UNIVERSIDADE ESTADUAL PAULISTA "Júlio de Mesquita Filho"
INSTITUTO DE QUÍMICA DE ARARAQUARA

Programa de Pós-Graduação em Biotecnologia

MARIANA MARCHI SANTONI BIASIOLI

Perfil transcricional comparativo de folha e raiz de *Monteverdia ilicifolia* para a biossíntese de terpenos

Araraquara

Dez/2022

MARIANA MARCHI SANTONI BIASIOLI

Perfil transcricional comparativo de folha e raiz de *Monteverdia ilicifolia* para a biossíntese de terpenos

Tese apresentada ao Instituto de Química da Universidade Estadual Paulista, como parte dos requisitos para obtenção do título de Doutora em Biotecnologia.

Orientadora: Prof°. Dr° Cleslei Fernando Zanelli

Co-orientadora: Profª. Drª. Maysa Furlan

B579p

Biasioli, Mariana Marchi Santoni
Perfil transcricional comparativo de folha e raiz de Monteverdia ilicifolia para a biossíntese de terpenos / Mariana Marchi Santoni Biasioli. -- Araraquara, 2022
72 f. : il., tabs.

Tese (doutorado) - Universidade Estadual Paulista (Unesp), Instituto de Química, Araraquara
Orientador: Cleslei Fernando Zanelli
Coorientadora: Maysa Furlan

1. Sequenciamento de nucleotídeo. 2. Metabólitos. 3. Maytenus. 4. Biossíntese. 5. Transcrição gênica. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Química, Araraquara. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: "Biossíntese de metabólitos secundários de *Monteverdea ilicifolia* por meio de perfil transcricional órgão-comparativo"

AUTORA: MARIANA MARCHI SANTONI BIASIOLI
ORIENTADOR: CLESLEI FERNANDO ZANELLI
COORIENTADORA: MAYSA FURLAN

Aprovada como parte das exigências para obtenção do Título de Doutora em Biotecnologia, pela Comissão Examinadora:

gov.br Documento assinado digitalmente
CINTIA DUARTE DE FREITAS MILAGRE
Data: 05/01/2023 15:07:37-0300
Verifique em https://verificador.iti.br

Profa. Dra. CINTIA DUARTE DE FREITAS MILAGRE (Participaçao Virtual)
Departamento de Bioquimica e Quimica Organica / Instituto de Quimica - UNESP - Araraquara

Prof. Dr DANILO TRABUCO DO AMARAL (Participaçao Virtual)
Universidade Federal do ABC - UFABC - Santo André

Prof. Dr. CAIO CESAR DE MELO FREIRE (Participaçao Virtual)
Departamento de Genética e Evolução / Universidade Federal de São Carlos - UFSCar - São Carlos

Drª. TATIANA MARIA DE SOUZA MOREIRA (Participaçao Virtual)
Departamento de Ciências Biológicas / Faculdade de Ciências Farmacêuticas - UNESP - Araraquara

Dra. BIANCA BACCILI ZANOTTO VIGNA (Participaçao Virtual)
Embrapa Pecuária Sudeste - EMBRAPA - São Carlos

Araraquara, 20 de dezembro de 2022

Dados Curriculares

Endereço Profissional
Laboratório de Biologia Molecular de Microrganismos, Departamento de Ciências Biológicas da FCF-UNESP.
Rodovia Araraquara-Jaú, Km 01, Campus Ville, S/N, Araraquara SP. CEP: 14800-903
e-mail: mariana.santoni@unesp.br

Formação Acadêmica:
2002 – 2006: Graduação
Licenciatura e Bacharel em Ciências Biológicas.
Universidade Federal de São Carlos – SP.

2006 – 2008: Mestrado
Genética e evolução.
Universidade Federal de São Carlos – SP
Dissertação: Biologia de nidificação e estrutura sociogenética intranidal em espécies de *Trypoxylon* (Hymenoptera:Crabronidae)

2019 – 2022: Doutorado
Biotecnologia. Área de concentração: Bioinformática.
Instituto de Química (IQ) - Universidade Estadual Paulista "Júlio de Mesquita
Tese: Biossíntese de metabólitos secundários de Monteverdia ilicifolia por meio de perfil transcricional órgão-comparativo

Produção Bibliográfica (trabalhos relacionados à presente tese):
Bicalho, Keylla U ; Santoni, Mariana M ; Arendt, Philipp ; Zanelli, Cleslei F; Furlan, Maysa ; Goossens, Alain ; Pollier, Jacob . CYP712K4 Catalyzes the C-29 Oxidation of Friedelin in the *Maytenus ilicifolia* Quinone Methide Triterpenoid Biosynthesis Pathway. **Plant and Cell Physiology**, v. 60, p. 2510-2522, 2019.

Santoni, Mariana Marchi; Lima, João Vitor Félix de; Bicalho, Keylla Utherdyany; Moreira, Tatiana Maria de Souza Moreira; Valentini, Sandro Roberto; Zanelli, Cleslei Fernando. Comparative Transcriptome Profiling of *Maytenus ilicifolia* Root and Leaf. **Lecture Notes in Computer Science, 2021**, v. 13063, p. 3-14

Participação em eventos (durante o período de doutoramento):
I Escola Latino-Americana de bioinformática para as ciências "ômicas", LNCC, 2019.
Brazilian Symposium on Bioinformatics, 2021.

Dedicatória

Dedico esse trabalho aos que me mostram as várias formas de viver, que me levam para outros lugares (físicos ou imagináveis) e me tiram do meu lugar comum, José Eduardo e Vitor.

Sabemos agora que muito do rico repertório comportamental de uma planta é difícil de observar porque é reproduzido numa arena química. As plantas superam as limitações da imobilidade aproveitando as suas proezas em sintetizar compostos orgânicos (Baldwin, 2015)

Agradecimentos

Ao meu orientador, Cleslei, e à minha co-orientadora, Maysa, pela oportunidade de realização desse trabalho e por dedicarem parte de seu tempo para a minha formação.

Aos meus familiares, pelo imenso apoio e ajuda.

Aos colegas e responsáveis do laboratório de Biologia Molecular e Celular de Microrganismos.

Aos funcionários e professores da faculdade de ciências farmacêuticas e do instituto de química da Unesp de Araraquara, que direta e indiretamente possibilitam a estrutura da pós-graduação.

Aos colaboradores dos trabalhos publicados.

Aos demais amigos e colegas que encontrei nessa trajetória.

Às agências de fomento, FAPESP, CAPES, INCT e PADC.

Resumo

As plantas produzem uma grande diversidade de compostos denominados metabólitos secundários (MS), importantes para a sua sobrevivência. Evolutivamente, originam-se como resposta ecológica a competição, defesa ou sinalização e sua biossíntese é um processo altamente regulado por enzimas específicas. Devido ao seu amplo espectro de bioatividades, os MS apresentam aplicações medicinais, mas, como a síntese química não é economicamente viável, a extração das plantas é única opção. Diferentes estratégias biotecnológicas são aplicadas para melhorar o rendimento da bioprodução desses compostos, no entanto, este processo limitado é pelo pouco conhecimento sobre vias biossintéticas e regulatórias. A espécie *Maytenus ilicifolia*, atualmente, *Monteverdia ilicifolia*, uma planta medicinal tradicional, é brasileira e pertence à família Celastraceae. Conhecida como "espinheira santa", apresenta três classes principais de MS: sesquiterpênicos, flavonóides e quinonametídeos, que são produzindo tanto em folhas como raízes. O objetivo deste trabalho foi identificar os genes responsáveis pela biossíntese dos MS de *M. ilicifolia* e fatores de regulação destas rotas, pelo sequenciamento *de novo* do transcriptoma desta espécie. Quatro bibliotecas de cDNA foram preparadas a partir de folhas e raízes. O transcriptoma *de novo* incluiu 109.982 sequências que capturou 92% dos ortólogos da base “BUSCO”. Os transcritos apresentaram um comprimento médio de 737pb e um conteúdo GC de aproximadamente 42%. As análises de anotação funcional identificaram homologia para 44,8% dos transcritos. Em termos de expressão comparativa, 67.625 sequencias foram expressas em ambos os órgãos, mas 1.044 e 1.717 sequências foram diferencialmente expressas na raiz e na folha, respectivamente. Quanto aos MS, genes codificadores para enzimas envolvidas na biossíntese de monoterpenos, isoflavonoides foram identificadas nas raízes e para as folhas, biossíntese de flavanoides e biossíntese de alcaloides. Em termos de regulação das vias, 191 transcritos foram anotados como fatores de transcrição, sendo 12 diferencialmente nas folhas e três nas raízes. Foram identificadas enzimas envolvidas na biossíntese de terpenos, incluindo as vias do mevalonato (8) e do metileritritol-fosfato (9). Adicionalmente, 126 transcritos foram atribuídos como codificadores para enzimas CYP450. Por fim, a escolha da raiz e da folha para a análise comparativa do transcriptoma facilitou a identificação dos genes envolvidos na biossíntese de MS, uma abordagem amplamente utilizada em plantas.

Palavras-chave: RNA-Seq, metabólitos secundários, *Maytenus* sp, transcritos

Abstract

Plants produce a wide variety of compounds called secondary metabolites (SMs), which are extremely important for their survival. SMs have also medicinal applications, but as chemical synthesis is not economically viable, plant extraction is the mainly option. Different biotechnology strategies are applied to improve the yield of bioproduction of these compounds, but commonly without the desired results due the limited knowledge of biosynthetic and regulatory pathways. *Maytenus ilicifolia*, a traditional Brazilian medicinal plant from Celastraceae family, produces in both root and leaves three main classes of SMs: sesquiterpenics, flavonoids and quinonemethides. In this study, four cDNA libraries were prepared from root and leaf tissues. The de novo transcriptome included 109,982 sequences that capture 92% of BUSCO orthologs, presented an average length of 737bp and a GC content about 42% of. Function annotation analysis identified homology for 44.8% of the transcripts. Moreover, 67,625 sequences were commonly expressed in both tissues, while 1,044 and 1,171 were differentially expressed in root and leaf, respectively. In terms of SM, enzymes involved in “monoterpenoid biosynthesis” and “isoflavonoid biosynthesis" were identified in root while “flavonoid biosynthesis” and “Biosynthesis of alkaloids” in leaf. In terms of pathway regulation, 191 transcripts were annotated as transcription factors, with 12 differentially in leaves and three in roots. Enzymes involved in terpene biosynthesis were identified, including the mevalonate (8) and methylerythritol-phosphate (9) pathways. Additionally, 126 transcripts were assigned as coding for CYP450 enzymes. Finally, the choice of root and leaf for comparative transcriptome analysis facilitated the identification of genes involved in MS biosynthesis, an approach widely used in plants.

Key-words: RNA-Seq, secundary metabolites, *Maytenus* sp, transcripts

Sumário

1. Introdução

Uma característica de plantas e outros organismos sésseis é a capacidade de biossintetizar uma variedade de compostos de baixo peso molecular, chamados metabólitos secundários (MS) (WINK, 2010a) . Estes compostos são componentes chave para a interação destes organismos com o ambiente e para a adaptação às condições de estresse biótico e abiótico (YANG et al., 2018). Sua biossíntese é um processo altamente regulado, catalisado por enzimas específicas (WINK, 2010a) que foram selecionadas para esta finalidade, sendo derivadas de ancestral comum a outras enzimas do metabolismo primário ou de genes importados pelos cloroplastos e mitocôndrias (WINK, 2010a).

Evolutivamente, o metabolismo secundário das plantas pode ser visto como um comportamento para sua adaptação e sobrevivência em resposta aos estímulos ambientais (METLEN; ASCHEHOUG; CALLAWAY, 2009). Origina-se como resposta ecológica da planta frente à competição por recursos como luz, água e nutrientes, defesa contra herbívoros ou agentes infectantes ou até mesmo atuam como compostos antioxidantes, protetores de UV ou de armazenamento de nitrogênio (DZIGGEL; SCHÃFER; WINK, 2017). Algumas plantas fizeram uso de MS para estabelecer relações entre outros organismos (MUSILOVA et al., 2016), como sinais de comunicação entre plantas e microrganismos simbióticos ou na atração polinizadores e dispersores de sementes (WINK, 2010a).

Com esta alta variedade de funções biológicas, os compostos secundários apresentam grande diversidade estrutural (DZIGGEL; SCHÃFER; WINK, 2017; YANG et al., 2018), apesar de serem provenientes de vias metabólicas básicas, como a glicólise e o ciclo ácido cítrico, também denominados de blocos construtores (GIWELI et al., 2013). Embora apenas 20-30% das plantas tenham sido investigadas, aponta-se que dezenas de milhares de MS foram isolados, sendo os mais abundantes os **alcaloides** - grupo estruturalmente amplo, que contém nitrogênio e são derivados de aminoácidos -, seguidos pelos **terpenos** - derivados de unidades C5 – e pelos **fenilpropanoides** - sintetizados a partir de aminoácidos aromáticos e unidades de acetil-CoA (DZIGGEL; SCHÃFER; WINK, 2017; YANG et al., 2018).

Em relação aos terpenos, as plantas empregam esse metabólito para uma variedade de funções no crescimento e desenvolvimento. Contudo, sua participação é majoritariamente desempenhar interações químicas e proteção em ambientes bióticos e abióticos (SILVA et al., 2020). Já o triterpenos, metabolitos secundários pertencentes a

classe dos terpenos e caracterizados quimicamente pela presença de seis unidades de isopreno, com um total de 30 átomos de carbono, têm mostrado um grande espectro de atividades biológicas, ais como: anti-inflamatória, antinociceptiva, hepatoprotetor, efeito sedativo, antioxidante, antialérgico, antiangiogênica, antimicrobiana e alta seletividade anticancerígena (SILVA et al., 2020).

Os MS são identificados em todos os órgãos e sua formação e regulação gênica é geralmente órgão, tecido, célula e, também, desenvolvimento específicos. Isso significa que uma bateria de fatores de transcrição precisa cooperar para ativar e transcrever genes do metabolismo secundário, controlando a maquinaria geral das vias biossintéticas na produção, transporte e armazenamento (UPADHYAY et al., 2014; WINK, 2010b). Portanto, a biossíntese de MS exibe uma complexidade notável: as enzimas são específicas para cada via e são altamente reguladas em termos de compartimentação, tempo e espaço (WINK, 2010b).

Em relação às enzimas envolvidas no metabolismo secundário, destaca-se a superfamília das enzimas citocromo P450 (CYP). Presentes em todos os táxons e apresentando uma particular diversidade em *Viridiplantae*, grupo que abrange as algas verdes e as plantas terrestres, as P450 constituem uma superfamília de enzimas que catalisam um expressivo arsenal de reações, que estão envolvidas na formação de esteróis de membrana, fito-hormônios, moléculas de sinalização, biopolímeros estruturais e de proteção e diversos compostos orgânicos voláteis e metabólitos especializados envolvidos em interações bióticas e abióticas (HANSEN et al., 2021).

Devido ao amplo espectro de bioatividades, os MS derivados de plantas são diversamente aplicados como compostos farmacêuticos, nutracêuticos ou aromatizantes (DZIGGEL; SCHÃFER; WINK, 2017), porém, em muitos casos, a síntese química não é economicamente viável, sendo o isolamento de plantas ainda a única opção (PAZ et al., 2017). Diferentes estratégias biotecnológicas têm sido aplicadas para melhorar o rendimento da produção desses compostos na planta e nas culturas de células vegetais, mas muitas vezes sem os resultados desejados (WINK, 2010b), pois o conhecimento sobre suas vias biossintéticas e de regulação ainda é limitado (PAZ et al., 2017).

A utilização de técnicas biotecnológicas como transcriptomas, proteomas, metabolomas e silenciamento de genes em plantas aumenta a busca por genes e sua função nas vias metabólicas das plantas, facilitando gradativamente a produção de MS. Recentemente, caminhos completos foram elucidados e, em alguns casos, foi possível sintetizar compostos secundários em hospedeiros microbianos recombinantes (bactérias

ou leveduras biologicamente modificados), a partir de fontes baratas de carbono (DZIGGEL; SCHÃFER; WINK, 2017).

O gênero *Monteverdia*, com aproximadamente 300 espécies, é amplamente distribuído nos trópicos e subtrópicos. Aproximadamente 50 espécies crescem em diferentes regiões do Brasil, incluindo Amazônia, Mata Atlântica, Caatinga e Cerrado (GROPPO et al., 2014). A espécie *M. ilicifolia* (Mart. ex Reissek) Biral é uma planta brasileira da família Celastraceae e seu uso como planta medicinal tradicional está descrito desde 1922, focada no tratamento da úlcera gástrica (CARLINI, 1988) mas suas propriedades terapêuticas são atualmente amplas, sendo também empregada em tratamentos de diabetes, infecções do trato urinário, problemas intestinais, doenças nervosas, doenças do rim e do sangue (MARIOT; BARBIERI, 2007a) e alguns tipos de tumores (PAZ et al., 2017) sendo, por esta razão, popularmente conhecida como "espinheira-santa" (PÉRICO et al., 2018).

M. ilicifolia contém três classes principais de compostos bioativos: alcaloides piridínicos sesquiterpênicos, flavonoides e triterpenos quinonametídeos. Suas folhas contêm variados MS como flavonoides, terpenos, triterpenos, glicosídeos e alcaloides, enquanto que sua raiz contém terpenos, triterpenos, alcaloides e especialmente os triterpenos quinonametídeos (PÉRICO et al., 2018). Os principais produtos são maitenina, friedelina, fridelanol, pristimerina, terpenos que exibem uma gama de atividade biológica, caracterizando a espécie como bioprodutora de MS (PÉRICO et al., 2018).

Com relação à localização da produção de metabólitos secundários em espinheira- santa, estudos identificaram que as folhas produzem 3b-friedelanol e friedelina (friedelano) e raízes acumulam maitenina e pristimerina (quinonametídicos) (FILHO et al., 2002). Os triterpenos derivados de friedelano, uma vez biossintetizados nas folhas, são translocados para as raízes e posteriormente transformados nos triterpenóides quinonametídicos, que apresentam ação antitumor. Estes triterpenóides não foram encontrados em folhas, somente em raízes. Os flavonóides são encontrados em todos os órgãos das plantas (MARIOT; BARBIERI, 2007b).

Atualmente, um estudo em proteômica desta espécie detectou muitas enzimas envolvidas no metabolismo secundário, incluindo as monooxigenases dependentes do citocromo P450, que potencialmente catalisam as etapas finais na biossíntese de triterpenos quinonemetídeos (PAZ et al., 2017), mas os genes que codificam as principais enzimas dessas vias, juntamente com sua regulação, devem ser entendidos com precisão.

2. Objetivos

Considerando que a biossíntese, o acúmulo e a regulação da expressão de metabólitos secundários (MS) apresentam padrão de expressão órgão-específicos e, baseando-se na hipótese de que genes diferencialmente expressos entre dois órgãos distintos pode fornecer informações sobre os transcritos e seus reguladores envolvidos em vias metabólicas, o objetivo do trabalho foi identificar os genes e fatores de transcrição responsáveis pela biossíntese dos MS em especial na via de terpenos de *M. ilicifolia* pela análise do sequenciamento do transcriptoma de órgãos foliares e radiculares desta espécie.

3. Material e Métodos

3.1 Aquisição de exemplares e extração de RNA

Para este estudo, foram utilizadas folhas de exemplar adulto de *M. ilicifolia* proveniente do horto de plantas medicinais da Faculdade de Ciências Farmacêuticas, e folhas e raízes de mudas adquiridas e identificadas, com aproximadamente 6 meses de plantio. O RNA total de 2 exemplares de raízes (2 mudas) e 2 exemplares de folhas (1 folha de muda, coincidente com um dos exemplares utilizado para a extração da raiz e 1 folha de indivíduo adulto) foi isolado de 500mg de material. Os órgãos foram homogeneizados em nitrogênio líquido e o isolamento do RNA foi realizado utilizando-se o kit RNeasy Plant mini kit (Qiagen,USA), seguindo as recomendações do fabricante e quantificados pelo espectrofotômetro NanoDrop 2000 (ThermoFisher Scientific, USA). A qualidade do RNA extraído foi avaliada por eletroforese em 2100 Bioanalyzer (Agilent Technologies, USA), por meio da identificação de bandas 18S e 28S intactas e pelo valor de RIN (RNA Integrity Number).

3.2 Preparo da biblioteca para sequenciamento de RNA

O preparo de biblioteca de transcritos seguiu o protocolo TruSeq RNA v3 kit (Illumina, USA) de acordo com as recomendações do fabricante. Resumidamente, depois da extração total, o RNA total é submetido à purificação dos RNAs mensageiros (mRNAs) por meio de partículas magnéticas contendo Oligo(dT) e, posteriormente, são

quimicamente fragmentados. Estes fragmentos de mRNAs são transcritos para cDNA de dupla fita e recebem em suas extremidades oligonucleotídeos adaptadores à lâmina de sequenciamento. Finalmente, as bibliotecas de cDNA são amplificadas e avaliadas quanto à qualidade, utilizando um chip de DNA de alta sensibilidade em 2100 bioanalyzer, e quanto à quantidade, por PCR quantitativa com o kit Kapa (Roche, USA). 20 pmol das bibliotecas foram submetidos a sequenciamento “single-read” em equipamento HiSeq 2000 (folha do indivíduo adulto) e sequenciamento "paired-end" em equipamento MiSeq (folha e raízes de mudas).

3.3 Análise de dados de sequenciamento

Os dados gerados pelo sequenciamento, arquivos FASTQ, foram avaliados pelo software *FastQC* (0.11.9) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) quanto à qualidade antes e depois da filtragem. Estas sequências, denominadas reads, foram filtradas pelo software *TrimGalore!* (0.6.7) (https://github.com/FelixKrueger/TrimGalore), removendo aquelas que se classificaram como sequências de adaptador e de baixa qualidade (qualidade média abaixo de 25). Também foram removidas bases iniciais e finais das sequências com valores de *q* menor que 25 e, finalmente, no arquivo FASTQ final de reads filtradas permaneceram aquelas com tamanho maior que 50 pares de base.

3.4 Montagem dos transcriptomas

Uma vez que *M. ilicifolia* não possui genoma sequenciado e depositado, a abordagem da montagem das reads seguiu uma abordagem do tipo *de novo* ou *ab initio*, utilizada quando não se há informação sobre o genoma da espécie. A partir de todos os arquivos de dados filtrados, as reads foram agrupadas e alinhadas pelo software *Trinity* (2.9.1) (GRABHERR et al., 2011) a fim de se obter os transcritos. Para esta montagem, foram considerados os parâmetros de alinhamento das reads como taxa de não correspondência de nucleotídeos (mismatch cost) e taxa de inserção e deleção de bases. Ao final deste processo, o arquivo gerado apresentou os possíveis transcritos da espécie, quantificados nos diferentes órgãos e em cada exemplar.

3.5 Avaliação de qualidade da montagem do transcriptoma

Aspectos do transcriptoma como conteúdo CG e N50 foram analisados pelo software *fasta-stats* (1.0.1) (https://github.com/raymondkiu/sequence-stats). A integridade da montagem do transcriptoma foi avaliada usando o software BUSCO (4.1.2) (SIMAO et al., 2015) para 57 espécies com 425 ortólogos do banco de dados *viridiplantae_odb10* e a validação dos transcritos foi avaliada pelo mapeamento do transcriptoma às reads filtradas. A partir da ferramenta *Salmon* (1.5.1) (PATRO et al., 2017), as reads filtradas por qualidade e utilizadas na montagem do transcriptoma foram remapeadas contra o transcriptoma montado a fim de se obter uma matriz de expressão em unidade de transcritos por milhões de kilobase (TPM) que foi utilizada para a análise de componentes principais (PCA). Esta matriz permitiu remoção de transcritos com baixa expressão, considerando apenas aqueles com no mínimo 1% de expressão da isoforma dominante, gerando, por fim, um transcriptoma filtrado.

3.6 Anotação funcional

A ferramenta *TransDecoder* (5.5.0) (HAAS et al., 2013) foi utilizada para encontrar possíveis regiões codificadoras dos transcritos e janelas abertas de leitura (ORFs) com tamanho mínimo de 100 aminoácidos. Posteriormente, a anotação funcional dos transcritos foi realizada utilizando o BLASTX pelo software *Diamond* (2.0.15) (BUCHFINK; XIE; HUSON, 2014) contra os bancos de dados da UniProtKB/SwissProt e Uniprot trEMBL plants (E-value < 1e-5). Além disso, a homologia também foi realizada a partir da busca com BLASTP com o mesmo software, utilizando as proteínas preditas contra a base de dados da UniProtKB/SwissProt (E-value < 1e-5). Termos de ontologia genética (GO) aos transcritos foram realizados com base no banco de dados do UniProtKB/SwissProt para atribuir categorias funcionais aos transcritos. Além disso, as proteínas com números da Comissão de Enzimas (EC) foram mapeadas no Banco de Dados de Via da Enciclopédia de Genes e Genomas de Kyoto (KEGG) usando o Servidor de Anotação Automática KEGG online (www.genome.jp/kegg/kaas) para atribuir informações de via aos transcritos.

3.7 Análise de diferença de expressão

A ferramenta *Salmon* (1.5.1) (PATRO et al., 2017) foi aplicada para estimar o nível de expressão dos transcritos. Cada arquivo FASTQ filtrado foi alinhado

separadamente ao transcriptoma filtrado. Em seguida, o nível de expressão de cada transcrição foi normalizado e reportado em TPM. Para resumir os resultados e fornecer testes estatísticos para comparação de órgãos, a análise de expressão diferencial foi realizada usando o pacote do *R DESeq2* (1.34.0) (LOVE; HUBER; ANDERS, 2014) e a diferença de expressão dos transcritos foi considerada significativa quando o valor de p ajustado apresentava-se menor que 0,05.

3.8 Enriquecimento ontológico e análise de vias

A análise de enriquecimento de ontologia genética (GO) para o processo biológico (BP) e função molecular (MF) para os transcritos diferencialmente expressos em cada órgão foi conduzida usando o pacote do *R topGO* (2.50.0) (https://bioconductor.org/packages/release/bioc/html/topGO.html). Os termos GO significativos (valor de p do teste exato de Fisher <0,01) foram visualizados usando *REViGO* (revigo.irb.hr) para redução do espaço semântico. As transcrições associadas aos números da Enzyme Commission (EC) foram mapeadas no banco de dados da via KEGG.

3.9 Identificação de potenciais Fatores de Transcrição

O método de análise dos fatores de transcrição (TF) seguiu o descrito por (YANG et al., 2018). Resumidamente, todos os genes diferencialmente expressos foram alinhados à base de dados de fatores de transcrição de plantas (Plant TFDB5.0: http://planttfdb.gao-lab.org/) para identificar potenciais TFs (valor E ≤ 1e-6).

3.10 Perfil dos genes relacionados à biossíntese de terpenos

Os dados de anotação funcional do transcriptoma de *M. ilicifolia* foram utilizados para identificar sequências de genes mapeados na via de biossíntese de terpenos. Adicionalmente, foram utilizados os valores de TPM para caracterizar a abundância de cada transcrito nos diferentes órgãos.

3.11 Perfil de genes de P450

Os dados de anotação funcional do transcriptoma de *M. ilicifolia* foram utilizados para identificar sequências de genes candidatos à transcritos de CYP450. Foram

utilizados os valores de TPM, calculados como descrito acima, para caracterizar a abundância de cada transcrito nas diferentes amostras.

4. Resultados e Discussão

4.1 Sequenciamento dos transcriptomas

A biblioteca single-read da folha e as bibliotecas paired-end de folha e raízes submetidas a sequenciamento do transcriptoma completo geraram cerca de 115 milhões de transcritos (Figura 1).

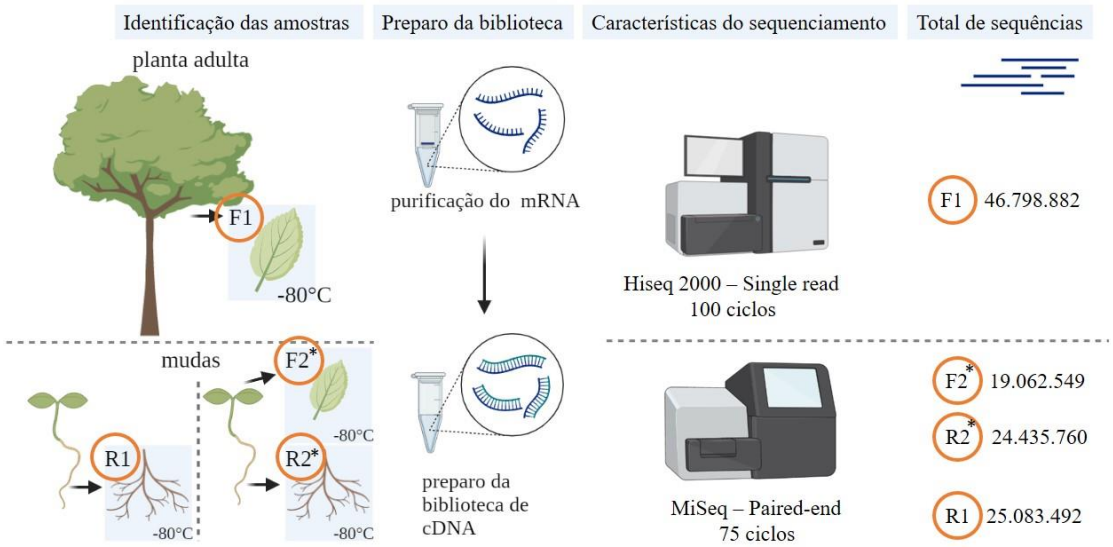

Figura 1. Características do sequenciamento e número total de sequências geradas por amostra de *Monteverdia ilicifolia* submetidas à análise do transcriptoma (*mesmo indivíduo).

Os dados originais de sequenciamento foram filtrados, mantendo apenas sequências com alta qualidade (valores de q iguais ou superiores a 25), resultando em 108.945.849 reads que foram utilizadas para a montagem. Uma vez que *M. ilicifolia* não apresenta genoma de referência, foi utilizada a montagem de transcriptomas *de novo*, originando 163.780 transcritos, com tamanho variando de 200 a 16.289 pares de bases e média de 737 pb (Figura 2).

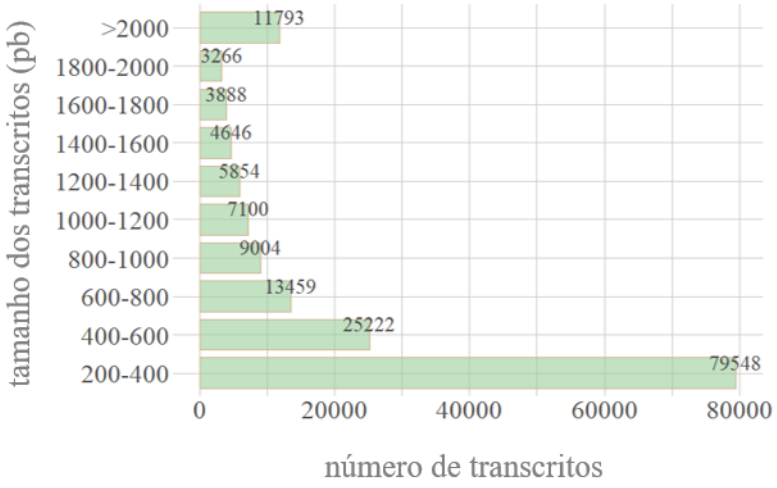

Figura 2. Distribuição do tamanho dos transcritos de *Monteverdia ilicifolia* montados por método *de novo*.

4.2 Avaliação de qualidade da montagem do transcriptoma

Em uma análise comparativa entre dados de folhas e raízes desta espécie, foi possível a identificação de transcritos exclusivos de cada órgão (Figura 3) e 67.625 (41,3%) isoformas foram e encontrados nos dois órgãos.

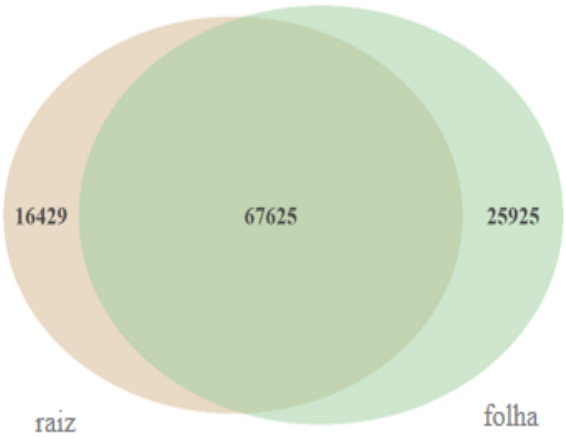

Figura 3. Diagrama de Venn apresentado o número de transcritos para cada situação experimental, folha e raíz, do transcriptoma de *Monteverdia Ilicifolia.*

Um dos aspectos qualitativos da arquitetura genômica é a composição de nucleotídeos, expressa pela proporção das bases guanina e citosina (conteúdo de GC) (ŠMARDA et al., 2014). Esta característica fornece uma visão sobre aspectos relacionados ao genoma de um organismo, incluindo evolução, estrutura do gene, estabilidade térmica e regulação do gene (DEVI et al., 2016). Valores altos de conteúdo GC podem estar associados à capacidade de as plantas crescerem em condições ambientais extremas (como climas frios e secos, e na presença de estresse de dessecação)

(ŠMARDA et al., 2014). Para *M. ilicifolia*, ambos transcriptomas apresentaram conteúdo GC próximos a 40% (Figura 4), valores similares aos reportados para arabidopsis (42.5%), soja (40.9%) e grão-de-bico (40.3%) (DEVI et al., 2016).

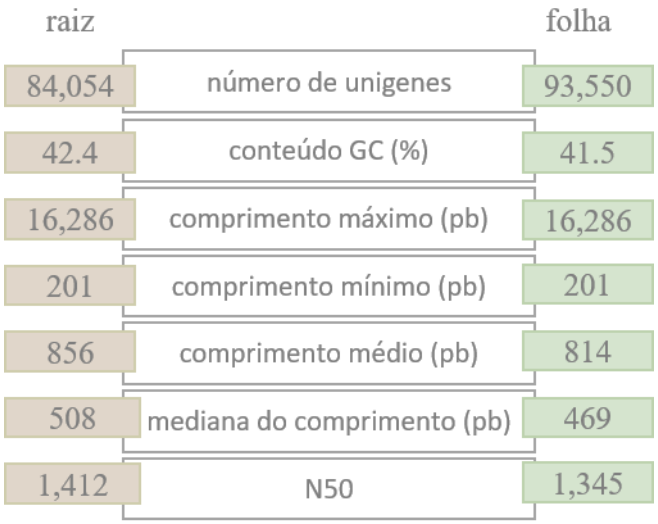

Figura 4. Características dos transcriptomas de folha e raiz de *Monteverdia Ilicifolia*.

Uma métrica estatística frequentemente usada para descrever a integridade de um genoma montado é a N50, a qual fornece informações sobre a distribuição de comprimentos de contigs. Este parâmetro é definido como o comprimento N para o qual 50% de todas as bases nas sequências estão em uma sequência de comprimento L <N. Em outras palavras, pega-se o valor total de bases sequenciadas e observa-se em qual comprimento estão 50% destas bases nos transcritos. Quanto maior o valor de N50, melhor a montagem. Para a montagem dos transcriptomas de *M. ilicifolia*, foram encontrados valores de 1.412 e 1.345 para raiz e folha, respectivamente (Figura 3), indicando uma boa métrica.

Considerando a expressão dos transcritos, 15.704 sequências representaram 90% do total de dados de expressão (Ex90) e apresentavam um N50 de 1487 pb (Ex90M50). Transcritos com baixa expressão foram removidos da montagem.

A comparação de anotações de genes com bancos de dados centrais existentes é uma outra maneira de avaliar a qualidade do transcriptoma. BUSCO (Benchmarking Universal Single-Copy Orthologs) é uma dessas ferramentas que compara os transcritos a conjuntos de genes essenciais para grupos específicos de organismos (eucariotos, bactérias, plantas, etc.). O transcriptoma de *M. ilicifolia* capturou 92,6% dos ortólogos descritos para a base de dados de Virdiplantae (atualizado: 2020-09-10) 53,0%, 39,6%, 4,2%, e 3,2% dos genes BUSCO foram classificados como completo cópia única,

completo duplicado, fragmentado e ausente, respectivamente. Após a filtragem por baixa expressão, o transcriptoma final incluiu 109.982 sequências. Ainda, resultados de análise de componentes principais revelaram diferenças entre padrões de expressão entre as amostras de diferentes órgãos. Os dois principais componentes apresentaram 69,12% da informação, agrupando os órgãos separadamente (Figura 5).

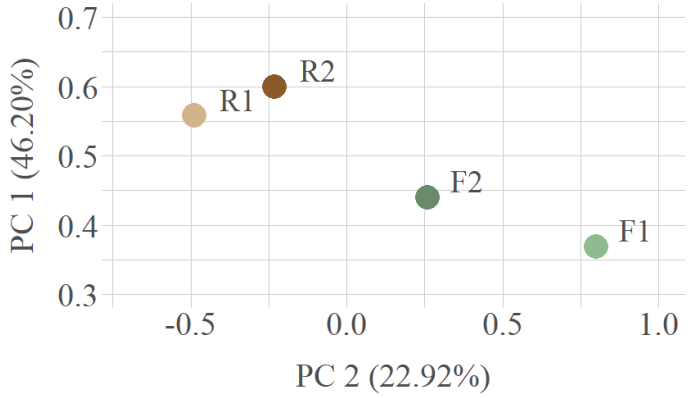

Figura 5. Análise de componentes principais a partir da matriz de número de transcritos de amostras de folhas e raízes de *Monteverdia ilicifolia* (F1, F2: amostras de folha e R1, R2: amostras de raiz).

Em resumo, o transcriptoma de *M. ilicifolia* apresentou conteúdo GC próximo a 40%, valor similar aos reportados em outras espécies da família Celastraceae como Celastrus (41,5%) (LI et al., 2019) e videira trovão de deus (37,2%) (PEI et al., 2021). Além disso, os resultados da análise de BUSCO caputraram mais de 90% dos ortólogos descritos para a base de dados escolhida. Juntos, estes resultados indicam uma alta integridade da montagem e sequenciamento de alta qualidade, que permitem análises futuras.

4.3 Anotação funcional

O alinhamento dos transcritos (BLASTX) contra o banco de dados de proteínas de plantas (uniprot_trEMBL_plants) encontrou 36.625 comparações significativas e revelou que os transcritos previstos de *M. ilicifolia* têm maior similaridade com um organismo classificado na mesma família, *Tripterygium sp* (47,3%) (Figura 6).

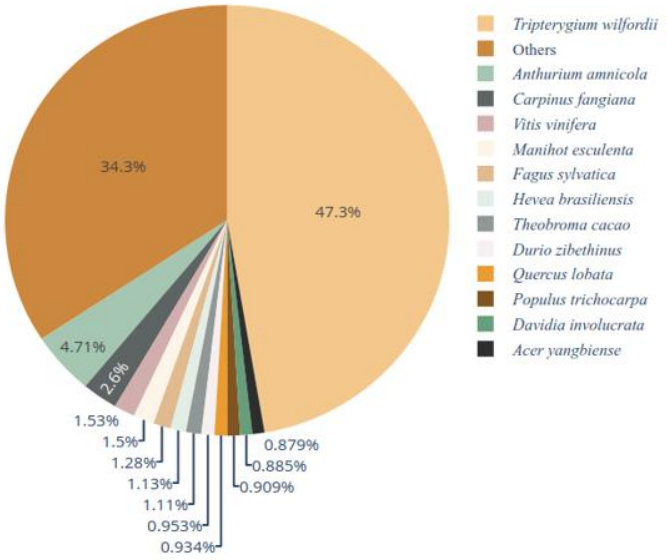

Figura 6. Homologia das sequências transcritos de *Monteverdia ilicifolia* contra a base de dados de proteína para diferentes espécies.

Homologia para outras famílias de organismos também foi observada (Figura 7).

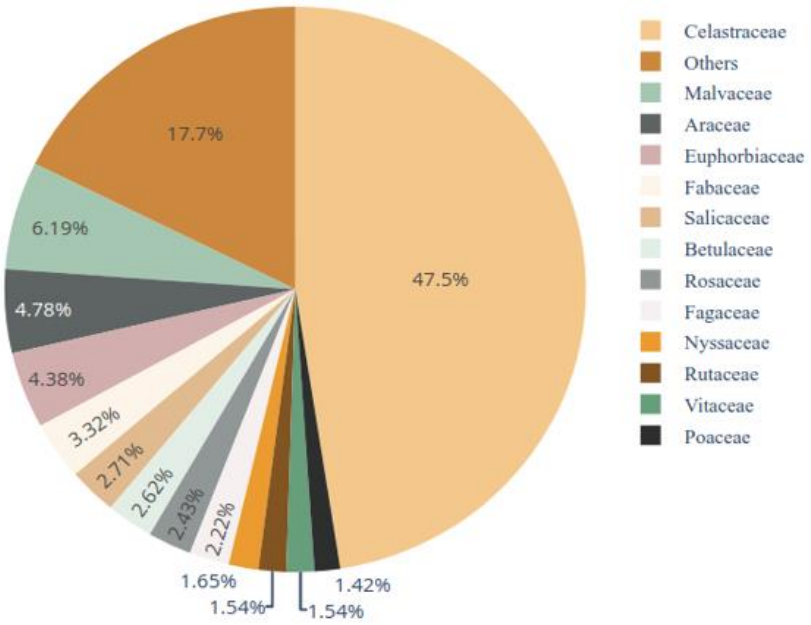

Figura 7. Homologia das sequências transcritos de *Monteverdia ilicifolia* contra a base de dados de proteína para diferentes famílias.

As regiões codificantes candidatas no transcriptoma de *M. ilicifolia* foram identificadas pelo TransDecoder e 65.533 ORFs e 46.282 sequências codificantes prováveis foram previstas. Os resultados da pesquisa de homologia de sequência contra o banco de dados UniprotKB / SwissProt por BLASTX (valor E< 1e − 5, para transcritos filtrados) e BLASTP (valor E< 1e − 5, para sequências de proteínas preditas) foram 49.319 (44,8%) e 36.344 (55,5%) transcritos alinhados, respectivamente.

Sequências completas e preditas como proteínas (36.344) foram mapeadas contra o banco de dados de enzimas KEGG e permitiu a identificação de 10.677 (29,4%) prováveis enzimas, sendo 2.506 (72%) delas envolvidas em vias de metabolismo secundário (Figura 8).

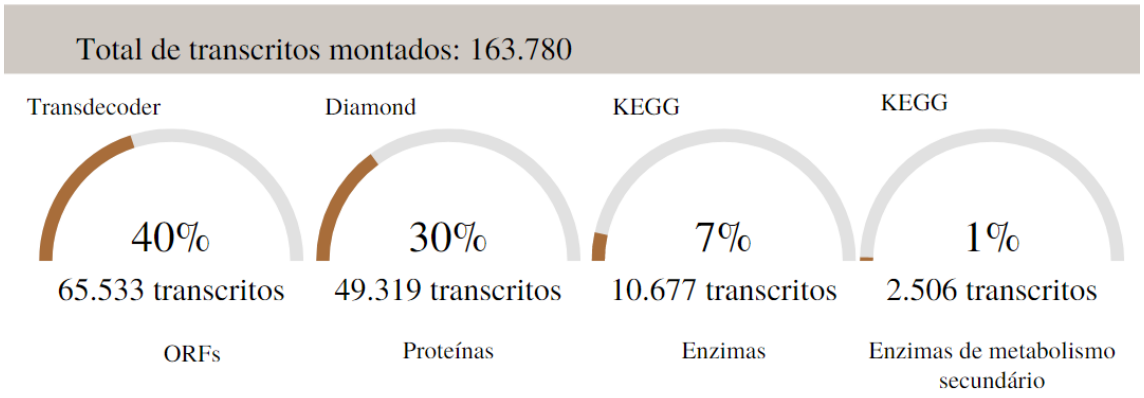

Figura 8. Número de transcritos identificados como ORFs ("Open reading frame": fase de leitura aberta) pelo software *Transdecoder*, proteínas, pelo software Diamond, enzimas, pela base de dados KEGG do transcriptoma de *Monteverdia ilicifolia*.

Enzimas na biossíntese de alcaloides, flavonoides e terpenos, dentre outras vias foram identificadas e mapeadas (Tabela S1).

A anotação funcional para o transcriptoma filtrado foi seguida pela análise de GO e 43.322 sequências anotadas (93,6% das sequências preditas) foram categorizadas em 9.989 IDs de GO. O número de transcritos em três categorias principais de função molecular (MF), processo biológico (BP) e componente celular (CC) foi 41.148, 39.404 e 39.582, respectivamente. Os termos GO mais dominantes na categoria MF foram "ligação de proteína", "ligação de ATP" e "ligação de ferro metálico" (Figura 9).

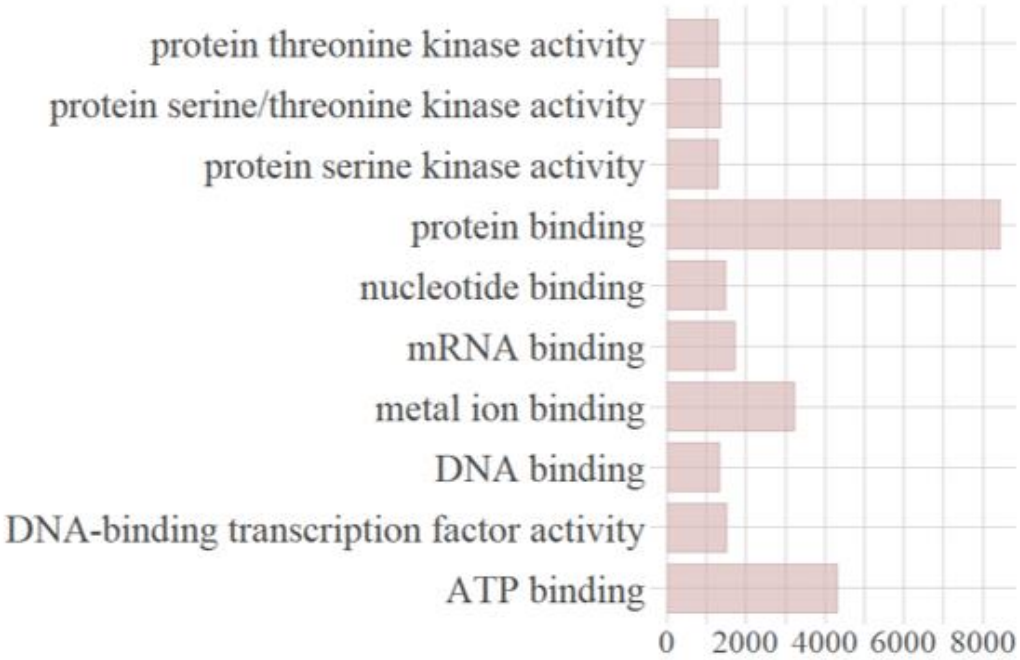

Figura 9. Número de transcritos associados ao termo ontológico dos dez principais termos na montagem do transcriptoma da função molecular de *Monteverdia ilicifolia*.

Na categoria BP, "regulação da transcrição", "fosforilação de proteínas" e "ubiquitinação de proteínas" foram as mais proeminentes (Figura 10).

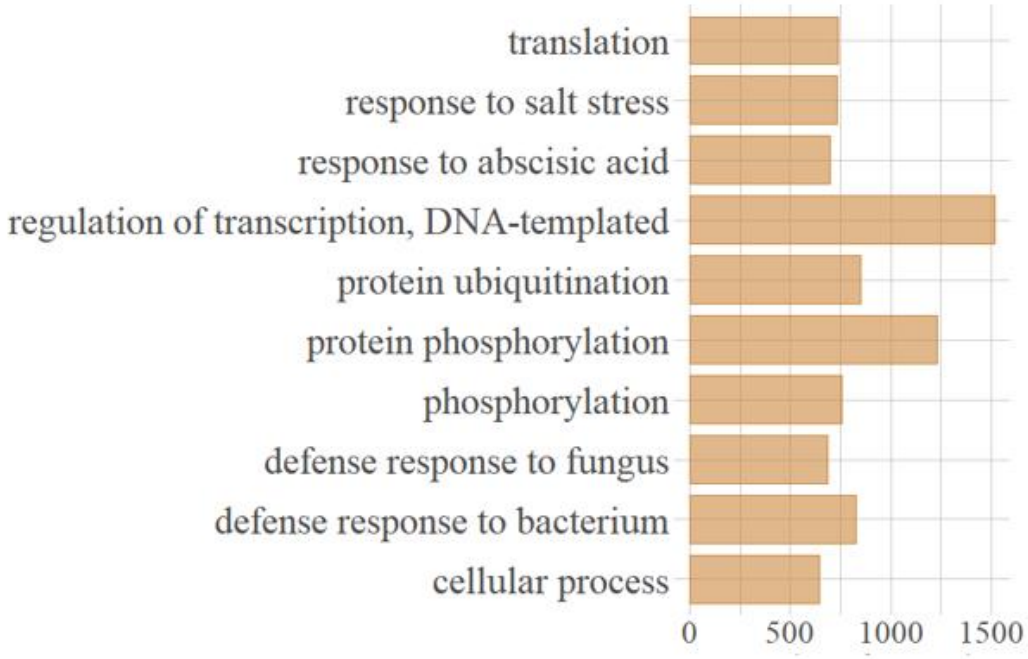

Figura 10. Número de transcritos associados ao termo ontológico dos dez principais termos na montagem do transcriptoma do processo biológico de *Monteverdia ilicifolia*.

Na categoria CC, "núcleo", "membrana plasmática" e "componente integral da membrana" foram os termos mais abundantes (Figura 11).

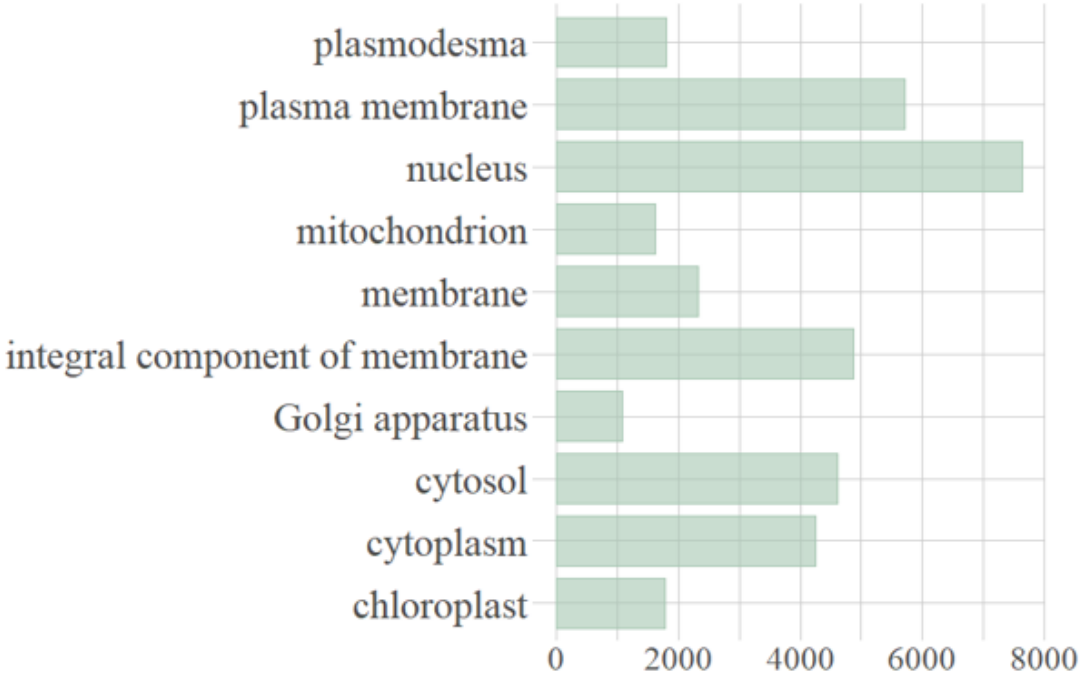

Figura 11. Número de transcritos associados ao termo ontológico dos dez principais termos na montagem do transcriptoma do componente celular de *Monteverdia ilicifolia*.

4.4 Identificação de transcritos diferencialmente expressos

A comparação do nível de abundância dos transcritos revelou expressão diferencial significativa de 2.215 transcritos (FDR <0,05) entre os transcriptomas de ambos os órgãos. Os níveis de expressão foram representados como razão $\log_2$ da abundância de transcritos entre as amostras de folha e raiz (Figura 12), mostrando 1.044 transcritos diferencialmente expressos na raiz e 1.171 na folha. Trabalhando em ambos os órgãos, foi observado que um número de transcritos foi expresso exclusivamente em qualquer um dos órgãos: entre os transcritos diferencialmente expressos, 424 foram expressos exclusivamente na folha e 298 na raiz.

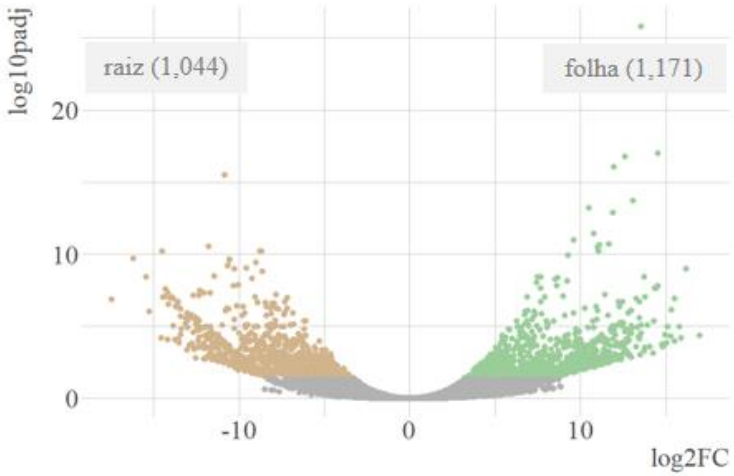

Figura 12. Diferenças na expressão gênica entre os órgãos radicular e foliar de *Monteverdia ilicifolia*. Os valores $-\log_{10}$ do p ajustado foram plotados de acordo com a expressão diferencial entre a raiz e a folha ($\log_2$ Fold Change). Os transcritos de raiz expressos diferencialmente estão destacados em marrom (esquerda) e os transcritos de folhas expressos diferencialmente, em verde (direita).

Para melhor caracterizar o perfil do transcriptoma enviesado pelo órgão, o pacote topGO foi usado para avaliar o enriquecimento de GO (p-valor <0,01) para os transcritos expressos diferencialmente e, posteriormente, os termos representativos foram resumidos na remoção do redundante usando REVIGO. Entre os 770 transcritos expressos diferencialmente expressos e anotados na raíz, 568 genes foram atribuídos a 260 termos GO, enquanto que na folha, dos 902 transcritos diferencialmente expressos e anotados, 610 foram classificados em 265 termos GO.

A GO revelou enriquecimento por processos biológicos (PB) na raiz para “resposta ao etileno”, “regulação do processo celular” e outros (Figura 12), enquanto na folha para "fotossíntese", "ligação proteína-cromóforo" e outros (Figura 12). De acordo com termos de análise funcional, folhas e raízes de *M. illicifoia* também diferem em níveis de função molecular (MF), com transcritos superexpressos nas raízes sendo principalmente associados com "ligação de íons de cálcio", "ligação de íons de ferro" e outros (Figura 12), enquanto o transcritos foliares superexpressos estão associados com “atividade oxidorredutase”, “ligação à clorofila” e outros (Figura 13).

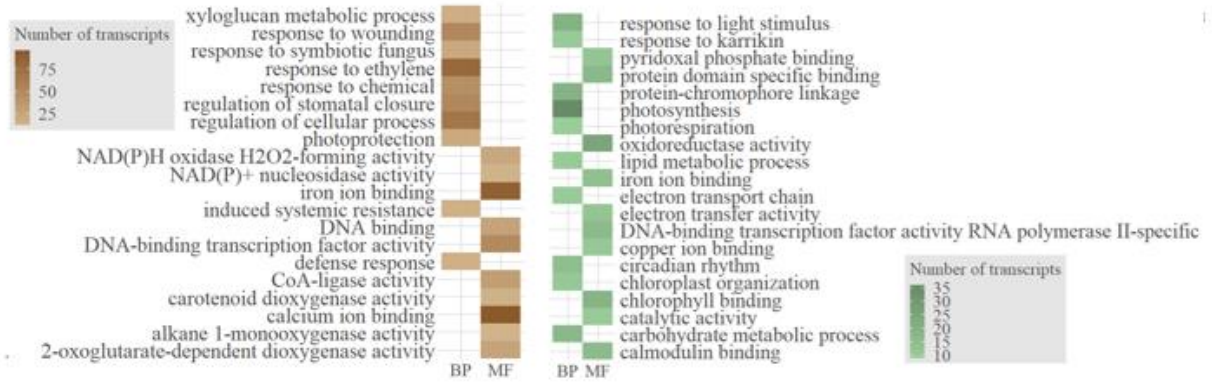

Figura 13. Os dez termos mais representados da análise de enriquecimento de ontologia genética em processo biológico (BP) e função molecular (MF) para transcritos diferencialmente expressos para raiz (esquerda) e folha (direita).

Termos de GO associados ao metabolismo secundário foram encontrados em 295 transcritos expressos diferencialmente, 164 na raiz e 131 na folha. Alguns termos foram encontrados enriquecidos em órgão específico, por exemplo, “atividade de dioxigenase dependente de 2-oxoglutarato” e “resposta ao herbívoro” na raiz e “atividade de beta-amirina sintase” e “processo biossintético de triterpenoide” na folha. Termos coincidentes como “atividade oxidorredutase” foram observados em transcritos superexpressos de ambos os órgãos (Tabela 1).

Tabela 1. Transcritos diferencialmente expressos na raiz ou folha de *Monteverdia ilicifolia* enriquecidos para termos de ontologia genética (MF: função molecular e BP: processo biológico) envolvidos no metabolismo secundário.

GO ID	Termos de GO	Número de transcritos na raiz (TPM)	Número de transcritos na folha (TPM)
GO:0016706 (MF)	Atividade 2-oxoglutarato-dependente de dioxigenase	26	-
GO:0080027 (BP)	Resposta à herbivoria	15	-
GO:0016491 (MF)	Atividade oxidorredutase	7	25
GO:0016709 (MF)	Atividade oxidorredutase, agindo em pares de doadores com a incorporação ou redução de moléculas de oxigênio, NAD(P)H como doador e incorporação de átomos de oxigênio	6	2
GO:0019742 (BP)	Processo metabólico de triterpenoide pentacíclico	3	-
GO:0016106 (BP)	Processo de biossíntese de sesquiterpenoides	3	-
GO:0042300 (MF)	Atividade de síntese de beta-amirina	-	6
GO:0016104 (BP)	Processo de biossíntese de triterpenoide	-	5

A análise comparativa do transcriptoma levou à identificação de 350 e 487 transcritos associados aos números da Enzyme Commission (EC) na raiz e na folha, respectivamente. Esses trancritos órgão-específicos foram mapeadas no banco de dados de vias do KEGG para o “Mapa de biossíntese de metabólitos secundários de plantas” (ko01060) e vias relacionados. As enzimas envolvidas na "biossíntese monoterpenoide" e na biossíntese de “isoflavonoides" foram identificadas em transcritos superexpressos de raízes, enquanto" biossíntese de “flavonoides" e "Biossíntese de alcaloides derivados de histidina e purina" na folha (Tabela 2).

Tabela 2. As enzimas mapearam as vias KEGG identificadas nas análises comparativas do transcriptoma da raiz e da folha de *Monteverdia ilicifolia*.

Via ID	Nome da via	EC number
	Folha	
map00230	Metabolismo de purina	EC:2.4.2.14 EC:2.7.6.5 EC:3.5.2.5
map00941	Biossíntese de flavonoides	EC:2.3.1.133
map00950	Biossíntese de alcaloides isoquinolano	EC:1.4.3.21 EC:2.6.1.5

map00960	Biossíntese de alcaloides trofanos, piperidanos e piridanos	EC:1.4.3.21 EC:2.6.1.5
map01064	Biossíntese de alcaloides derivados de ornitina, lisina e ácido nicotinico	EC:1.4.3.21 EC:2.6.1.5
map01065	Biosíntese de alcalóides derviados de histidina e purina	EC:4.1.2.13 EC:1.2.1.9 EC:2.4.2.14 EC:2.7.6.5 EC:5.1.3.1 EC:1.1.1.49 EC:2.2.1.1 EC:3.5.2.5
Raiz		
map00902	Biossíntese de monoterpenos	EC:2.1.1.50
map00943	Biossíntese de isoflavonoides	EC:2.1.1.46

Juntos, os resultados de enriquecimento GO e mapeamento KEGG de transcritos superexpressos na raiz ou folha de *M. ilicifolia* confirmaram o acúmulo de SMs já relatados por outros procedimentos metodológicos, incluindo flavonoides, triterpenos e sesquiterpenos nas folhas (DE SOUZA et al., 2008), enquanto as raízes contêm terpenos, triterpenos, alcaloides e especialmente os triterpenos quinonometídeos (FILHO et al., 2002; PAZ et al., 2017; PÉRICO et al., 2018).

4.5 Identificação de potenciais Fatores de Transcrição

No total, 191 transcritos foram anotados como fatores de transcrição (TFs). Deste total, 12 TFs pertencentes a 9 famílias distintas foram expressos diferencialmente nas folhas de *M. ilicifolia* e três, pertencentes à família *ERF*, nas raízes (Tabela 3).

Tabela 3. Lista dos fatores de transcrição diferencialmente expressos em folhas de *Monteverdia ilicifolia*.

Transcrito	Família	Número de transcritos na folha (TPM)	Número de transcritos na raiz (TPM)
Folha			
DN36518_c0_g2_i1	bHLH	1742	5
DN47776_c0_g1_i1		7761	9
DN48531_c1_g1_i1		1086	0
DN41721_c3_g1_i1	GATA	2182	0
DN44541_c6_g5_i1	SBP	1775	0
DN46860_c2_g1_i2	WRKY	788	0
DN39766_c0_g2_i1	DBB	18278	6
DN45471_c0_g1_i1	CO-like	9668	10
DN46597_c0_g1_i1		1554	0

DN41570_c1_g1_i1	MIKC_MADS	1849	5
DN51316_c0_g2_i2	HB-PHD	1387	0
DN52840_c2_g1_i5	MYB_related	13213	4
Raiz			
DN33972_c0_g1_i1	ERF	1	338
DN40539_c3_g1_i1		0	187
DN40539_c3_g1_i2		0	103

Vários fatores de transcrição são relatados desempenhando papéis importantes na síntese e no metabolismo de metabólitos secundários. Estes fatores de transcrição altamente expressos nas folhas e raízes de *M. ilicifolia* pertencem às famílias de bHLH, GATA, SBP, WRKY, DBB, CO-like, MIKC_MADS, HB-PHD, MYB_related e ERF. Sendo os membro da família DBB mais abundantes nas folhas e de ERF nas raízes. Fatores de transcrição das famílias CO-like, DBB, ERF, GATA, HB-other, já foram identificadas em gêneros da família Solanaceae (TRIPATHI et al., 2017) . Além disso, já foi relatado que membros da família WRKY participam na regulação da biossíntese dos triterpenóides (TRIPATHI et al., 2017). A identificação dos fatores de transcrição diferencilamente expressos constituem um valioso recurso genético para estudos adicionais das suas funções reguladoras na biossíntese de metabólitos secundários.

4.6 Identificação de genes relacionados à biossíntese do esqueleto de terpenoides

A análise dos transcritos de *M. ilicifolia* também possibilitou a identificação de enzimas envolvidas na biossíntese de terpenos, incluindo as vias do mevalonato (MEV) e do metileritritol-fosfato (MEP) (Figura 14).

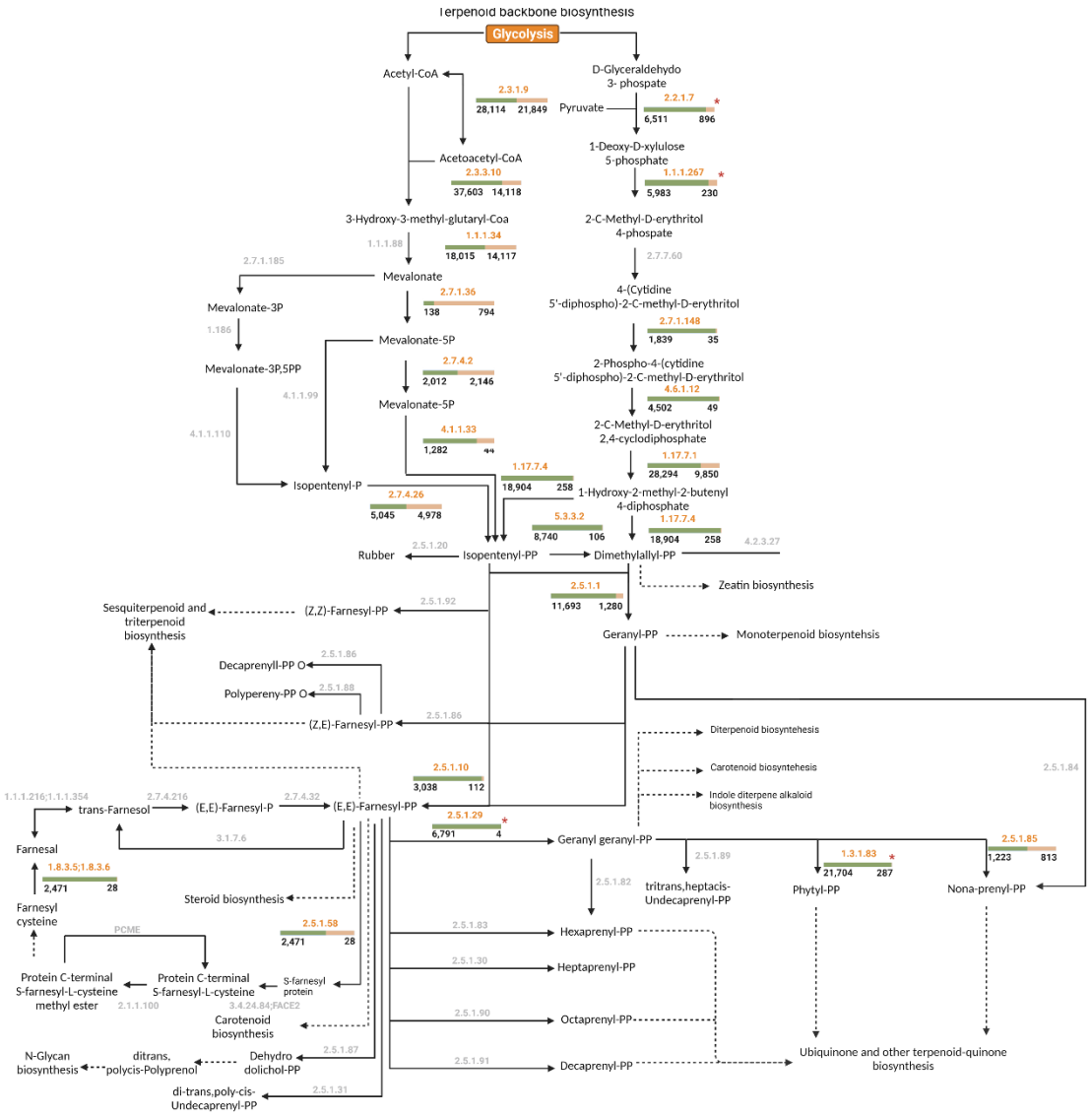

Figura 14. Transcritos de *Monteverdia ilicifolia* envolvidos na biossíntese de terpenos. A barra abaixo do código representa a porcentagem de reads normalizadas (TPM) identificadas nas amostras de folha (verde) e raiz (marrom); os números abaixo da barra representam o valor absoluto de número de reads. *Valor de p ajustado <0,05 para teste *t* para a comparação estatística entre o número de reads de cada órgão.

Oito transcritos foram mapeados na via do MEP, incluindo enzimas responsáveis pela síntese dos blocos contrutores pirofosfato de isopentenilo (IPP) e difosfato de dimetilalilo (DMAPP). São: 1-desoxi-D-xilulose-5-fosfato sintase (DXS), 1-desoxi-D-xilulose-5-fosfato redutoisomerase (DXR), 4-(citidina 5'-difosfo)-2-C-metil-D-eritritol quinase (ispE), 2-C-metil-D-eritritol 2,4- ciclodifosfato sintase (ispF) e (E)-4-hidroxi-3-metilbut-2-enil-difosfato sintase (ispG), 4-hidroxi-3-metilbut-2-enil difosfato redutase (ispH) e isopentenil-difosfato delta-isomerase (IDI). As duas últimas são comuns às vias do MEP e do MEP na formação de IPP e DMAPP (Figura 12).

Para a via do MEV, além dos dois transcritos apontados anteriormente, outros sete foram anotados como enzimas pertencentes às etapas iniciais, como acetil-CoA C-acetiltransferase (ACAT), hidroximetilglutaril-CoA sintase (HMGCS), hidroximetilglutaril-CoA redutase (NADPH) (HMGCR), mevalonato quinase (MVK), fosfomevalonato quinase (PMVK), difosfomevalonato descarboxilase (MVD), isopentenil fosfato quinase (IPK) (Figura 12).

Após a formação do IPP e DMAPP, essas unidades podem ser combinadas por um ligação cabeça-cauda para a formação de monoterpenos (C10), sesquiterpenos (C15) e diterpenos (C20) (Dewick, 2009). No caso da via MEP, a enzima geranil-difosfato sintase (GPS) responsável pela formação de geranil-PP e, portanto, pela biossíntese de monoterpenóides, foi identificada junto com outras duas enzimas relacionadas à ubiquinona e outros quinonoterpenoides, a geranilgeranil difosfato redutase (chlP) e all-trans-nonaprenil-difosfato sintase (SPS).

Quatro transcritos foram anotados como enzimas na via MEV, a qual que leva à formação de diterpenos e triterpenos. A geranilgeranil difosfato sintase (GGPS) é responsável pela síntese de geranil geranil-PP e, portanto, à biossíntese de diterpenos. Farnesil difosfato sintase (FDPS), prenilcisteína oxidase (PCYOX1) e proteína farnesiltransferase (FNTB) estão relacionadas à síntese de farnesil-PP, um precursor de sesquiterpenos e triterpenos. Uma ligação cauda a cauda de dois farnesil-PP gera o 2,3-oxidosqualeno, um precursor de triterpenoides e esteroides.

Com base na análise de expressão diferencial, todos os transcritos mapeados na via MEP e a maioria dos mapeados para a via MVA, exceto MVK e PMVK, apresentaram nível de expressão muito maior nas folhas do que nas raízes, corroborando com estudos anteriores sobre identificação de terpeno em tecidos foliares de *M. ilicifolia*. Esta análise comparativa do transcriptoma é uma abordagem amplamente utilizada para identificar novos genes específicos de tecidos em vias biossintéticas de metabólitos secundários (GUO et al., 2021; ZHU et al., 2022).

4.7 Identificação de genes codificadores de P450

A partir da anotação do transcriptoma contra a base de dados, foram identificados 126 transcritos completos como codificadores para enzimas CYP450 em *M. ilicifolia*. Considerando a expressão gênica diferencial, um total de cinco transcritos apresentaram maior expressão na folha e 31 na raiz (tabela 4).

Tabela 4. Lista dos transcritos diferencialmente expressos em raízes de *Monteverdia ilicifolia* potencialmente identificados como codificadores de P450.

Transcrito	Número de transcritos na folha (TPM)	Número de transcritos na raiz (TPM)	Proteína alinhada contra (NCBI) [espécie]	Identidade
		Folha		
DN42163_c0_g1_i9	266	0	laccase-7-like isoform X2 [*T. wilfordii*]	51%
DN45165_c1_g2_i3	4925	7	soflavone reductase homolog [*T. wilfordii*]	92%
DN51328_c5_g1_i4	2168	15	cytochrome P450 CYP82D47-like [*T. wilfordii*]	81%
DN52089_c3_g1_i4	1529	46	cytochrome P450 81Q32-like [*T. wilfordii*]	84%
DN52089_c3_g1_i5	1710	48	cytochrome P450 81Q32-like [*T. wilfordii*]	84%
		Raiz		
DN29044_c0_g1_i1	0	136	iridoid oxidase-like [*T. wilfordii*]	88%
DN33873_c0_g1_i1	64	3486	peroxidase 55-like [*T. wilfordii*]	90%
DN35625_c0_g2_i1	1	65	laccase-7-like [*T. wilfordii*]	92%
DN37736_c2_g1_i1	0	937	probable 2-oxoglutarate-dependent dioxygenase AOP1 [*T. wilfordii*]	91%
DN37736_c2_g2_i1	0	851	probable 2-oxoglutarate-dependent dioxygenase AOP1 [*T. wilfordii*]	86%
DN37736_c2_g2_i2	2	3649	probable 2-oxoglutarate-dependent dioxygenase AOP1 [*T. wilfordii*]	87%
DN37803_c0_g1_i1	0	325	cytochrome P450 81Q32-like [*T. wilfordii*]	90%
DN40427_c0_g1_i1	0	4607	cytochrome P450 CYP749A22-like isoform X1 [*T. wilfordii*]	90%
DN40575_c2_g1_i1	0	620	cytochrome P450 CYP749A22-like [*T. wilfordii*]	92%
DN41232_c0_g1_i1	5	682	laccase-7-like [*T. wilfordii*]	90%

DN43662_c0_g1_i6	557	3631	eugenol synthase 1-like [*T. wilfordii*]	90%
DN44545_c0_g1_i1	0	145	(+)-neomenthol dehydrogenase-like [*T. wilfordii*]	91%
DN44545_c0_g1_i2	5	126	(+)-neomenthol dehydrogenase-like [*T. wilfordii*]	92%
DN46557_c0_g1_i1	10	615	laccase-7-like isoform X2 [*T. wilfordii*]	89%
DN46557_c0_g1_i5	23	703	laccase-7-like isoform X2 [*T. wilfordii*]	90%
DN46557_c0_g1_i6	6	283	laccase-7-like isoform X2 [*T. wilfordii*]	90%
DN46557_c0_g2_i1	31	654	laccase-7-like [*T. wilfordii*]	90%
DN46557_c0_g2_i2	0	116	laccase-7-like [*T. wilfordii*]	87%
DN46557_c0_g2_i4	0	124	laccase-7-like [*T. wilfordii*]	90%
DN46557_c0_g3_i1	5	163	laccase-7-like [*T. wilfordii*]	89%
DN46793_c0_g1_i1	21	1317	cytochrome P450 71D9-like [*T. wilfordii*]	82%
DN46879_c0_g1_i1	5	1922	abscisic acid 8'-hydroxylase 3-like [*T. wilfordii*]	92%
DN46964_c1_g1_i1	16	240	perakine reductase-like [*T. wilfordii*]	94%
DN47569_c1_g1_i3	5	274	(+)-neomenthol dehydrogenase-like [*T. wilfordii*]	81%
DN47829_c1_g1_i1	2	1360	iridoid oxidase-like [*T. wilfordii*]	92%
DN48389_c0_g1_i1	3	162	S-norcoclaurine synthase 1-like [*T. wilfordii*]	84%
DN48651_c0_g1_i1	0	181	Cupredoxin superfamily protein [*T. wilfordii*]	82%
DN49544_c0_g1_i1	2	479	laccase-7-like [*T. wilfordii*]	87%
DN49544_c0_g1_i2	1	3506	laccase-7-like [*T. wilfordii*]	87%
DN50012_c0_g1_i3	0	108	isoeugenol synthase 1-like [*T. wilfordii*]	90%
DN53120_c2_g1_i1	0	72	berberine bridge enzyme-like 15 [*T. wilfordii*]	95%

Os prováveis genes codificantes para enzimas P450 apresentaram, como esperado, uma maior expressão na raiz. Além disso, observou-se alta identidade com proteínas identificadas em *T. wilfordii*.

Em angiospermas, o CYPoma, como é denominado o número total de CYPs em um determinado organismo, constitui cerca de 1% do transcriptoma (Nelson and Werck-Reichhart, 2011). A exemplo, em uva (*Vitis vinífera*) foram identificadas 279 genes codificadores para CYPs (Ilc et al., 2018), em arroz (*Oryza sativa*), 329 (Nelson and Werck- Reichhart, 2011), em eucalipto (*Eucalyptus grandis*) e em trigo (*Triticum aestivum*) 443 (HANSEN et al., 2021).

Finalmente, a partir do alinhamento do transcriptoma de raiz de *M. ilicifolia* deste trabalho contra a sequência do mRNA da CYP712K de *T. wilfordii* (MN621243.1) depositada na base de dados, foi identificado o gene candidato em *M. ilicifolia* que codifica a CYP712K (DN44265_c0_g1_i1). Essa identificação da sequência do transcrito possibilitou a caracterização funcional da respectiva proteína, mostrando que essa enzima é responsável por três etapas oxidativas da friedelina na posição C-29 (BICALHO et al., 2019). Em termos de expressão relativa, curiosamente esse transcrito não apresentou diferença de expressão entre as duas situações amostrais desse estudo.

As P450 estão envolvidas na biossíntese de triterpenoides em plantas, metabólitos exclusivamente presentes nas raízes. O primeiro estágio de biossíntese consiste na ciclização do 2,3-oxidosqualeno à fridelina, catalisado pela friedelina sintase (SOUZA-MOREIRA et al., 2016). Na segunda etapa da biossíntese dos triterpenóides nas plantas, os blocos triterpênicos gerados sofrem uma ou mais modificações oxidativas em várias posições do esqueleto, que são catalisadas por várias enzimas do citocromo P450 e aumentam drasticamente a diversidade estrutural dos triterpenóides (THIMMAPPA et al., 2014). Três membros da subfamília CYP712K da planta medicinal *T. wilfordii* foram sugeridos a catalisar a oxidação na friedelina em C-29 (HANSEN et al., 2017) sendo a CYP712K4 identificada como a enzima que cataliza a oxidação de friedelina em *M. ilicifolia* na via dos quinonomedídeos triterpenoides (BICALHO et al., 2019).

5 Conclusões

O presente estudo apresenta um extenso conjunto de dados do transcriptoma gerado a partir de análises de sequenciamento *de novo* de *M. ilicifolia*. Foram utilizados dois órgãos distintos da planta (folhas e raízes), de duas situações de desenvolvimento

diferentes (mudas e adulto), além de envolver dois processos diferentes de sequenciamento (HiSeq e MiSeq). A partir da qualidade do sequenciamento e de métricas envolvidas com a identificação de genes ortólogos foi possível verificar que a cobertura dos dados do transcriptoma é consistente para inferir sobre a sequência de genes envolvidos nas vias metabólicas de *M. ilicifolia*. A análise comparativa do transcriptoma de folhas e raízes facilitou a identificação dos genes envolvidos no metabolismo secundário especificamente expressos nesses órgãos, uma abordagem amplamente utilizada para minerar e identificar novos genes na biossíntese de SMs em plantas. No entanto, para a identificação de alguns transcritos envolvidos nesse metabolismo, é importante não se restringir à diferença de expressão diferencial. Por fim, a apresentação de sequências completas dos prováveis genes codificantes para proteínas envolvidas no metabolismo secundário de *M. ilicifolia* permite uma gama de trabalhos futuros na elucidação de suas vias.

6 Material Suplementar

Tabela S1 Lista dos transcritos de *Monteverdia ilicifolia* identificados como enzimas e pertencentes à uma via de biossíntese de metabolismo secundário.

Código enzimático	Nome da enzima	Transcrito
Map00100 - Biossíntese de esteroides		
EC:1.14.19.20	Delta(7)-sterol 5(6)-desaturase	DN45589_c3_g1_i2
EC:1.3.1.21	7-dehydrocholesterol reductase	DN42082_c0_g1_i1
EC:1.3.1.72	Delta(24)-sterol reductase	DN50161_c0_g1_i2
EC:2.1.1.143	24-methylenesterol C-methyltransferase	DN35734_c0_g1_i1
EC:2.1.1.41	Sterol 24-C-methyltransferase	DN43896_c0_g1_i1
EC:5.4.99.7	Lanosterol synthase	DN20261_c0_g1_i1
EC:5.5.1.9	Cycloeucalenol cycloisomerase	DN45511_c1_g1_i1
Map00230 - Metabolismo de purina		
EC:1.1.1.205	IMP dehydrogenase	DN49612_c0_g1_i1
EC:1.17.4.1	Ribonucleoside-diphosphate reductase	DN41598_c0_g1_i1
EC:1.7.3.3	Factor independent urate hydroxylase	DN45663_c0_g1_i1
EC:2.4.2.14	Amidophosphoribosyltransferase	DN44971_c0_g1_i1
EC:2.4.2.7	Adenine phosphoribosyltransferase; Translocases	DN34554_c0_g1_i1
EC:2.7.1.20	Adenosine kinase	DN45934_c0_g1_i3
EC:2.7.1.25	Adenylyl-sulfate kinase	DN42588_c0_g1_i1
EC:2.7.1.40	Pyruvate kinase	DN42331_c0_g1_i1
EC:2.7.4.6	Nucleoside-diphosphate kinase	DN39664_c1_g1_i2
EC:2.7.6.1	Ribose-phosphate diphosphokinase	DN35369_c0_g1_i1

EC:2.7.6.5	GTP diphosphokinase	DN42531_c0_g1_i1
EC:2.7.7.4	Sulfate adenylyltransferase; Adenylyl-sulfate kinase	DN45794_c0_g1_i1
EC:3.1.3.5	5'-nucleotidase	DN42284_c1_g2_i1
EC:3.2.2.8	Ribosylpyrimidine nucleosidase; Adenosine nucleosidase; Inosine nucleosidase; Purine nucleosidase; Uridine nucleosidase	DN44821_c0_g1_i2
EC:3.5.1.5	Urease	DN47269_c1_g1_i2
EC:3.5.3.26	(S)-ureidoglycine aminohydrolase	DN46681_c0_g2_i1
EC:3.5.3.9	Allantoate deiminase	DN44617_c0_g1_i1
EC:3.6.1.13	ADP-ribose diphosphatase; Nucleotide diphosphatase	DN37728_c0_g1_i1
EC:3.6.1.5	Apyrase; Nucleoside diphosphate phosphatase	DN49219_c1_g1_i1
EC:3.6.1.6	Nucleoside diphosphate phosphatase	DN43603_c0_g1_i1
EC:3.6.1.9	Nucleotide diphosphatase; Guanosine-3',5'-bis(diphosphate) 3'-diphosphatase	DN40418_c0_g1_i1
EC:4.1.1.21	Phosphoribosylaminoimidazole carboxylase	DN50486_c0_g1_i1
EC:4.6.1.1	Adenylate cyclase	DN32415_c0_g2_i1
EC:4.6.1.2	Guanylate cyclase	DN45942_c0_g1_i1
EC:5.4.2.2	Phosphoglucomutase (alpha-D-glucose-1,6-bisphosphate-dependent)	DN26548_c0_g1_i1
EC:6.3.2.6	Phosphoribosylaminoimidazolesuccinocarboxamide synthase	DN50546_c0_g3_i1
EC:6.3.3.1	Phosphoribosylformylglycinamidine cyclo-ligase	DN47361_c0_g1_i2
EC:6.3.4.4	Adenylosuccinate synthase	DN44785_c0_g1_i1
EC:6.3.5.2	GMP synthase (glutamine-hydrolyzing)	DN45265_c0_g1_i3
	Map00902 - Biossíntese de monoterpenos	
EC:1.1.1.208	(+)-neomenthol dehydrogenase	DN39902_c0_g1_i1
EC:2.1.1.50	Loganate O-methyltransferase	DN39137_c0_g2_i1
	Map00904 - Biossíntese de diterpenos	
EC:1.14.11.13	Gibberellin 2-beta-dioxygenase	DN39056_c0_g1_i1
EC:1.14.11.15	Gibberellin 3-beta-dioxygenase	DN40445_c0_g1_i1
EC:1.14.14.86	Ent-kaurene monooxygenase; Acting on paired donors, with incorporation or reduction of molecular oxygen. The oxygen incorporated need not be derived from O(2)	DN18948_c0_g1_i1
	Map00908 - Biossíntese de zeatina	
EC:1.5.99.12	Cytokinin dehydrogenase	DN37944_c0_g1_i1
EC:2.4.2.40	Zeatin O-beta-D-xylosyltransferase; Trans-zeatin O-beta-D-glucosyltransferase	DN23670_c0_g1_i1
EC:2.5.1.27	Adenylate dimethylallyltransferase (AMP-dependent); tRNA dimethylallyltransferase	DN44349_c2_g1_i1
EC:2.5.1.75	tRNA dimethylallyltransferase	DN48466_c0_g1_i1
	Map00940 - Biossíntese de fenilpropanoides	
EC:1.1.1.195	Cinnamyl-alcohol dehydrogenase	DN22797_c0_g1_i1
EC:1.14.14.91	Trans-cinnamate 4-monooxygenase; Acting on paired donors, with incorporation or reduction of molecular oxygen. The oxygen incorporated need not be derived from O(2)	DN46509_c1_g1_i1

EC:1.2.1.44	Cinnamoyl-CoA reductase	DN41977_c0_g1_i1
EC:1.2.1.68	Coniferyl-aldehyde dehydrogenase; Aldehyde dehydrogenase (NAD(P)(+)); Aldehyde dehydrogenase (NAD(+))	DN44246_c1_g1_i1
EC:2.1.1.104	Caffeoyl-CoA O-methyltransferase	DN43249_c0_g1_i1
EC:2.1.1.68	Caffeate O-methyltransferase	DN39256_c0_g1_i1
EC:2.3.1.133	Shikimate O-hydroxycinnamoyltransferase	DN21324_c0_g1_i1
EC:2.4.1.111	Coniferyl-alcohol glucosyltransferase	DN43846_c0_g1_i1
EC:2.4.1.128	Scopoletin glucosyltransferase	DN43792_c1_g1_i2
EC:3.2.1.126	Coniferin beta-glucosidase; Beta-glucosidase; Beta-galactosidase; Beta-D-fucosidase	DN50830_c0_g1_i2
EC:3.2.1.21	Beta-glucosidase; Glucan endo-1,3-beta-D-glucosidase	DN39445_c0_g1_i1
EC:4.3.1.24	Phenylalanine ammonia-lyase; Phenylalanine/tyrosine ammonia-lyase	DN23934_c0_g1_i1
EC:6.2.1.12	4-coumarate--CoA ligase	DN37352_c0_g1_i1
Map00941 - Biossíntese de flavonoides		
EC:1.1.1.219	Dihydroflavanol 4-reductase; Flavanone 4-reductase	DN41004_c0_g1_i1
EC:1.14.11.9	Flavanone 3-dioxygenase	DN48516_c0_g1_i1
EC:1.14.14.81	Flavonoid 3',5'-hydroxylase; Acting on paired donors, with incorporation or reduction of molecular oxygen. The oxygen incorporated need not be derived from O(2)	DN39300_c0_g1_i1
EC:1.14.14.82	Flavonoid 3'-monooxygenase; Acting on paired donors, with incorporation or reduction of molecular oxygen. The oxygen incorporated need not be derived from O(2)	DN4171_c0_g1_i1
EC:1.14.14.91	Trans-cinnamate 4-monooxygenase; Acting on paired donors, with incorporation or reduction of molecular oxygen. The oxygen incorporated need not be derived from O(2)	DN46509_c1_g1_i1
EC:1.17.1.3	Leucoanthocyanidin reductase	DN46952_c0_g1_i3
EC:2.1.1.104	Caffeoyl-CoA O-methyltransferase	DN43249_c0_g1_i1
EC:2.3.1.133	Shikimate O-hydroxycinnamoyltransferase	DN21324_c0_g1_i1
EC:2.3.1.74	Chalcone synthase	DN43762_c1_g2_i1
EC:5.5.1.6	Chalcone isomerase	DN41769_c1_g1_i1
Map00943 - Biossíntese de isoflavonoides		
EC:2.1.1.46	Isoflavone 4'-O-methyltransferase	DN1836_c0_g1_i1
EC:4.2.1.105	2-hydroxyisoflavanone dehydratase; Carboxylesterase	DN47821_c0_g1_i3
Map00944 - Biossíntese de flavonoides de flavonol		
EC:1.1.1.366	L-idonate 5-dehydrogenase (NAD(+))	DN47541_c1_g1_i1
Map00950 - Biossíntese de alcaloide isoquinolino		
EC:1.14.11.32	Codeine 3-O-demethylase	DN41817_c0_g1_i1
EC:1.4.3.21	Primary-amine oxidase	DN45587_c1_g1_i3
EC:2.6.1.1	Aspartate transaminase	DN39436_c0_g2_i2
EC:2.6.1.5	Tyrosine transaminase	DN49229_c2_g1_i2
Map00960 - Biossíntese de alcaloides peridinos, trofanos e piperidinos		

EC:1.1.1.206	Tropinone reductase I	DN50091_c0_g1_i1
EC:1.1.1.237	Hydroxyphenylpyruvate reductase	DN41320_c0_g1_i2
EC:1.4.3.21	Primary-amine oxidase	DN45587_c1_g1_i3
EC:2.6.1.1	Aspartate transaminase	DN39436_c0_g2_i2
EC:2.6.1.5	Tyrosine transaminase	DN49229_c2_g1_i2
EC:2.6.1.9	Histidinol-phosphate transaminase	DN46775_c0_g1_i1
Map1061 - Biossíntese de fenilpropanoides		
EC:1.1.1.195	Cinnamyl-alcohol dehydrogenase	DN22797_c0_g1_i1
EC:1.14.14.91	Trans-cinnamate 4-monooxygenase; Acting on paired donors, with incorporation or reduction of molecular oxygen. The oxygen incorporated need not be derived from O(2)	DN46509_c1_g1_i1
EC:1.2.1.44	Cinnamoyl-CoA reductase	DN41977_c0_g1_i1
EC:1.2.1.68	Coniferyl-aldehyde dehydrogenase; Aldehyde dehydrogenase (NAD(P)(+)); Aldehyde dehydrogenase (NAD(+))	DN44246_c1_g1_i1
EC:2.1.1.104	Caffeoyl-CoA O-methyltransferase	DN43249_c0_g1_i1
EC:2.1.1.68	Caffeate O-methyltransferase	DN39256_c0_g1_i1
EC:2.3.1.133	Shikimate O-hydroxycinnamoyltransferase	DN21324_c0_g1_i1
EC:2.4.1.111	Coniferyl-alcohol glucosyltransferase	DN43846_c0_g1_i1
EC:2.4.1.128	Scopoletin glucosyltransferase	DN43792_c1_g1_i2
EC:3.2.1.126	Coniferin beta-glucosidase; Beta-glucosidase; Beta-galactosidase; Beta-D-fucosidase	DN50830_c0_g1_i2
EC:3.2.1.21	Beta-glucosidase; Glucan endo-1,3-beta-D-glucosidase	DN39445_c0_g1_i1
EC:4.3.1.24	Phenylalanine ammonia-lyase; Phenylalanine/tyrosine ammonia-lyase	DN23934_c0_g1_i1
EC:6.2.1.12	4-coumarate--CoA ligase	DN37352_c0_g1_i1
EC:1.1.1.219	Dihydroflavanol 4-reductase; Flavanone 4-reductase	DN41004_c0_g1_i1
EC:1.14.11.9	Flavanone 3-dioxygenase	DN48516_c0_g1_i1
EC:1.14.14.81	Flavonoid 3',5'-hydroxylase; Acting on paired donors, with incorporation or reduction of molecular oxygen. The oxygen incorporated need not be derived from O(2)	DN39300_c0_g1_i1
EC:1.14.14.82	Flavonoid 3'-monooxygenase; Acting on paired donors, with incorporation or reduction of molecular oxygen. The oxygen incorporated need not be derived from O(2)	DN4171_c0_g1_i1
EC:1.17.1.3	Leucoanthocyanidin reductase	DN46952_c0_g1_i3
EC:2.3.1.74	Chalcone synthase	DN43762_c1_g2_i1
EC:5.5.1.6	Chalcone isomerase	DN41769_c1_g1_i1
EC:2.1.1.46	Isoflavone 4'-O-methyltransferase	DN1836_c0_g1_i1
EC:4.2.1.105	2-hydroxyisoflavanone dehydratase; Carboxylesterase	DN47821_c0_g1_i3
EC:3.2.1.31	Beta-glucuronidase	DN39979_c0_g1_i1
Map1063 - Biossíntese de alcaloides derivados da via de shikimato		

EC:1.1.1.35	3-hydroxyacyl-CoA dehydrogenase; Delta(3)-Delta(2)-enoyl-CoA isomerase; Enoyl-CoA hydratase; 3-hydroxybutyryl-CoA epimerase	DN44825_c0_g1_i1
EC:1.11.1.6	Catalase	DN25355_c0_g1_i1
EC:1.2.3.1	Aldehyde oxidase	DN45199_c0_g1_i1
EC:1.4.3.22	Diamine oxidase; Primary-amine oxidase	DN30246_c0_g1_i1
EC:1.8.1.4	Dihydrolipoyl dehydrogenase	DN25751_c0_g1_i1
EC:2.1.1.4	Acetylserotonin O-methyltransferase	DN39217_c0_g1_i1
EC:2.3.1.61	Dihydrolipoyllysine-residue succinyltransferase	DN43049_c0_g1_i1
EC:2.6.1.27	Tryptophan transaminase	DN48211_c1_g2_i1
EC:3.5.1.9	Arylformamidase	DN44614_c0_g1_i1
EC:4.2.1.84	Nitrile hydratase; 3-cyanoalanine hydratase; Cyanoalanine nitrilase; Nitrilase	DN47991_c1_g1_i1
EC:1.14.11.32	Codeine 3-O-demethylase	DN41817_c0_g1_i1
EC:1.4.3.21	Primary-amine oxidase	DN45587_c1_g1_i3
EC:2.6.1.1	Aspartate transaminase	DN39436_c0_g2_i2
EC:2.6.1.5	Tyrosine transaminase	DN49229_c2_g1_i2
EC:1.1.1.1	Alcohol dehydrogenase; Alcohol dehydrogenase (NAD(P)(+))	DN42576_c0_g1_i1
EC:1.14.13.8	Flavin-containing monooxygenase; Indole-3-pyruvate monooxygenase	DN35718_c1_g2_i1
EC:1.2.1.5	Aldehyde dehydrogenase (NAD(P)(+)); Aldehyde dehydrogenase (NAD(+))	DN43514_c0_g1_i1
EC:2.4.1.17	Glucuronosyltransferase	DN12103_c0_g1_i1
EC:2.5.1.18	Glutathione transferase	DN29115_c0_g1_i1
Map1064 - Biossíntese de alcaloides derivados da ornitina, lisina e ácido nicotínico		
EC:1.1.1.206	Tropinone reductase I	DN50091_c0_g1_i1
EC:1.1.1.237	Hydroxyphenylpyruvate reductase	DN41320_c0_g1_i2
EC:1.4.3.21	Primary-amine oxidase	DN45587_c1_g1_i3
EC:2.6.1.1	Aspartate transaminase	DN39436_c0_g2_i2
EC:2.6.1.5	Tyrosine transaminase	DN49229_c2_g1_i2
EC:2.6.1.9	Histidinol-phosphate transaminase	DN46775_c0_g1_i1
Map1065 - Biossíntese de alcaloides derivados da histidina e purina		
EC:1.1.1.43	Phosphogluconate 2-dehydrogenase; Phosphogluconate dehydrogenase (NADP(+)-dependent, decarboxylating)	DN35609_c0_g1_i1
EC:1.1.1.44	Phosphogluconate dehydrogenase (NADP(+)-dependent, decarboxylating)	DN31272_c0_g1_i1
EC:1.1.1.49	Glucose-6-phosphate dehydrogenase (NADP(+))	DN44829_c0_g1_i1
EC:1.2.1.9	Glyceraldehyde-3-phosphate dehydrogenase (NADP(+))	DN42165_c0_g1_i1
EC:2.2.1.1	Transketolase	DN22520_c0_g1_i1
EC:2.7.1.11	6-phosphofructokinase	DN36230_c0_g1_i2
EC:2.7.1.12	Gluconokinase	DN34712_c0_g1_i1
EC:2.7.1.90	Diphosphate--fructose-6-phosphate 1-phosphotransferase; 6-phosphofructokinase	DN36580_c0_g1_i1
EC:2.7.6.1	Ribose-phosphate diphosphokinase	DN35369_c0_g1_i1

EC:3.1.1.31	6-phosphogluconolactonase; Carboxylesterase	DN42946_c0_g1_i1
EC:4.1.2.13	Fructose-bisphosphate aldolase	DN21907_c0_g1_i1
EC:5.1.3.1	Ribulose-phosphate 3-epimerase	DN46752_c0_g1_i1
EC:5.3.1.6	Ribose-5-phosphate isomerase	DN45258_c0_g1_i1
EC:5.3.1.9	Glucose-6-phosphate isomerase	DN47734_c0_g1_i1
EC:5.4.2.2	Phosphoglucomutase (alpha-D-glucose-1,6-bisphosphate-dependent)	DN26548_c0_g1_i1
EC:1.1.1.205	IMP dehydrogenase	DN49612_c0_g1_i1
EC:1.17.4.1	Ribonucleoside-diphosphate reductase	DN41598_c0_g1_i1
EC:1.7.3.3	Factor independent urate hydroxylase	DN45663_c0_g1_i1
EC:2.4.2.14	Amidophosphoribosyltransferase	DN44971_c0_g1_i1
EC:2.4.2.7	Adenine phosphoribosyltransferase; Translocases	DN34554_c0_g1_i1
EC:2.7.1.20	Adenosine kinase	DN45934_c0_g1_i3
EC:2.7.1.25	Adenylyl-sulfate kinase	DN42588_c0_g1_i1
EC:2.7.1.40	Pyruvate kinase	DN42331_c0_g1_i1
EC:2.7.4.6	Nucleoside-diphosphate kinase	DN39664_c1_g1_i2
EC:2.7.6.1	Ribose-phosphate diphosphokinase	DN35369_c0_g1_i1
EC:2.7.6.5	GTP diphosphokinase	DN42531_c0_g1_i1
EC:2.7.7.4	Sulfate adenylyltransferase; Adenylyl-sulfate kinase	DN45794_c0_g1_i1
EC:3.1.3.5	5'-nucleotidase	DN42284_c1_g2_i1
EC:3.2.2.8	Ribosylpyrimidine nucleosidase; Adenosine nucleosidase; Inosine nucleosidase; Purine nucleosidase; Uridine nucleosidase	DN44821_c0_g1_i2
EC:3.5.1.5	Urease	DN47269_c1_g1_i2
EC:3.5.3.26	(S)-ureidoglycine aminohydrolase	DN46681_c0_g2_i1
EC:3.5.3.9	Allantoate deiminase	DN44617_c0_g1_i1
EC:3.6.1.13	ADP-ribose diphosphatase; Nucleotide diphosphatase	DN37728_c0_g1_i1
EC:3.6.1.5	Apyrase; Nucleoside diphosphate phosphatase	DN49219_c1_g1_i1
EC:3.6.1.6	Nucleoside diphosphate phosphatase	DN43603_c0_g1_i1
EC:3.6.1.9	Nucleotide diphosphatase; Guanosine-3',5'-bis(diphosphate) 3'-diphosphatase	DN40418_c0_g1_i1
EC:4.1.1.21	Phosphoribosylaminoimidazole carboxylase	DN50486_c0_g1_i1
EC:4.6.1.1	Adenylate cyclase	DN32415_c0_g2_i1
EC:4.6.1.2	Guanylate cyclase	DN45942_c0_g1_i1
EC:5.4.2.2	Phosphoglucomutase (alpha-D-glucose-1,6-bisphosphate-dependent)	DN26548_c0_g1_i1
EC:6.3.2.6	Phosphoribosylaminoimidazolesuccinocarboxamide synthase	DN50546_c0_g3_i1
EC:6.3.3.1	Phosphoribosylformylglycinamidine cyclo-ligase	DN47361_c0_g1_i2
EC:6.3.4.4	Adenylosuccinate synthase	DN44785_c0_g1_i1
EC:6.3.5.2	GMP synthase (glutamine-hydrolyzing)	DN45265_c0_g1_i3
EC:2.7.7.27	Glucose-1-phosphate adenylyltransferase	DN20458_c0_g1_i1
EC:3.2.2	Glycosylases	DN13756_c0_g1_i1
	Map00900 - Biossíntese do esqueleto de terpenos	
EC:2.2.1.7	1-deoxy-D-xylulose-5-phosphate synthase	DN34606_c0_g1_i1
EC:1.1.1.267	1-deoxy-D-xylulose-5-phosphate reductoisomerase	DN46714_c1_g1_i1

EC:4.6.1.12	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	DN38205_c0_g1_i1
EC:2.7.1.148	4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase	DN46628_c0_g1_i2
EC:2.3.3.10	Hydroxymethylglutaryl-CoA synthase	DN17612_c0_g1_i1
EC:1.1.1.34	Hydroxymethylglutaryl-CoA reductase (NADPH)	DN27565_c0_g1_i1
EC:2.7.1.36	Mevalonate kinase	DN46537_c0_g1_i1
EC:2.7.4.2	Phosphomevalonate kinase	DN49445_c0_g2_i1
EC:2.7.4.26	Isopentenyl phosphate kinase	DN42983_c0_g1_i1
EC:5.3.3.2	Isopentenyl-diphosphate Delta-isomerase	DN40052_c3_g1_i1
EC:2.5.1.1	Dimethylallyltranstransferase	DN48171_c0_g1_i1
EC:2.5.1.10	(2E,6E)-farnesyl diphosphate synthase	DN28631_c0_g1_i1
EC:2.5.1.29	Geranylgeranyl diphosphate synthase; 15-cis-phytoene synthase; (2E,6E)-farnesyl diphosphate synthase	DN19040_c0_g1_i1
EC:1.3.1.83	Geranylgeranyl diphosphate reductase	DN44762_c0_g1_i1
EC:2.5.1.58	Protein farnesyltransferase	DN44890_c0_g1_i1
EC:2.5.1.59	Protein geranylgeranyltransferase type I; Protein farnesyltransferase	DN43654_c0_g1_i1
EC:1.8.3.5	Prenylcysteine oxidase; Farnesylcysteine lyase	DN50894_c0_g1_i1
Map00999 - Biossíntese de variados metabólitos secundários de plantas		
EC:1.21.3.7	Tetrahydrocannabinolic acid synthase	DN41420_c0_g1_i1
Map00909 - Biossíntese de sesquiterpenos e triterpenos		
EC:4.2.3.57	(-)-beta-caryophyllene synthase; Alpha-humulene synthase	DN36769_c1_g1_i1

7 Anexos - Trabalhos completos publicados ou aceitos para publicação

7.1 SANTONI, M. M.; DE LIMA, J. V. F.; BICALHO, K. U.; DE SOUZA MOREIRA, T. M.; VALENTINI, S. R.; FURLAN, M.; ZANELLI, C. F. Comparative Transcriptome Profiling of *Maytenus ilicifolia* Root and Leaf. Lecture Notes in Computer Science [s. l.], p. 3–14, 2021.

Comparative Transcriptome Profiling of *Maytenus ilicifolia* Root and Leaf

Mariana Marchi Santoni[1(✉)], João Vítor Félix de Lima[1], Keylla Utherdyany Bicalho[2], Tatiana Maria de Souza Moreira[1], Sandro Roberto Valentini[1], Maysa Furlan[3], and Cleslei Fernando Zanelli[2]

[1] Department of Biological Sciences, School of Pharmaceutical Sciences, São Paulo State University (UNESP), Araraquara, SP 14800-903, Brazil
mariana.santoni@unesp.br

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

[3] Department of Organic Chemistry, Institute of Chemistry, São Paulo State University (UNESP), Araraquara, SP 14800-900, Brazil

Abstract. Plants produce a wide variety of compounds called secondary metabolites (SMs), which are extremely important for their survival. SMs have also medicinal applications, but as chemical synthesis is not economically viable, plant extraction is the mainly option. Different biotechnology strategies are applied to improve the yield of bioproduction of these compounds, but commonly without the desired results due the limited knowledge of biosynthetic and regulatory pathways. *Maytenus ilicifolia*, a traditional Brazilian medicinal plant from Celastraceae family, produces in both root and leaves three main classes of SMs: sesquiterpenics, flavonoids and quinonemethides. In this study, four cDNA libraries were prepared from root and leaf tissues. The *de novo* transcriptome included 109,982 sequences that capture 92% of BUSCO orthologs, presented an average length of 737bp and a GC content about 42% of. Function annotation analysis identified homology for 44.8% of the transcripts. Moreover, 67,625 sequences were commonly expressed in both tissues, while 1,044 and 1,171 were differentially expressed in root and leaf, respectively. In terms of SM, enzymes involved in "monoterpenoid biosynthesis" and "isoflavonoid biosynthesis" were identified in root while "flavonoid biosynthesis" and "Biosynthesis of alkaloids" in leaf.

Keywords: RNA-Seq · *De novo* assembly · Metabolic pathways

1 Introduction

Compounds produced by plants are categorized into primary and secondary metabolites (SMs). Primary metabolites, such as carbohydrates, lipids, and proteins, are involved in plant development [9] and essential for cell growth [17]. In

Supported by São Paulo Research Foundation (FAPESP) [2013/07600-3; 2016/16970-7]; National Council for Scientific and Technological Development (CNPq) [303757/2017-5]; National Institute for Science and Technology (INCT).

P. F. Stadler et al. (Eds.): BSB 2021, LNBI 13063, pp. 3–14, 2021.
https://doi.org/10.1007/978-3-030-91814-9_1

contrast, SMs (low molecular weight compounds) are multifunctional metabolites produced as an evolutionary adaptation [4,23]. They are involved in plant defense and environmental communication [4,8,23], plant color, taste, and scent [9] and responses to biotic and abiotic stress [15,16,19].

The high variety of biological functions of the SMs is explained by diversified chemical structure [4,24] originated from a restricted and distinct number of metabolic pathways such as the acetate, shikimic acid, mevalonic acid or methylerythritol phosphate pathways [9,21]. SM are grouped in three classes: terpenes, alkaloids and phenylpropanoids, each one with its respective and unique properties [4,24]. These compounds are identified in all plant tissues and their formation and gene regulation is usually organ, tissue, cell and also development specific, indicating that a range of transcription factors must cooperate to transcribe secondary metabolism genes, controlling the general machinery of biosynthetic pathways in production, transport and storage [18,22].

Many SMs are sources of drugs however, as chemical synthesis is uneconomical, isolation from plants still represents the only option [4,13]. Different biotechnological strategies have been applied to improve the production of these compounds, but often without the desired results due to the lack of knowledge about the biosynthetic routes [13,21]. Biotechnology techniques such as transcriptome, proteome or metabolomics are used to identify genes and their functions in plant metabolic pathways in order to clarify the mechanisms involved in SMs synthesis [4].

Maytenus ilicifolia Mart ex Reissek (Celastraceae) is a Brazilian native plant known for its variety of therapeutic properties. It has been used as a treatment of several diseases such as gastric ulcer, dyspepsia, stomach acidity, diabetes and cancer [12,13,20]. This species produces three main classes of bioactive compounds: alkaloids sesquiterpene pyridines, flavonoids and quinonemethide triterpenes [13] and the mainly products are maitenin, friedelin, fridelanol, pristimerine and terpenes [14]. Additionally, like other members of Celastraceae family, some compounds are synthetized in a specific tissue: quinone methide triterpenoids are accumulated in root bark [1,13] and flavonoids in leaves [2].

The analysis of differentially expressed transcripts between two tissues can provide a better understanding of genes involved in secondary metabolic pathways [3,10,11]. In this context, the aim of the present study was to analyze whole transcriptome of *M. ilicifolia* and identify genes involved in biosynthesis of SMs by a comparative profiling of root and leaf. This study is the first report of high-throughput analysis (*de novo* RNA-Seq) of *M. ilicifolia transcriptome* that provides new insights at molecular knowledge.

2 Methods

2.1 Plant Material and Total RNA Isolation

Leaves of adult specimen of *M. ilicifolia* from the medicinal plant garden of the Faculty of Pharmaceutical Sciences and leaves and roots of identified seedlings, with approximately 6 months of planting, were harvested and stored in −80 °C

(Fig. 1A). The total RNA from two specimens of roots (from two seedlings) and two specimens of leaves (one leaf from seedling, coinciding with one of the specimens used for root extraction and one leaf from adult specimen) was isolated from 500 mg of material using RNeasy Plant mini kit (Qiagen, USA) according to the manufacturer's protocol. RNA quantity and quality were evaluated using Nanodrop 1000 spectrophotometer and Agilent 2100 Bioanalyzer. RNA samples with quality ratios greater than 1.8 (260/280 nm and 28S/18S) and RNA integrity number (RIN) greater than 7 were selected for subsequent processes.

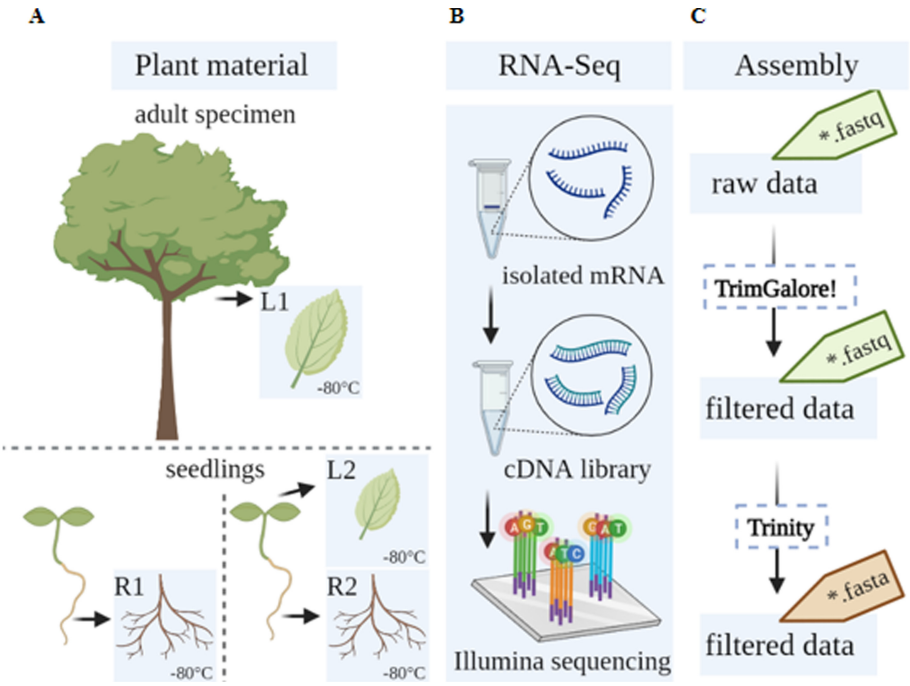

Fig. 1. Experimental approaches for *Maytenus ilicifolia* transcriptome study. **A.** Two samples of leaves (one leaf from adult specimen and one leaf from seedling) - L1 and L2 - and two samples of roots from two independent seedlings (one root sample coinciding with the same specimen of the leaf sample) - R1 and R2 - were collected and stored in −80 °C for posterior RNA isolation. **B.** Library preparation and transcriptome sequencing **C.** Pipeline used for *de novo* assembly.

2.2 Library Preparation and Sequencing

After isolated from total RNA with magnetic Oligo (dT) particles, mRNA was chemically fragmented. Subsequently, cDNA libraries were prepared using Illumina TruSeq RNA sample preparation v3 kit (Illumina, USA) (Fig. 1B). Quantification and quality assessment of resulting libraries were performed on Agilent 2100 Bioanalyzer. A total of 20 pmol of the libraries was submitted to

"single-read" sequencing in HiSeq 2000 platform (leaf of the adult specimen) - FCAV/Unesp - to generate 100bp reads or sequencing in MiSeq equipment (leaf and roots of seedlings) - LAB Multi-FCFAR/Unesp - to generate 75bp paired-end reads (Table 1).

Table 1. RNA-Seq traits of four *Maytenus ilicifolia* (*same specimen).

Sample ID	Plant material	Seq platform	Protocol	Read length	No. of reads
L1	Leaf (adult)	HiSeq 2000	Single read	100	46,798,882
L2*	Leaf (seedling)	MiSeq	Paired-end	75	19,062,549
R1	Root (seedling)	MiSeq	Paired-end	75	24,435,760
R2*	Root (seedling)	MiSeq	Paired-end	75	25,083,492

2.3 Quality Control and *de novo* Assembly

The public server Galaxy (usegalaxy.org) was used to process the high-throughput data. The raw data generated by the sequencing, FASTQ files, were evaluated by the FastQC tool (v0.11.8) for quality before and after filtering and for GC content. Reads were filtered by TrimGalore! (v0.6.3), removing adapter contamination and low-quality sequences (average quality below 25). Initial and final bases were also removed from sequences with "q" value lower than 25 and, finally, in the final FASTQ file of filtered reads, those with a size greater than 50 base pairs remained.

The high-quality data of roots and leaves samples was assembled using Trinity (v2.9.1) on default parameters. The *de novo* assembly was evaluated by different quality metrics including N50 length and BUSCO v4.1.2 analysis using OrthoDB v10 'embryophyta' database as a reference to access the assembly and annotation completeness. Filtered reads were remapped to the assembled transcriptome in order to obtain, using Salmon tool, an expression matrix reported in transcripts per kilobase million (TPM). This matrix allowed the filtering of transcripts by low expression, considering only those with at minimum 1% of dominant isoform expression, generating the filtered transcriptome.

2.4 Functional Annotation

TransDecoder tool was used to find the probable coding regions of transcripts and the open reading frames (ORFs) with a minimum length of 100 amino acids. Then, functional annotation of the transcripts was performed using BLASTX against Uni-ProtKB/SwissProt databases and uniprot _trEMBL _plants database (E-value <1e−5). Moreover, a homology search based on the BLASTP was performed using the predicted proteins as query against UniProtKB/SwissProt databases (E-value <1e−5). The assignments of Gene Ontology (GO) terms to transcripts were performed based on UniProtKB/SwissProt database to assign unigenes to functional categories. Additionally, the proteins

with Enzyme Commission (EC) numbers were mapped onto the Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway Database using online KEGG Automatic Annotation Server (www.genome.jp/kegg/kaas) to assign pathway information to the transcripts.

2.5 Differential Expression Analysis

Salmon tool was applied to estimate the expression level of transcripts. Each filtered FASTQ file was separately aligned to the filtered transcriptome. Then, the expression level of each transcript was normalized and reported in TPM. To summarize the results and provide statistical tests for tissue comparison, the differential expression analysis was performed using DESeq2 R package and transcript expression difference was considered significant when the adjusted p-value < 0.05.

2.6 Gene Ontology Enrichment and KEGG Analysis

Gene ontology (GO) enrichment analysis for biological process (BP) and molecular function (MF) for the differentially expressed transcripts in each tissue was conducted using topGO R package. Significant GO terms (Fisher's exact test p-value < 0.01) were visualized using REViGO (revigo.irb.hr) for semantic space reduction. Transcripts associated with Enzyme Commission (EC) numbers were mapped onto the KEGG pathway database.

3 Results and Discussion

3.1 *De novo* Assembly and Functional Annotation of *M. ilicifolia*

The single-read leaf cDNA library and the paired-end leaf and root cDNA libraries subjected to full transcriptome sequencing generated about 115 million of raw reads. The detailed information of the read numbers in different samples is provided in Table 1.

High quality sequencing data, 112,609,211 reads, was used for assembly. The de novo transcriptome generated included 163,780 transcripts (isoforms) with a GC content of 41.8% and the N50 resulting in 1,222bp. The average transcript size was 737 and 22% of them presented more than 1,000bp (Fig. 2A). By considering transcript expression, 15,704 transcripts represented 90% of the total expression data (Ex90) and had an N50 of 1,487bp (Ex90N50). In addition, the assembled transcriptome of *M. ilicifolia* captured 92.6% of the 1,614 orthologs described for the Virdiplantae database (updated 2020-09-10): 53.0%, 39.6%, 4.2%, and 3.2% of the BUSCO genes were respectively classified as complete single copy, complete duplicate, fragmented and absent. After filtering by low expression, the final transcriptome included 109,982 sequences. These results indicate that the integrity of assembly was high, and the sequencing quality had met the requirements of further analysis.

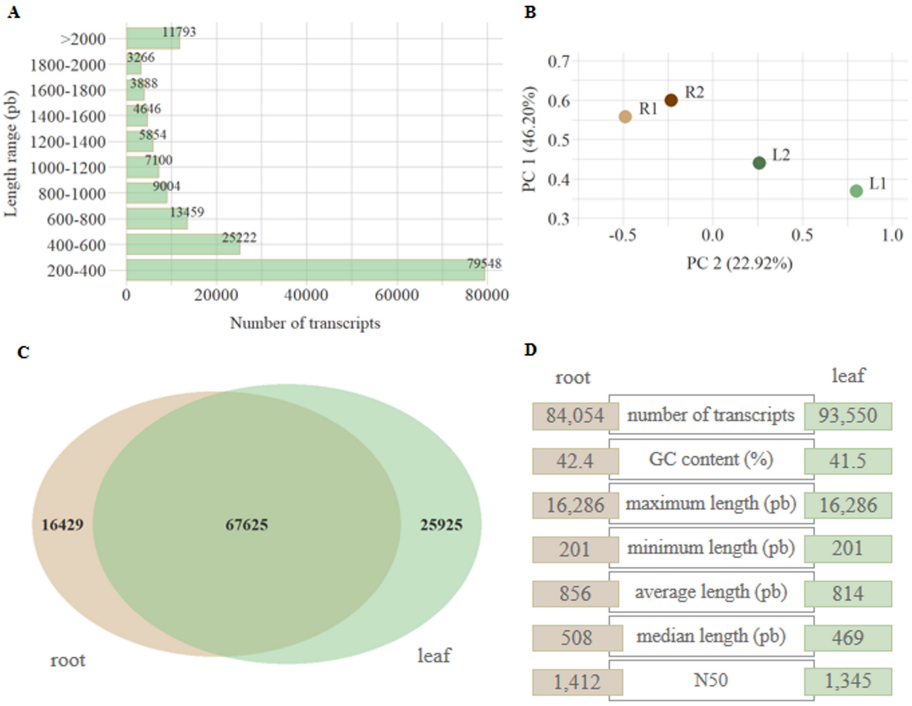

Fig. 2. Aspects of *Maytenus ilicifolia* transcriptome assembly. A. Size distribution of assembled data. **B.** Principal component analysis (PCA) on the read counts of root and leaf samples. (L1, L2, R1 and R2 - sample identification described in Fig. 1) **C.** Venn diagram showing the number of transcripts for each source of sample **D.** transcriptome traits for each tissue.

Results of PCA analysis revealed the distinct differences in transcript expression patterns among the samples. The first two principal components contain 69.12% of the information grouping different tissues in separate clusters (Fig. 2B). Considering transcripts identified in leaf and root individually, 67,625 isoforms were found in both tissues (Fig. 2C) and showed similar aspects in respect to transcriptome traits (Fig. 2D).

In summary, M. ilicifolia transcriptome had GC content close to 40%, similar values to those reported for Celastraceae family species like staff vine (41.5%) [20] and thunder god vine (37.2%) [22]. Moreover, results of BUSCO analysis captured more than 90% of the orthologs described for the chosen database and the PCA results allowed the confirmation of expression differences in both tissues, root and leaf.

The BLASTX against the uniprot_trEMBL _plants database found 36,625 alignments and revealed that *M. ilicifolia* predicted transcripts have highest similarity with an organism classified in the same family, Tripterygium sp (47.3%) (Fig. 3A), but homology was find for other family organisms (Fig. 3B). Candi-

date coding regions in *M. ilicifolia* transcriptome were identified by TransDecoder and 65,533 ORFs and 46,282 probable coding sequences were predicted. Sequence homology search results against the UniprotKB/SwissProt database by BLASTX (E-value <1e−5, for filtered transcripts) and BLASTP (E-value <1e−5, for predicted protein sequences) were 49,319 (44.8%) and 36,344 (55.5%) aligned transcripts, respectively.

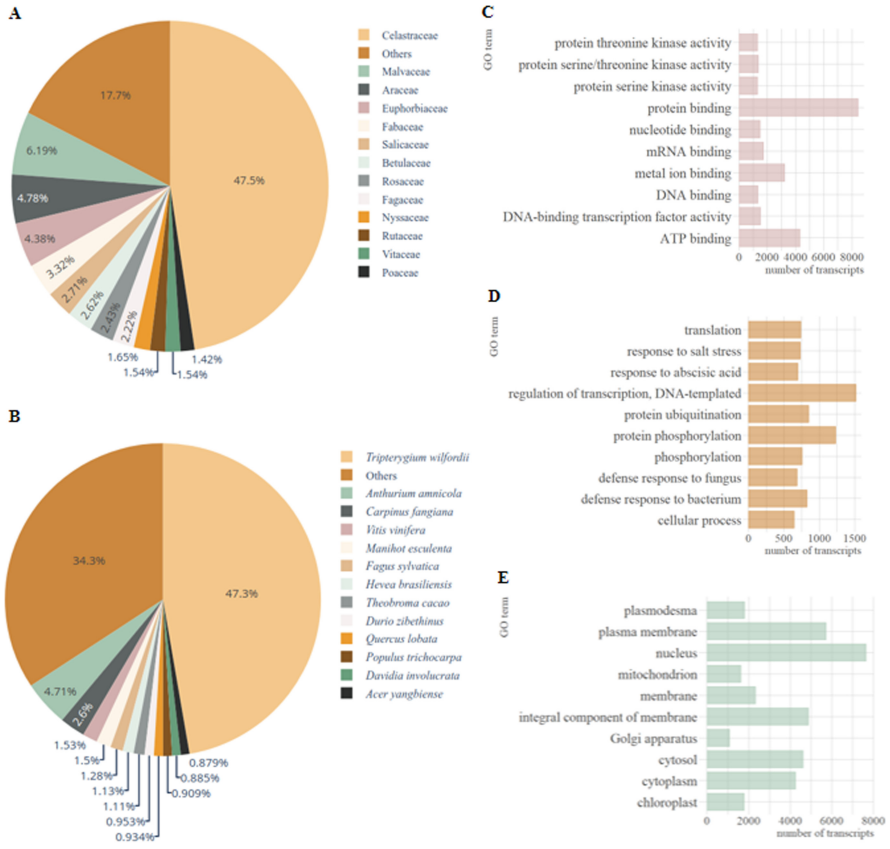

Fig. 3. Functional annotation for *Maytenus ilicifolia* transcriptome. Similarity frequency distribution of different **A.** families and **B.** species. The BLASTx was performed against the trEMBL plants database. Top ten GO terms in the transcriptome assembly from **C.** molecular function, **D.** biological process and

Functional annotation for filtered transcriptome was followed by GO analysis and 43,322 annotated transcripts were categorized into 9,989 GO IDs. The number of transcripts in three main categories of molecular function (MF), biological process (BP) and cellular component (CC) was 41,148, 39,404 and 39,582, respectively. The most dominant GO terms in the MF category were “protein binding,”

"ATP binding" and "metal iron binding" (Fig. 3C). In the BP category, "regulation of transcription", "protein phosphorylation" and "protein ubiquitination" were the most prominent (Fig. 3D). In the CC category, "nucleus", "plasma membrane" and "integral component of membrane" were the most abundant terms (Fig. 3E).

KEGG annotation analysis was performed to identify active metabolic processes in *M. ilicifolia* transcriptome. In conclusion, 2,326 transcripts were assigned to 428 KEGG pathways. Considering "Metabolism of terpenoids and polyketides", the most representative pathway was "Terpenoid backbone biosynthesis (ko00900)", followed by "Sesquiterpenoid and triterpenoid biosynthesis (ko00909)", with 216 and 127 mapped sequences that represent 50% and 10% of the orthologous for each pathway, respectively.

In conclusion, the identification of about 40,000 protein accessions indicates that in this study the de novo RNA-Seq and assembly could generate substantial information about *M. ilicifolia* genes. The functional annotation of transcripts covered a broad range of GO categories and KEGG allowed the identification of transcripts involved in biosynthesis of triterpenoid backbone, as expected for this species.

3.2 Identification of Differentially Expressed Transcripts in Both Tissues

Comparative transcript abundance level revealed significant differential expression of 2,215 transcripts (FDR <0.05) between the transcriptome of both tissues. Levels of expression were represented as $\log_2$ ratio of transcripts abundance between leaf and root samples (Fig. 4A), showing the 1,044 differentially expressed transcripts in root and 1,171 in leaf. Working on both tissues, it was observed that a number of transcripts was expressed uniquely in either of the tissues: among differentially expressed transcripts, 424 were exclusively expressed in leaf and 298 in root.

To better characterize the tissue-biased transcriptome profile, topGO package were used to evaluated GO enrichment (p-value <0.01) for the differentially expressed transcripts and further the representative terms were summarized upon removal of redundant using REVIGO. Among the 770 roots differentially expressed annotated transcripts, 568 genes were assigned to 260 GO terms, while in the leaf, from the 902 differentially expressed annotated transcripts, 610 were classified in 265 GO terms.

The GO analysis revealed enrichment for biological processes (BP) in root for "response to ethylene", "regulation of cellular process" and others (Fig. 3B), while in leaf for "photosynthesis", "protein-chromophore linkage" and others (Fig. 3C). According to functional analysis terms, leaves and roots of *M. illicifoia* also differ at levels of molecular function (MF), with transcripts overexpressed in roots being mainly associated with "calcium ion binding", "iron ion binding" and others (Fig. 4B), while the overexpressed leaf transcripts are associated with "oxidoreductase activity", "chlorophyll binding" and others (Fig. 4C).

Significant GO terms linked to secondary methabolism were found in 295 differentially expressed transcripts, 164 in root and 131 in leaf. Some terms were found enriched in specific tissue, for example, "2-oxoglutarate-dependent dioxygenase activity" and "response to herbivore" in root and "beta-amyrin synthase activity" and "triterpenoid biosynthetic process" in leaf. Coincident terms like "oxidoreductase activity" were observed in overexpressed transcripts from both tissues (Table 2).

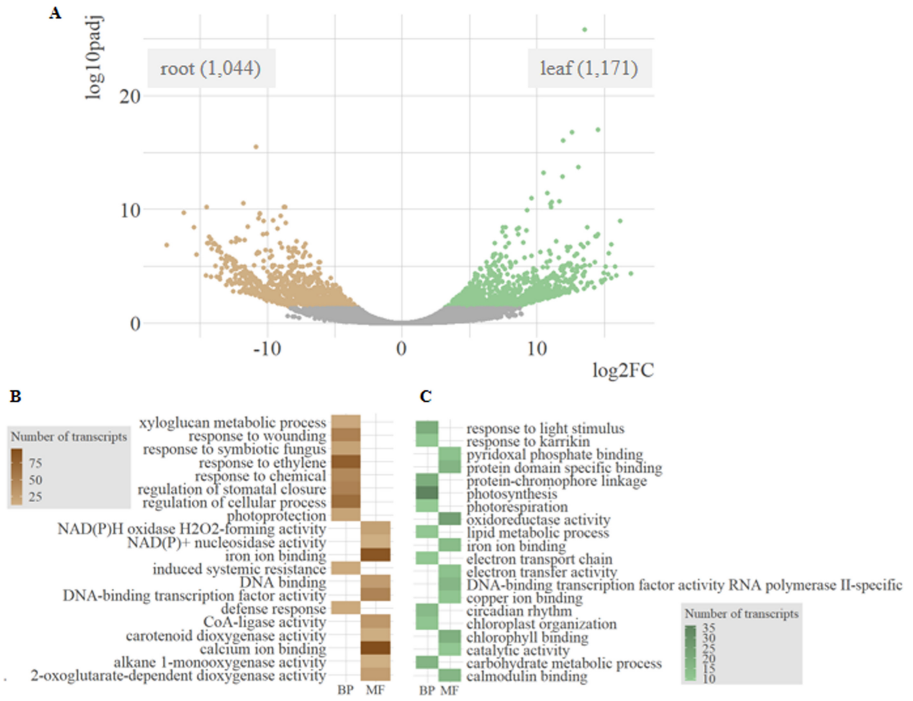

Fig. 4. Gene expression differences between root and leaf tissues of *Maytenus ilicifolia*. A. The values of -$\log_{10}$ adjusted p-value were plotted according to the differential expression between root and leaf ($\log_2$ fold change). Differentially expressed root transcripts are high-lighted in brown (left) and differentially expressed leaf transcripts, in green (right). Top ten most represented terms of gene ontology enrichment analysis in biological process (BP) and molecular function (MF) for differentially expressed transcripts for **B.** root and **C.** and leaf (Color figure online)

The comparative transcriptome analysis led to the identification of 350 and 487 transcripts associated with Enzyme Commission (EC) numbers in root and leaf, respectively. These tissue-biased transcripts were mapped onto the KEGG pathway database for the "Biosynthesis of plant secondary metabolites map" (ko01060) and related pathways. Enzymes involved in "monoterpenoid biosynthesis" and isoflavonoid biosynthesis" were identified in root overexpressed tran-

scripts while “flavonoid biosynthesis” and “Biosynthesis of alkaloids derived from histidine and purine” in leaf (Table 3).

Taking together, the results of GO enrichment analysis and KEGG mapping of transcripts overexpressed in root or leaf of *M. ilicifolia* confirmed the well-

Table 2. Number of transcripts overexpressed in *Maytenus ilicifolia* root or leaf characterized according to enriched GO terms involved in secondary metabolism.

GO ID	GO term	Root	Leaf
GO:0016706 (MF)	2-oxoglutarate-dependent dioxygenase activity	26	–
GO:0080027 (BP)	Response to herbivore	15	–
GO:0016491 (MF)	Oxidoreductase activity	7	25
GO:0016709 (MF)	Oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, NAD(P)H as one donor, and incorporation of one atom of oxygen	6	2
GO:0019742 (BP)	Pentacyclic triterpenoid metabolic process	3	–
GO:0016106 (BP)	Sesquiterpenoid biosynthetic process	3	–
GO:0042300 (MF)	Beta-amyrin synthase activity	–	6
GO:0016104 (BP)	Triterpenoid biosynthetic process	–	5

Table 3. Enzymes mapped KEGG pathways identified in the comparative transcriptome analyses of root and leaf of *Maytenus ilicifolia*.

Root		
map00902	Monoterpenoid biosynthesis	EC:2.1.1.50
map00943	Isoflavonoid biosynthesis	EC:2.1.1.46
Leaf		
map00230	Purine metabolism	EC:2.4.2.14 EC:2.7.6.5 EC:3.5.2.5
map00941	Flavonoid biosynthesis	EC:2.3.1.133
map00950	Isoquinoline alkaloid biosynthesis	EC:1.4.3.21 EC:2.6.1.5
map00960	Tropane, piperidine and pyridine alkaloid biosynthesis	EC:1.4.3.21 EC:2.6.1.5
map01064	Biosynthesis of alkaloids derived from ornithine, lysine and nicotinic acid	EC:1.4.3.21 EC:2.6.1.5
map01065	Biosynthesis of alkaloids derived from histidine and purine	EC:4.1.2.13 EC:1.2.1.9 EC:2.4.2.14 EC:2.7.6.5 EC:5.1.3.1 EC:1.1.1.49 EC:2.2.1.1 EC:3.5.2.5

reported SMs accumulation reveled by other methodological procedures, including flavonoids, triterpenes, and sesquiterpenes in leaves [2], while roots contain terpenes, triterpenes, alkaloids and especially the quinonemethide triterpenes [5,13,14].

Finally, from the present study, an extensive transcriptome dataset has been generated from *de novo* sequencing analyses of *M. ilicifolia*. The coverage of the transcriptome data is consistent to discover genes involved in the secondary metabolic pathways. Therefore, choosing the root and the leaf for comparative transcriptome analysis facilitated the identification of the genes involved in the organ-specific biosynthesis, an approach widely used for mining and identifying novel genes in biosynthesis of SMs in plants[3,6,7,18,25,26].

References

1. Coppede, J.S., et al.: Cell cultures of Maytenus ilicifolia Mart. Are richer sources of quinone-methide triterpenoids than plant roots in natura. Plant Cell Tissue Organ Cult. (PCTOC) **118**(1), 33–43 (2014). 10/f56kq9
2. De Souza, L.M., Cipriani, T.R., Iacomini, M., Gorin, P.A.J., Sassaki, G.L.: HPLC/ESI-MS and NMR analysis of flavonoids and tannins in bioactive extract from leaves of Maytenus ilicifolia. J. Pharm. Biomed. Anal. **47**(1), 59–67 (2008). 10/c4vp7v
3. Devi, K., Mishra, S.K., Sahu, J., Panda, D., Modi, M.K., Sen, P.: Genome wide transcriptome profiling reveals differential gene expression in secondary metabolite pathway of Cymbopogon winterianus OPEN, **6**(21026), 1–11 (2016). 10/f79vzf. Nature Publishing Group
4. Dziggel, C., Schäfer, H., Wink, M.: Tools of pathway reconstruction and production of economically relevant plant secondary metabolites in recombinant microorganisms. Biotechnol. J. **12**(1), 1–14 (2017). 10/f3tn87
5. Filho, W.B., Corsino, J., Bolzani, V.d.S., Furlan, M., Pereira, A.M.S., França, S.C.: Quantitative determination of cytotoxicFriedo-nor-oleanane derivatives from five morphological types of Maytenus ilicifolia (celastraceae) by reverse-phase high-performance liquid chromatography. Phytochem. Anal. Int. J. Plant Chem. Biochem. Tech. **13**(2), 75–78 (2002). https://doi.org/10.1002/PCA.626
6. Guo, D., Kang, K., Wang, P., Li, M., Huang, X.: Transcriptome profiling of spike provides expression features of genes related to terpene biosynthesis in lavender. Sci. Rep. **10**(1), 1–13 (2020). https://doi.org/10.1038/s41598-020-63950-4
7. Hansen, N.L., et al.: The terpene synthase gene family in Tripterygium wilfordii harbors a labdane-type diterpene synthase among the monoterpene synthase TPS-b subfamily. Plant J. **89**(3), 429–441 (2017). 10/f9qghw
8. Hartmann, T.: 10/bvxmg2
9. Jan, R., Asaf, S., Numan, M., Lubna, Kim, K.M.: Plant secondary metabolite biosynthesis and transcriptional regulation in response to biotic and abiotic stress conditions. Agronomy **11**(5), 1–31 (2021). 10/gmk7dd
10. Li, W., et al.: De novo leaf and root transcriptome analysis to explore biosynthetic pathway of Celangulin v in Celastrus angulatus maxim. BMC Genomics **20**(1), 1–15 (2019). 10/gz2c
11. Liu, M.H., et al.: Transcriptome analysis of leaves, roots and flowers of Panax notoginseng identifies genes involved in ginsenoside and alkaloid biosynthesis. BMC Genomics **16**(1), 1–12 (2015). 10/f69r7v

12. Mariot, M.P., Barbieri, R.L.: Metabólitos secundários e propriedades medicinais da espinheira-santa (Maytenus ilicifolia Mart. ex Reiss. e M. aquifolium Mart.). Revista Brasileira de Plantas Medicinais **9**(3), 89–99 (2007)
13. Paz, T.A., et al.: Proteome profiling reveals insights into secondary metabolism in Maytenus ilicifolia (Celastraceae) cell cultures producing quinonemethide triterpenes. Plant Cell Tissue Organ Cult. **130**(2), 405–416 (2017). 10/gbpzrf
14. Périco, L.L., Rodrigues, V.P., de Almeida, L.F.R., Fortuna-Perez, A.P., Vilegas, W., Hiruma-Lima, C.A.: Maytenus ilicifolia Mart. ex Reissek pp. 323–335 (2018). https://doi.org/10.1007/978-94-024-1552-0_29
15. Pradhan, J., Sahoo, S., Lalotra, S., Sarma, R.: Positive impact of abiotic stress on medicinal and aromatic plants. Int. J. Plant Sci. **12**(2), 309–313 (2017). 10/gz2g
16. Ramakrishna, A., Ravishankar, G.A.: Influence of abiotic stress signals on secondary metabolites in plants. Plant Signal. Behav. **6**(11), 1720–1731 (2011). 10/fx4rjw
17. Saddique, M., Kamran, M., Shahbaz, M.: Differential Responses of Plants to Biotic Stress and the Role of Metabolites. Elsevier Inc. (2018). 10/gz2h
18. Upadhyay, S., Phukan, U.J., Mishra, S., Shukla, R.K.: De novo leaf and root transcriptome analysis identified novel genes involved in Steroidal sapogenin biosynthesis in Asparagus racemosus. BMC Genomics **15(**1), 1–13 (2014). 10/gb3gr4
19. Van Loon, L.C., Rep, M., Pieterse, C.M.: Significance of inducible defense-related proteins in infected plants. Ann. Rev. Phytopathol. **44**, 135–162 (2006). 10/csvjsr
20. Vellosa, J.C., et al.: Antioxidant activity of Maytenus ilicifolia root bark. Fitoterapia **77**(3), 243–244 (2006). 10/dzm4t9
21. Wink, M.: Introduction: biochemistry, physiology and ecological functions of secondary metabolites. Biochem. Plant Second. Metab. Second Ed. **40**, 1–19 (2010). 10/b8sdms
22. Wink, M.: Secondary metabolites: deterring herbivores. In: Encyclopedia of Life Sciences, pp. 1–9, March 2010. 10/c65zd8
23. Wink, M., Schimmer, O.: Molecular modes of action of defensive secondary metabolites, vol. 39 (2010). 10/cpz4j7
24. Yang, L., Wen, K.S., Ruan, X., Zhao, Y.X., Wei, F., Wang, Q.: Response of plant secondary metabolites to environmental factors. Molecules **23**(4), 1–26 (2018). 10/gdrnqc
25. Younesi-Melerdi, E., Nematzadeh, G.A., Pakdin-Parizi, A., Bakhtiarizadeh, M.R., Motahari, S.A.: De novo RNA sequencing analysis of Aeluropus littoralis halophyte plant under salinity stress. Sci. Rep. **10**(1), 1–14 (2020). 10/gz2m
26. Zhang, C., Yao, X., Ren, H., Chang, J., Wang, K.: RNA-Seq reveals flavonoid biosynthesis-related genes in pecan (Carya illinoinensis) kernels. J. Agric. Food Chem. **67**, 148–158 (2018). 10.gz2n

7.2 BICALHO, K. U.; SANTONI, M. M.; ARENDT, P.; ZANELLI, C. F.; FURLAN, M.; GOOSSENS, A.; POLLIER, J. CYP712K4 Catalyzes the C-29 Oxidation of Friedelin in the *Maytenus ilicifolia* Quinone Methide Triterpenoid Biosynthesis Pathway. Plant and Cell Physiology, [s. l.], v. 60, n. 11, p. 2510–2522, 2019.

PCP
PLANT & CELL PHYSIOLOGY

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]Department of Plant Biotechnology and Bioinformatics, Ghent University, 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
*Corresponding authors: Jacob Pollier, E-mail, jacob.pollier@ugent.vib.be; Fax, +32 9 331 38 09; Alain Goossens, E-mail, alain.goossens@ugent.vib.be; Fax, +32 9 331 38 09.
(Received May 10, 2019; Accepted July 13, 2019)

The native Brazilian plant *Maytenus ilicifolia* accumulates a set of quinone methide triterpenoids with important pharmacological properties, of which maytenin, pristimerin and celastrol 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 (FRS). In this study, we produced heterologous friedelin by the expression of *M. ilicifolia* FRS in *Nicotiana benthamiana* leaves and in 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 discovery of the pathway 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 (Niero et al. 2011, Biral et al. 2017, 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, etc. (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 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 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 (Kutney et al. 1981, Corsino et al. 2000). 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, the 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 toward the biosynthesis of specialized triterpenoids in plants. This reaction is catalyzed by specific oxidosqualene cyclases and >100 different triterpene scaffolds have been described (Moses et al. 2013, Thimmappa et al. 2014, Miettinen et al. 2018). Similarly, the first committed step in the quinone methide triterpenoid biosynthesis is the

Plant Cell Physiol. 0(0): 1–13 (2019) doi:10.1093/pcp/pcz144, Advance Access publication on 26 July 2019, available online at https://academic.oup.com/pcp

cyclization of 2,3-oxidosqualene to friedelin, which is catalyzed by friedelin synthase (FRS, **Fig. 1A**). This oxidosqualene cyclase has been characterized in a few plant species, including *M. ilicifolia* (Souza-Moreira et al. 2016, Alves et al. 2018) 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 (Moses et al. 2013, Thimmappa et al. 2014, Seki et al. 2015, Miettinen et al. 2018). 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, multienzyme pathways can be reconstituted. *Nicotiana 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 et al. 2017, Reed and Osbourn 2018). 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. 2013, Moses et al. 2014). 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* FRS (*Mi*FRS) 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, *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

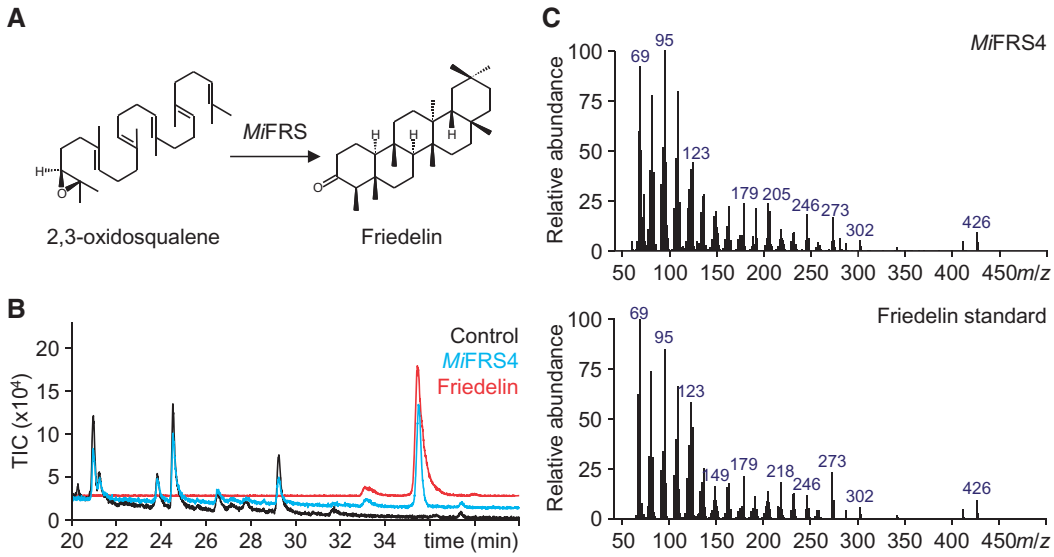

Fig. 1 Heterologous production of friedelin in *N. benthamiana* leaves by expression of *MiFRS4*. (A) Cyclization of 2,3-oxidosqualene by *Mi*FRS. (B) Overlay of the GC-MS chromatograms from a friedelin standard (red) and extracts of *N. benthamiana* leaves infiltrated with *tHMGR1* alone (Control, black)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

Plant Cell Physiol. 0(0): 1–13 (2019) doi:10.1093/pcp/pcz144

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 4 d postinfiltration were analyzed using gas chromatography-mass spectrometry (GC-MS). The GC-MS chromatogram of leaves infiltrated with *tHMGR1* and *MiFRS4* revealed a single unique peak eluting at 35.5 min compared with the GC-MS chromatogram of leaves infiltrated with *tHMGR1* only (**Fig. 1B**). The retention time and electron ionization (EI)-MS fragmentation spectrum of this peak corresponded well with those of an authentic friedelin standard (**Fig. 1B, C**), confirming the 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 ortholog(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 (2 × 76 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 (Supplementary Fig. 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 (**Fig. 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 (**Fig. 2B**). Furthermore, the MS^2 fragmentation spectra of the friedelin produced in *N. benthamiana* and the friedelin standard are nearly identical (**Fig. 2C**), confirming that the peak eluting at 24.1 min corresponds to friedelin.

Compared with 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 (**Fig. 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 (**Fig. 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 (**Fig. 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 (**Fig. 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

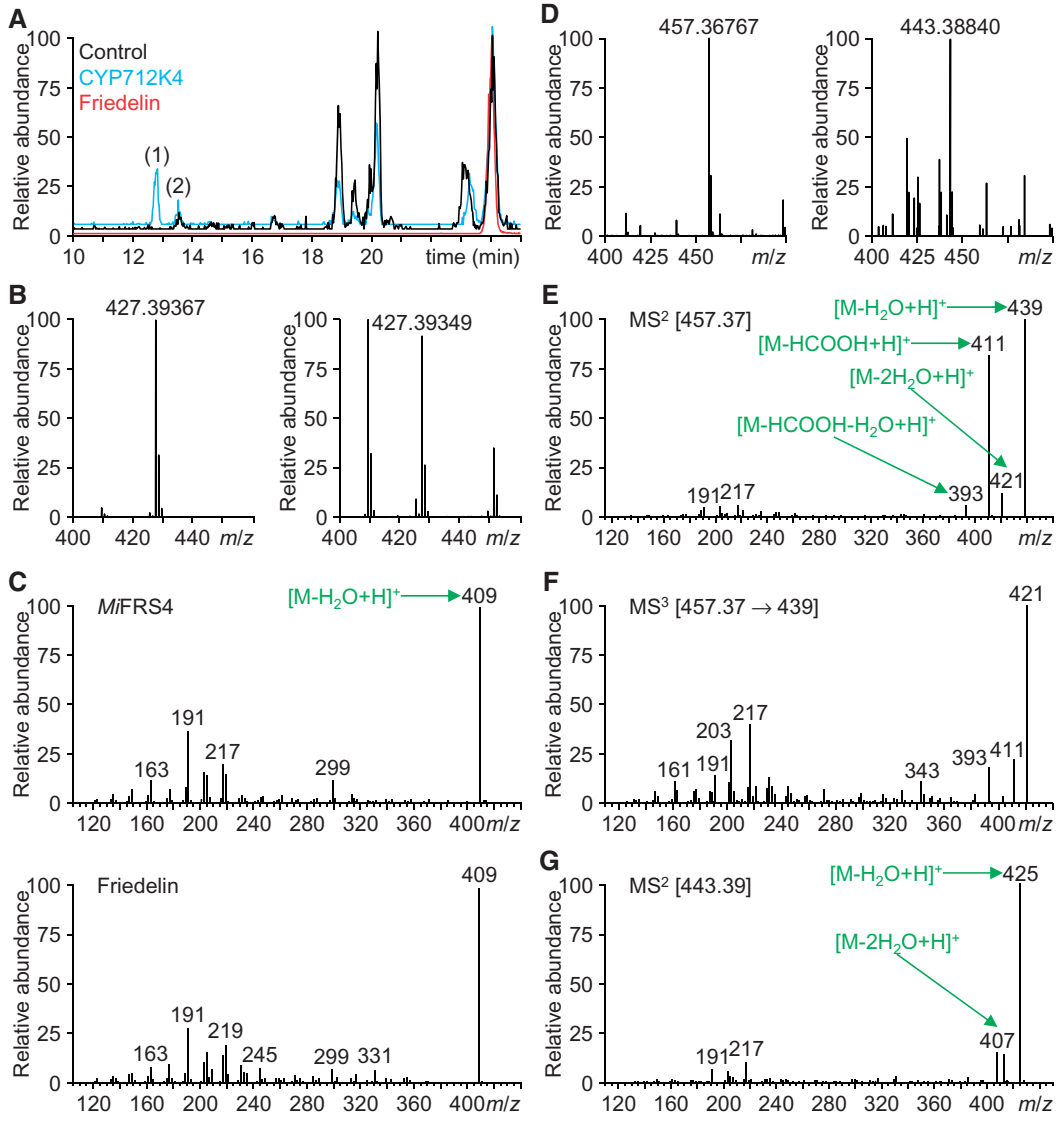

Fig. 2 Expression of *CYP712K4* in *N. benthamiana* leaves. (A) Overlay of the LC-APCI-FT-MS chromatograms from a friedelin standard (red) and extracts of *N. benthamiana* leaves infiltrated with *tHMGR1* and *MiFRS4* with (CYP712K4, blue) or without (Control, black) *CYP712K4*. The two peaks unique 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*.

0.90 ± 0.34 mg/l of friedelin (**Fig. 3A–C**). Compared with the production in its wild-type strain CEN.PK113-5D (0.30 ± 0.01 mg/l), friedelin production in strain KB1 was about 3-fold 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 (**Fig. 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 of constitutive expression (**Fig. 3D**), but the variation between individual colonies was considerably reduced and the overall production did not decrease over time.

Plant Cell Physiol. 0(0): 1–13 (2019) doi:10.1093/pcp/pcz144

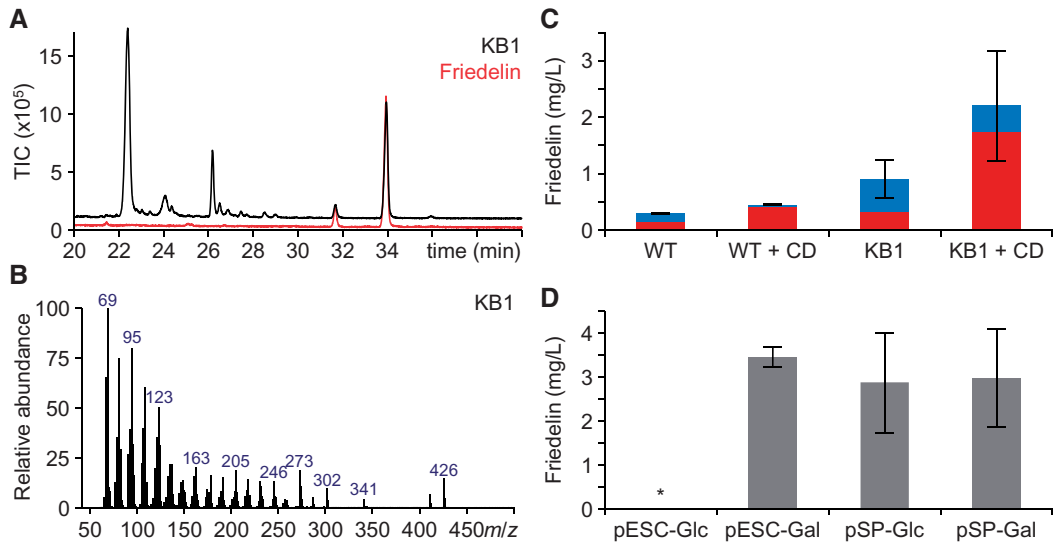

Fig. 3 Evaluation of the friedelin production in *S. cerevisiae*. (A) Overlay of the GC-MS chromatograms from an extract of yeast strain KB1 expressing *MiFRS* from the high-copy number plasmid pSP[P_{TEF1}-*MiFRS*;P_{PGK1}-*tHMG1*] (black) and a friedelin standard (red). (B) EI-MS spectrum of the friedelin produced in yeast strain KB1, which compares well with the EI-MS spectrum of the friedelin standard presented in **Fig. 1C**. (C) Friedelin production in yeast strains CEN.PK113-5D (WT) and KB1 in the presence (+CD) or absence of MβCD. 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.

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 (**Fig. 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 (**Fig. 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 (**Fig. 4C**), which revealed that the 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 with 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 (**Fig. 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 *M. 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 with 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 the extracts of the empty vector control strain (**Fig. 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* (**Fig. 5B–D**). The third peak, eluting at 14.8 min with an observed mass of 441.37258 Da (**Fig. 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 (**Fig. 5F**). These data confirm 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 alcohol and no detectable levels of aldehyde. By contrast, yeast produces significant amounts of both alcohol and aldehyde intermediates in addition to the acid of friedelin.

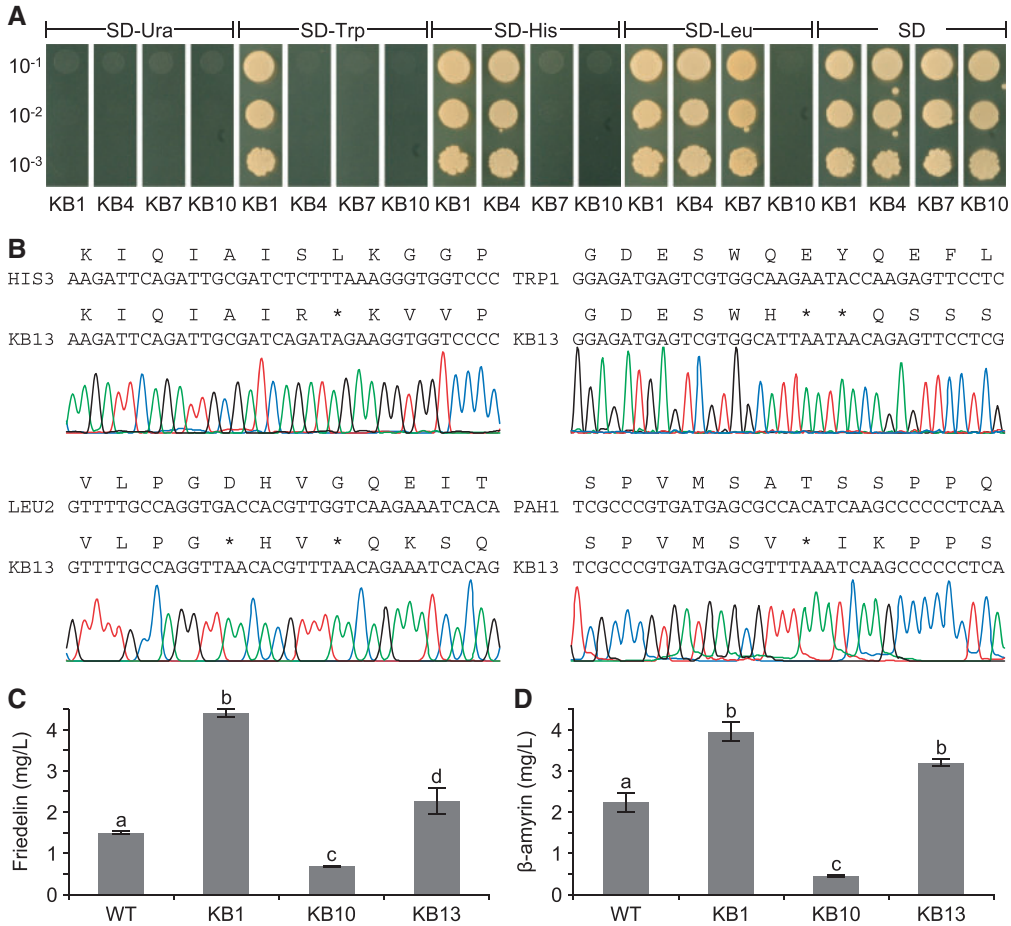

Fig. 4 Engineering of *S. cerevisiae* using CRISPR/Cas9. (A) Drop-tests confirming the auxotrophy of the generated strains. A serial dilution of yeast strains KB1, KB4, KB7 and KB10 was dropped on plates containing synthetic defined (SD) medium lacking uracil (-Ura), tryptophan (-Trp), histidine (-His) or leucine (-Leu). Yeast precultures were diluted 10-fold (10^{-1}), 100-fold (10^{-2}) or 1,000-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 with 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).

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*

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 using 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 (Supplementary Figs. 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. On the basis of this information and in analogy with, for instance, CYP716A12 (Carelli et al. 2011) or CYP72A154 (Seki et al. 2011), we concluded that the alcohol

Plant Cell Physiol. 0(0): 1–13 (2019) doi:10.1093/pcp/pcz144

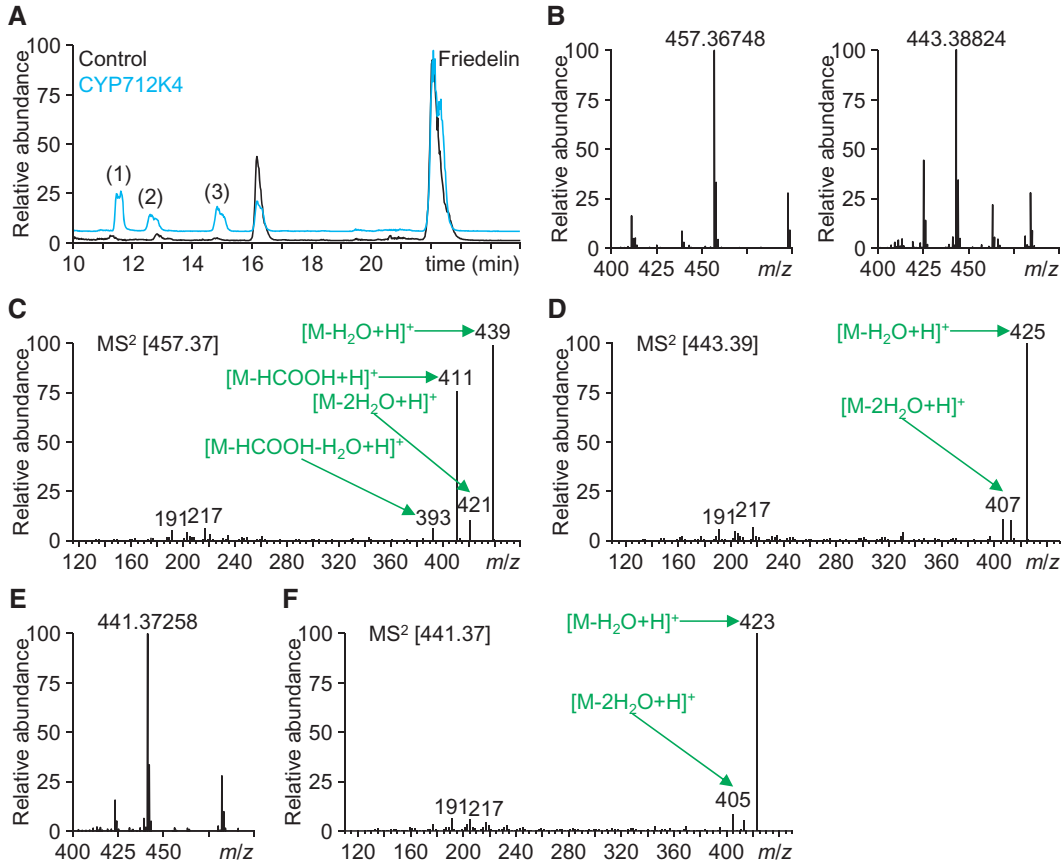

Fig. 5 Evaluation of CYP712K4 in *S. cerevisiae*. (A) Overlay of LC-APCI-FT-MS chromatograms from extracts of the spent medium of yeast strain KB13 expressing *tHMGR1*, *MTR1* and codon-optimized *MiFRS* with (CYP712K4, blue) or without (Control, black) *CYP712K4*. Three peaks unique to the extracts 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*.

and aldehyde intermediates correspond to 29-hydroxyfriedelin and 3-oxo-friedelan-29-al (**Fig. 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 (**Fig. 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 (**Fig. 1A**; Souza-Moreira et al. 2016, Alves et al. 2018). The subsequent steps in the biosynthesis of celastrol and ultimately maytenin and pristimerin remain speculative. A possible biosynthetic intermediate is cangoronine (**Fig. 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 (**Fig. 6B**), metabolites that have been isolated from plants producing quinone methide triterpenoids (Morota et al. 1995, Li et al. 1997). 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

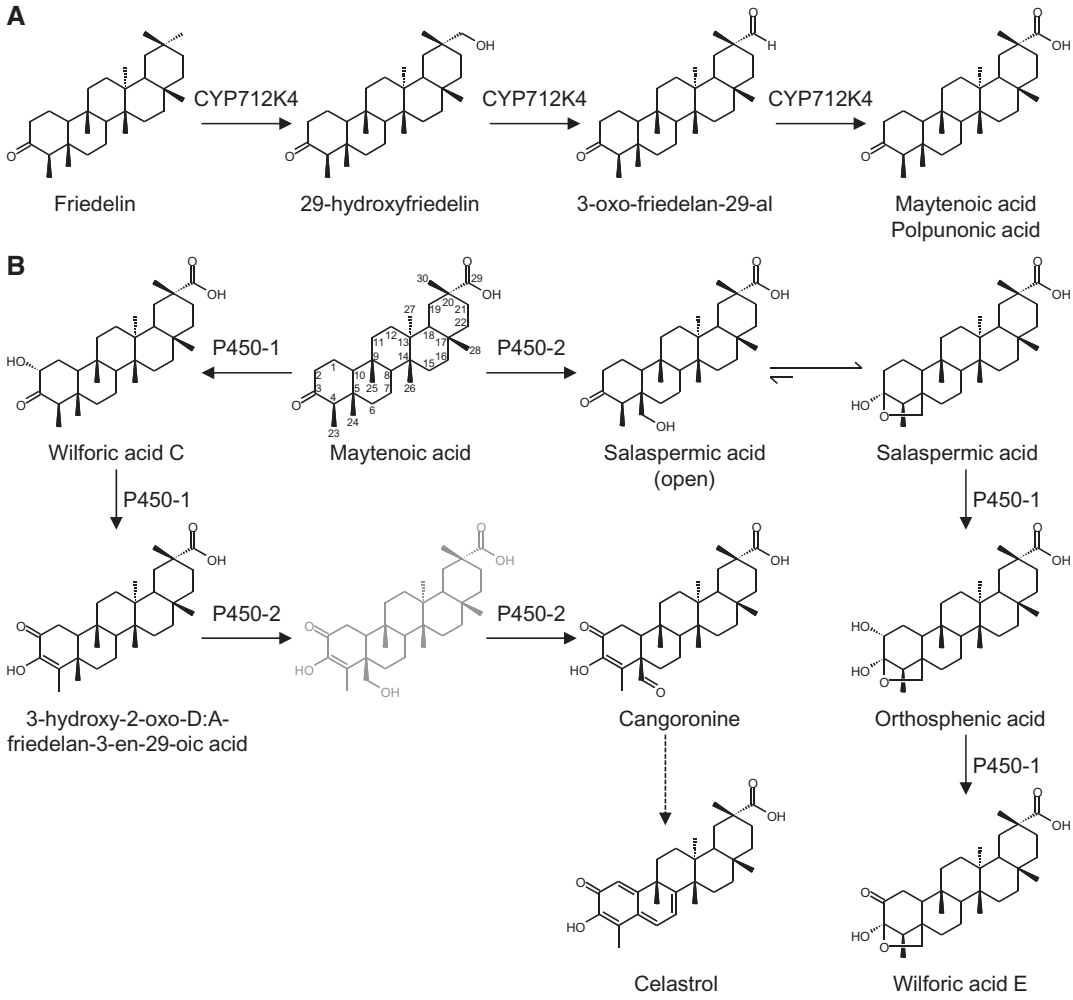

Fig. 6 Biosynthesis of quinone methide triterpenoids. (A) Three-step oxidation of friedelin catalyzed by CYP712K4. (B) Proposed biosynthetic pathway to cangoronine by the combined activity of two P450s. Apart from the biosynthetic intermediate in gray, all metabolites have been shown to be present in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

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 (**Fig. 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 (Supplementary Fig. S4). Salaspermic acid was isolated from *Salacia macrosperma*, a member of the Celastraceae family that also accumulates maytenin and pristimerin (Viswanathan 1979, Kutney et al. 1981), 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 to be the P450 that also catalyzes the two-step oxidation at the C-2 position of maytenoic acid (**Fig. 6B**). However it cannot be excluded, further oxidation of wilforic acid E to cangoronine seems unlikely as the 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 (**Fig. 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

Plant Cell Physiol. 0(0): 1–13 (2019) doi:10.1093/pcp/pcz144

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 the 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 toward 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 multienzyme 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 (Supplementary Fig. S5).

After the 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 (Supplementary Fig. S6). In *M. ilicifolia*, the produced celastrol may be the substrate for 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 (Supplementary Fig. S7).

Toward heterologous production of quinone methide triterpenoids

Despite being a promising natural product for the development of anti-cancer drugs, the commercial production of quinone methide triterpenoids is hampered by the slow growth of the producing plant species and low production amounts observed in plants (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 FRS compared with β-amyrin synthase when expressed in *N. benthamiana* leaves. However, we did not observe such a difference in the 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 with 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 FRS 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* FRS (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, Germantown, MD, USA). The RNA quality was checked using a Bioanalyzer instrument (Agilent Technologies, Santa Clara, CA, USA), and the RNA concentration was determined using a Nanodrop-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Subsequently, mRNA was isolated from the total RNA by 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, San Diego, CA, USA).

RNA-Seq and transcriptome assembly

The paired-end cDNA sequencing library was prepared using the Illumina TruSeq RNA sample preparation v3 kit (Illumina). 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 were 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 (2 × 76 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 (Supplementary 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 Laboratories Inc., Hercules, CA, USA) 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 (Supplementary 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 (Supplementary 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

Nicotiana 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 Technologies) as described (Moses et al. 2014). For quantification, calibration curves were generated for authentic friedelin and β-amyrin standards (Sigma-Aldrich, Saint Louis, MO, USA). Peak areas were calculated with the default settings of the ChemStation software (Agilent Technologies).

LC-APCI-FT-MS analysis

LC-APCI-FT-MS was performed using a Hypersil BDS C18 HPLC column (100 × 2.1 mm, 3 µm; Thermo Fisher Scientific) mounted on an Accela UPLC system (Thermo Fisher Scientific). The LC system was coupled to a LTQ FT Ultra (Thermo Fisher Scientific) 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 and 1,400 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

The 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 3 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 10 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 (Bruker Corporation, Billerica, MA, USA) 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 (Sigma Aldrich). 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 spectroscopy (gCOSY), ^1H-{^1H} nuclear Overhauser spectroscopy, ^1H-{^{13}C} multiplicity-edited gradient heteronuclear single quantum correlation (g-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 Corporation).

First, the ^1H NMR analysis (Supplementary Fig. 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 (Supplementary Fig. S3A). Besides those, other high field peaks, corresponding to three hydrogens, were observed. The structural limitation

Plant Cell Physiol. 0(0): 1–13 (2019) doi:10.1093/pcp/pcz144

of the friedelin backbone results in considerable resonance overlapping of the aliphatic peaks. In order to overcome this, ^{1}H-{^{1}H}-TOCSY experiments (Supplementary Fig. 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 ^{1}H-{^{1}H}-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 (Supplementary Fig. S3B). ^{1}H-{^{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 ^{1}H-{^{13}C}-HMBC correlation between C-29 and the nonassigned high-field protons, along with the observation of individual coupling systems for peaks at δ_H 2.34 and 2.16, on ^{1}H-{^{1}H}-TOCSY, allowed the assignment of these peaks as H-19 and H-21, respectively, completing ring E proton assignment (Supplementary Fig. 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 (Supplementary 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; Supplementary 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, CA, USA). 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 microgram 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 microgram 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, the yeast strain KB13 was derived from strain KB10 by knocking out the *PAH1* gene using CRISPR/Cas9. One microgram 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 Supplementary 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 Laboratories, Mountain View, CA, USA). 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.

Supplementary Data

Supplementary data are available at PCP online.

Funding

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.

Acknowledgments

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.

Footnotes: The nucleotide sequence reported in this article has been submitted to the GenBank database with accessionnumber MK829814.

Disclosures

The authors have no conflicts of interest to declare.

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, D., 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. Tiss. 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. p. 188. *Ph.D. Thesis*. Department of Plant and Environmental Sciences, 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.

Plant Cell Physiol. 0(0): 1–13 (2019) doi:10.1093/pcp/pcz144

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. Des.* 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, V., Jr., 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: 225–257.

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.

8 Referências Bibliográficas

BICALHO, K. U.; SANTONI, M. M.; ARENDT, P.; ZANELLI, C. F.; FURLAN, M.; GOOSSENS, A.; POLLIER, J. CYP712K4 Catalyzes the C-29 Oxidation of Friedelin in the Maytenus ilicifolia Quinone Methide Triterpenoid Biosynthesis Pathway. **Plant and Cell Physiology**, [s. l.], v. 60, n. 11, p. 2510–2522, 2019.

BUCHFINK, B.; XIE, C.; HUSON, D. H. Fast and sensitive protein alignment using DIAMOND. **Nature Methods**, [s. l.], v. 12, n. 1, p. 59–60, 2014.

CARLINI, E. Toxicologia clinica (Fase I) da Espinheira-Santa (Maytenus Ilicifolia). **Estudo de acao antiulcera gastrica de plantas brasileiras (maytevirus ilicifolia "espinheira-santa" e outras).**, [s. l.], 1988.

DE SOUZA, L. M.; CIPRIANI, T. R.; IACOMINI, M.; GORIN, P. A. J.; SASSAKI, G. L. HPLC/ESI-MS and NMR analysis of flavonoids and tannins in bioactive extract from leaves of Maytenus ilicifolia. **Journal of Pharmaceutical and Biomedical Analysis**, [s. l.], v. 47, p. 59–67, 2008. Disponível em: <www.sciencedirect.com>. Acesso em: 26 ago. 2021.

DEVI, K.; MISHRA, S. K.; SAHU, J.; PANDA, D.; MODI, M. K.; SEN, P. Genome wide transcriptome profiling reveals differential gene expression in secondary metabolite pathway of Cymbopogon winterianus OPEN. **Nature Publishing Group**, [s. l.], v. 6, n. 21026, p. 1–11, 2016.

DZIGGEL, C.; SCHÃFER, H.; WINK, M. Tools of pathway reconstruction and production of economically relevant plant secondary metabolites in recombinant microorganisms. **Biotechnology Journal**, [s. l.], v. 12, n. 1, p. 1–14, 2017.

FILHO, W. B.; CORSINO, J.; BOLZANI, V. da S.; FURLAN, M.; PEREIRA, A. M. S.; FRANÇA, S. C. Quantitative determination of cytotoxic-Friedo-nor-oleanane derivatives from five morphological types of Maytenus ilicifolia (celastraceae) by reverse-phase high-performance liquid chromatography. **Phytochemical Analysis**, [s. l.], v. 13, n. 2, p. 75–78, 2002.

GIWELI, A. A.; DŽAMIĆ, A. M.; SOKOVIĆ, M.; RISTIĆ, M. S.; JANAĆKOVIĆ, P.; MARIN, P. D. The chemical composition, antimicrobial and antioxidant activities of the essential oil of salvia fruticosa growing wild in Libya. **Archives of Biological Sciences**,

[s. l.], v. 65, n. 1, p. 321–329, 2013.

GRABHERR, M. G.; HAAS, B. J.; YASSOUR, M.; LEVIN, J. Z.; THOMPSON, D. A.; AMIT, I.; ADICONIS, X.; FAN, L.; RAYCHOWDHURY, R.; ZENG, Q.; CHEN, Z.; MAUCELI, E.; HACOHEN, N.; GNIRKE, A.; RHIND, N.; DI PALMA, F.; BIRREN, B. W.; NUSBAUM, C.; LINDBLAD-TOH, K.; FRIEDMAN, N.; REGEV, A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. **Nature Biotechnology**, [s. l.], v. 29, n. 7, p. 644–652, 2011.

GROPPO, M.; SIMMONS, M. P.; CAPPA, J. J.; BIRAL, L.; LOMBARDI, J. A. A new species of maytenus (Celastraceae) with fleshy fruits from eastern brazil, with notes on the delimitation of maytenus. **Systematic Botany**, [s. l.], v. 39, n. 2, p. 478–484, 2014. Disponível em: <https://www.jstor.org/stable/24546156>. Acesso em: 7 set. 2021.

GUO, J.; HUANG, Z.; SUN, J.; CUI, X.; LIU, Y. Research Progress and Future Development Trends in Medicinal Plant Transcriptomics. **Frontiers in Plant Science**, [s. l.], v. 12, n. July, p. 1–10, 2021.

HAAS, B. J.; PAPANICOLAOU, A.; YASSOUR, M.; GRABHERR, M.; BLOOD, P. D.; BOWDEN, J.; COUGER, M. B.; ECCLES, D.; LI, B.; LIEBER, M.; MACMANES, M. D.; OTT, M.; ORVIS, J.; POCHET, N.; STROZZI, F.; WEEKS, N.; WESTERMAN, R.; WILLIAM, T.; DEWEY, C. N.; HENSCHEL, R.; LEDUC, R. D.; FRIEDMAN, N.; REGEV, A. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. **Nature Protocols 2013 8:8**, [s. l.], v. 8, n. 8, p. 1494–1512, 2013. Disponível em: <https://www.nature.com/articles/nprot.2013.084>. Acesso em: 27 dez. 2022.

HANSEN, C. C.; NELSON, D. R.; MØLLER, B. L.; WERCK-REICHHART, D. Plant cytochrome P450 plasticity and evolution. **Molecular Plant**, [s. l.], v. 14, n. 8, p. 1244–1265, 2021. Disponível em: <https://doi.org/10.1016/j.molp.2021.06.028>

HANSEN, N. L.; HESKES, A. M.; HAMBERGER, B.; OLSEN, C. E.; HALLSTRÖM, B. M.; ANDERSEN-RANBERG, J.; HAMBERGER, B. The terpene synthase gene family in Tripterygium wilfordii harbors a labdane-type diterpene synthase among the monoterpene synthase TPS-b subfamily. **Plant Journal**, [s. l.], v. 89, n. 3, p. 429–441, 2017.

LI, W.; XU, R.; YAN, X.; LIANG, D.; ZHANG, L.; QIN, X.; CAIYIN, Q.; ZHAO, G.; XIAO, W.; HU, Z.; QIAO, J. De novo leaf and root transcriptome analysis to explore biosynthetic pathway of Celangulin v in Celastrus angulatus maxim. **BMC Genomics**, [s. l.], v. 20, n. 1, p. 1–15, 2019.

LOVE, M. I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome Biology**, [s. l.], v. 15, n. 12, p. 1–21, 2014.

MARIOT, M. P.; BARBIERI, R. L. Metabólitos secundários e propriedades medicinais da espinheira-santa (Maytenus ilicifolia Mart. ex Reiss. e M. aquifolium Mart.). **Revista Brasileira de Plantas Medicinais**, [s. l.], v. 9, n. 3, p. 89–99, 2007. a.

MARIOT, M. P.; BARBIERI, R. L. Secondary metabolites and medicinal properties of espinheira-santa (Maytenus ilicifolia Mart. ex Reiss. e Maytenus aquifolium Mart.). **Revista Brasileira de Plantas Medicinais**, [s. l.], v. 9, n. 3, 2007. b.

METLEN, K. L.; ASCHEHOUG, E. T.; CALLAWAY, R. M. Plant behavioural ecology: dynamic plasticity in secondary metabolites. **Plant, cell & environment**, [s. l.], v. 32, n. 6, p. 641–653, 2009. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/19021888/>. Acesso em: 7 set. 2021.

MUSILOVA, L.; RIDL, J.; POLIVKOVA, M.; MACEK, T.; UHLIK, O. Effects of Secondary Plant Metabolites on Microbial Populations: Changes in Community Structure and Metabolic Activity in Contaminated Environments. **International Journal of Molecular Sciences**, [s. l.], v. 17, n. 8, 2016. Disponível em: </pmc/articles/PMC5000603/>. Acesso em: 7 set. 2021.

PATRO, R.; DUGGAL, G.; LOVE, M. I.; IRIZARRY, R. A.; KINGSFORD, C. Salmon provides fast and bias-aware quantification of transcript expression. **Nature Methods**, [s. l.], v. 14, n. 4, p. 417–419, 2017.

PAZ, T. A.; DOS SANTOS, V. A. F. F. M.; INÁCIO, M. C.; DIAS, N. B.; PALMA, M. S.; PEREIRA, A. M. S.; FURLAN, M. Proteome profiling reveals insights into secondary metabolism in Maytenus ilicifolia (Celastraceae) cell cultures producing quinonemethide triterpenes. **Plant Cell, Tissue and Organ Culture**, [s. l.], v. 130, n. 2, p. 405–416, 2017.

PEI, T.; YAN, M.; KONG, Y.; FAN, H.; LIU, J.; CUI, M.; FANG, Y.; GE, B.; YANG, J.; ZHAO, Q.; YAN, M.; KONG, Y.; FAN, H.; LIU, J.; CUI, M.; FANG, Y.; GE, B.; YANG, J.; ZHAO, Q. The genome of Tripterygium wilfordii and characterization of the celastrol biosynthesis pathway. **Gigabyte**, [s. l.], v. 2021, p. 1–32, 2021.

PÉRICO, L. L.; RODRIGUES, V. P.; ALMEIDA, L. F. R. De; FORTUNA-PEREZ, A. P.; VILEGAS, W.; HIRUMA-LIMA, C. A. Maytenus ilicifolia Mart. ex Reissek. [s. l.], p. 323–335, 2018.

SILVA, F. C. O.; FERREIRA, M. K. A.; DA SILVA, A. W.; MATOS, M. G. C.; MAGALHÃES, F. E. A.; DA SILVA, P. T.; BANDEIRA, P. N.; DE MENEZES, J. E. S.

A.; SANTOS, H. S. Bioativities of plant-isolated triterpenes: A brief review. **Revista Virtual de Quimica**, [s. l.], v. 12, n. 1, p. 234–247, 2020.

SIMAO, F. A.; WATERHOUSE, R. M.; IOANNIDIS, P.; KRIVENTSEVA, E. V; ZDOBNOV, E. M. BUSCO online supplementary information: assessing genome assembly and annotation completeness with single-copy orthologs. **Bioinformatics**, [s. l.], v. 31, n. 19, p. Supplementary information, 2015.

ŠMARDA, P.; BUREŠ, P.; HOROVÁ, L.; LEITCH, I. J.; MUCINA, L.; PACINI, E.; TICHÝ, L.; GRULICH, V.; ROTREKLOVÁ, O. Ecological and evolutionary significance of genomic GC content diversity in monocots. **Proceedings of the National Academy of Sciences of the United States of America**, [s. l.], v. 111, n. 39, p. E4096–E4102, 2014.

SOUZA-MOREIRA, T. M.; ALVES, T. B.; PINHEIRO, K. A.; FELIPPE, L. G.; DE LIMA, G. M. A.; WATANABE, T. F.; BARBOSA, C. C.; SANTOS, V. A. F. F. M.; LOPES, N. P.; VALENTINI, S. R.; GUIDO, R. V. C.; FURLAN, M.; ZANELLI, C. F. Friedelin Synthase from Maytenus ilicifolia: Leucine 482 Plays an Essential Role in the Production of the Most Rearranged Pentacyclic Triterpene. **Scientific Reports**, [s. l.], v. 6, n. May, p. 1–13, 2016.

THIMMAPPA, R.; GEISLER, K.; LOUVEAU, T.; O'MAILLE, P.; OSBOURN, A. Triterpene biosynthesis in plants. **Annual Review of Plant Biology**, [s. l.], v. 65, p. 225–257, 2014.

TRIPATHI, S.; SANGWAN, R. S.; NARNOLIYA, L. K.; SRIVASTAVA, Y.; MISHRA, B.; SANGWAN, N. S.; SINGH, A. K.; KUMAR, S. R.; DWIVEDI, V.; RAI, A.; PAL, S.; SHASANY, A. K.; NAGEGOWDA, D. A. Transcription factor repertoire in Ashwagandha (Withania somnifera) through analytics of transcriptomic resources: Insights into regulation of development and withanolide metabolism. **Scientific Reports**, [s. l.], v. 215, n. 1, p. 1115–1131, 2017. Disponível em: <http://dx.doi.org/10.1038/s41598-017-14657-6>

UPADHYAY, S.; PHUKAN, U. J.; MISHRA, S.; SHUKLA, R. K. De novo leaf and root transcriptome analysis identified novel genes involved in Steroidal sapogenin biosynthesis in Asparagus racemosus. **BMC Genomics**, [s. l.], v. 15, n. 1, p. 1–13, 2014.

WINK, M. Secondary Metabolites: Deterring Herbivores. **Encyclopedia of Life Sciences**, [s. l.], n. March, p. 1–9, 2010. a.

WINK, M. Introduction: Biochemistry, Physiology and Ecological Functions of Secondary Metabolites. **Biochemistry of Plant Secondary Metabolism: Second**

Edition, [s. l.], v. 40, p. 1–19, 2010. b.

YANG, L.; WEN, K. S.; RUAN, X.; ZHAO, Y. X.; WEI, F.; WANG, Q. Response of plant secondary metabolites to environmental factors. **Molecules**, [s. l.], v. 23, n. 4, p. 1–26, 2018.

ZHU, Y.; ZHANG, X.; ZHANG, Y.; L. CHANGFU LI, S. Identification of Genes Involved in Celastrol Biosynthesis by Comparative Transcriptome Analysis in Tripterygium wilfordii. **Phyton**, [s. l.], v. 91, n. 2, p. 279–291, 2022.