CYP712K4 Catalyzes the C-29 Oxidation of Friedelin in the *Maytenus ilicifolia* Quinone Methide Triterpenoid Biosynthesis Pathway

Short title: CYP712K4, a *Maytenus* P450 that oxidizes friedelin

Corresponding authors:

A. Goossens and J. Pollier

Center for Plant Systems Biology

VIB - Ghent University

Technologiepark 71

B-9052 Gent (Belgium)

Subject areas: (4) proteins, enzymes and metabolism; (9) natural products

Black and white figures: 1

Color figures: 5

Tables: 1

Supplemental figures: 7

Supplemental tables: 2

CYP712K4 Catalyzes the C-29 Oxidation of Friedelin in the *Maytenus ilicifolia* Quinone Methide Triterpenoid Biosynthesis Pathway

Keylla U. Bicalho[1,2,3,4], Mariana M. Santoni[3,4], Philipp Arendt[1,2,], Cleslei F. Zanelli[4], Maysa Furlan[3], Alain Goossens[1,2,*] and Jacob Pollier[1,2,5,*]

[1] Ghent University, Department of Plant Biotechnology and Bioinformatics, 9052 Ghent, Belgium

[2] VIB Center for Plant Systems Biology, 9052 Ghent, Belgium

[3] Department of Organic Chemistry, Institute of Chemistry, São Paulo State University (UNESP), Araraquara, São Paulo, Brazil

[4] Department of Biological Sciences, School of Pharmaceutical Sciences, São Paulo State University (UNESP), Araraquara, São Paulo, Brazil

[5] VIB Metabolomics Core, 9052 Ghent, Belgium

* Correspondence (Tel. +32 9 331 38 54; fax, +32 9 331 38 09; email jacob.pollier@ugent.vib.be and alain.goossens@ugent.vib.be)

Corresponding authors:

Alain Goossens and Jacob Pollier

Center for Plant Systems Biology

VIB - Ghent University

Technologiepark 71

B-9052 Gent (Belgium)

Short title: CYP712K4, a *Maytenus* P450 that oxidizes friedelin

Footnotes: The nucleotide sequence reported in this paper has been submitted to the GenBank database with accession number MK829814.

Abstract

The native Brazilian plant *Maytenus ilicifolia* accumulates a set of quinone methide triterpenoids with important pharmacological properties, among which maytenin, pristimerin and celastrol that accumulate exclusively in the root bark of this medicinal plant. The first committed step in the quinone methide triterpenoid biosynthesis is the cyclization of 2,3-oxidosqualene to friedelin, catalyzed by the oxidosqualene cyclase friedelin synthase. Here, we achieved heterologous friedelin production by expression of *M. ilicifolia* friedelin synthase in *Nicotiana benthamiana* leaves and a *Saccharomyces cerevisiae* strain engineered using CRISPR/Cas9. Furthermore, friedelin-producing *N. benthamiana* leaves and *S. cerevisiae* cells were used for the characterization of CYP712K4, a cytochrome P450 from *M. ilicifolia* that catalyzes the oxidation of friedelin at the C-29 position, leading to maytenoic acid, an intermediate of the quinone methide triterpenoid biosynthesis pathway. Maytenoic acid produced in *N. benthamiana* leaves was purified and its structure was confirmed using high-resolution mass spectrometry and nuclear magnetic resonance analysis. The three-step oxidation of friedelin to maytenoic acid by CYP712K4 can be considered as the second step of the quinone methide triterpenoid biosynthesis pathway, and may form the basis for further pathway discovery and heterologous production of friedelanes and ultimately quinone methide triterpenoids.

Keywords: Celastrol, CYP712K4, Friedelin, Maytenoic acid, *Maytenus ilicifolia*, Quinone methide triterpenoids.

Introduction

The native Brazilian medicinal plant *Maytenus ilicifolia* Mart. ex Reissek (espinheira santa, *Monteverdia ilicifolia*) is a shrub or small tree of which the leaves are traditionally used as a folk medicine for the treatment of gastric ulcer (Biral et al. 2017; Niero et al. 2011; Périco et al. 2018). Like other members of the Celastraceae family, *M. ilicifolia* accumulates quinone methide triterpenoids in its root bark. These chemotaxonomic markers of the Celastraceae family have been shown to exhibit anti-inflammatory, antifeedant, antioxidant, trypanocidal, antimicrobial, and antitumor activities, among others (Buffa Filho et al. 2002; Coppede et al. 2014). The major quinone methide triterpenoids produced by *M. ilicifolia* are maytenin and pristimerin (Buffa Filho et al. 2002), the latter being a promising natural product for the treatment of breast cancer (Cevatemre et al. 2018) or to be used as a contraceptive agent (Mannowetz et al. 2017). Both metabolites are likely synthesized from celastrol, another quinone methide triterpenoid and promising agent for the treatment of obesity (Liu et al. 2015) that also accumulates to significant amounts in the root bark of *M. ilicifolia* (Buffa Filho et al. 2002).

Like most triterpenoids, the quinone methide triterpenoids are synthesized from 2,3-oxidosqualene, which is derived from isopentenyl pyrophosphate generated through the mevalonate (MVA) pathway (Corsino et al. 2000; Pina et al. 2016). As friedelanes and quinone methide triterpenoids co-occur in various plant species, a common biogenesis was suspected (Corsino et al. 2000; Kutney et al. 1981). Remarkably, the biosynthesis of these metabolites occurs in different tissues of the plant. Whereas friedelanes accumulate in the leaves, the quinone methide triterpenoids are exclusively present in the root bark. Furthermore, feeding of radiolabeled mevalonolactone leads to the accumulation of radiolabeled friedelin in the leaves, and radiolabeled quinone methide triterpenoids in the roots, implying transport of a biosynthetic intermediate from the leaves to the roots of the plant (Corsino et al. 2000).

The cyclization of 2,3-oxidosqualene constitutes the first committed step towards the biosynthesis of specialized triterpenoids in plants. This reaction is catalyzed by specific oxidosqualene cyclases and more than 100 different triterpene scaffolds have been described (Miettinen et al. 2018; Moses et al. 2013; Thimmappa et al. 2014). Similarly, the first committed step in the quinone methide triterpenoid biosynthesis is the cyclization of 2,3-oxidosqualene to friedelin, which is catalyzed by friedelin synthase (FRS, Figure 1A). This oxidosqualene cyclase has been characterized in a few plant species, including *M. ilicifolia* (Alves et al. 2018; Souza-Moreira et al. 2016) and the Chinese medicinal plant *Tripterygium wilfordii* (Zhou et al. 2019). In the second step of the triterpenoid biosynthesis in plants, the generated triterpene scaffolds undergo one or more oxidative modifications at various positions of the backbone, which are catalyzed by various cytochrome P450 enzymes and dramatically increase the triterpenoid structural diversity (Miettinen et al. 2018; Moses et al. 2013; Seki et al. 2015; Thimmappa et al. 2014). So far, no cytochromes P450 involved in quinone methide triterpenoid biosynthesis have been described.

P450s involved in triterpenoid biosynthesis have been mostly characterized by heterologous expression in *Nicotiana benthamiana* leaves or *Saccharomyces cerevisiae* cells producing the triterpene substrates. *Agrobacterium*-mediated transient expression in *N. benthamiana* leaves is a process in which a suspension of *Agrobacterium tumefaciens* cells carrying a binary vector with the gene of interest are introduced in *N. benthamiana* leaves by vacuum infiltration or by direct injection using a needle-less syringe (Reed and Osbourn 2018). Because this system allows for the simultaneous expression of multiple genes by co-infiltration of

A. tumefaciens strains carrying different expression constructs, multi-enzyme pathways can be reconstituted. *N. benthamiana* is becoming increasingly popular for the characterization of P450s involved in triterpenoid biosynthesis and as a heterologous host for the production of triterpenoids in gram-scale quantities (Reed and Osbourn 2018; Reed et al. 2017). In spite of the increasing use of *N. benthamiana*, heterologous expression in the microbial host *S. cerevisiae* remains a well-applied system for the characterization of P450s involved in triterpenoid biosynthesis and the heterologous production of triterpenoids. To this end, *S. cerevisiae* strains engineered for the accumulation of 2,3-oxidosqualene are often used (Moses et al. 2014; Moses et al. 2013). In addition, heterologous triterpenoid production can be further enhanced by the use of cyclodextrins, cyclic oligosaccharides that sequester triterpenes (Moses et al. 2014), and by knocking out *PAH1*, which leads to a dramatic expansion of the endoplasmic reticulum and thereby stimulates heterologous triterpenoid production (Arendt et al. 2017).

In this study, we expressed *M. ilicifolia* friedelin synthase in *N. benthamiana* leaves and a sterol-engineered *S. cerevisiae* strain for the production of friedelin. Friedelin-producing *N. benthamiana* leaves and engineered *S. cerevisiae* cells were subsequently used for the characterization of CYP712K4, a cytochrome P450 from *M. ilicifolia* that was shown to catalyze the three-step oxidation of friedelin at the C-29 position, leading to the heterologous production of maytenoic acid in yeast and *N. benthamiana*.

Results

Expression of *MiFRS4* in *N. benthamiana* leads to heterologous friedelin production

In a pilot study, *M. ilicifolia* FRS (*Mi*FRS) was shown to catalyze the cyclization of 2,3-oxidosqualene to friedelin in *S. cerevisiae* (Souza-Moreira et al. 2016). Recently, three additional *MiFRS* isoforms, *MiFRS2*, *MiFRS3*, and *MiFRS4* were cloned from *M. ilicifolia* cDNA and characterized in yeast (Alves et al. 2018). As expression of *MiFRS4* (GenBank accession number MG677554) in yeast led to the highest friedelin production (Alves et al. 2018), this FRS isoform was chosen for friedelin biosynthesis in *N. benthamiana* leaves by transient expression using *A. tumefaciens*-mediated leaf infiltration. To enhance the heterologous triterpene production capacity of the *N. benthamiana* leaves, a truncated version of *Medicago truncatula HMGR1* (*tHMGR1*) was co-infiltrated with *MiFRS4* (Pollier et al. 2013; Reed and Osbourn 2018). In addition, gene silencing was suppressed by co-infiltration with the *35S:p19* strain (Voinnet et al. 2003). Organic extracts of *N. benthamiana* leaves harvested four days post-infiltration were analyzed by gas chromatography-mass spectrometry (GC-MS). Compared to the GC-MS chromatogram of leaves infiltrated with *tHMGR1* only, the GC-MS chromatogram of leaves infiltrated with *tHMGR1* and *MiFRS4* revealed a single unique peak eluting at 35.5 min (Figure 1B). The retention time and electron ionization (EI)-MS fragmentation spectrum of this peak corresponded well with those of an authentic friedelin standard (Figure 1B, C), confirming functionality of *MiFRS4* upon heterologous expression in *N. benthamiana*. Quantification of the friedelin produced by the infiltrated leaves indicated a production of 0.43 ± 0.16 mg of friedelin per gram (dry weight) of *N. benthamiana* leaves.

A *M. ilicifolia* RNA-Seq database for the identification of biosynthesis genes

Previously, three members of the CYP712K subfamily from the Chinese medicinal plant *T. wilfordii* were

suggested to catalyze the C-29 oxidation of friedelin, thereby yielding maytenoic acid and its precursor 29-hydroxyfriedelin (Hansen 2017). To identify the gene(s) encoding the CYP712K orthologue(s) from *M. ilicifolia*, we first created a *M. ilicifolia* root transcriptome via RNA-Seq. RNA extracted from *M. ilicifolia* roots was used for paired-end RNA sequencing with an Illumina MiSeq system, yielding a total of 24,435,760 paired-end reads (2x76 bp). *De novo* transcriptome assembly generated 38,256 contigs with an average length of 760 bp. The minimum and maximum contig length were 173 and 6,585 bp, respectively. A TBLASTX search with the nucleotide sequence of *T. wilfordii CYP712K1* in the generated *M. ilicifolia* RNA-Seq assembly led to the identification of a single full-length candidate *CYP712K* sequence in addition to several partial gene sequences that correspond to different *CYP712K* isoforms. The full-length coding sequence of the candidate *CYP712K* gene was cloned from *M. ilicifolia* cDNA, and the nucleotide sequence obtained was submitted to GenBank (accession number MK829814) and the P450 naming committee, which annotated the encoded protein as CYP712K4.

CYP712K4 oxidizes the friedelin backbone in *N. benthamiana* leaves

To investigate the functionality of *CYP712K4* in *N. benthamiana* leaves, the gene was transiently expressed in combination with *tHMGR1* and *MiFRS4*. As the volatility of oxidized friedelin is too low for GC-MS analysis and as trimethylsilylation of friedelin and its derivatives leads to multiple peaks, likely due to stabilization of the enol tautomer by the derivatization reagent (Figure S1), metabolite profiling of infiltrated *N. benthamiana* leaves was carried out using liquid chromatography atmospheric pressure chemical ionization Fourier-transform mass spectrometry (LC-APCI-FT-MS). The LC-APCI-FT-MS chromatogram of an extract of leaves infiltrated with *tHMGR1* and *MiFRS4* only showed a peak eluting at 24.1 min that co-elutes with the friedelin standard (Figure 2A). The molecular ion of friedelin produced in *N. benthamiana* had a mass of 427.39349 Da (determined by FT-MS), which corresponds well (δ ppm = 0.111) with the predicted mass of friedelin and the accurate mass (427.39367 Da) observed for the friedelin standard (Figure 2B). Furthermore, the MS^2 fragmentation spectra of the friedelin produced in *N. benthamiana* and the friedelin standard are nearly identical (Figure 2C), confirming that the peak eluting at 24.1 min corresponds to friedelin.

Compared to leaves infiltrated with *tHMGR1* and *MiFRS4* only, the LC-APCI-FT-MS chromatogram of an extract of leaves infiltrated with *tHMGR1*, *MiFRS4*, and *CYP712K4* contained two unique peaks (Figure 2A). The first new peak, eluting at 12.8 min, corresponds to a protonated metabolite with a molecular ion at *m*/*z* 457.36767 Da (Figure 2D). The calculated chemical formula corresponding to this mass, $C_{30}H_{48}O_3$ (δ ppm = 0.105), indicates that the first peak might correspond to a metabolite with a friedelin backbone and a carboxylic acid moiety. The presence of a carboxylic acid moiety was further confirmed by the MS^n fragmentation spectra of the molecule, in which a neutral loss of a formic acid molecule was observed in addition to the loss of a water molecule (Figure 2E, F). The second peak, eluting at 13.5 min with an observed mass of 443.38840 Da and a calculated chemical formula of $C_{30}H_{50}O_2$ (δ ppm = 0.096) likely corresponds to an alcohol of friedelin, as is also corroborated with the sequential loss of two water molecules upon MS^2 fragmentation of the metabolite (Figure 2G). The production of an alcohol and an acid of friedelin upon infiltration of *CYP712K4* in *N. benthamiana* leaves suggests that CYP712K4 catalyzes a three-step oxidation of a free methyl group of the friedelin backbone, yielding an acid, in analogy with for instance CYP716A12 that catalyzes the three-step oxidation of β-amyrin to oleanolic acid

(Carelli et al. 2011) and CYP72A154 or CYP72A63 that catalyze the three-step oxidation of β-amyrin and 11-oxo-β-amyrin to 11-deoxoglycyrrhetinic acid and glycyrrhetinic acid, respectively (Seki et al., 2011).

Engineering of *S. cerevisiae* for the production of friedelanes

To produce friedelanes in *S. cerevisiae* we opted for a yeast strain that was engineered and successfully used for the production of friedelin. In this strain, which we named KB1 (Table 1), the native *ERG7* promoter was replaced with the weak kexin (*KEX2*) promoter, leading to the production of friedelin upon constitutive expression of codon-optimized *MiFRS* and *tHMG1* from the high-copy number plasmid pSP[P_{TEF1}-*MiFRS*;P_{PGK1}-*tHMG1*]. To further improve the friedelin production of this strain, we evaluated its production capacity in the presence of methyl-β-cyclodextrins (MβCD), cyclic oligosaccharides that sequester triterpenes, leading to increased triterpene production in yeast (Moses et al. 2014). When cultivated in the absence of MβCD, strain KB1 produced 0.90 ± 0.34 mg/L of friedelin (Figure 3A-C). Compared to the production in its wild-type strain CEN.PK113-5D (0.30 ± 0.01 mg/L), friedelin production in strain KB1 was about threefold higher, confirming improved friedelin production in strain KB1. When cultivated in the presence of 5 mM of MβCD, the total friedelin production of strain KB1 increased to 2.20 ± 0.98 mg/L. Moreover, as reported for β-amyrin (Moses et al. 2014), about half of the friedelin was sequestered in the spent medium, whereas the other half remained within the cells (Figure 3C). However, upon repeated cultivation of these strains, we noticed that the friedelin production was highly variable between individual colonies derived from the same transformation event and that the overall production of the created strains decreased over time. We reasoned that this might be due to toxicity of the constitutively produced friedelin, leading to decreased fitness and outcompetition of high-producing cells. To overcome this, we expressed the codon-optimized *MiFRS* from the high-copy number, galactose-inducible plasmid pESC-URA-tHMG1-DEST (Fiallos-Jurado et al. 2016). With a production of 3.45 ± 0.23 mg/L of friedelin in the presence of MβCD by galactose-induced yeast cells expressing *tHMG1* and *MiFRS*, the friedelin production was similar to that following constitutive expression (Figure 3D), but the variation between the individual colonies was considerably reduced and the overall production did not decrease over time.

To use this strain for the production of friedelanes, simultaneous expression of *MiFRS*, *CYP712K4* and a gene encoding a plant P450 reductase needs to be achieved. However, KB1 has only one available auxotrophic marker, *URA3* (Table 1). Hence, additional auxotrophic markers were introduced into this strain. Sequential CRISPR/Cas9-mediated knockout of *TRP1*, *HIS3* and *LEU2* led to strains KB4 (KB1; *trp1Δ0*); KB7 (KB4; *his3Δ0*) and KB10 (KB7; *leu2Δ0*), respectively, the latter strain having four available auxotrophic markers (Figure 4A, B, Table 1). To further improve friedelane production, the phosphatidic acid phosphatase-encoding *PAH1* was also disrupted using CRISPR/Cas9 in strain KB10, leading to strain KB13 (Figure 4B). Knockout of *PAH1* has been reported to lead to a dramatic expansion of the endoplasmic reticulum, which stimulates triterpenoid production (Arendt et al. 2017). Next, we evaluated friedelin production in these strains (Figure 4C), which revealed that introduction of additional auxotrophic markers has a negative effect on friedelin production, whereas disruption of *PAH1* improved friedelin production. The negative effect on friedelin production by the introduction of additional auxotrophic markers was anticipated as it has been observed that nutritional auxotrophy supplementation leads to lower yeast growth rates compared to genetic auxotrophy complementation (Pronk 2002). The final friedelin yield in strain KB13 was 2.27 ± 0.32 mg/L. These results were confirmed by quantification of

the β-amyrin production upon galactose-induced expression of *tHMG1* and *GgbAS* (Moses et al. 2014) in the same strains (Figure 4D). β-amyrin production in strain KB13 was 3.20 ± 0.09 mg/L.

Evaluation of CYP712K4 in *S. cerevisiae*

To verify the results obtained with *CYP712K4* in *N. benthamiana*, we expressed this gene in yeast strain KB13, in combination with *tHMG1*, codon-optimized *MiFRS* and the *Medicago truncatula* P450 reductase *MTR1*, all from galactose-inducible plasmids. This yeast strain was cultivated in the presence of MβCD, and extracts of the spent medium of this strain were compared to those of a control KB13 strain expressing an empty vector instead of *CYP712K4*. The LC-APCI-FT-MS chromatograms of extracts of the strain expressing *CYP712K4* showed three unique peaks that did not occur in extracts of the empty vector control strain (Figure 5A), suggesting functionality of *CYP712K4* in yeast. The accurate mass and fragmentation spectra of the first two peaks matched well with the accurate mass and fragmentation spectra of the acid and alcohol of friedelin produced in *N. benthamiana* leaves infiltrated with *tHMGR1*, *MiFRS4*, and *CYP712K4* (Figure 5B-D). The third peak, eluting at 14.8 min with an observed mass of 441.37258 Da (Figure 5E) and a calculated chemical formula of $C_{30}H_{48}O_2$ (δ ppm = -0.288) likely corresponds to an aldehyde of friedelin, which is also corroborated with the sequential loss of two water molecules upon MS^2 fragmentation of the metabolite (Figure 5F). This data confirms that CYP712K4 likely catalyzes the three-step oxidation of a free methyl group of the friedelin backbone. In *N. benthamiana*, the acid of friedelin is the major product, with only trace amounts of the alcohol and no detectable levels of the aldehyde. In contrast, yeast produces significant amounts of both the alcohol and aldehyde intermediates in addition to the acid of friedelin.

CYP712K4 catalyzes the three-step oxidation of friedelin at the C-29 position to yield maytenoic acid

To unambiguously confirm the identity of the metabolites produced upon expression of *tHMGR1*, *MiFRS4*, and *CYP712K4* in *N. benthamiana* and *S. cerevisiae*, we carried out a large-scale leaf infiltration of over 100 *N. benthamiana* plants. Seven days after infiltration, the leaves were harvested, frozen in liquid nitrogen and extracted with dichloromethane. After solvent evaporation, the crude leaf extract was used for compound isolation by silica column chromatography, which yielded 10 mg of the main CYP712K4 product. The identity of the metabolite was determined via NMR analysis, which confirmed the presence of a carboxyl group at the C-29 position of the friedelin backbone (Figure S2, S3). These data are consistent with the molecular mass and MS^2 fragmentation patterns determined via LC-APCI-FT-MS and suggest that CYP712K4 catalyzes the oxidation of friedelin at the C-29 position, thereby yielding maytenoic acid, when expressed in yeast or *N. benthamiana* leaves. Based on this information and in analogy with, for instance, CYP716A12 (Carelli et al. 2011) or CYP72A154 (Seki et al., 2011), we concluded that the alcohol and aldehyde intermediates correspond to 29-hydroxyfriedelin and 3-oxo-friedelan-29-al (Figure 6A).

Discussion

Like other members of the Celastraceae family, the South-American medicinal plant *M. ilicifolia* accumulates a set of quinone methide triterpenoids with important pharmacological properties. The major quinone methide triterpenoids present in *M. ilicifolia* root bark are maytenin and pristimerin, metabolites that are likely derived

from celastrol, a promising agent for the treatment of obesity (Liu et al. 2015) that also accumulates to significant amounts in the root bark of this medicinal plant (Buffa Filho et al. 2002). Via heterologous expression in yeast and *N. benthamiana*, we characterized CYP712K4, a cytochrome P450 from *M. ilicifolia* that catalyzes the three-step oxidation of friedelin at the C-29 position (Figure 6A), leading to maytenoic acid, an intermediate of the quinone methide triterpenoid biosynthesis pathway.

The quinone methide triterpenoid biosynthesis pathway

The three-step oxidation of friedelin by CYP712K4 can be considered the second step of the *M. ilicifolia* quinone methide triterpenoid biosynthesis pathway, the first committed step being the cyclization of 2,3-oxidosqualene to friedelin, catalyzed by *Mi*FRS (Figure 1A)(Alves et al. 2018; Souza-Moreira et al. 2016). The subsequent steps in the biosynthesis of celastrol and ultimately maytenin and pristimerin remain speculative. A possible biosynthetic intermediate is cangoronine (Figure 6B), which was also shown to accumulate in the root bark of *M. ilicifolia* (Itokawa et al. 1991). This metabolite may be derived from maytenoic acid by oxidation at the C-2 and C-24 positions. Oxidation of maytenoic acid at the C-2 position would lead to wilforic acid C and 3-hydroxy-2-oxo-D:A-friedelan-3-en-29-oic acid (Figure 6B), metabolites that have been isolated from plants producing quinone methide triterpenoids (Li et al. 1997; Morota et al. 1995). In *Centella asiatica*, CYP716C11 catalyzes the C-2α oxidation of oleanolic acid to maslinic acid (Miettinen et al. 2017), and thus also the oxidation of maytenoic acid at the C-2 position may be catalyzed by a P450. The produced 3-hydroxy-2-oxo-D:A-friedelan-3-en-29-oic acid may be further converted to cangoronine by the activity of a second P450 (Figure 6B). This P450 may also accept maytenoic acid as a substrate. A single oxidation of maytenoic acid at its C-24 position may lead to a metabolite with a C-24 hydroxyl group, which will be quickly interconverted to its hemiketal salaspermic acid by intramolecular nucleophilic addition of the C-24 hydroxyl to the C-3 carbonyl (Figure S4). Salaspermic acid was isolated from *Salacia macrosperma*, a member of the Celastraceae family that also accumulates maytenin and pristimerin (Kutney et al. 1981; Viswanathan 1979), indicating the possibility of this reaction. Furthermore, the isolation of orthosphenic acid from *Orthosphenia mexicana* Standley (González et al. 1983) and wilforic acid E from *T. wilfordii* (Duan et al. 2000), both plant species that accumulate quinone methide triterpenoids, suggests salaspermic acid may serve as a substrate for an enzyme that catalyzes a two-step oxidation at the C-2 position. This enzyme is likely the P450 that also catalyzes the 2-step oxidation at the C-2 position of maytenoic acid (Figure 6B). However it cannot be excluded, further oxidation of wilforic acid E to cangoronine seems unlikely as opening of the intramolecular hemiketal ring is energetically disfavored. Together, these observations suggest that maytenoic acid may be the substrate for the competitive action of two P450s. If oxidation happens at the C-24 position, a stable hemiketal is produced that cannot be further converted to cangoronine and downstream quinone methide triterpenoids. On the other hand, if oxidation occurs first at the C-2 position, the resulting 3-hydroxy-2-oxo-D:A-friedelan-3-en-29-oic acid carries a hydroxyl group at C-3 position, which prevents intramolecular nucleophilic addition and allows further oxidation at the C-24 position leading to cangoronine (Figure 6B).

Ultimately, for the biosynthesis of celastrol, the C-5β formyl moiety of cangoronine needs to be removed. The occurrence of salaspermic acid and cangoronine, with respectively a hydroxyl and a carbonyl group at the C-24 position, suggests oxidative elimination of the C-5β methyl group from the friedelane backbone. Several mechanisms for oxidative demethylation of triterpenes exist in nature. CYP51 family members are involved in

sterol synthesis and demethylate 2,3-oxidosqualene cyclization products like lanosterol and cycloartenol in three oxidative steps. During the first two steps, the 14α-methyl group is oxidized into an aldehyde via an alcohol intermediate. In the final step, the aldehyde group is released as formic acid and a double bond is simultaneously introduced in the sterol backbone (Lepesheva and Waterman 2007). As introduction of multiple double bonds in the friedelane backbone is essential in the quinone methide biosynthesis, such a P450-catalyzed oxidative C-5β demethylation may occur. However, alternative demethylation mechanisms exist. In triterpene biosynthesis, two distinct C-4 demethylation processes have been described. In sterol synthesis, C-4α demethylation is achieved by the successive action of C-4 sterol methyloxidase, sterol C4-decarboxylase and 3-keto reductase (Nes 2011), whereas in the biosynthesis of helvolic acid, C-4β demethylation is achieved by successive action of a cytochrome P450 and a promiscuous short-chain dehydrogenase/reductase enzyme (Lv et al. 2017). For both mechanisms, decarboxylation occurs only when a carbonyl group is present at the C-3 position (Lv et al. 2017), as is the case for maytenoic acid, but not for cangoronine. Furthermore, demethylation according to these reaction mechanisms would not lead to the introduction of an additional double bond and would imply further oxidation of the C-24 aldehyde towards a carboxylic acid. However, no friedelanes with a C-24 carboxylic acid group have been described (Shan et al. 2013). Because of these reasons, the multi-enzyme reaction mechanisms seem less likely, and thus we propose C-5β demethylation of the friedelane backbone via three oxidative steps catalyzed by a single P450 enzyme (Figure S5).

After demethylation of the friedelane backbone, two additional double bonds need to be introduced into the resulting phenolic 24-norfriedelane backbone to obtain the quinone methide triterpenoid celastrol. The presence of several phenolic 24-norfriedelanes with a hydroxyl or a carbonyl group at the C-6 and/or C-7 position (Shan et al. 2013) suggests the additional double bonds are introduced via oxidation and dehydration at these positions (Figure S6). In *M. ilicifolia*, the produced celastrol may be the substrate of two additional enzymes, a methyltransferase that methylates the C-29 carboxylic acid group of celastrol, leading to pristimerin, and an oxidase, likely a P450, that oxidizes celastrol at the C-21 position, resulting in another oxidative demethylation of the friedelane backbone and ultimately the formation of maytenin (Figure S7).

Towards heterologous production of quinone methide triterpenoids

In spite of being promising natural products for the development of anti-cancer drugs, the commercial production of quinone methide triterpenoids is hampered by slow growth of the producing plant species and low production amounts observed *in planta* (Coppede et al. 2014). As an alternative to their natural plant sources, heterologous production of these valuable compounds may be envisaged. Quantification of the friedelin extracted from the *N. benthamiana* leaves indicated a production of 0.43 ± 0.16 mg/g (dry weight) of leaves. This is considerably lower than the production amounts previously reported for β-amyrin (3.3 mg/g dw) purified from vacuum-infiltrated *N. benthamiana* leaves (Reed et al. 2017) and may be due to a lower catalytic efficiency of friedelin synthase compared to β-amyrin synthase when expressed in *N. benthamiana* leaves. However, we did not observe such a difference in production amounts upon expression of *MiFRS* or *GgbAS* in yeast.

The *S. cerevisiae* strain that we engineered for the production of friedelanes, KB13, produced 2.27 ± 0.32 mg of friedelin per liter. This is roughly the double of its parent strain KB1 that achieved a production of 0.90 ± 0.34 mg/L. Increased production was mainly achieved by using MβCD in the cultivation medium and by knocking

out *PAH1*. The introduction of additional auxotrophic markers, necessary for the use of multiple plasmids, drastically decreased the production capacity of our yeast strain, which may be due to lower yeast growth rates upon nutritional auxotrophy supplementation compared to genetic auxotrophy complementation (Pronk 2002). Recently, Zhou et al. (2019) reported a heterologous friedelin production of 37.07 mg/L in a yeast strain expressing a friedelin synthase from *T. wilfordii*. These high production amounts were achieved by extensive engineering of a diploid BY4741 starter strain, optimization of the cultivation medium, and by mutation of the *T. wilfordii* friedelin synthase (Zhou et al. 2019). This underscores the potential of yeast to reach high titers of friedelin and downstream biosynthesis products required for the economical production of valuable quinone methide triterpenoids.

A prerequisite for the reconstitution of the complete quinone methide triterpenoid biosynthesis pathway and heterologous production of these compounds is the identification of the remaining biosynthetic enzymes. Given the specific location of their biosynthesis and accumulation, with friedelin being produced in the leaves and subsequently translocated to the roots where it is further oxidized (Corsino et al. 2000), a co-expression analysis with the identified *CYP712K4* in transcriptome datasets of different *M. ilicifolia* tissues, including roots, may lead to the identification of candidate biosynthesis genes. The functionality of these genes could subsequently be checked in maytenoic acid-producing *N. benthamiana* leaves or *S. cerevisiae* cells, which could allow for the step-wise reconstitution of the quinone methide triterpenoid biosynthesis pathway and ultimately may provide a sustainable heterologous source of these pharmacologically important metabolites.

Materials and Methods

RNA extraction

Total RNA was extracted from frozen *M. ilicifolia* roots using the RNeasy Plant mini kit (Qiagen, USA). The RNA quality was checked using a Bioanalyzer instrument (Agilent Technologies), and the RNA concentration was determined using a Nanodrop-1000 spectrophotometer (Thermo Scientific, USA). Subsequently, mRNA was isolated from the total RNA with using magnetic Oligo(dT) beads. Finally, the mRNA was chemically fragmented according to the instructions of the Illumina TruSeq RNA sample preparation v3 kit (Illumina, USA).

RNA-Seq and transcriptome assembly

The paired-end cDNA sequencing library was prepared using the Illumina TruSeq RNA sample preparation v3 kit (Illumina, USA). Quantification and quality assessment of the resulting libraries was carried out using a Bioanalyzer instrument (Agilent Technologies). A total of 0,3 pmol of library was loaded into the reagent Cartridge of the MiSeq sequencer (Illumina). Paired-end sequencing, in which the template fragments were sequenced in both the forward and reverse directions, was carried out for 76 cycles (2x76 bp). The quality of the raw data was checked using FastQC 0.11.7 and trimmed using Trimmomatic 0.36 (Bolger et al. 2014). Read trimming included adapter trimming, quality trimming (threshold of 25), removing low-quality bases at the start and the end of the reads (Score <25), and removing the reads below 50 bp length. The trimmed reads were used for *de novo* transcriptome assembly using Trinity 2.2.0 (Grabherr et al. 2011) with default parameters, except for a mismatch cost of 2, and an insertion and deletion cost of 3.

Full-length cloning and generation of expression clones

The full-length *MiFRS4* coding sequence (Alves et al. 2018) was PCR-amplified using primer pair COMBI7292-COMBI7293 (Table S1) and Gateway recombined into pDONR207. For cloning of the full-length *CYP712K4* coding sequence, single-stranded cDNA was prepared using the iScript™ cDNA Synthesis kit (Bio-Rad) from 500 ng of total RNA isolated from *M. ilicifolia* roots. The prepared cDNA was used as PCR template for the amplification of the full-length *CYP712K4* coding sequence using the primer pair COMBI6598-COMBI6599 (Table S1). The obtained PCR product was purified using the GeneJet Gel Extraction kit (Thermo Fisher Scientific), reamplified using the AttB primer pair COMBI6536-COMBI6537, and finally Gateway recombined into pDONR207. Sequence-verified *MiFRS4* and *CYP712K4* entry clones were Gateway recombined into pK7WG2D (Karimi et al. 2002) and pEAQ-*HT*-DEST1 (Sainsbury et al. 2009), respectively. *M. truncatula tHMGR1* (Pollier et al. 2013) was cloned in pEAQ-*HT*-DEST1. For expression in *N. benthamiana*, sequence-verified expression clones were transformed into *A. tumefaciens* C58C1 RifR (pMP90). For expression in *S. cerevisiae*, the codon-optimized *MiFRS* (Souza-Moreira et al. 2016) was PCR-amplified using the primer pair COMBI6933-COMBI6934 (Table S1), Gateway recombined into pDONR207, sequence-verified and finally Gateway recombined into the high-copy number yeast destination vector pESC-URA-tHMG1-DEST (Fiallos-Jurado et al. 2016). *CYP712K4* was Gateway recombined into the high-copy number yeast destination vector pAG423GAL-ccdB (Addgene plasmid 14149; (Alberti et al. 2007). As P450 reductase, *MTR1* (Miettinen et al. 2017) cloned in the low-copy number yeast destination vector pAG415GAL-ccdB (Addgene plasmid 14145; (Alberti et al. 2007) to ensure an optimal P450 reductase:P450 ratio (Moses et al. 2014) was used.

Agroinfiltration of *N. benthamiana* leaves

N. benthamiana leaf infiltrations were carried out as described (Moses et al. 2015). To enhance triterpene yield, *A. tumefaciens* containing *M. truncatula tHMGR1* (Pollier et al. 2013) was infiltrated together with the strains containing the triterpene biosynthesis genes (Reed and Osbourn 2018). In addition, the 35S:p19 strain was co-infiltrated to suppress gene silencing (Voinnet et al. 2003) when no *A. tumefaciens* strains containing the pEAQ-HT-DEST1 vector were used.

Metabolite extractions for GC-MS or LC-APCI-FT-MS analysis

Four days after Agroinfiltration, *N. benthamiana* leaves were harvested and ground to a fine powder in liquid nitrogen. For each sample, 100 mg of leaf material was extracted with 1 mL of methanol. The resulting organic extract was evaporated to dryness under vacuum. For GC-MS analysis of friedelin, the obtained residue was dissolved in 200 µL of heptane. For GC-MS analysis of β-amyrin, the residue was trimethylsilylated using 10 µL of pyridine and 50 µL of *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide. For LC-APCI-FT-MS analysis, the residue was dissolved in 200 µL of acetonitrile.

GC–MS analysis

GC–MS analysis was carried out using a GC model 6890 and MS model 5973 (Agilent, Santa Clara, United States) as described (Moses et al. 2014). For quantification, calibration curves were generated for authentic friedelin and

β-amyrin standards (Sigma-Aldrich). Peak areas were calculated with the default settings of the ChemStation software (Agilent).

LC–APCI–FT–MS analysis

LC–APCI–FT-MS was performed using a Hypersil BDS C18 HPLC column (100 x 2.1 mm, 3 μm; Thermo Fisher Scientific) mounted on an Accela UPLC system (Thermo Electron Corporation). The LC system was coupled to a LTQ FT Ultra (Thermo Electron Corporation) *via* an atmospheric pressure chemical ionization (APCI) system operated in positive mode. The following gradient was run using water/acetonitrile (99:1, v/v) (solvent A) and acetonitrile/water (99:1, v/v) (solvent B): time 0 min, 50% B; 19 min, 99.5% B; 25 min, 99.5% B. The injection volume was 10 μL, flow rate 400 μL/min, and column temperature 40°C. Positive ionization using the APCI source was obtained with the following parameter values: capillary temperature 200°C, APCI probe 350°C, sheath gas 40 (arbitrary units), aux. gas 5 (arbitrary units), spray voltage 6.0°kV, discharge current 5 μA, capillary voltage 4.0 V, and tube lens 70 V. Full MS spectra between *m*/*z* 250–1400 were recorded at a resolution of 100,000. Full MS spectra were interchanged with a dependent MS^2 scan event in which the most abundant ion in the previous full MS scan was fragmented, two dependent MS^3 scan events in which the two most abundant daughter ions were fragmented, and a dependent MS^4 scan event in which the most abundant granddaughter ion of the first MS^3 scan event was fragmented. The collision energy was set at 35%.

Purification of maytenoic acid

Frozen leaf material from large-scale infiltrations (108 plants in total) was ground using a mixer in a total of 2 L of cold dichloromethane and allowed to extract for three days at room temperature. The organic phase was dried using a rotavapor, yielding 3 g of crude extract. Half of this crude extract was submitted to a first purification by silica gel chromatography eluted with a hexane:ethyl acetate (0-100%) gradient, resulting in ten fractions. Fraction F7 was further purified by silica gel chromatography eluted with a chloroform:acetone (0-100%) gradient. From one of the resulting fractions, 10 mg of pure maytenoic acid was obtained.

NMR analysis

All NMR spectra were measured on a Bruker Avance III HD spectrometer operating at a ^1H and ^{13}C frequency of 600 and 150 MHz, respectively. The dried sample was dissolved in 650 μL of deuterated chloroform ($CDCl_3$) with 99.96% atom-D. The spectra recorded on the sample included 1D ^1H, ^{13}C DEPT-135, ^1H-{^1H} total correlation spectroscopy (TOCSY) and 2D ^1H-{^1H} gradient selected correlation spectrometry (COSY), ^1H-{^1H} nuclear Overhauser spectroscopy (NOESY), ^1H-{^{13}C} multiplicity-edited gradient heteronuclear single quantum correlation (HSQC), and ^1H-{^{13}C} heteronuclear multiple bond correlation (HMBC). Chemical shifts were recorded in ppm (δ_H for ^1H and δ_C for ^{13}C) and referenced to the residual $CDCl_3$ solvent signals at 7.26 and 77.2 ppm for ^1H and ^{13}C frequency, respectively. Coupling constants are given in Hertz (Hz) and the spectra were processed by TopSpin 3.6.0 (Bruker).

First, the analysis of the ^1H NMR (Figure S2A) in combination with ^1H-{^{13}C}-gHSQC, allowed the immediate identification of seven aliphatic methyl groups at δ_H 1.26(s), 1.10(s), 1.0(s), 0.88(s), 0.87(s), 0.86(d) and 0.71(s),

which indicated the loss of one methyl group in comparison with the eight methyl groups observed for friedelin (Ragasa et al. 2015). However, sufficient information could be extracted from the data to prove the friedelin backbone, which is characterized by the presence of a carbonyl group on ring A, which was observed at δ_C 213.5 ppm on the ^{13}C DEPT-135 spectrum, and further confirmed by correlation of this carbonyl group with specific peaks of hydrogens from this ring at δ_H 2.40 (ddd, H-2β), 2.24 (dd, H-2α), 2.21 (q, H-4), 1.97 (m, H-1α) and methyl group Me23, δ_H 0.86 (d) on the ^1H-{^{13}C}-HMBC experiment (Figure S3A). Besides those, other high field peaks, corresponding to three hydrogens, were observed. The structural limitation of the friedelin backbone results in considerable resonance overlapping of the aliphatic peaks. In order to overcome this, ^1H-{^1H}-TOCSY experiments (Figure S2B), irradiating at δ_H 2.40, 2.34, 2.16 and 1.75 ppm, showed individual spin coupling systems that allowed the complete assignment of rings A and B. The ^1H-{^1H}-TOCSY experiment showed that the three other high field peaks corresponding to protons around a second polar moiety on the structure. High-resolution MS data of this compound indicated the presence of a carboxyl group on the structure, suggesting that the missing methyl group from the friedelin backbone was oxidized to an acid, which was further confirmed by the observation of a quaternary carbon at δ_C 184.6 ppm on the ^{13}C DEPT-135 spectrum (Figure S3B). ^1H-{^{13}C}-HMBC correlation between the carboxyl carbon with Me30 at δ_H 1.26, locates the carboxyl group at the C-29 position, on ring E. In addition, combined information of ^1H-{^{13}C}-HMBC correlation between C-29 and the non-assigned high field protons, along with the observation of individual coupling systems for peaks at δ_H 2.34 and 2.16, on ^1H-{^1H}-TOCSY, allowed the assignment of these peaks as H-19 and H-21, respectively, completing ring E proton assignment (Figure S2B). Together, these data indicate maytenoic acid (3-oxofriedelan-29-oic acid or polpunonic acid) as the oxidation product of CYP712K4. Full assignment of protons and carbons for maytenoic acid (Table S2) was achieved by detailed analysis of all NMR experiments and comparison with data from the literature (Lindsey et al. 2006).

Yeast engineering

Yeast strain KB4 was derived from strain KB1 by knocking out the *TRP1* gene using CRISPR/Cas9. To this end, 1 μg of pCAS-TRP1 (Calegario et al. 2016) and 10 μmol of homologous recombination (HR) donor (prepared by annealing the single-stranded oligonucleotides CRISPR059 and CRISPR060; Table S1) were co-transformed in yeast strain KB1. The resulting colonies were analyzed for positive CRISPR events by replica plating on SD medium with or without tryptophan. Tryptophan auxotrophs were confirmed by Sanger sequencing and finally cured of pCAS-TRP1 by counterselection on YPD plates containing 1 mg/mL 5-fluoroorotic acid (5-FOA; Zymo Research, Irvine, United States). The resulting KB1-derived *trp1Δ0* strain was named KB4.

Yeast strain KB7 was derived from strain KB4 by CRISPR/Cas9-mediated knock-out of the *HIS3* gene. First, pCASmGG2 that contains an optimized sgRNA structure to improve CRISPR-Cas9 knockout efficiency (Dang et al. 2015) was created from the advanced CRISPR vector pCASmGG (Arendt et al. 2017). Mutation of the sgRNA sequence was achieved by PCR amplification of two overlapping fragments using primer pairs COMBI3245-COMBI4867 and COMBI4866-COMBI3246 with pCASmGG as template. The resulting PCR fragments were joined by overlap-extension PCR using primer pair COMBI3245-COMBI3246, subcloned into pDONR221 and finally recombined into pCASm-ccdB (Arendt et al. 2017), leading to pCASmGG2. The HIS3 CRISPR plasmid was generated by GoldenGate cloning of the oligonucleotides CRISPR587 and CRISPR588 in the created

advanced CRISPR vector pCASmGG2 as described (Arendt et al. 2017). One μg of the resulting pCASmGG2-HIS3 plasmid and 10 μmol of HR donor (annealed oligonucleotides CRISPR585 and CRISPR586) were co-transformed in yeast strain KB7 already containing the p414-TEF1p-Cas9-CYC1t (DiCarlo et al. 2013) plasmid. The resulting colonies were analyzed for positive CRISPR events by replica plating on SD medium with or without histidine. The identified histidine auxotrophs were confirmed by Sanger sequencing and cured of pCASmGG2-HIS3 and p414-TEF1p-Cas9-CYC1t by counterselection on YPD plates containing 1 mg/mL 5-FOA and 0.5 mg/mL 5-fluoroanthranilic acid (5-FAA; Sigma-Aldrich). The resulting KB4-derived *his3Δ0* strain was named KB7.

Yeast strain KB10 was derived from strain KB7 by knocking out the *LEU2* gene using CRISPR/Cas9. pCAS-LEU2 was prepared like pCAS-TRP1 (Calegario et al. 2016), using primer pair COMBI3245-CRISPR009 to amplify SNR52p and COMBI3246 and CRISPR026 to amplify sgRNA-CYC1t. One μg of the resulting pCAS-LEU2 plasmid and 10 μmol of HR donor (annealed oligonucleotides CRISPR049 and CRISPR050) were co-transformed in yeast strain KB7 already containing the p414-TEF1p-Cas9-CYC1t plasmid. The resulting colonies were analyzed by replica plating on SD medium with or without leucine and identified auxotrophs were confirmed by Sanger sequencing and cured of pCAS-LEU2 and p414-TEF1p-Cas9-CYC1t by counterselection on YPD plates containing 1 mg/mL 5-FOA and 0.5 mg/mL 5-FAA. The resulting KB7-derived *leu2Δ0* strain was named KB10.

Finally, yeast strain KB13 was derived from strain KB10 by knocking out the *PAH1* gene using CRISPR/Cas9. One μg of the pCAS-PAH1 (Arendt et al. 2017) plasmid and 10 μmol of HR donor (annealed oligonucleotides CRISPR459 and CRISPR460) were co-transformed in yeast strain KB10 already containing the p414-TEF1p-Cas9-CYC1t plasmid. The resulting colonies were analyzed by yeast colony PCR (primer pair CRISPR461-CRISPR462) and *pah1Δ0* strains were identified by restriction fragment length polymorphism (RFLP) in which a successful mutation led to the gain of a *Dra*I restriction site. Positive knock-out events were confirmed by Sanger sequencing (with RFLP primers) and the resulting strain was cured of pCAS-PAH1 and p414-TEF1p-Cas9-CYC1t by counterselection on YPD plates containing 1 mg/mL 5-FOA and 0.5 mg/mL 5-FAA. The resulting KB10-derived *pah1Δ0* strain was named KB13. All oligonucleotides used to create novel yeast strains and all yeast strains created and/or used in this study are listed in Table S1 and Table 1, respectively.

Production of friedelanes in yeast

To assess friedelin production in yeast, codon-optimized *MiFRS* (Souza-Moreira et al. 2016) was cloned in the high-copy number yeast destination vector pESC-URA-tHMG1-DEST (Fiallos-Jurado et al. 2016). For analysis of maytenoic acid production in yeast strain KB13, pESC-URA-tHMG1-*MiFRS* was used in combination with pAG423GAL-*CYP712K4* and pAG415GAL-*MTR1*. Yeast transformation was carried out using the lithium acetate/single-stranded carrier DNA/polyethylene glycol method (Gietz and Woods 2002) and transformed cells were selected on SD medium supplemented with appropriate dropout supplements (Clontech). The yeast cells were cultivated in the presence of methyl-β-cyclodextrin (MβCD) as described (Moses et al. 2014). For GC-MS analysis, 10 mL of the spent medium was extracted thrice with 0.5 volumes of hexane. The organic extracts were pooled, evaporated to dryness and the obtained residue was dissolved in 200 μL of heptane. For LC-APCI-FT-MS analysis, 10 mL of the spent medium was extracted thrice with ethyl acetate, the organic extracts were pooled,

evaporated to dryness and the obtained residue was dissolved in 200 μL of acetonitrile. GC-MS and LC-APCI-FT-MS analyses were carried out as described for the *N. benthamiana* extracts.

Funding

This work was supported by the São Paulo Research Foundation (FAPESP) [CIBFar-2013/07600-3; 2016/16970-7], the National Council for Scientific and Technological Development (CNPq) [303757/2017-5], the National Institute for Science and Technology (INCT)-FAPESP [2014/50926-0], INCT-CNPq [2014/465637-0], and the Coordination for the Improvement of Higher Education Personnel (CAPES) [Finance Code 001]. K.U.B. is a postdoctoral fellow of the São Paulo Research Foundation (FAPESP) [2015/25175-3; 2017/10911-1]. P.A. is indebted to the VIB International PhD Fellowship Program for a predoctoral fellowship. J.P. is a postdoctoral fellow of the Research Foundation-Flanders.

Disclosures

Conflicts of interest: No conflicts of interest declared

Acknowledgements

We thank David Nelson for naming the P450, Nivaldo Borale, Juliana Rodrigues, Bianca Ribeiro and Seowoo Kim for technical assistance, Nikolaj Hansen for sharing the *CYP712K1* sequence, Tatiana Souza-Moreira for the pSP[P_{TEF1}-*MiFRS*;P_{PGK1}-*tHMG1*] plasmid and yeast strain KB1, and Annick Bleys for help with preparing the manuscript.

References

Alberti, S., Gitler, A.D. and Lindquist, S. (2007) A suite of Gateway® cloning vectors for high-throughput genetic analysis in *Saccharomyces cerevisiae*. *Yeast* 24: 913-919.
Alves, T.B., Souza-Moreira, T.M., Valentini, S.R., Zanelli, C.F. and Furlan, M. (2018) Friedelin in *Maytenus ilicifolia* is produced by friedelin synthase isoforms. *Molecules* 23: 700.
Arendt, P., Miettinen, K., Pollier, J., De Rycke, R., Callewaert, N. and Goossens, A. (2017) An endoplasmic reticulum-engineered yeast platform for overproduction of triterpenoids. *Metab. Eng.* 40: 165-175.
Biral, L., Simmons, M.P., Smidt, E.C., Tembrock, L.R., Bolson, M., Archer, R.H., et al. (2017) Systematics of New World *Maytenus* (Celastraceae) and a new delimitation of the genus. *Syst. Bot.* 42: 680-693.
Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114-2120.
Buffa Filho, W., Corsino, J., Bolzani da Silva, V., Furlan, M., Pereira, A.M.S. and Castro França, S. (2002) Quantitative determination for cytotoxic *friedo-nor*-oleanane derivatives from five morphological types of *Maytenus ilicifolia* (Celastraceae) by reverse-phase high-performance liquid chromatography. *Phytochem. Anal.* 13: 75-78.
Calegario, G., Pollier, J., Arendt, P., de Oliveira, L.S., Thompson, C., Soares, A.R., et al. (2016) Cloning and functional characterization of cycloartenol synthase from the red seaweed *Laurencia dendroidea*. *PLoS ONE* 11: e0165954.
Carelli, M., Biazzi, E., Panara, F., Tava, A., Scaramelli, L., Porceddu, A., et al. (2011) *Medicago truncatula* CYP716A12 is a multifunctional oxidase involved in the biosynthesis of hemolytic saponins. *Plant Cell* 23: 3070-3081.
Cevatemre, B., Erkısa, M., Aztopal, N., Karakas, D., Alper, P., Tsimplouli, C., et al. (2018) A promising natural product, pristimerin, results in cytotoxicity against breast cancer stem cells *in vitro* and xenografts *in vivo* through apoptosis and an incomplete autopaghy in breast cancer. *Pharmacol. Res.* 129: 500-514.

Coppede, J.S., Pina, E.S., Paz, T.A., Fachin, A.L., Marins, M.A., Bertoni, B.W., et al. (2014) Cell cultures of *Maytenus ilicifolia* Mart. are richer sources of quinone-methide triterpenoids than plant roots *in natura*. *Plant Cell Tissue Organ Cult.* 118: 33-43.
Corsino, J., de Carvalho, P.R.F., Kato, M.J., Ribeiro Latorre, L., Oliveira, O.M.M.F., Araújo, A.R., et al. (2000) Biosynthesis of friedelane and quinonemethide triterpenoids is compartmentalized in *Maytenus aquifolium* and *Salacia campestris*. *Phytochemistry* 55: 741-748.
Dang, Y., Jia, G., Choi, J., Ma, H., Anaya, E., Ye, C., et al. (2015) Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. *Genome Biol.* 16: 280.
DiCarlo, J.E., Norville, J.E., Mali, P., Rios, X., Aach, J. and Church, G.M. (2013) Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.* 41: 4336-4343.
Duan, H., Takaishi, Y., Momota, H., Ohmoto, Y., Taki, T., Jia, Y., et al. (2000) Triterpenoids from *Tripterygium wilfordii*. *Phytochemistry* 53: 805-810.
Fiallos-Jurado, J., Pollier, J., Moses, T., Arendt, P., Barriga-Medina, N., Morillo, E., et al. (2016) Saponin determination, expression analysis and functional characterization of saponin biosynthetic genes in *Chenopodium quinoa* leaves. *Plant Sci.* 250: 188-197.
Gietz, R.D. and Woods, R.A. (2002) Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.* 350: 87-96.
González, A.G., Fraga, B.M., González, P., González, C.M., Ravelo, A.G. and Ferro, E. (1983) Crystal structure of orthosphenic acid. *J. Org. Chem.* 48: 3759-3761.
Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., et al. (2011) Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* 29: 644-652.
Hansen, N.L. (2017) Terpenoid pathway discovery in *Tripterygium wilfordii*: a Chinese medicinal plant. In *Department of Plant and Environmental Sciences* p. 188. University of Copenhagen, Copenhagen.
Itokawa, H., Shirota, O., Ikuta, H., Morita, H., Takeya, K. and Iitaka, Y. (1991) Triterpenes from *Maytenus ilicifolia*. *Phytochemistry* 30: 3713-3716.
Karimi, M., Inzé, D. and Depicker, A. (2002) GATEWAY™ vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.* 7: 193-195.
Kutney, J.P., Beale, M.H., Salisbury, P.J., Stuart, K.L., Worth, B.R., Townsley, P.M., et al. (1981) Isolation and characterization of natural products from plant tissue cultures of *Maytenus buchananii*. *Phytochemistry* 20: 653-657.
Lepesheva, G.I. and Waterman, M.R. (2007) Sterol 14α-demethylase cytochrome P450 (CYP51), a P450 in all biological kingdoms. *Biochim. Biophys. Acta* 1770: 467-477.
Li, K., Duan, H., Kawazoe, K. and Takaishi, Y. (1997) Terpenoids from *Tripterygium wilfordii*. *Phytochemistry* 45: 791-796.
Lindsey, K.L., Budesinsky, M., Kohout, L. and van Staden, J. (2006) Antibacterial activity of maytenonic acid isolated from the root-bark of *Maytenus senegalensis*. *S. Afr. J. Bot.* 72: 473-477.
Liu, J., Lee, J., Salazar Hernandez, M.A., Mazitschek, R. and Ozcan, U. (2015) Treatment of obesity with celastrol. *Cell* 161: 999-1011.
Lv, J.-M., Hu, D., Gao, H., Kushiro, T., Awakawa, T., Chen, G.-D., et al. (2017) Biosynthesis of helvolic acid and identification of an unusual C-4-demethylation process distinct from sterol biosynthesis. *Nat. Commun.* 8: 1644.
Mannowetz, N., Miller, M.R. and Lishko, P.V. (2017) Regulation of the sperm calcium channel CatSper by endogenous steroids and plant triterpenoids. *Proc. Natl. Acad. Sci. USA* 114: 5743-5748.
Miettinen, K., Iñigo, S., Kreft, L., Pollier, J., De Bo, C., Botzki, A., et al. (2018) The TriForC database: a comprehensive up-to-date resource of plant triterpene biosynthesis. *Nucleic Acids Res.* 46: D586-D594.
Miettinen, K., Pollier, J., Buyst, D., Arendt, P., Csuk, R., Sommerwerk, S., et al. (2017) The ancient CYP716 family is a major contributor to the diversification of eudicot triterpenoid biosynthesis. *Nat. Commun.* 8: 14153.
Morota, T., Yang, C.-X., Sasaki, H., Qin, W.-Z., Sugama, K., Miao, K.-L., et al. (1995) Triterpenes from *Tripterigium wilfordii*. *Phytochemistry* 39: 1153-1157.
Moses, T., Pollier, J., Almagro, L., Buyst, D., Van Montagu, M., Pedreño, M.A., et al. (2014) Combinatorial biosynthesis of sapogenins and saponins in *Saccharomyces cerevisiae* using a C-16α hydroxylase from *Bupleurum falcatum*. *Proc. Natl. Acad. Sci. USA* 111: 1634-1639.
Moses, T., Pollier, J., Shen, Q., Soetaert, S., Reed, J., Erffelinck, M.-L., et al. (2015) OSC2 and CYP716A14v2 catalyze the biosynthesis of triterpenoids for the cuticle of aerial organs of *Artemisia annua*. *Plant Cell* 27: 286-301.
Moses, T., Pollier, J., Thevelein, J.M. and Goossens, A. (2013) Bioengineering of plant (tri)terpenoids: from metabolic engineering of plants to synthetic biology *in vivo* and *in vitro*. *New Phytol.* 200: 27-43.
Nes, W.D. (2011) Biosynthesis of cholesterol and other sterols. *Chem. Rev.* 111: 6423-6451.
Niero, R., Faloni de Andrade, S. and Cechinel Filho, V. (2011) A review of the ethnopharmacology, phytochemistry and pharmacology of plants of the *Maytenus* genus. *Curr. Pharm. Design* 17: 1851-1871.

Périco, L.L., Rodrigues, V.P., de Almeida, L.F.R., Fortuna-Perez, A.P., Vilegas, W. and Hiruma-Lima, C.A. (2018) *Maytenus ilicifolia* Mart. ex Reissek. In *Medicinal and aromatic plants of South America (Medicinal and Aromatic Plants of the World 5)*. Edited by Albuquerque, U.P., Patil, U. and Máthé, Á. pp. 323-335. Springer, Dordrecht.
Pina, E.S., Silva, D.B., Teixeira, S.P., Coppede, J.S., Furlan, M., França, S.C., et al. (2016) Mevalonate-derived quinonemethide triterpenoid from *in vitro* roots of *Peritassa laevigata* and their localization in root tissue by MALDI imaging. *Sci. Rep.* 6: 22627.
Pollier, J., Moses, T., González-Guzmán, M., De Geyter, N., Lippens, S., Vanden Bossche, R., et al. (2013) The protein quality control system manages plant defence compound synthesis. *Nature* 504: 148-152.
Pronk, J.T. (2002) Auxotrophic yeast strains in fundamental and applied research. *Appl. Environ. Microbiol.* 68: 2095-2100.
Ragasa, C.Y., Ebajo Jr., V., De Los Reyes, M.M., Mandia, E.H., Brkljača, R. and Urban, S. (2015) Triterpenes from *Calophyllum inophyllum* Linn. *Int. J. Pharmacol. Phytochem. Res.* 7: 718-722.
Reed, J. and Osbourn, A. (2018) Engineering terpenoid production through transient expression in *Nicotiana benthamiana*. *Plant Cell Rep.* 37: 1431-1441.
Reed, J., Stephenson, M.J., Miettinen, K., Brouwer, B., Leveau, A., Brett, P., et al. (2017) A translational synthetic biology platform for rapid access to gram-scale quantities of novel drug-like molecules. *Metab. Eng.* 42: 185-193.
Sainsbury, F., Thuenemann, E.C. and Lomonossoff, G.P. (2009) pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol. J.* 7: 682-693.
Seki, H., Sawai, S., Ohyama, K., Mizutani, M., Ohnishi, T., Sudo, H., et al. (2011) Triterpene functional genomics in licorice for identification of CYP72A154 involved in the biosynthesis of glycyrrhizin. *Plant Cell* 23: 4112-4123.
Seki, H., Tamura, K. and Muranaka, T. (2015) P450s and UGTs: key players in the structural diversity of triterpenoid saponins. *Plant Cell Physiol.* 56: 1463-1471.
Shan, W.-G., Zhang, L.-W., Xiang, J.-G. and Zhan, Z.-J. (2013) Natural friedelanes. *Chem. Biodivers.* 10: 1392-1434.
Souza-Moreira, T.M., Alves, T.B., Pinheiro, K.A., Felippe, L.G., De Lima, G.M.A., Watanabe, T.F., et al. (2016) Friedelin synthase from *Maytenus ilicifolia*: leucine 482 plays an essential role in the production of the most rearranged pentacyclic triterpene. *Sci. Rep.* 6: 36858.
Thimmappa, R., Geisler, K., Louveau, T., O'Maille, P. and Osbourn, A. (2014) Triterpene biosynthesis in plants. *Annu. Rev. Plant Biol.* 65.
Viswanathan, N.I. (1979) Salaspermic acid, a new triterpene acid from *Salacia macrosperma* Wight. *J. Chem. Soc., Perkin Trans. 1* 1979: 349-352.
Voinnet, O., Rivas, S., Mestre, P. and Baulcombe, D. (2003) An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant J.* 33: 949-956.
Zhou, J., Hu, T., Gao, L., Su, P., Zhang, Y., Zhao, Y., et al. (2019) Friedelane-type triterpene cyclase in celastrol biosynthesis from *Tripterygium wilfordii* and its application for triterpenes biosynthesis in yeast. *New Phytol.* 223:722-735.

Tables

Table 1: Yeast strains used in this study.

Strain	Genotype
CEN.PK113-5D	*MATa*, *MAL2-8c*, *SUC2*, *ura3-52*
KB1	CEN.PK113-5D; P_{ERG7}::P_{KEX2}-*ERG7*
KB4	KB1; *trp1Δ0*
KB7	KB4; *his3Δ0*
KB10	KB7; *leu2Δ0*
KB13	KB10; *pah1Δ0*

Legends to figures

Figure 1 Heterologous production of friedelin in *N. benthamiana* leaves by expression of *MiFRS4*. (A) Cyclization of 2,3-oxidosqualene by *Mi*FRS. (B) Overlay of the GC-MS chromatograms from a friedelin standard (red) and extracts of *N. benthamiana* leaves infiltrated with *tHMGR1* alone (Control, black) or co-infiltrated with *tHMGR1*

and *MiFRS4* (blue). (C) Comparison of the EI-MS spectra of the friedelin produced in *N. benthamiana* leaves (*Mi*FRS4, top) and the friedelin standard (bottom).

Figure 2 Expression of *CYP712K4* in *N. benthamiana* leaves. (A) Overlay of the LC-APCI-FT-MS chromatograms from a friedelin standard (red) and extracts of *N. benthamiana* leaves infiltrated with *tHMGR1* and *MiFRS4* with (CYP712K4, blue) or without (Control, black) *CYP712K4*. The two peaks unique to the leaves infiltrated with *CYP712K4* are indicated. (B) Comparison of the accurate mass observed for the friedelin standard (left) and the friedelin produced in *N. benthamiana* leaves (right). (C) Comparison of the MS^2 fragmentation spectra of the friedelin produced in *N. benthamiana* leaves (top) and the friedelin standard (bottom). (D) Accurate masses observed for the first (left) and the second (right) peak unique to the leaves infiltrated with *CYP712K4*. (E) MS^2 fragmentation of the $[M + H]^+$ ion at *m*/*z* 457.37 of the first peak unique to leaves infiltrated with *CYP712K4*. (F) MS^3 fragmentation of the daughter ion at *m*/*z* 439 of the first peak unique to leaves infiltrated with *CYP712K4*. (G) MS^2 fragmentation of the $[M + H]^+$ ion at *m*/*z* 443.39 of the second peak unique to leaves infiltrated with *CYP712K4*.

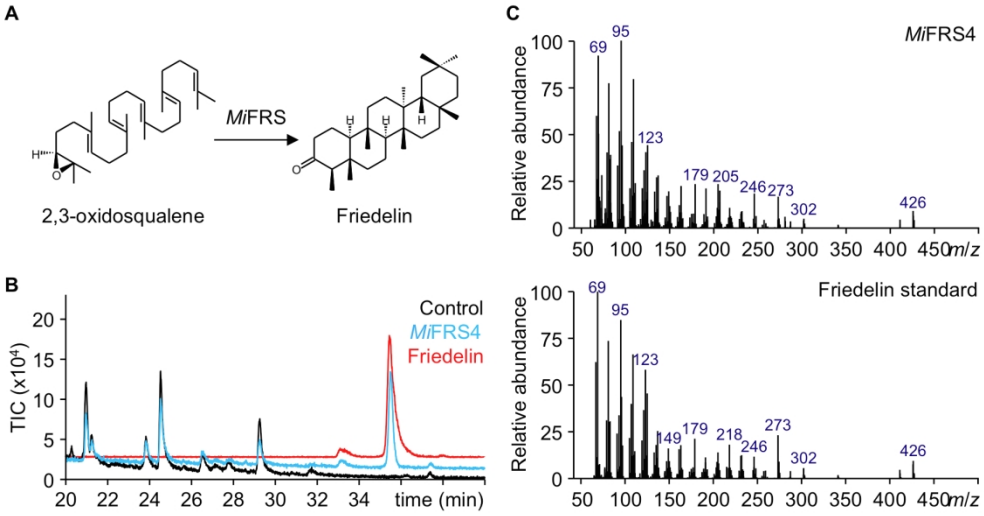
Figure 3 Evaluation of the friedelin production in *S. cerevisiae*. (A) Overlay of the GC-MS chromatograms from an extract of yeast strain KB1 expressing *MiFRS* from the high-copy number plasmid pSP[P_{TEF1}-*MiFRS*;P_{PGK1}-*tHMG1*] (black) and a friedelin standard (red). (B) EI-MS spectrum of the friedelin produced in yeast strain KB1, which compares well with the EI-MS spectrum of the friedelin standard presented in Figure 1C. (C) Friedelin production in yeast strains CEN.PK113-5D (WT) and KB1 in the presence (+ CD) or absence of methyl-β-cyclodextrins. Blue and red colored bars indicate friedelin was extracted from the cells or the medium, respectively. Error bars represent the standard error ($n = 5$ biological replicates). No statistically significant differences ($P > 0.01$; one-way ANOVA with Tukey's test) were observed. (D) Friedelin production in yeast strain KB1 expressing *MiFRS* from the high-copy number plasmid pSP[P_{TEF1}-*MiFRS*;P_{PGK1}-*tHMG1*] (pSP) or pESC-URA-tHMG1-DEST (pESC) in medium containing glucose (Glc) or galactose (Gal) as sugar source. Error bars represent the standard error ($n = 5$ biological replicates). No statistically significant differences ($P > 0.01$; one-way ANOVA with Tukey's test) were observed. * Friedelin was produced, but below the quantification limit.

Figure 4 Engineering of *S. cerevisiae* using CRISPR/Cas9. (A) Drop-tests confirming the auxotrophy of the generated strains. A serial dilution of yeast strains KB1, KB4, KB7, and KB10 was dropped on plates containing synthetic defined (SD) medium lacking uracil (-Ura), tryptophan (-Trp), histidine (-His), or leucine (-Leu). Yeast pre-cultures were diluted 10-fold (10^{-1}), 100-fold (10^{-2}), or 1000-fold (10^{-3}) in sterile water prior to dropping on the plates. (B) Sequence analysis of yeast strain KB13 for the mutations introduced by CRISPR/Cas9. For each introduced auxotrophic marker and the *pah1* mutation, the wild-type amino acid and nucleotide sequences are compared to the mutated sequences in strain KB13. (C) Production of friedelin in yeast strain CEN.PK113-5D (WT) and the engineered strains KB1, KB10 and KB13. Error bars represent the standard error ($n = 5$ biological replicates). (D) Production of β-amyrin in yeast strain CEN.PK113-5D (WT) and the engineered strains KB1, KB10 and KB13. Error bars represent the standard error ($n = 5$ biological replicates). Different letters indicate statistically significant differences with $P < 0.01$ (one-way ANOVA with Tukey's test).

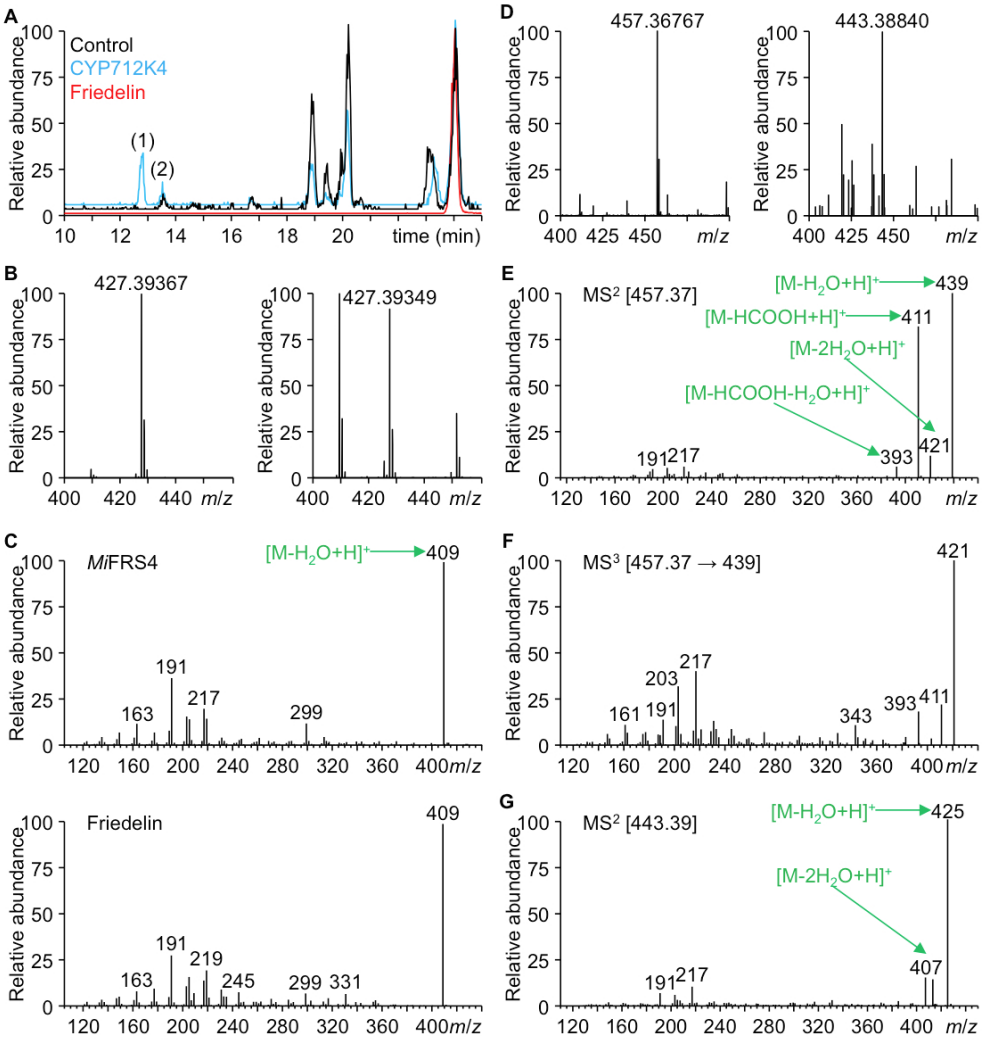
Figure 5 Evaluation of CYP712K4 in *S. cerevisiae*. (A) Overlay of LC-APCI-FT-MS chromatograms from extracts of the spent medium of yeast strain KB13 expressing *tHMGR1*, *MTR1*, and codon-optimized *MiFRS* with

(CYP712K4, blue) or without (Control, black) *CYP712K4*. Three peaks unique to the extracts of the strain expressing *CYP712K4* are indicated. (B) Accurate mass observed for the acid (left) and alcohol (right) produced in yeast. (C) MS^2 fragmentation of the $[M + H]^+$ ion at *m*/*z* 457.37 of the first peak unique to yeast cells expressing *CYP712K4*. (D) MS^2 fragmentation of the $[M + H]^+$ ion at *m*/*z* 443.39 of the second peak unique to yeast cells expressing *CYP712K4*. (E) Accurate mass observed for the third peak unique to yeast cells expressing *CYP712K4*. (F) MS^2 fragmentation of the $[M + H]^+$ ion at *m*/*z* 441.37 of the third peak unique to yeast cells expressing *CYP712K4*.

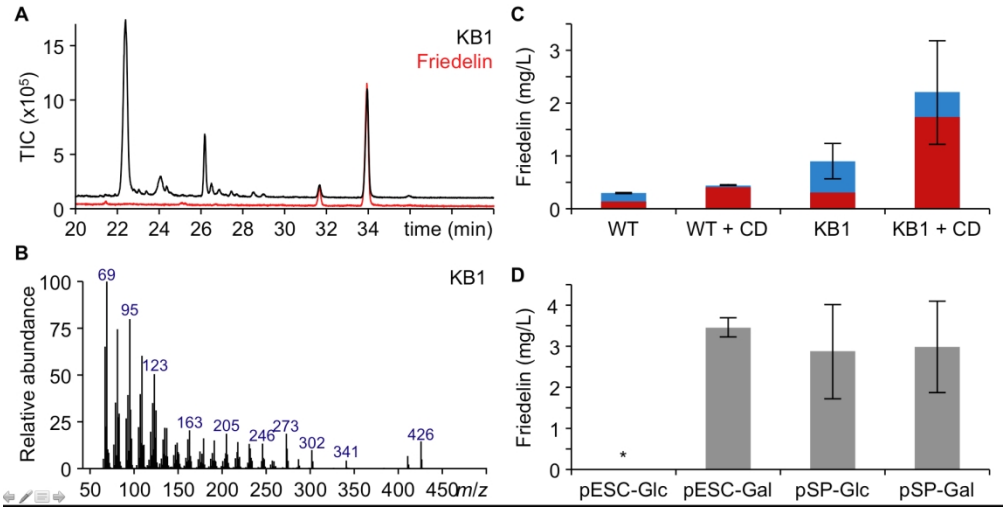
Figure 6 Biosynthesis of quinone methide triterpenoids. (A) Three-step oxidation of friedelin catalyzed by CYP712K4. (B) Proposed biosynthetic pathway to cangoronine by the combined activity of two P450s. Apart from the biosynthetic intermediate in grey, all metabolites have been shown to be present in plants producing quinone methide triterpenoids (Shan et al. 2013). The dashed arrow indicates multiple enzymatic steps.



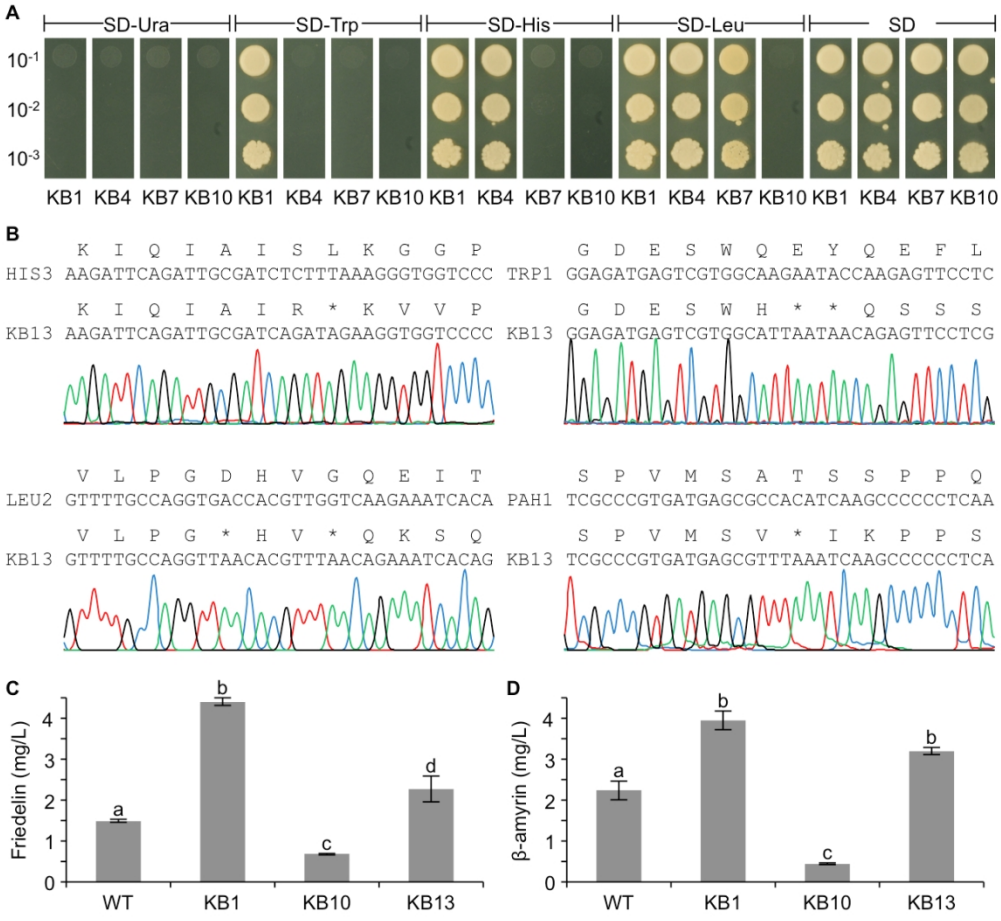
673x359mm (72 x 72 DPI)



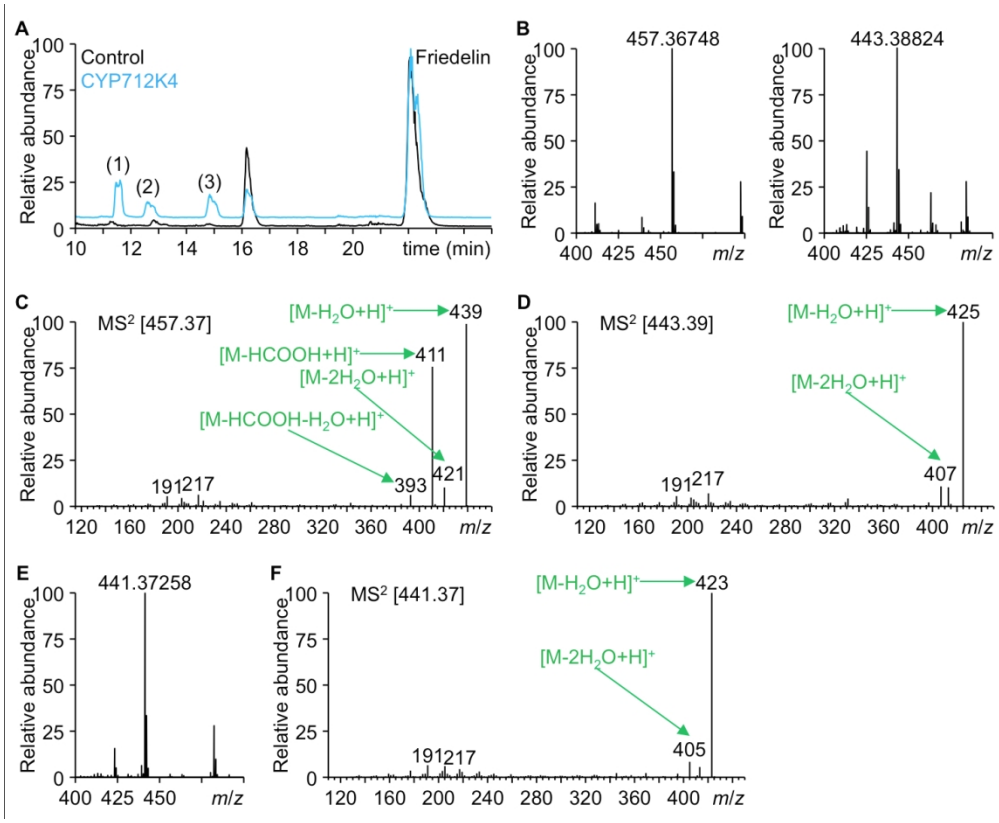
395x422mm (72 x 72 DPI)



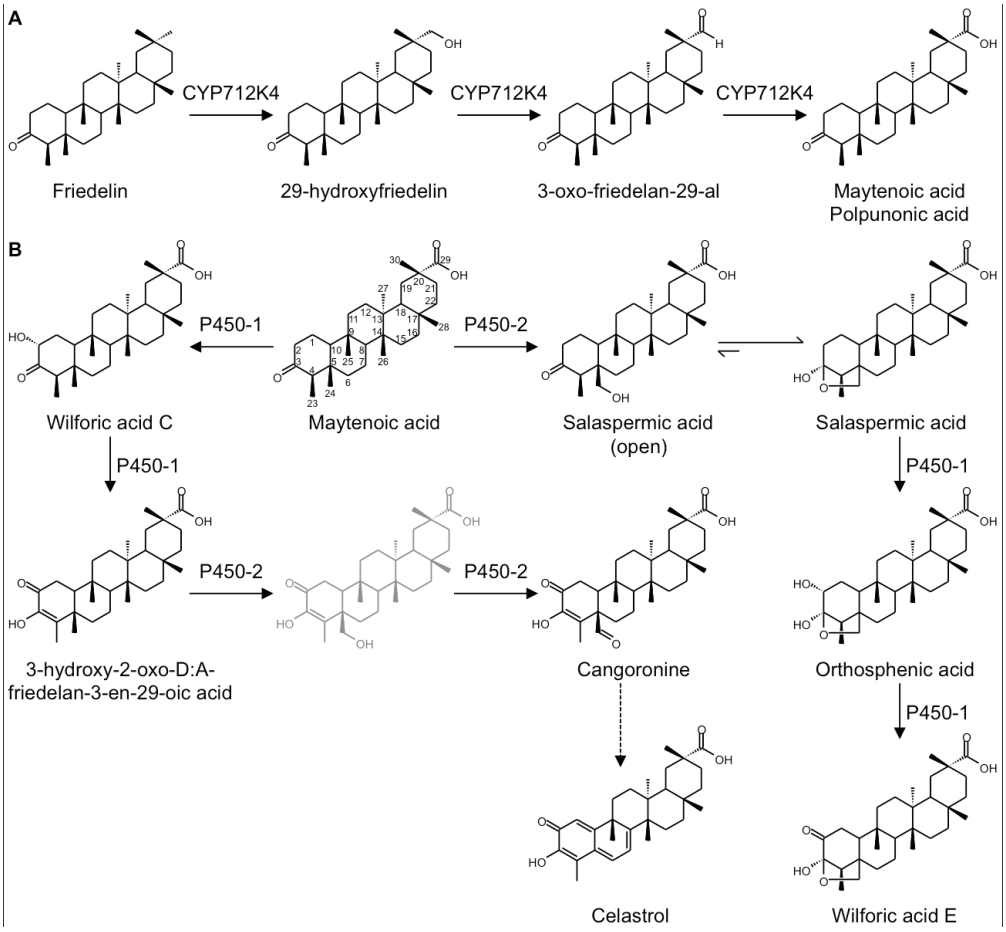
674x342mm (72 x 72 DPI)



456x421mm (72 x 72 DPI)



518x422mm (72 x 72 DPI)



457x421mm (72 x 72 DPI)