin Y, Wilde NC, Baran PS. 2016. Short, Enantioselective Total Synthesis of Highly Oxidized Taxanes. *Angewandte Chemie International Edition* **55:** 8280-8284.
Zerbe P, Hamberger B, Yuen MMS, Chiang A, Sandhu HK, Madilao LM, Nghuen A, Hamberger B, Spanner Bach S, Bohlmann J. 2013. Gene discovery of modular diterpene metabolism in non-model systems. *Plant Physiology*, 10.1104/pp.113.218347
Zhang WW, Luo JG, Wang JS, Lu YY, Kong LY. 2009. LC–DAD–ESI-MS-MS for characterization and quantitative analysis of diterpenoids from *Coleus forskohlii*. *Chromatographia* **70:** 1635-1643.
Zhou K, Qiao K, Edgar S, Stephanopoulos G. 2015. Distributing a metabolic pathway among a microbial consortium enhances production of natural products. *Nature Biotechnology* **33:** 377-383.
Zi J, Peters RJ. 2013. Characterization of CYP76AH4 clarifies phenolic diterpenoid biosynthesis in the Lamiaceae. *Organic and Biomolecular Chemistry* **11:** 7650-7652.

870

Figures

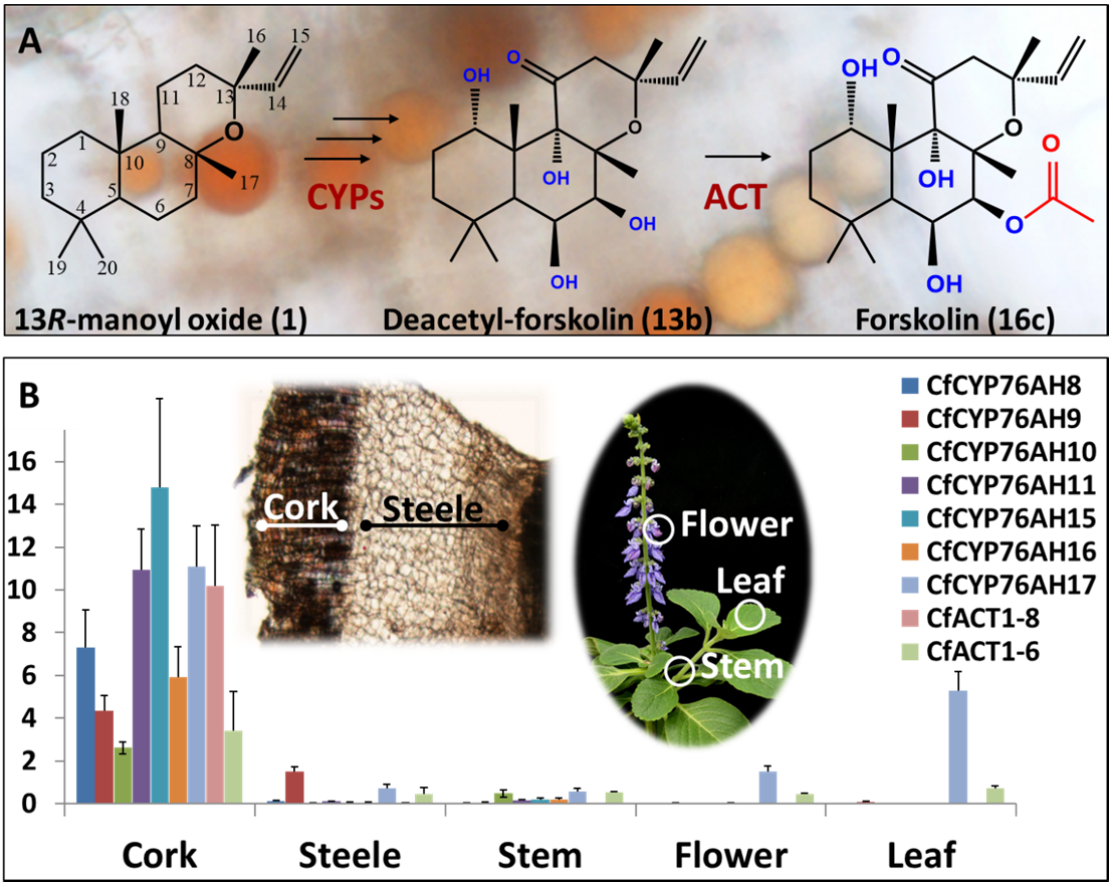

Figure 1. Biosynthesis of forskolin in the root cork cells of *C. forskohlii.* (**A**) Scheme showing the structures of the diterpene precursor 13*R*-manoyl oxide, deacetylforskolin and forskolin on a background of root cork cells with forskolin containing oil bodies. (**B**) Transcript profiles of biosynthetic candidate genes in selected tissues of *C. forskohlii* as shown on the illustrations.

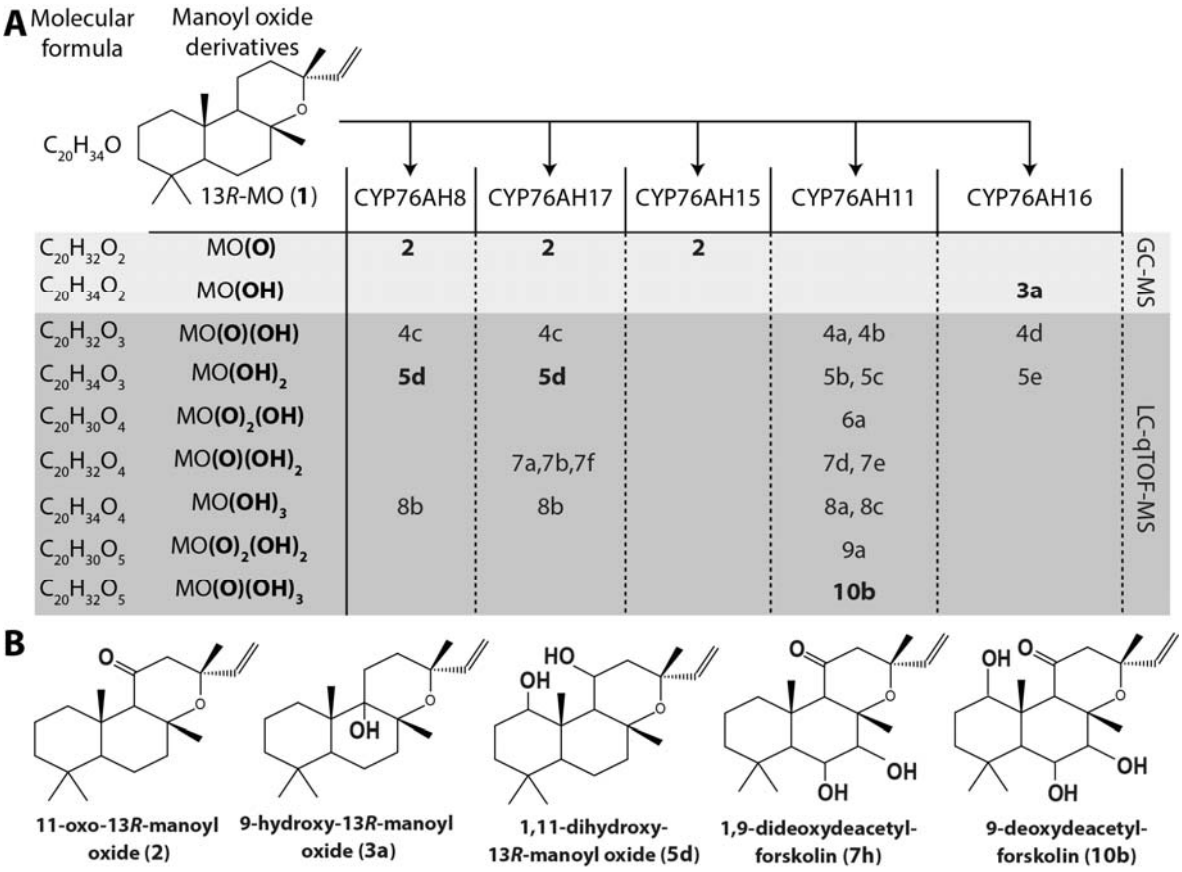

Figure 2. 13*R*-manoyl oxide oxide-derived hydroxylated products formed following transient expression of *Cf*CYP76AHs in *N. benthamiana*. (**A**) Molecular formulas of the hydroxylated products obtained using different *Cf*CYP76AHs. The number of hydroxylations of each compound was deduced from its accurate molecular mass (<5ppm, ***Table 3***) as determined by LC-qTOF-MS or NMR. Each different compound is marked by a number. (**B**) Chemical structures of the compounds marked with numbers in bold in A as determined by NMR (***Table 1 and 2***).

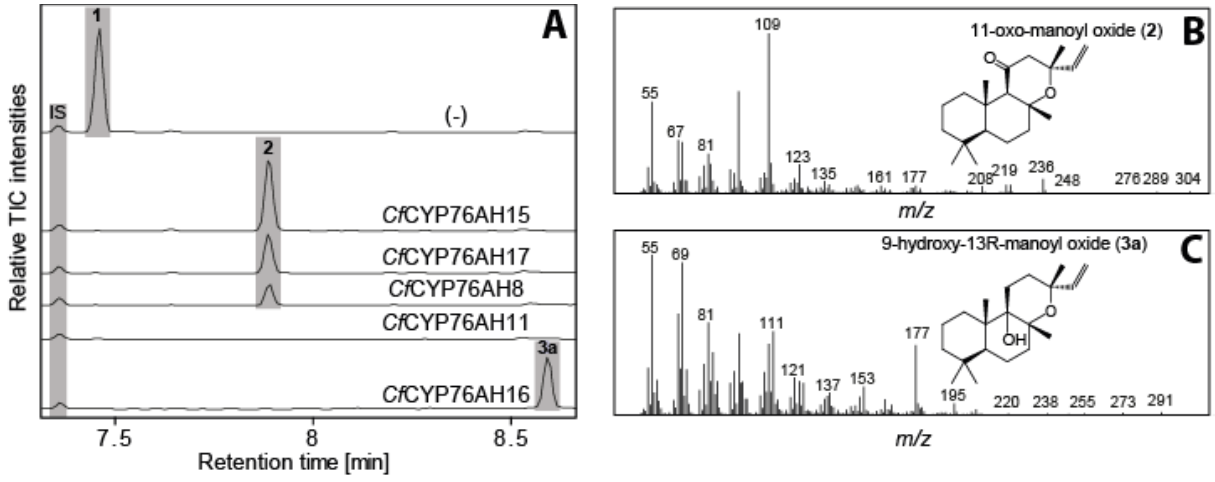

Figure 3. GC-MS analysis of 13*R*-manoyl oxide (**1**) derived diterpenoids obtained by transient expression of CYP76AHs from *C. forskohlii* in *N. benthamiana*. (**A**) GC-MS total ion chromatograms (TIC) of extracts from *N. benthamiana* transiently expressing *CfCXS*, *CfGGPPS*, *CfTPS2* and *CfTPS3* (13*R*-manoyl oxide biosynthesis) genes in combination with water (-), *CfCYP76AH15, CfCYP76AH17, CfCYP76AH8, CfCYP76AH11* or *CfCYP76AH16*. 1-eicosene was used as internal standard (IS). 13*R*-manoyl oxide (**1**) was identified only in (-), indicating that it is further metabolized by the *Cf*CYP76AH15, *Cf*CYP76AH17, *Cf*CYP76AH8, *Cf*CYP76AH11 and *Cf*CYP76AH16 enzymes. (**B**) *m/z* spectrum of 11-oxo-13*R*-manoyl oxide (2). (**C**) *m/z* spectrum of 9-hydroxy-13*R*-manoyl oxide (**3a)**. The structure of both compounds was verified by NMR analysis (***Table 1 and 2***). The compounds have been identified previously in *C. forskohlii* as putative intermediates in the *in planta* biosynthesis of forskolin (*Asada et al., 2012*). Extracts from leaves of three different *N. benthamiana* plants have been analyzed and representative chromatograms are shown.

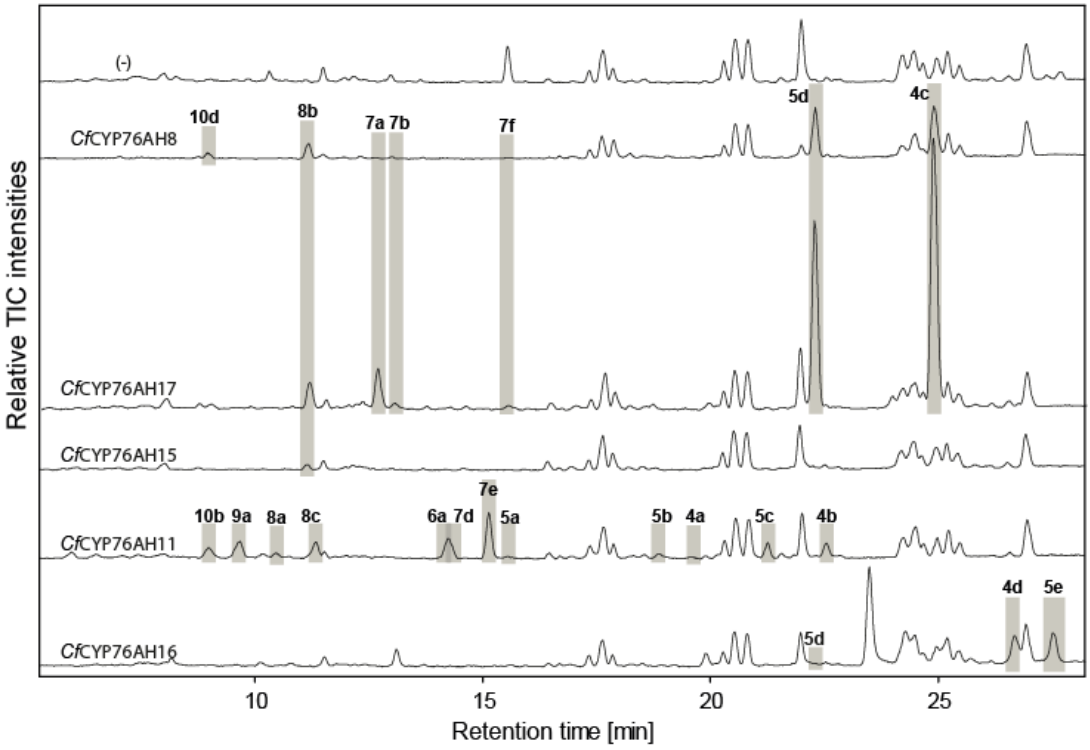

Figure 3-Figure Supplement 1. LC-qTOF-MS analysis of 13*R*-manoyl oxide-derived diterpenoids obtained by transient expression of *C. forskohlii* CYP76AH encoding genes in *N. benthamiana*. Total ion chromatograms (TIC) of extracts expressing the 13*R*-manoyl oxide biosynthesis genes (*CfCXS*, *CfGGPPS*, *CfTPS2*, *CfTPS3)* in combination with water (-), *CfCYP76AH8, CfCYP76AH17, CfCYP76AH15, CfCYP76AH11* or *CfCYP76AH16* are shown. 13*R*-manoyl oxide-derived oxygenated compounds formed (marked with grey bars) and their identity including their molecular formulas was confirmed by their accurate mass (5 ppm tolerance, ***Table 3***). The identity of 1,11-dihydroxy-13*R*-manoyl oxide (**5d**) and 9-deoxydeactylforskolin (**10b**) was confirmed by NMR analysis (***Figure 4 and Table 1 and 2***). No 13*R*-manoyl oxide-derived diterpenoids were detected in the water control (-). Extracts from leaves of three different *N. benthamiana* plants have been analyzed and representative chromatograms are shown.

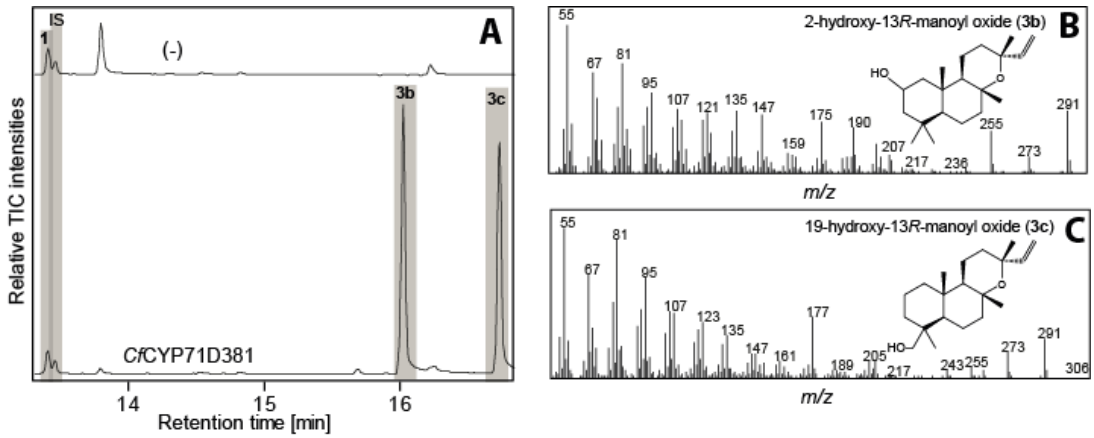

Figure 3-Figure Supplement 2. GC-MS analysis of 13*R*-manoyl oxide-derived diterpenoids following transient expression in *N. benthamiana* of the *C. forskohlii* gene encoding *Cf*CYP71D281 together with genes encoding the required enzymes for biosynthesis of 13*R*-manoyl oxide (*CfCXS*, *CfGGPPS*, *CfTPS2*, *CfTPS3)*. A: GC-MS total ion chromatograms (TIC) of extracts from *N. benthamiana* transiently expressing 13*R*-manoyl oxide biosynthesis genes in combination with water (-) or *CfCYP71D381*, respectively. 1-eicosene was used as internal standard (IS) and 13*R*-manoyl oxide (**1**) was identified in both (-) and the *Cf*CYP71D381 samples. Compounds **3b** and **3c** were identified in extracts from *N. benthamiana* expressing *Cf*CYP71D381 together with the genes in 13*R*-manoyl oxide biosynthesis. *Cf*CYP71D381 efficiently converted **1** to a mixture of two mono-hydroxylated 13*R*-manoyl oxide derivatives (**3b** and **3c**). Structural elucidation by NMR (***Figure 4 and Table 1 and 2***) showed hydroxylation of **1** at positions C-2 (**3b**) and C-19 (**3c**). These hydroxylation positions do not coincide with those found in forskolin and to our knowledge have not been observed in other diterpenoids known from *C. forskohlii.* B: *m/z* spectrum of 2-hydroxy-13*R*-manoyl oxide (**3b**). C: *m/z* spectrum of 19-hydroxy-13*R*-manoyl oxide (**3c**). Extracts from leaves of three different *N. benthamiana* plants have been analyzed and representative chromatograms are shown.

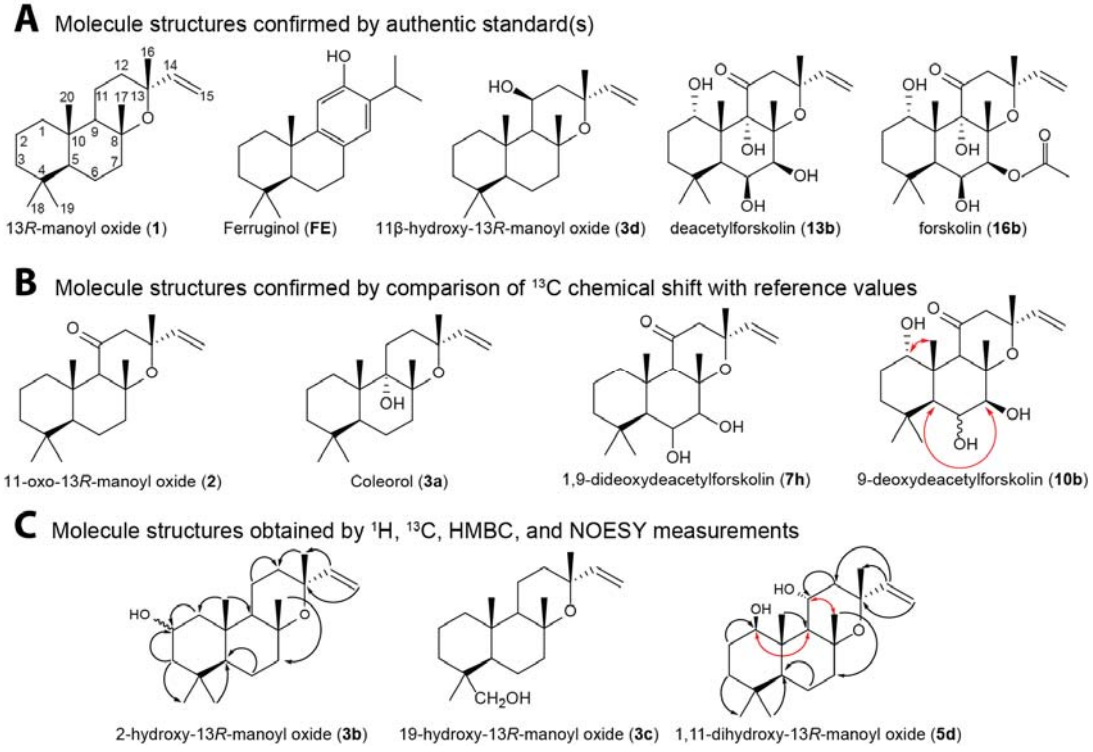

Figure 4. Structures of key compounds presented in this work. A: compounds confirmed using authentic standards. B: compounds which structure was confirmed/identified by comparison of ^{13}C NMR data with existing literature. C: compounds which structure was confirmed/identified by HMBC and NOE correlations for assigning position of OH-groups (marked in red), whereas couplings identified in the previously uncharacterized compounds **3b**, **3c** and **5d** are marked in black. All other molecular structures were confirmed by ^{13}C chemical shifts in comparisons to reference values (***Table 1, Figure 4-source data 1***).

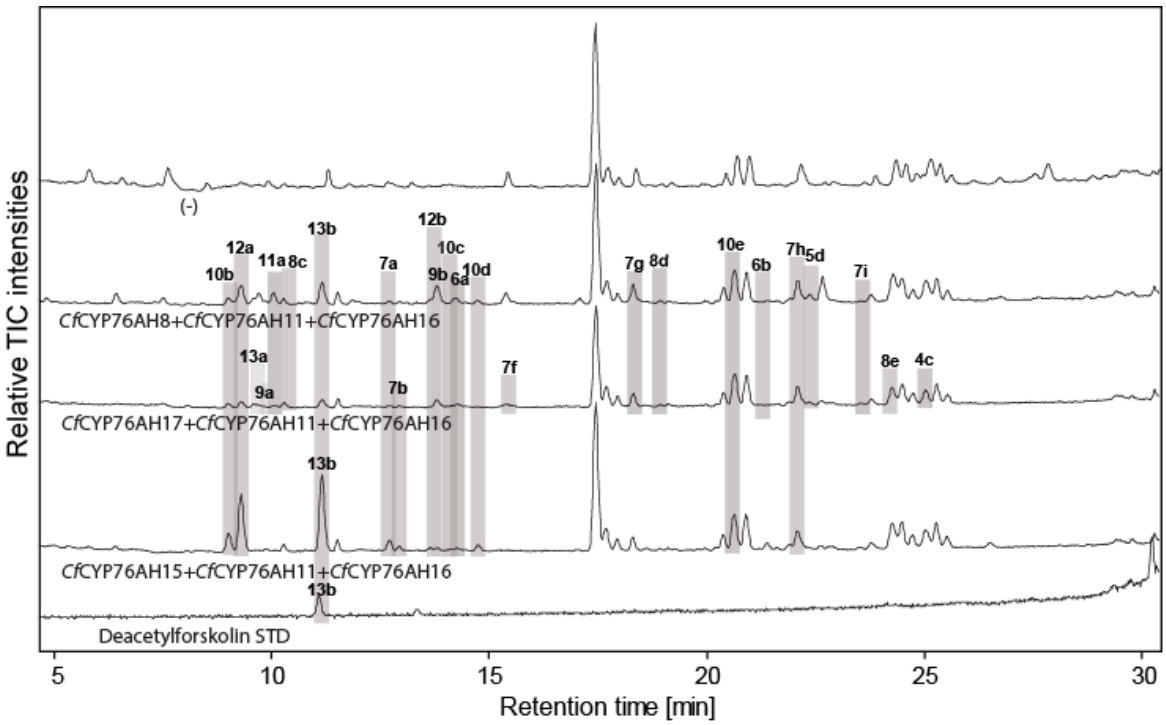

Figure 5. LC-qTOF-MS analysis of 13*R*-manoyl oxide-derived diterpenoids obtained by transient expression of combinations of *C. forskohlii* CYP encoding genes, together with genes encoding the required enzymes for biosynthesis of 13*R*-manoyl oxide in *N. benthamiana*. Total ion chromatograms (TIC) of extracts expressing the 13*R*-manoyl oxide biosynthesis genes (*CfCXS*, *CfGGPPS*, *CfTPS2*, *CfTPS3)*, in combination with (from the top) water (-), *CfCYP76AH8 + CfCYP76AH11 + CfCYP76AH16, CfCYP76AH17 + CfCYP76AH11 + CfCYP76AH16,* and *CfCYP76AH15 + CfCYP76AH11 + CfCYP76AH16* are shown. Hydroxylated 13*R*-manoyl oxide-derived diterpenoids (marked with grey bars) and their identity including their molecular formulas were confirmed by accurate mass (5 ppm tolerance, ***Table 3***). Compounds present in trace amounts are not marked. The identity of 1,11-dihydroxy-13*R*-manoyl oxide (**5d**), 9-deoxydeacetylforskolin (**10b**) and 1,9-dideoxydeacetylforskolin (**7h**) was confirmed by NMR analysis ***(Figure 4 and Table 1 and 2)***, whereas the identity of deacetylforskolin (**13b**) was confirmed by comparison to an authentic chemically synthesized standard. No 13*R*-manoyl oxide-derived diterpenoids were identified in the water control (-). Extracts from leaves of three different *N. benthamiana* plants have been analyzed and representative chromatograms are shown.

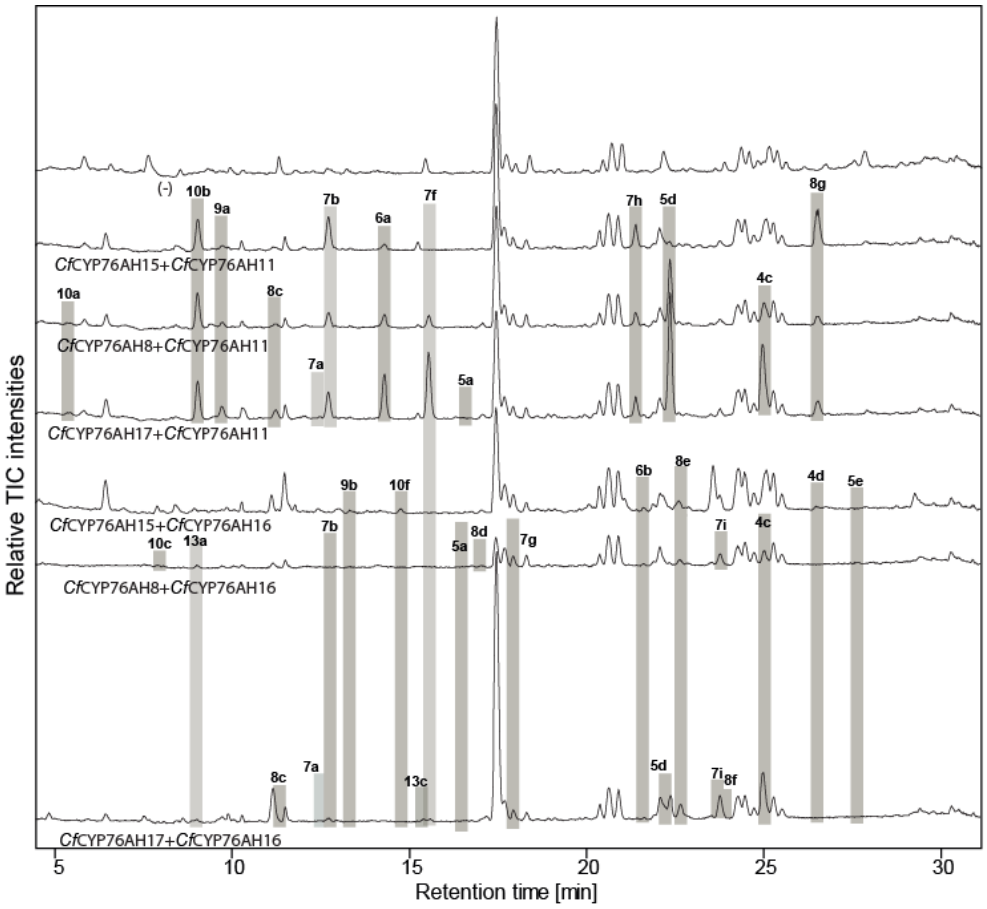

Figure 5-Figure Supplement 1. LC-qTOF-MS analysis of 13*R*-manoyl oxide-derived diterpenoids obtained by transient expression of combinations of *C. forskohlii* CYP76AH encoding genes in *N. benthamiana*. Total ion chromatograms (TIC) of extracts expressing the 13*R*-manoyl oxide biosynthesis genes (*CfCXS*, *CfGGPPS*, *CfTPS2*, *CfTPS3)* in combination with (from the top) water (-), *CfCYP76AH15 + CfCYP76AH11, CfCYP76AH8 + CfCYP76AH11, CfCYP76AH17 + CfCYP76AH11, CfCYP76AH15 + CfCYP76AH16, CfCYP76AH8 + CfCYP76AH16* and *CfCYP76AH17 + CfCYP76AH16* are shown. Oxygenated 13*R*-manoyl oxide-derived diterpenoids (marked with grey bars) and their identity including their molecular formulas were confirmed by their accurate mass (5 ppm tolerance, ***Table 3***). Compounds present in trace amounts are not marked. The identity of 1,11-dihydroxy-13*R*-manoyl oxide (**5d**), 9-deoxydeacetylforskolin (**10b**), 1,9-dideoxydeacetylforskolin (**7h**) was confirmed by NMR analysis (***Figure 4 and Table 1 and 2***). No 13*R*-manoyl oxide-derived diterpenoids were detected in the water control (-). Extracts from leaves of three different *N. benthamiana* plants have been analyzed and representative chromatograms are shown.

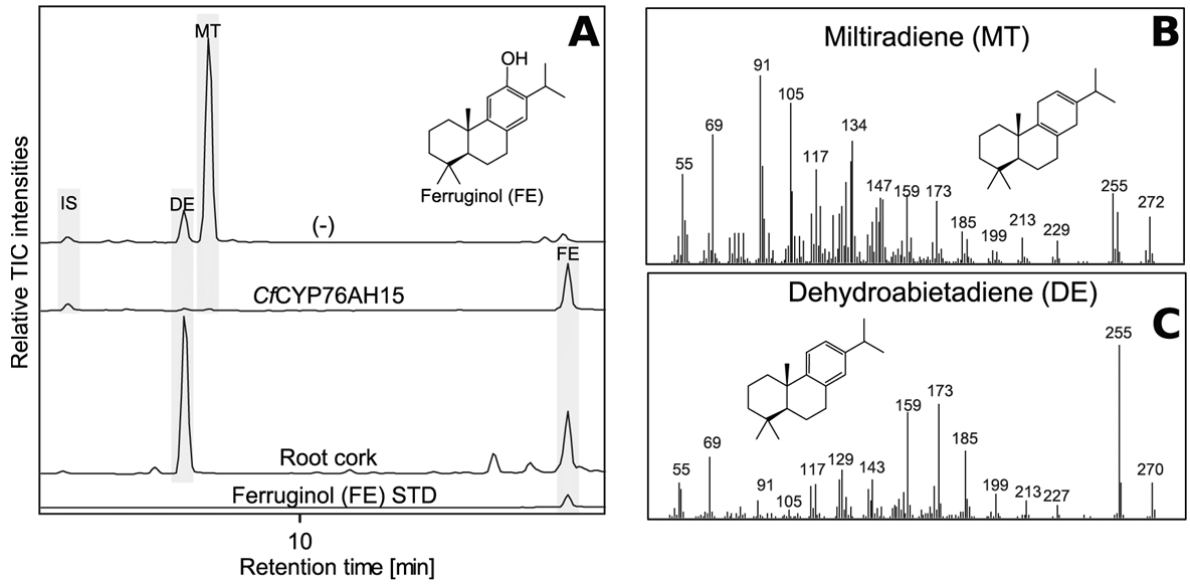

Figure 6. GC-MS analysis of miltiradiene-derived diterpenoids obtained by transient expression of *CfCYP76AH15* in *N. benthamiana*. (**A**) Total ion chromatograms (TIC) of extracts transiently expressing *CfCXS, CfGGPPS, CfTPS1* and *CfTPS3* (miltiradiene biosynthesis genes) in combination with water (-) or *CfCYP76AH15*. Dehydroabietadiene (DE) and miltiradiene (MT) were observed in the (-) extract, whereas ferruginol was observed in extracts from tissue expressing the miltiradiene biosynthesis genes together with *CfCYP76AH15.* In root cork extracts ferruginol was detected together with dehydroabietadiene. Presence of ferruginol was confirmed by comparison to an authentic standard (***Ignea et al., 2016a***), while presence of miltiradiene (**B**) and dehydroabietadiene (**C**) were confirmed by comparison of *m/z* spectra with previously characterized compounds (***Andersen-Ranberg et al., 2016***). Extracts from leaves of three different *N. benthamiana* plants have been analyzed and representative chromatograms are shown.

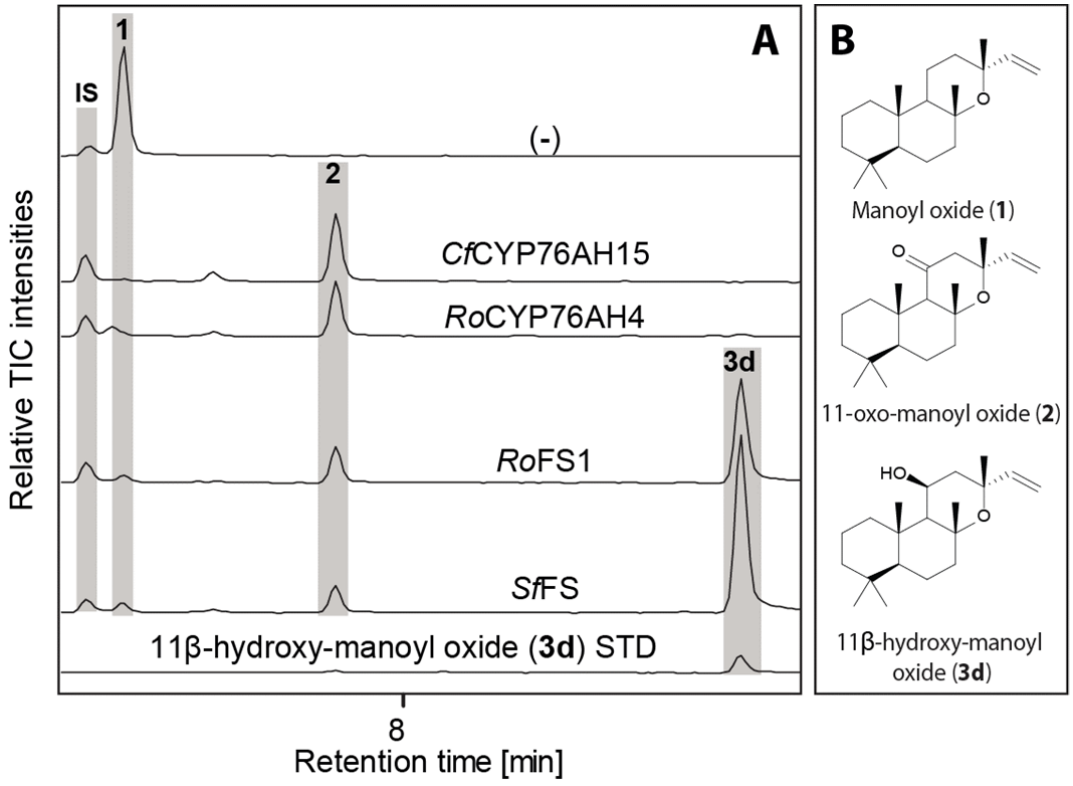

Figure 7. GC-MS analysis of 13*R*-manoyl oxide-derived diterpenoids obtained by transient expression of *CYP76AHs* in *N. benthamiana*. (**A**) Total ion chromatograms (TIC) of extracts transiently expressing *CfCXS, CfGGPPS, CfTPS2* and *CfTPS3* (13*R*-manoyl oxide biosynthesis genes) in combination with water (-), *CfCYP76AH15, RoCYP76AH4, RoFS1* and *SpFS* are shown. 13*R*-manoyl oxide was observed in the (-) extracts, while 11-oxo-13*R*-manoyl oxide (**2**) was observed in the *CfCYP76AH15, RoCYP76AH4, RoFS1* and *SfFS* extracts. 11-Hydroxy-13*R*-manoyl oxide (**3d**) is observed only in extracts expressing the *RoFS1* and *SfFS1* genes. Presence of 11-hydroxy-13*R*-manoyl oxide was verified by comparison to an authentic standard (***Ignea et al., 2016b***) while identification of 11-oxo-13*R*-manoyl oxide was confirmed by comparison of *m/z* spectra with a previously characterized compound (**2**). The results show *Ro*CYP76AH4 has an activity similar to *Cf*CYP76AH15, able to convert efficiently and specifically 13*R*-manoyl oxide to **2**. *Ro*FS1, as well as *Sf*FS, can also convert 13*R*-manoyl oxide to **2** but they catalyze the synthesis of an additional product, 11-hydroxy-13*R*-manoyl oxide (**3d**). Extracts from leaves of three different *N. benthamiana* plants have been analyzed and representative chromatograms are shown.

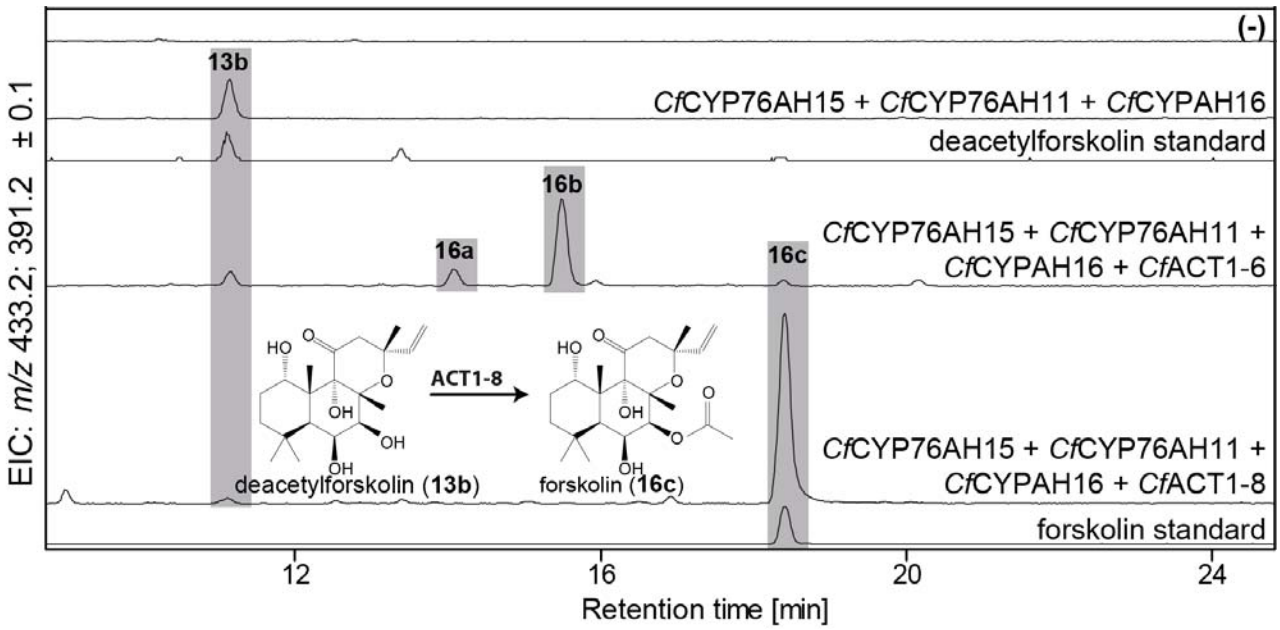

Figure 8. *De novo* biosynthesis of forskolin by transient expression of *C. forskohlii* genes in *N. benthamiana* as monitored by LC-MS based extracted ion chromatograms (EIC). To monitor both deacetylforskolin (**13b**) and forskolin (**16c**), the EIC were selected as the sum of *m/z* 433.2 ± 0.1 and m/z 391.2 ± 0.1. Chromatograms represent LC-MS analysis of extracts from leaves expressing the 13*R*-manoyl oxide biosynthesis genes (*CfDXS*, *CfGGPPS*, *CfTP2* and *CfTPS3*) in combination with (from the top): water (-); *CfCYP76AH15*, *CfCYP76AH11* and *CfCYPAH16; CfCYP76AH15*, *CfCYP76AH11, CfCYPAH16* and *CfACT1-6; CfCYP76AH15*, *CfCYP76AH11, CfCYPAH16* and *CfACT1-8,* shown together with authentic standards (deacetylforskolin and forskolin). From leaves expressing *CfACT1-6* together with forskolin specific CYPs, forskolin (**16c**) was identified together with two other acetylated compounds (e.g. **16a**, **16b**) with the same molecular mass (***Table 3***). When *CfACT1-8* was expressed instead, a predominant accumulation of forskolin was observed, with a drastic reduction of non-specific acetylated products.

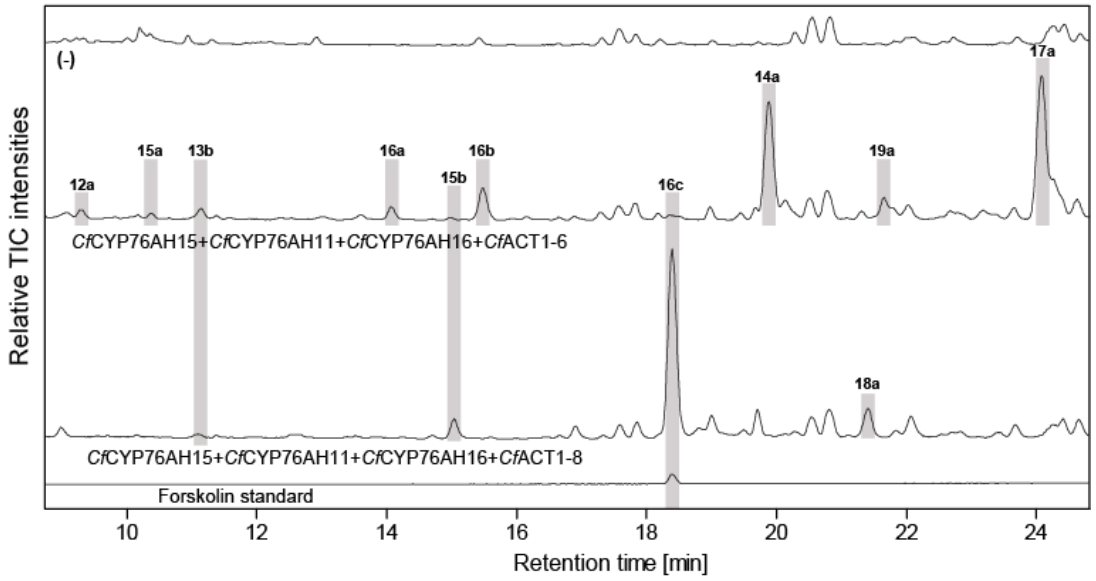

Figure 9. LC-qTOF-MS analysis of 13*R*-manoyl oxide-derived diterpenoids obtained by transient expression of combinations of *C. forskohlii* CYP and ACT encoding genes in *N. benthamiana*. Total ion chromatograms (TIC) from extracts expressing the 13*R*-manoyl oxide biosynthesis genes (*CfCXS*, *CfGGPPS*, *CfTPS2*, *CfTPS3)* in combination with (from the top) water (-), *CfCYP76AH15* + *CfCYP76AH11+* + *CfCYP76AH16* + *CfACT1-6*, and *CfCYP76AH15* + *CfCYP76AH11* + *CfCYP76AH16* + *CfACT1-8* are shown. Hydroxylated and acetylated 13*R*-manoyl oxide-derived diterpenoids (marked with grey bars) and their identity, including their molecular formulas, was confirmed by their accurate mass (5 ppm tolerance, ***Table 3***). Compounds present in trace amounts are not marked. The identity of deacetylforskolin (**13b**) and forskolin (**16c**) was confirmed by comparison to authentic standards. No 13*R*-manoyl oxide-derived diterpenoids were detected in the water control (-). Extracts from leaves of three different *N. benthamiana* plants have been analyzed and representative chromatograms are shown.

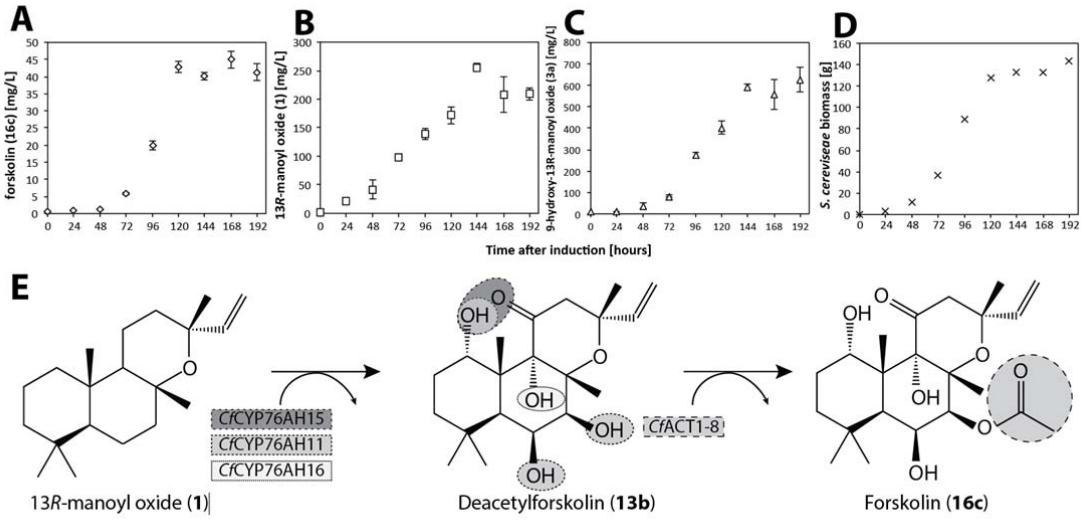

Figure 10. Forskolin production in *Saccharomyces cerevisiae* following stable genomic integration of codon-optimized *C. forskohlii* genes. (**A**) Forskolin (**16c**) accumulation in a fermenter batch using the EVST21543 strain (expressing *Cf*CYP76AH15, *Cf*CYP76HA11, *Cf*CYP76AH16 and *Cf*ACT1-8 encoding genes in the EFSC4498 *S. cerevisiae* strain, optimized for the production of 13*R*-manoyl oxide (***Andersen-Ranberg et al., 2016***)). (**B**) 13*R*-manoyl oxide (**1**) accumulation in EVST21543 strain. (**C**) 9-hydroxy-13*R*-manoyl oxide (**3a**) accumulation in EVST21543 strain. (**D**) EVST21543 strain biomass monitored during the fermentation process. (**E**) The biosynthetic pathway used for the production of forskolin in yeast. The fermentation event occurred once, and a triplicate of samples have been analysed from each time course.

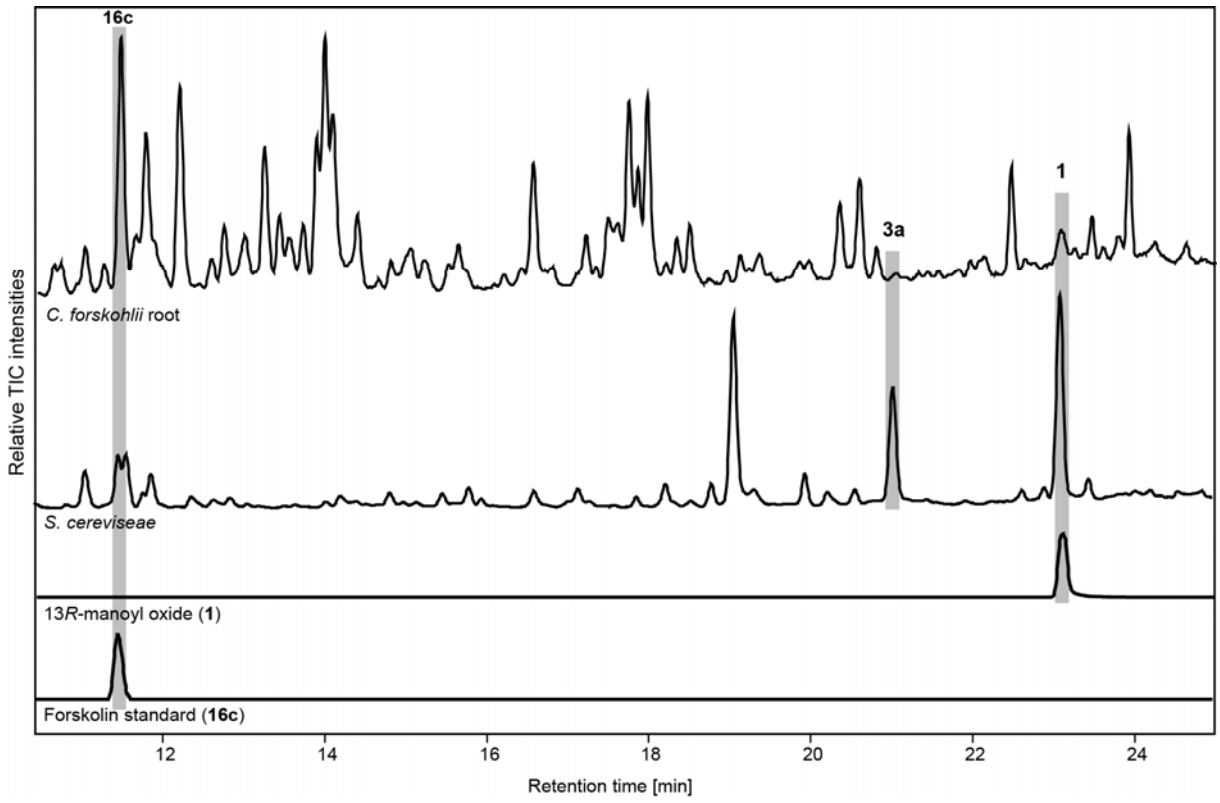

Figure 10-Figure Supplement 1. Comparison of metabolite profiles between fermenter grown yeast culture of the EVST21543 strain and *C. forskohlii* root extract analyzed by LC-MS. Forskolin (**16c**) and 13*R*-manoyl oxide (**1**) were identified based on co-elution with standards and 9-hydroxy-13*R*-manoyl oxide (**3a**) was identified based on the presence of the $[M+Na]^+$ ion, 329.2457 ($C_{20}H_{34}O_2Na^+$, Δ 0.7 ppm).

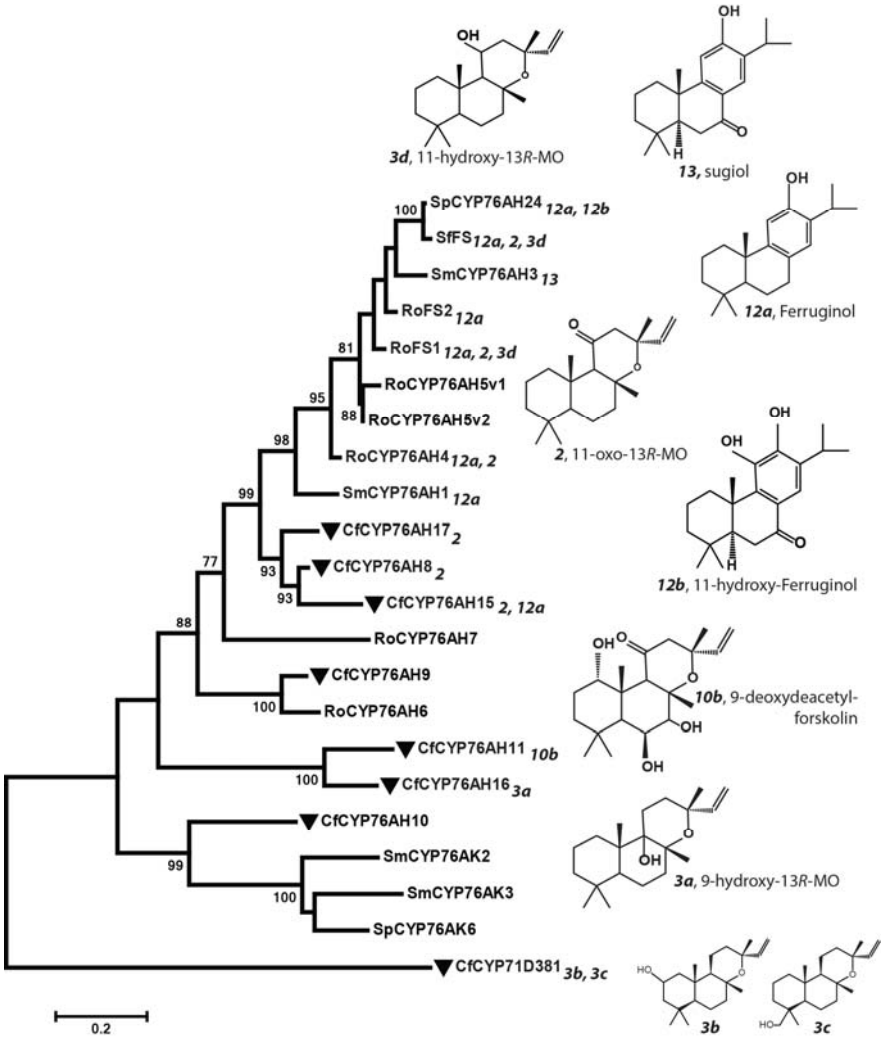

Figure 11. Phylogeny of known full-length CYP76AHs. The enzymes used are listed below with their accession numbers or source of publication: *Cf*CYP76AH15, KT382358; *Cf*CYP76AH17, KT382360; *Cf*CYP76AH8, KT382348; *Cf*CYP76AH11, KT382349; *Cf*CYP76AH16, KT382359; *Cf*CYP76AH9, KT382347; *Cf*CYP76AH10, KT382346; *Cf*CYP71D381, KT382342; *Ro*FS1, AJQ30187 (***Božić et al., 2015***); *Sm*CYP76AH3, KR140168 (***Guo et al., 2015***); *Ro*FS2, AJQ30188 (***Božić et al., 2015***); *Sf*FS, AJQ30186 (***Božić et al., 2015***); *Ro*CYP76AH4, (***Zi et al., 2013***); *Ro*CYP76AH5v1, (***Zi et al., 2013***); *Ro*CYP76AH5v2, (***Zi et al., 2013***); *Ro*CYP76AH6, (***Zi et al., 2013***); *Ro*CYP76AH7, (***Zi et al., 2013***); *Sm*CYP76AH1, AGN04215 (***Guo et al., 2013***); *Sp*CYP76AH24, ALM25796 (***Ignea et al., 2016a***). *Coleus forskohlii* enzymes are indicated by a solid black triangle. *Cf*CYP71D381 was chosen as a root because it can accept 13*R*-manoyl oxide as a substrate, but does not catalyze the synthesis of forskolin-related products. The number subscripts indicated at each enzyme refer to their respective enzymatic products, the structures of which are given on the right. Only the main products of each enzymes are mentioned. MO stands for manoyl oxide.

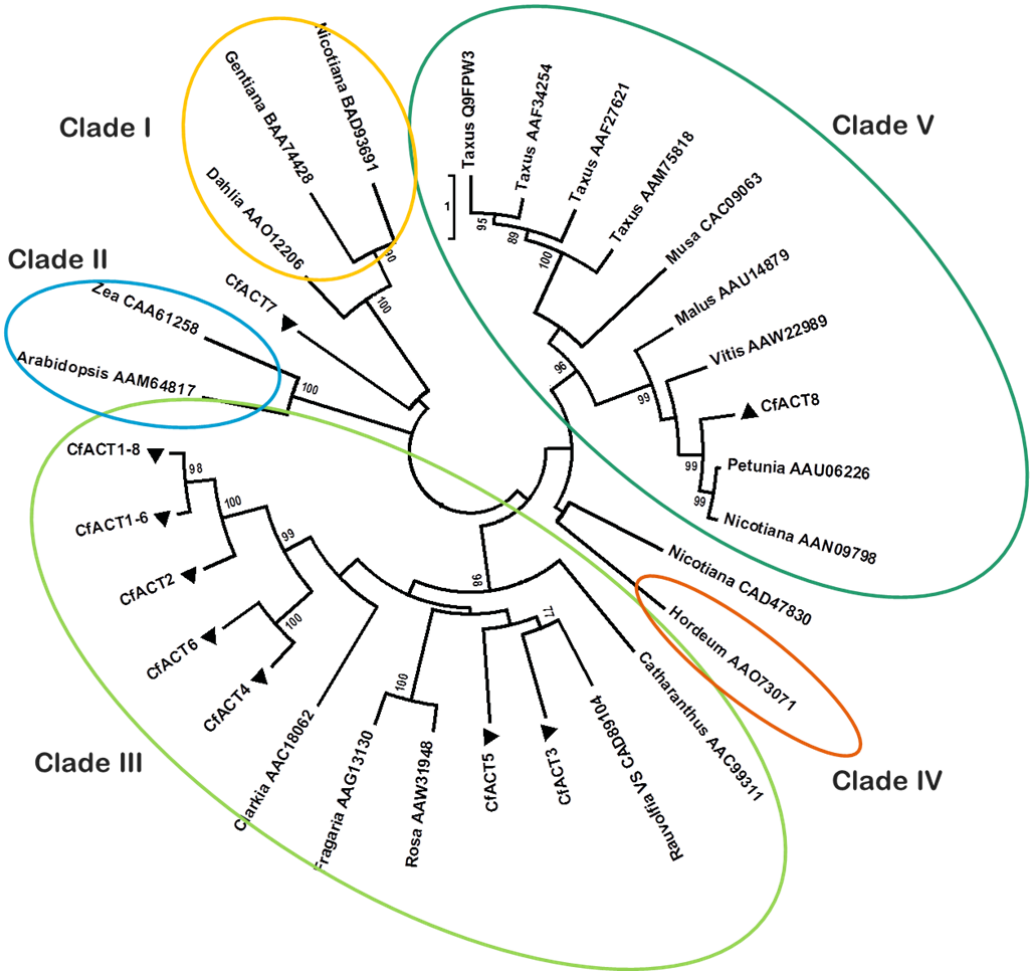

Figure 12. Phylogenetic tree of *Cf*ACT encoding candidate genes together with BAHD family acyltransferase representatives from all clades according to D'Auria (2006). Accession numbers of the non-*Coleus forskohlii* selected protein sequences are shown next to the tree taxon names, while *C. forskohlii* peptide accession numbers are provided in ***Figure 1-source data 1***. The analysis only includes functionally characterized members. *Coleus forskohlii* enzymes are indicated by a solid black triangle. The majority of the selected *Cf*ACTs belong to Clade III, which includes mainly members which accept a diverse range of hydroxylated substrates and use acetyl-CoA as the main acyl donor (***D'Auria 2006***). Interestingly, the ACTs known to be involved in Taxol biosynthesis belong to Clade V.

Data source figure legends

Figure 1-source data 1.

cDNAs identified in the *C. forskohlii* root cork transcriptome and cloned during this work, with the GeneBank accession numbers

Figure 1-source data 2.

Table of FPKM (Fragments Per Kilobase of transcript per Million mapped reads) values of the first 20 most abundant cDNAs identified in the root cork transcriptome library. cDNAs involved in terpenoid metabolism are marked in bold.

Figure 1-source data 3.

Table of primers used in this study. Construction of plasmids for expression of *Cf*TPS2, *Cf*TPS3, *Cf*TPS1 is described in (***Andersen-Ranberg et al., 2016***). U (uracil, marked in bold), represents the cleavage site, used in the USER cloning (***Nour-Eldin et al., 2006***)

Figure 4-source data 1.

NMR spectra's of selected 13*R*-manoyl oxide derived molecules

Table 1.

^{1}H-NMR and ^{13}C-NMR chemical shifts (***Figure 4-source data 1***) of novel oxygenated 13*R*-(+)-manoyl oxide-derived diterpenoids formed following transient expression of CYP encoding genes from *C. forskohlii*.

	19-hydroxy-13*R*-manoyl oxide (**3c**)[a]		2-hydroxy-13*R*-manoyl oxide (**3b**)[a]		1,11-dihydroxy-13*R*-manoyl oxide (**5d**)[a]	
Pos.	^{1}H (*n*H; *m*; *J*(*Hz*))	^{13}C	^{1}H (*n*H; *m*; *J*(*Hz*))	^{13}C	^{1}H (*n*H; *m*; *J*(*Hz*))	^{13}C
1	0.89 (1H; *m*) 1.63 (1H; *m*)	39.1	1.10 (1H; *t*(*br*); 11.9, 11.9) 1.77 (1H; *m*)	51.3	3.49 (1H; *dd*;11.1, 4.5)	79.0
2	1.44 (1H; *m*) 1.56 (1H; *m*)	18.1	3.92 (1H; *m*)	65.3	1.75 (1H; *td*; 13.5, 11.1, 3.9) 1.60 (1H; *m*)	29.0
3	0.95 (1H; *m*) 1.78 (1H; *m*)	35.8	0.76 (1H; *t*(*br*); 11.9, 11.9) 1.99 (1H; *d*(*br*); 11.9)	48.2	1.47 (1H; *dd*; 13.6, 3.9) 1.39 (1H; *td*; 13.5, 3.6)	39.6
4		38.5		34.9		33.4
5	1.10 (1H; *dd*; 2.3, 12.6)	56.9	0.95 (1H; *dd*; 2.2, 12.4)	55.9	0.84 (1H; *dd*; 11.3, 2.0)	55.6
6	1.36 (1H; *dd*; 3.6, 12.6) 1.75 (1H; *m*)	20.1	1.68 (1H; *m*) 1.27 (1H; *m*)	19.7	1.47 (1H; *m*) 1.64 (1H; *m*)	20.2
7	1.42 (1H; *m*) 1.83 (1H; *dt*; 3.3, 12.2)	43.6	1.45 (1H; *dd*(*br*); 3.6, 12.5) 1.85 (1H; *dt*(*br*); 2.9, 12.5)	43.2	1.48 (1H; *m*) 1.85 (1H; *m*)	44.0
8		75.1		75.1		75.3
9	1.35 (1H; *dd*; 4.3, 12.0)	55.7	1.40 (1H; *dd*; 4.2, 11.9)	55.4	1.54 (1H; *d*; 5.8)	55.8
10		37.3		38.7		43.8
11	1.48 (1H; *m*) 1.58 (1H; *m*)	15.4	1.53 (1H; *m*) 1.61 (1H; *m*)	15.6	4.38 (1H; *br q*; ≈8.6)	65.6
12	1.78 (1H; *m*) 1.64 (1H; *m*)	35.7	1.78 (1H; *m*) 1.66 (1H; *m*)	35.5	2.02 (1H; *dd*; 14.3, 8.7) 2.27 (1H; *dd*; 14.3, 8.7)	35.8
13		73.4		73.4		72.8
14	5.87 (1H; *dd*; 10.8, 17.4)	147.7	5.87 (1H; *dd*; 10.8, 17.4)	147.7	5.90 (1H; *dd*; 17.4, 10.8)	147.1
15	4.92 (1H; *dd*; 1.5, 10.8) 5.14 (1H; *dd*; 1.5, 17.4)	110.2	4.92 (1H; *d*; 10.8) 5.14 (1H; *d*; 17.4)	110.3	4.94 (1H; *dd*; 10.7, 1.5) 5.17 (1H; *dd*; 17.4, 1.5)	111.2
16	1.27 (3H; *s*)	28.5	1.27 (3H; *s*)	28.7	1.27 (3H; *s*)	32.1
17	1.28 (3H; *s*)	25.3	1.29 (3H; s)	25.7	1.49 (3H; *s*)	27.8
18	0.97 (3H; *s*)	26.8	0.93 (3H; *s*)	33.5	0.78 (3H; *s*)	13.5
19	3.70 (1H; *d*; 10.9) 3.46 (1H; *d*; 10.9)	65.4	0.85 (3H; *s*)	22.2	0.85 (3H; *s*)	32.8
20	0.78 (3H; *s*)	15.7	0.84 (3H; *s*)	16.5	0.79 (3H; *s*)	21.1

[a] ^{1}H and ^{13}C NMR data acquired at 600 and 150 MHz, respectively, in methanol-d_4, at 300 K. *s* = singlet, *d* = doublet, *t* = triplet, *m* = multiplet, *br* = broad

Table 2.

Structural identification of four oxygenated 13*R*-manoyl oxide-derived diterpenoids formed following transient expression of CYP encoding genes from *C. forskohlii* based on comparison of their ^{1}H-NMR and ^{13}C-NMR ***(Figure 4-source data 1)*** chemical shifts to literature data. Chemical shifts for reference compounds marked with * have not been assigned to a specific carbon. The ^{13}C chemical shifts of 9-deoxyforskolin (***Gabetta et al., 1989***) were used as reference for 6,7-dihydroxy-11-oxo-13*R*-manoyl oxide (**7h**).

	9-deoxydeacetylforskolin (**10b**)[a]			1,9-dideoxydeacetylforskolin (**7h**)[a]			11-oxo-13*R*-manoyl oxide (**2**)[a]		Coleorol (**3a**)[a]	
Pos.	^{1}H (*n*H; *m*; *J*(*Hz*))	^{13}C	(***Gabetta et al., 1989***)	^{1}H (*n*H; *m*; *J*(*Hz*))	^{13}C	(***Gabetta et al., 1989***)	^{13}C	(***Gabetta et al., 1989***)	^{13}C	(***Asada et al., 2012***)
1	4.38 (1H; *t*; 2.8)	71.6	71.2	2.45 (1H, *d*(*br*); 13.1) 0.78 (H; *m*)	41.5	43.1	42.1	41.9	31.7	31.6
2	1.47 (1H; *m*) 2.14 (1H; *m*)	25.8	25.6	1.78 (H; *m*) 1.40 (H; *m*)	18.7	18.4	18.5	18.4	18.6	18.4
3	1.12 (1H; *dt*; 3.4, 13.2) 1.62 (1H; *dt*; 3.5, 13.5)	36.4	36.3	1.36 (H; *m*) 1.15 (H; *m*)	43.8	43.7	43.4	43.3	41.9	41.8
4		34.2	34.1		34.4	34.1	33.4	33.2	33.3	33.2
5	1.34 (1H; *d*; 2.1)	47.5	47.4	n.d.	55.7	55.2	56.0	55.8	45.7	45.5
6	4.44 (1H; *t*; 2.6)	70.8	70.2	4.39 (1H; *m*)	70.4	70.2	19.8	19.7	19.5	19.4
7	3.68 (1H; *d*; 3.6)	80.7	81.1	3.71 (1H; *d*; 3.8)	81.0	80.7	39.6	39.4	36.6	36.4
8		80.0	78.5		80.1	79.9	77.5	77.2	78.0	77.8
9	3.32 (1H; *s*)	58.0	58.2	2.59 (1H; *s*)	65.5	65.4	66.9	66.7	75.3	75.2
10		42.2	41.7		38.0	37.8	37.3	37.1	41.1	40.9
11		207.7	207.6		206.3	205.7	207.7	207.1	21.1	21.0
12	2.63 (1H; *d*; 18.0) 2.69 (1H; *d*; 18.0)	49.8	49.9	2.60 (1H; *d*; 18.1) 2.66 (1H; *d*; 18.1)	50.0	49.8	50.4	50.2	31.6	31.5
13		75.1	74.8		75.1	75.1	75.1	74.4	72.9	72.8
14	5.94 (1H; *dd*; 10.8, 17.4)	146.2	145.8	5.95 (1H; *dd*; 10.7, 17.4)	146.9	146.4	146.9	146.0	147.4	147.3
15	5.04 (1H; *d*; 10.8) 5.14 (1H; *d*; 17.4)	112.4	112.7	5.04 (1H; *d*; 10.7) 5.17 (1H; *d*; 17.4)	112.3	112.1	112.3	111.9	110.1	110.0
16	1.30 (3H; *s*)	31.5	31.5*	1.28 (3H; *s*)	31.6	33.2*	31.4	31.2*	28.9	28.8
17	1.54 (3H; s)	24.1	24.5*	1.50 (3H; s)	23.5	31.4*	28.1	27.9*	27.0	29.9
18	1.38 (3H; *s*)	33.1	18.2*	0.97 (3H; *s*)	33.4	23.9*	15.6	15.5*	33.7	33.6
19	1.21 (3H; *s*)	23.7	23.6*	1.21 (3H; *s*)	24.0	23.7*	21.8	21.6*	21.5	21.4
20	1.01 (3H; *s*)	18.5	32.8*	1.30 (3H; *s*)	17.2	16.7*	33.6	33.5*	17.0	16.8

[a] ^{1}H and ^{13}C NMR data acquired at 600 and 150 MHz, respectively,in methanol-d_4, at 300 K. *s* = singlet, *d* = doublet, *t* = triplet, *m* = multiplet, *br* = broad

1141 **Table 3.**

1142 Overview of 13*R*-manoyl oxide-derived diterpenoids identified in *N. benthamiana,* expressing
1143 combinations of *C. forskohlii* genes encoding CYPs and acetyltransferases together with genes
1144 encoding the required enzymes for biosynthesis of 13*R*-manoyl oxide (*Cf*DXS, *Cf*GGPPs, *Cf*TP2 and
1145 *Cf*TPS3). GC-MS and LC-qTOF-MS chromatograms of the identified diterpenoids are shown in
1146 previous figures.
1147

							Single CYPs							Two CYPs						Three CYPs			ACTs		Standards			
	#					(-)	CYP76AH15	CYP76AH8	CYP76AH17	CYP76AH11	CYP76AH16	CYP71D781	CYP76AH15+CYP76AH11	CYP76AH8+CYP76AH11	CYP76AH17+CYP76AH11	CYP76AH15+CYP76AH16	CYP76AH8+CYP76AH16	CYP76AH17+CYP76AH16	CYP76AH15+CYP76AH11+CYP76AH16	CYP76AH8+CYP76AH11+CYP76AH16	CYP76AH17+CYP76AH11+CYP76AH16	CYP76AH15+CYP76AH11+CYP76AH16+ACT1-6	CYP76AH15+CYP76AH11+CYP76AH16+ACT1-8	11-oxo-13*R*-manoyl oxide STD (2)	9-hydroxy-13*R*-manoyl oxide STD (3a)	1,9-dideoxydeacetylforskolin (7h) STD mix	1,11-dihydroxy-13*R*-manoyl oxide(5d)	9-deoxydeacetylforskolin STD (10b)	Deacetylforskolin STD (13b)	Forskolin STD (16c)
GC-MS	1	13*R*-(+)-manoyl oxide				x																							
	2	11-oxo-13*R*-manoyl oxide					x	x	x																x					
	3a	9-hydroxy-13*R*-manoyl oxide									x															x				
	3b	2-hydroxy-13*R*-manoyl oxide										x																		
	3c	19-hydroxy-13*R*-manoyl oxide										x																		
			MO deriv.	MF	RT [min]	Mw [Da]	*m/z*[M+Na]	ΔM (amu.)																						
LC-qTOF-MS	4a	MO(O)(OH)		$C_{20}H_{32}O_3$	19.66	###	###				#																				
	4b	MO(O)(OH)		$C_{20}H_{32}O_3$	22.50	###	###				#																				
	4c	MO(O)(OH)		$C_{20}H_{32}O_3$	24.90	###	###	tr	#	3.5				#5	#5			1.5	1.5		#5										
	4d	MO(O)(OH)		$C_{20}H_{32}O_3$	26.63	###	###					1.5																			
	5a	MO(OH)$_2$		$C_{20}H_{34}O_3$	16.46	###	###				#				#5			#5	#5												
	5b	MO(OH)$_2$		$C_{20}H_{34}O_3$	18.90	###	###				#																				
	5c	MO(OH)$_2$		$C_{20}H_{34}O_3$	21.20	###	###				3.1																				
	5d	MO(OH)$_2$		$C_{20}H_{34}O_3$	22.28	###	###	tr	#	#			#	1.1	#5			#5			#5	#5						1.2			
	5e	MO(OH)$_2$		$C_{20}H_{34}$	27.50	###	###					3.7				#5	#5	#5													

		O_3																	
6a	MO(O)$_2$(OH)	$C_{20}H_{30}O_4$	14.30	###	###			1.5	1.1	-1.1	-1.1				#	#	#		
6b	MO(O)$_2$(OH)	$C_{20}H_{30}O_4$	21.10	###	###							3.1	#	1.4	tr	#	#		
7a	MO(O)(OH)$_2$	$C_{20}H_{32}O_4$	12.50	###	###	#	1.3		tr	tr	#			#	#	#	1.5		
7b	MO(O)(OH)$_2$	$C_{20}H_{32}O_4$	12.75	###	###	1.6	#		#	#	#		#	#	#	#	#		
7d	MO(O)(OH)$_2$	$C_{20}H_{32}O_4$	14.30	###	###			1.8											
7e	MO(O)(OH)$_2$	$C_{20}H_{32}O_4$	15.10	###	###			#											
7f	MO(O)(OH)$_2$	$C_{20}H_{32}O_4$	15.60	###	###	#	1.7		#	#	#	#	#	#	tr		1.4		
7g	MO(O)(OH)$_2$	$C_{20}H_{32}O_4$	18.20	###	###								#	1.2		#	#		
7h	MO(O)(OH)$_2$	$C_{20}H_{32}O_4$	21.30	###	###				-1.1	1.2	#				#	3.7	#		1.3
7i	MO(O)(OH)$_2$	$C_{20}H_{32}O_4$	23.67	###	###								#	#		#	#		
8a	MO(OH)$_3$	$C_{20}H_{34}O_4$	10.69	###	###			1.6											
8b	MO(OH)$_3$	$C_{20}H_{34}O_4$	11.10	###	###	#	1.1												
8c	MO(OH)$_3$	$C_{20}H_{34}O_4$	11.30	###	###			#		#	0.7			2.1		#	3.5		
8d	MO(OH)$_3$	$C_{20}H_{34}O_4$	18.35	###	###											#	#		
8e	MO(OH)$_3$	$C_{20}H_{34}O_4$	22.58	###	###												#		
8f	MO(OH)$_3$	$C_{20}H_{34}O_4$	24.00	###	###									#					
8g	MO(OH)$_3$	$C_{20}H_{34}O_4$	26.60	###	###				#	#	#								
9a	MO(O)$_2$(OH)$_2$	$C_{20}H_{30}O_5$	9.70	###	373.19855			#	#	0.5	#				tr		1.5		
9b	MO(O)$_2$(OH)$_2$	$C_{20}H_{30}O_5$	13.79	###	373.19855							#	1.0	#	#	#	#		
10a	MO(O)(OH)$_3$	$C_{20}H_{32}O_5$	5.50	###	###					#	#								
10b	MO(O)(OH)$_3$	$C_{20}H_{32}O_5$	9.00	###	###			0.5	-1.1	#	#				#	#	#		1.6
10c	MO(O)(OH)$_3$	$C_{20}H_{32}O_5$	14.25	###	###										0.5	#	4.7		
10	MO(O)(OH)$_3$	C_{20}	14.77	###	###							#	4.	#	0.	#	#		

d		$H_{32}O_5$						5		5						
10e	MO(O)(OH)$_3$	$C_{20}H_{32}O_5$	20.80	######	######					#	#	#				
11a	MO(OH)$_4$	$C_{20}H_{34}O_5$	9.70	######	######			#	1.7		#	#				
12a	MO(O)$_2$(OH)$_3$	$C_{20}H_{30}O_6$	9.30	######	######					#	#	#	#			
12b	MO(O)$_2$(OH)$_3$	$C_{20}H_{30}O_6$	13.63	######	######					1.6	#	5.0				
13a	MO(O)(OH)$_4$	$C_{20}H_{32}O_6$	9.65	######	391.20911			#	1.8			#				
13b	MO(O)(OH)$_4$	$C_{20}H_{32}O_6$	11.16	######	391.20911			tr	tr	#	#	#	#	#	1.5	
14a	MO(O)(OH)$_2$(Ac)	$C_{22}H_{34}O_5$	19.87	######	######								#			
15a	MO(O)(OH)$_3$(Ac)	$C_{22}H_{34}O_6$	10.39	######	417.22475								#			
15b	MO(O)(OH)$_3$(Ac)	$C_{22}H_{34}O_6$	15.01	######	417.22475								#	#		
16a	MO(O)(OH)$_4$(Ac)	$C_{22}H_{34}O_7$	14.07	######	######								#			
16b	MO(O)(OH)$_4$(Ac)	$C_{22}H_{34}O_7$	15.47	######	######								#			
16c	MO(O)(OH)$_4$(Ac)	$C_{22}H_{34}O_7$	18.30	######	######								#	#		0.5
17a	MO(O)(OH)$_2$(Ac)$_2$	$C_{24}H_{36}O_6$	24.04	######	######								#			
18a	MO(O)(OH)$_4$(Ac)$_2$	$C_{24}H_{36}O_8$	21.26	######	######									#		
19a	MO(O)(OH)$_3$(Ac)$_3$	$C_{24}H_{38}O_8$	21.66	######	######								0.7			