
**PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS
(BIOLOGIA VEGETAL)**

**PASSOS INICIAIS DO METABOLISMO DE ISOPRENOIDES EM
ANGIOSPERMAS: GENÔMICA EVOLUTIVA E ANÁLISES MOLECULARES EM
*COFFEA SPP***

NATACHA SILVA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS
(BIOLOGIA VEGETAL)

PASSOS INICIAIS DO METABOLISMO DE ISOPRENOIDES EM
ANGIOSPERMAS: GENÔMICA EVOLUTIVA E ANÁLISES MOLECULARES EM
COFFEA SPP

NATACHA SILVA

Tese apresentada ao Instituto de Biociências do Câmpus de Rio Claro, Universidade Estadual Paulista, como parte dos requisitos para obtenção do título de doutor em Ciências Biológicas (Biologia Vegetal).

Orientador: Prof. Dr. Douglas Silva Domingues

Coorientadora: Prof. Dra. Suzana Tiemi Ivamoto-Suzuki

Rio Claro - SP
2020

S586p Silva, Natacha
Passos iniciais do metabolismo de isoprenoides em angiospermas:
genômica evolutiva e análises moleculares em Coffea spp / Natacha
Silva. -- Rio Claro, 2020
75 p.

Tese (doutorado) - Universidade Estadual Paulista (Unesp),
Instituto de Biociências, Rio Claro
Orientador: Douglas Silva Domingues
Coorientadora: Suzana Tiemi Ivamoto-Suzuki

1. Botânica. 2. Biologia molecular. 3. Genômica evolutiva. I.
Título.

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

Essa ficha não pode ser modificada.

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: PASSOS INICIAIS DO METABOLISMO DE ISOPRENOIDES EM ANGIOSPERMAS:
GENÔMICA EVOLUTIVA E ANÁLISES MOLECULARES EM COFFEA SPP

AUTORA: NATACHA SILVA

ORIENTADOR: DOUGLAS SILVA DOMINGUES

COORIENTADORA: SUZANA TIEMI IVAMOTO

Aprovada como parte das exigências para obtenção do Título de Doutora em CIÊNCIAS BIOLÓGICAS (BIOLOGIA VEGETAL), área: Biologia Vegetal pela Comissão Examinadora:



Prof. Dr. DOUGLAS SILVA DOMINGUES
Departamento de Biodiversidade / IB / UNESP - Rio Claro

Prof. Dr. JORGE MAURICIO COSTA MONDEGO
Centro de Pesquisa e Desenvolvimento em Recursos Genéticos Vegetais / Instituto Agrônômico de Campinas

Prof. Dr. LUIZ EDUARDO VIEIRA DEL BEM
Departamento de Botânica / Instituto de Ciências Biológicas (ICB), Universidade Federal de Minas Gerais (UFMG)

Dr^a LUISA FERNANDA BERMUDEZ SALAZAR
Instituto de Biotecnología Instituto Nacional de Tecnología Agropecuaria (INTA) - Argentina

Prof. Dr. IGOR CESARINO
Departamento de Botânica Instituto de Biociências Universidade de São Paulo (USP)

Rio Claro, 21 de setembro de 2020

*“Ensine-a a questionar a linguagem.
A linguagem é o repositório de nossos preconceitos,
de nossas crenças, de nossos pressupostos”
Chimamanda Ngozi Adichie*

*“Não precisamos de mais tempo,
precisamos de um tempo que seja nosso”
Mia Couto*

*“Reconhece a queda e não desanima.
Levanta, sacode a poeira, dá volta por cima”
Paulo Emílio Vanzolini*

Dedico este trabalho à minha família

AGRADECIMENTOS

Ao apoio e financiamento da FAPESP:

Processo nº 2016/10896-0, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

À agência de fomento que tornou este trabalho possível: Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq pela concessão de bolsa durante o período do doutorado.

Ao meu orientador Douglas, por acreditar no meu trabalho e profissionalismo. Por ter aberto, gentilmente, as portas para me orientar. Inúmeras vezes foi paciente com as minhas dificuldades e nunca desistia de me ensinar. Levarei, com muito carinho, o seu exemplo como educador para minha atuação profissional. “Anda com fé eu vou. Café não costuma faiá”

À minha coorientadora Suzana, por passar dias a fio, com extrema educação e paciência, me ensinando grande parte do desenvolvimento do trabalho. Por ser meu exemplo de força, dedicação e perseverança. Levarei de você “a marca Maria de ter fé na vida”. Você é especial!

Aos meus lindos e queridíssimos amigos de laboratório, Gian (o amuleto da sorte), Paula (a mãezona guerreira), Daniel (o papis da Disney) e Raíssa (a fada delicada), com vocês os meus dias de trabalho eram mais leves. Obrigada pelas risadas, pelas confidências trocadas e por me manter em pé em dias de muita tempestade. Com vocês pude ouvir aquela “canção do rádio, em que um antigo compositor baiano me dizia: tudo é divino, tudo é maravilhoso”.

À minha família, a minha riqueza maior do mundo. Mamis Neila, papai Aguinaldo, irmão Fili, talvez mais 70 páginas de tese descreveriam o que vocês sempre fazem por mim, mas não posso essa ousadia. Com vocês eu “Vou mastigando o mundo e ruminando. E assim vou tocando essa vida marvada. Pra todo aquele que só fala que não sabemos viver, chega lá em casa pro uma visitinha. Que num verso

ou num reverso da vida inteirinha, há de nos encontrar num caterete”. Obrigada, obrigada e obrigada. Amo vocês.

Ao meu maravilhoso parceiro de vida, Gabriel. Por ser o homem que me ampara e cuida! Assim como a música, você é vital em mim. Junto com você e com meu diploma de doutora, espero um dia poder dizer que “de quimeras mil um castelo ergui”. Amo você!

Às minhas amigas de infância, Marina, Vanessa, Amanda e Letícia (prima) por me darem total apoio e conforto nas voltas pra São Joaquim da Barra. Vocês fizeram a maior diferença!

Aos meus amigos da Unesp Rio Claro, Luís Felipe, Ana Angélica, Marina Gavassi, Artur “fininho” e Nayara por me ajudarem no “traz a saideira?”.

Aos meus amigos de Rio Claro, Teresa, Paola e Nuno, Burts, Leonel e Ju, Mari e Ralf, Ferfura e Samuca, Lara e Lincoln + Cacázinha, Clarinha, as irmãs Riani, Dani e todo o Núcleo de Samba Rio Clareense, por me oferecerem a mais pura musicalidade. Com vocês o doutorado foi bem mais tranquilo.

À Unesp de Ilha Solteira, por me apresentar as minhas joias raras, Desirée, Paola Trama e família, Felipe Negão, Lagostinha, Marilaine, Alysson, João do Milho, Vera Andrade e Felipe Andrade.

Ao professor Gustavo Habermann, por disponibilizar parte do seu laboratório de pesquisa.

À professora Alessandra Coan e a querida Celinha, por me ajudar em momentos burocráticos delicados.

Ao pesquisador Luís Filipe por disponibilizar o laboratório de pesquisa do IAPAR – Londrina.

E a todos que de alguma forma estiveram presentes neste momento...

Obrigada!

SILVA, Natacha. **Passos iniciais do metabolismo de isoprenoides em angiospermas: genômica evolutiva e análises moleculares em *Coffea* spp.** 2020. 73 p. Tese (Doutorado em Ciências Biológicas – Biologia Vegetal) – Universidade Estadual Paulista “Julio de Mesquita Filho”, Rio Claro. 2020.

RESUMO

Os isoprenoides compõem uma classe de metabólitos especializados em plantas, com funções biológicas, ecológicas e fisiológicas. As plantas sintetizam milhares de compostos derivados dos isoprenos, com função associada a atração de polinizadores, proteção contra pragas, desenvolvimento e crescimento das plantas. Além disso, eles são muito utilizados pelas indústrias farmacêutica e de cosméticos. Existem duas vias metabólicas capazes de produzir isoprenoides em plantas: a via citoplasmática do mevalonato (MVA) e a via plastidial do metileritritol fosfato (MEP). Os objetivos do trabalho foram: i) identificar os genes iniciais da via de biossíntese de isoprenoides em *Coffea* spp.; ii) avaliar se o elicitor ácido hexanoico (Hx) é capaz de modular o perfil transcricional dos genes da via MVA e MEP em *C. arabica*; iii) analisar a evolução dos genes chave da via MVA (*HMGR* e *MVK*) e MEP (*DXS* e *DXR*). Plantas de *C. arabica* cv. Catuaí Vermelho (CaCV) e cv. Obatã (CaOB) foram tratadas com aplicação de ácido hexanoico (Hx - 0,55 mM) em solução nutritiva aerada, e após 48 h foram coletadas folhas e raízes para extração de RNA e análises de perfil transcricional via RNA-seq e RT-qPCR. Foram identificados mais de uma cópia para cada gene da via MVA (*HMGS*, *HMGR*, *MVK*, *PMK*, *MVD*) e MEP (*DXS*, *DXR*, *CMK*, *MDS*, *HDS*, *HDR*) em *C. arabica*. *C. eugenioides* apresentou mais de uma cópia para alguns genes da via MVA (*AACT*, *HMGR*, *PMK*, *MVD*) e da via MEP (*DXS* e *HDR*). Em *C. canephora*, os genes da MVA (*AACT*, *HMGS*, *MVK*, *PMK*, *MVD*) e alguns da MEP (*DXR*, *MCT*, *CMK*, *MDS*, *HDS*) foram de cópia única. Os genes *AACT*, *HMGS*, *HMGR*, *MVK*, *PMK*, *MVD*, *DXS*, *DXR*, *CMK* e *MDS* foram induzidos em folhas tratadas com Hx em CaCV. Em CaOB, os genes *MVK*, *DXS*, *DXR*, *CMK*, *MDS* e *HDS* foram induzidos em folhas tratados com Hx. Os genes *AACT*, *HMGS*, *HMGR*, *MVK*, *PMK*, *DXS* e *DXR* foram reprimidos em raízes de CaCV e os genes *AACT*, *HMGS*, *HMGR*, *MVK*, *MVD*, *DXS*, *DXR* e *MDS* reprimidos em raízes de CaOB tratadas com Hx. Os nossos resultados indicam que o Hx é capaz de modular a expressão dos genes das vias MVA e MEP, em folhas e raízes de café de maneira genótipo e tecido-específica, evidenciando o poder reprogramador de Hx no metabolismo de isoprenoides.

Palavras-chave: Coffee, MVA, MEP, terpenoides, ácido hexanoico

SILVA, Natacha. **Initial steps of isoprenoids biosynthesis in angiosperms: genomic evolution and molecular analysis in *Coffea* spp.** 2020. 73 p. Thesis (PhD in Biological Sciences – Plant Biology) – Universidade Estadual Paulista “Julio de Mesquita Filho”, Rio Claro. 2020.

ABSTRACT

Isoprenoids comprise a class of specialized metabolites from plants comprising important biological, ecological and physiological functions. Plants synthesize thousands of isoprene compounds, which are related to pollinators attraction, anti-herbivory properties, plant growth and development. In addition, they are widely used by the pharmaceutical and cosmetic industries. There are two metabolic pathways responsible for isoprenoid biosynthesis in plants: the cytoplasmic mevalonate pathway (MVA) and the plastidic methylerythritol phosphate pathway (MEP). In this study our main goals were: i) to identify genes related to the initial steps of isoprenoid biosynthesis in *Coffea* spp.; ii) evaluate if hexanoic acid (Hx) can transcriptionally modulate genes of MVA and MEP pathway in *C. arabica*; iii) analyze the evolution of selected MVA (*HMGR* and *MVK*) and MEP (*DXS* and *DXR*) genes. *C. arabica* cv. Catuaí Vermelho (CaCV) and cv. Obatã (CaOB) plants were treated with exogenous application of hexanoic acid (Hx - 0.55 mM) in an aerated nutrient solution, and after 48 h, leaves and roots were collected for RNA extraction and transcriptional analyses using RNA-seq and RT-qPCR. More than one copy were observed for several MVA pathway genes (*HMGS*, *HMGR*, *MVK*, *PMK*, *MVD*) and MEP pathway genes (*DXS*, *DXR*, *CMK*, *MDS*, *HDS*, *HDR*) in *C. arabica*. *C. eugenioides* presented more than one copy of some genes from MVA (*AACT*, *HMGR*, *PMK*, *MVD*) and MEP (*DXS*, *HDR*) pathways. In *C. canephora*, some genes from MVA (*AACT*, *HMGS*, *MVK*, *PMK*, *MVD*) and MEP (*DXR*, *MCT*, *CMK*, *MDS*, *HDS*) pathways were classified as single copy. *AACT*, *HMGS*, *HMGR*, *MVK*, *PMK*, *MVD*, *DXS*, *DXR*, *CMK* and *MDS* genes were induced in leaves under Hx-treatment in CaCV. In CaOB, *MVK*, *DXS*, *DXR*, *CMK*, *MDS* and *HDS* genes were induced in leaves by Hx exogenous application. *AACT*, *HMGS*, *HMGR*, *MVK*, *PMK*, *DXS* and *DXR* genes were repressed in CaCV roots and *AACT*, *HMGS*, *HMGR*, *MVK*, *MVD*, *DXS*, *DXR* e *MDS* repressed in CaOB roots under Hx treatment. Our results indicated that Hx can modulate the expression of genes involved in MVA and MEP pathways in coffee in a genotype and tissue-specific way, emphasizing that Hx is able to reprogram isoprenoid metabolism.

Keywords: Coffee, MVA, MEP, terpenoids, hexanoic acid

SUMÁRIO

1. INTRODUÇÃO	12
1.1 A via de biossíntese de isoprenoides	12
1.2. A importância dos genes <i>HMGR</i> e <i>DXS</i>	16
1.2. A importância dos genes <i>MVK</i> e <i>DXR</i>	16
1.3. Os isoprenoides no gênero <i>Coffea</i> spp	17
1.4. A utilização de elicitores para a indução da biossíntese de compostos especializados	18
1.5. Referencial bibliográfico	19
2. OBJETIVO GERAL	23
2.1. Objetivos específicos	23
3. CAPÍTULO 1: Análise genômica das vias de biossíntese de isoprenoides em <i>Coffea</i> spp.	24
3.1. Introdução	25
3.2 Material e métodos	27
3.2.1. <i>Material vegetal, tratamento com ácido hexanóico e extração de RNA total</i>	27
3.2.2. <i>Análise do perfil transcricional dos genes das vias MVA e MEP em <i>C. arabica</i>.</i>	28
3.2.3. <i>Identificação dos genes das vias MVA e MEP nos genomas de <i>Coffea</i> spp</i>	29
3.2.4. <i>Análise filogenética dos genes <i>HMGR</i> e <i>DXS</i></i>	30
3.2.5. <i>Determinação das estruturas gênicas de <i>HMGR</i> e <i>DXS</i> em plantas</i>	30
3.3 Resultados e discussão	31
3.3.1 <i>Identificação do número de cópias dos genes das vias MVA e MEP em <i>Coffea</i> spp</i>	31
3.3.3 <i>Número de cópias e análise filogenética dos genes <i>HMGR</i> e <i>DXS</i></i>	37
3.3.4. <i>Estrutura éxon-íntron dos genes <i>HMGR</i> e <i>DXS</i></i>	42
3.4. Conclusão	44
4. CAPÍTULO 2: Low-Copy Genes in Terpenoid Metabolism: The Evolution and Expression of <i>MVK</i> and <i>DXR</i> Genes in Angiosperms	48
4.1. Introduction	48
4.2. Results	50
4.2.1. <i>Copy Number of <i>MVK</i> and <i>DXR</i> Genes in Plants</i>	50
4.2.2. <i>Phylogenetic Analyses</i>	51
4.2.3. <i>Exon–Intron Structure of <i>MVK</i> and <i>DXR</i> Genes</i>	54
4.2.4. <i>Nucleotide Substitution and Selection in <i>MVK</i> and <i>DXR</i> Genes</i>	55
4.2.5. <i>Estimated Time of Duplication of <i>MVK</i> and <i>DXR</i> Genes</i>	55

4.2.6. Recombination Events of MVK and DXR Genes	56
4.2.7. RNA-seq-Based Expression Profiles	57
4.2.8. RT-qPCR-Based Transcriptional Profile of MVK and DXR Genes in <i>C. arabica</i>	59
4.3. Discussion	61
4.4. Materials and Methods	62
4.4.1. Identification and Annotation of MVK and DXR Gene Families in Plants	62
4.4.2. Multiple Sequence Alignments and Phylogenetic Analysis	63
4.4.3. Determination of Gene Structures	63
4.4.4. Selection Pressure and Evolutionary Analyses	63
4.4.5. Gene Duplication Analysis	63
4.4.6. Gene Recombination Analysis	63
4.4.7. RNA-seq-Based Expression Profile Analysis of MVK and DXR Genes	64
4.4.8. Transcriptional Profiles of MVK and DXR Genes in <i>C. arabica</i> by RT-qPCR	64
4.5. Conclusions	65
4.6. References	65
5. CONSIDERAÇÕES FINAIS	75

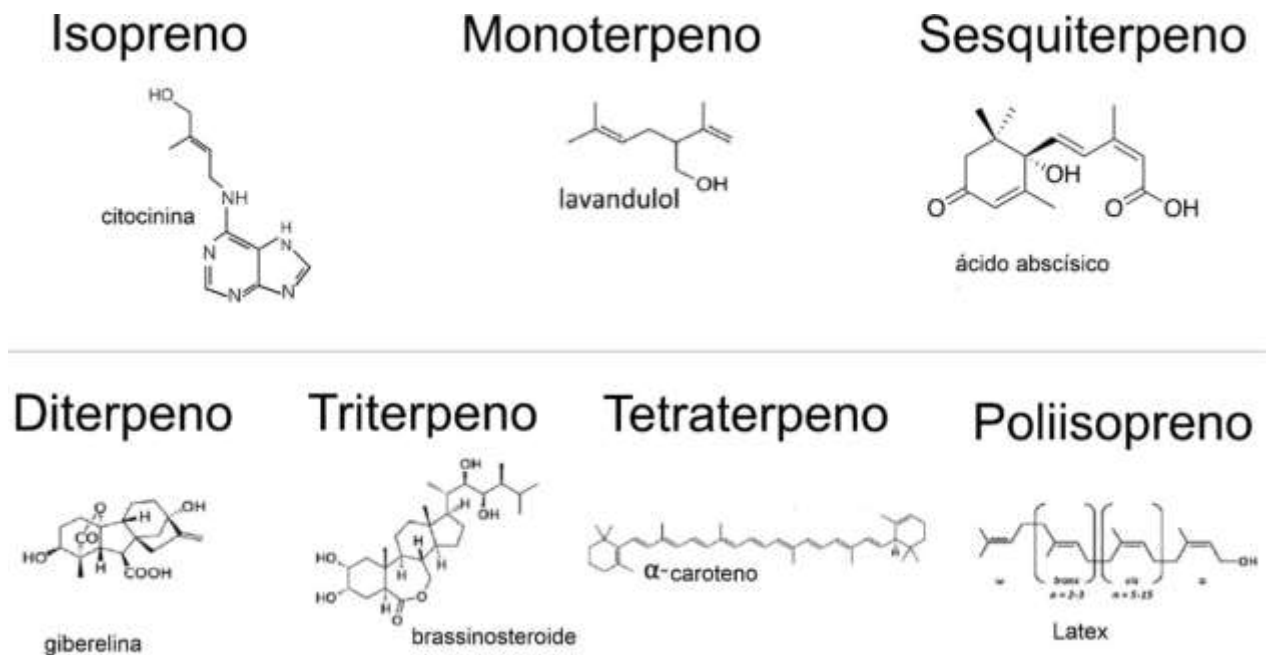
1. INTRODUÇÃO

1.1 A via de biossíntese de isoprenoides

Os isoprenoides ou terpenoides compõem uma classe de metabólitos especializados em plantas, com funções biológicas, ecológicas e fisiológicas (TAIZ et al., 2017). Os vegetais sintetizam mais de 40 mil compostos de terpenos relatados na literatura (GALATA; MAHMOUD, 2012; ZHOU; PICHERSKY, 2020). Estes compostos nos vegetais têm funções de proteção contra predadores, atração de polinizadores, controle da taxa transpiratória, proteção à luz ultravioleta, desenvolvimento e crescimento (TAIZ et al., 2017; PICHERSKY; RAGUSO, 2018).

Todos os tipos de isoprenoides derivam de precursores de cinco carbonos (C_5). Essas unidades geram moléculas de comprimento de cadeia de carbono definido (C_{10} , C_{15} , C_{20} , ... C_{5n}), que servem como ponto de partida para a produção de toda a variedade de isoprenoides (ZI et al., 2014; PICHERSKY; RAGUSO, 2018). Assim, isoprenoides são classificados como: os isoprenos - C_5 , que podem originar o hormônio citocinina (VRANOVÁ et al., 2013); os monoterpenos (C_{10}), que formam os compostos voláteis dos óleos essenciais de algumas plantas (hortelã, lavanda e citronela) e os piretroides presente nos crisântemos (repelente natural de insetos – PICHERSKY; RAGUSO, 2018); os sesquiterpenos (C_{15}), que originam alguns fitormônios como o ácido abscísico (ABA) e a estrigolactona (AL-BABILI; BOUWMEESTER, 2015); os diterpenos (C_{20}), que formam o hormônio giberelina (GA), o fitol presente na clorofila e estão presentes nas resinas em pinus (CHEN et al., 2011; ZI et al., 2014); os triterpenos (C_{30}), que fazem parte da composição dos hormônios brassinosteroides, dos esteroides (lipídios) presentes nas membranas vegetais e as saponinas relacionadas à defesa aos insetos (VRIET et al., 2013; ZHOU et al., 2015); os tetraterpenos (C_{40}), que originam os carotenoides, pigmentos fotossintético, presentes em cenoura, morango com propriedades antioxidantes (PICHERSKY; RAGUSO, 2018); os poliisoprenos (C_n) possuem mais de 40 carbonos em sua composição e são encontrados no látex da borracha, presente nas seringueiras (Figura 1. QU et al., 2015).

Figura 1. Estruturas químicas de exemplos de isoprenoides.



Fonte: Adaptado de Pichersky e Raguso (2018).

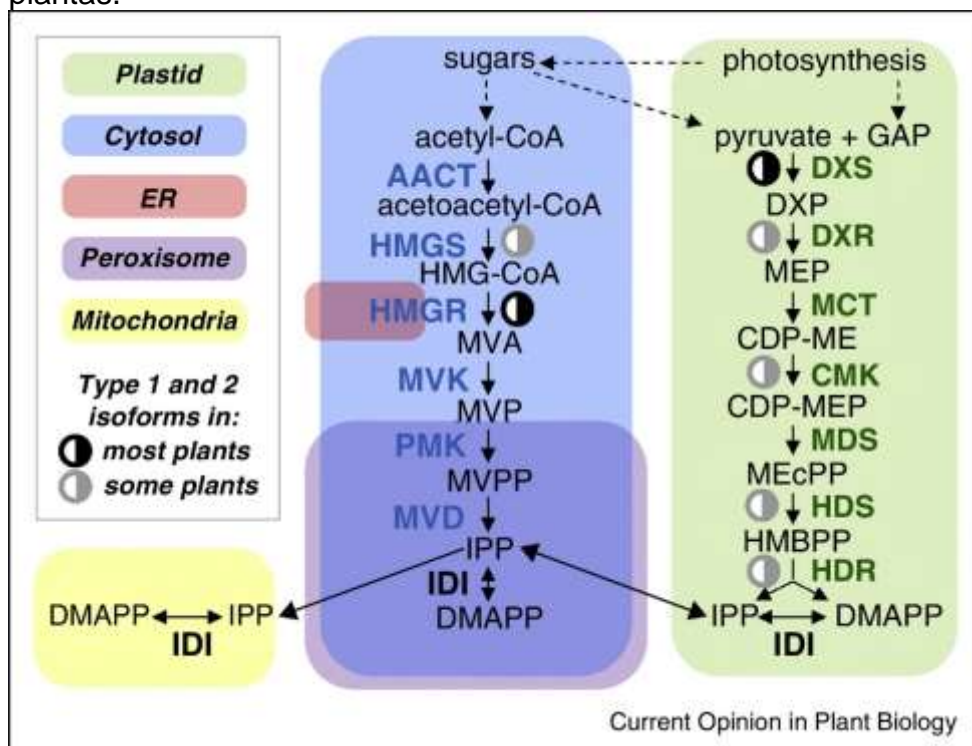
A formação dos variados tipos de isoprenoides pode ocorrer por duas rotas metabólicas: a via do ácido mevalônico (MVA), que ocorre predominantemente no citosol e a via metileritritol 4-fosfato (MEP), que ocorre nos plastídios (RODRÍGUEZ-CONCEPCIÓN; BORONAT, 2015). As duas vias são distintas e compartimentadas separadamente para construir unidades de isopreno (HEMMERLIN et al., 2012). A coexistência dessas vias no citosol e nos plastídios pode permitir a síntese de muitos tipos de isoprenoides essenciais para um organismo sésil como as plantas, além disso as duas vias podem estar ativas em paralelo ou separadamente (HEMMERLIN et al., 2012). Por exemplo, a via MVA produz terpenos que estão diretamente relacionados ao crescimento e desenvolvimento vegetal, e a via MEP está relacionada aos pigmentos que realizam a fotossíntese, estas funções podem ser realizadas por ambas as vias em paralelo ou ativadas separadamente dependendo da necessidade fisiológica vegetal (HEMMERLIN et al., 2012).

Assim, o início da via MVA (Figura 2) se dá pela enzima AACT (Acetoacetil-CoA tialose) responsável pela condensação de duas moléculas de acetil-CoA, formando uma molécula intermediária, a acetoacetil-CoA. Em seguida, a enzima 3-hidroxi-3-metilglutaril-CoA sintase (HMGS) realiza a condensação de uma molécula de acetoacetil-CoA com uma molécula de acetil-CoA, formando o 3-hidroxi-3-metil-

glutaril-CoA (HMG-CoA). A enzima HMG-CoA redutase (HMGR) reduz a molécula de HMG-CoA e a converte em ácido mevalônico (MVA) a partir da oxidação de duas moléculas de NADPH. As enzimas mevalonato quinase (MVK), fosfomevalonato quinase (PMK) e mevalonato difosfato descarboxilase (MVD) realizam a conversão de MVA em isopentenil difosfato (IPP). A MVK fosforila o MVA, com o consumo de ATP, e produz o fosfomevalonato (MVP) e adenosina trifosfato (ADP). Em seguida, a PMK catalisa a fosforilação de MVP a partir de ATP, produzindo difosfato de mevalonato (MVPP) e ADP. Finalmente, a MVD catalisa uma reação descarboxilativa, e elimina um grupo carboxila do mevalonato com a hidrólise de ATP, formando o IPP. Posteriormente, o IPP é convertido em dimetilalil difosfato (DMAPP) pela enzima isopentenil difosfato isomerase (IDI) e, em seguida, são biossintetizados os sesquiterpenos e triterpenos, específicos da via MVA (THOLL et al., 2011; RODRÍGUEZ-CONCEPCIÓN; BORONAT, 2015; ZHOU; PICHERSKY, 2020).

No início da via MEP (Figura 2) a enzima 1-Desoxi-D-xilulose 5-fosfato sintase (DXS) é responsável pela condensação de gliceraldeído 3-fosfato (GAP) e piruvato, para produzir uma molécula 1-Desoxi-D-xilulose 5-fosfato (DXP). Em seguida, a enzima DXP redutoisomerase (DXR) catalisa a conversão de DXP em 2-C-metil-D-eritritol 4-fosfato (MEP), por uma redução utilizando uma molécula de NADPH. A enzima MEP citidiltransferase (MCT) então catalisa a reação de conversão do MEP à 4-difosfocitidil-2-C-metil-D-eritritol (CDP-ME). A enzima CDP-ME quinase (CMK) é responsável, em seguida, pela reação de fosforilação de CDP-ME com o consumo de ATP, e produz a molécula 4- difosfocitidil-2-C-metil-D-eritritol 2-fosfato (CDP-MEP) e ADP. A enzima 2-C-metil-D-eritritol-2,4-ciclodifosfato sintase (MDS) converte CDP-MEP à 2-C-metil-D-eritritol-2,4-ciclodifosfato (MEcPP) e CMP. A enzima 4-hidroxi-3-metilbut-2-enil difosfato sintase (HDS) catalisa a conversão de MEcPP em 4-hidroxi-3-metilbut-2-enil difosfato (HMBPP) com a oxidação de um NADPH. Logo depois, a enzima HMBPP reductase (HDR) é responsável pela reação de redução da HMBPP em IPP e/ou DMAPP com oxidação de um NADPH. IPP e DMAPP são as moléculas precursoras da biossíntese dos monoterpenos, diterpenos e tetraterpenos, específicos da via MEP (THOLL et al., 2011; RODRÍGUEZ-CONCEPCIÓN; BORONAT, 2015; ZHOU; PICHERSKY, 2020).

Figura 2. Via de biossíntese de isoprenoides citosólica (MVA) e plastídica (MEP) em plantas.



As enzimas da via MVA são mostradas em azul e as enzimas da via MEP em verde. As setas marcam as etapas das vias. AACT, acetoacetyl-CoA tialose. HMGS, HMG-Coa sintase. HMGR, HMG-CoA redutase. MVA, ácido mevalônico. MVK, mevalonato quinase. MVP, fosfomevalonato. PMK, fosfomevalonato quinase. MVPP, difosfato mevalonato. MVD, mevalonato difosfato descarboxilase. IPP, isopentenil difosfato. DMAPP, dimetilalil difosfato. GAP, gliceraldeído 3-fosfato. DXS, DXP sintase. DXP, 1-desoxi-D-xilulose 5-fosfato. DXR, DXP redutoisomerase. MEP, 2-C-metil-D-eritritol 4-fosfato. MCT, MEP citidiltransferase. CDP-ME, 4- difosfocitidil-2-C-metil-D-eritritol 2-fosfato. CMK, CDP-ME quinase. CDP-MEP, 4- difosfocitidil-2-C-metil-D-eritritol 2-fosfato. MDS, 2-C-metil-D-eritritol-2,4-ciclodifosfato sintase. MEcPP, 2-C-metil-D-eritritol-2,4-ciclodifosfato. HDS, 4-hidroxi-3-metilbut-2-enil difosfato sintase. HMBPP, 4-hidroxi-3-metilbut-2-enil difosfato. HDR, HMBPP reductase. Fonte: Rodríguez-Concepción e Boronat (2015).

1.2. A importância dos genes *HMGR* e *DXS*

Os genes que codificam as enzimas *HMGR* e *DXS* são considerados genes chave para a regulação da biossíntese de isoprenoides e essenciais para as vias MVA e MEP (LI et al., 2014; PAN et al., 2019). Por exemplo, em *Artemisia annua*, gene chave *HMGR* apresentou maiores níveis de atividade transcricional em folha jovem e flor comparado aos níveis vistos em folhas velhas, caule e raízes (OLOFSSON et al., 2011), o que demonstra que a via MVA está ativa no desenvolvimento dos tecidos da planta jovem. Além do elevado número de transcritos detectados para esse gene, análises evolutivas sugeriram que a função de *HMGR* são conservadas em todo o reino vegetal (LI et al., 2014), isso mostra que *HMGR* na via MVA permanece com as mesmas funções na biossíntese de terpenos.

A expressão gênica de *DXS* foi caracterizada em *Oryza sativa*, e observou-se um elevado número de transcritos em folhas, sugerindo que esse gene tem função relevante nos tecidos fotossintéticos (YOU et al., 2020). Análises realizadas em linhagens transgênicas de *Arabidopsis thaliana* mostraram que a indução na expressão do gene *DXS*, aumentou os níveis de compostos derivados de moléculas de isoprenos, como por exemplo, clorofila, tocoferol, carotenoides, ácido abscísico (ABA) e giberelina (GA) (CARRETERO-PAULET et al., 2013). Os mesmos autores observaram que o oposto ocorreu quando os níveis de *DXS* foram suprimidos, houve uma redução na quantidade desses mesmos compostos.

1.2. A importância dos genes *MVK* e *DXR*

Os genes *MVK* e *DXR* também são considerados genes importantes na via de biossíntese de isoprenoides (THOLL, 2015). Entretanto, diferentemente de outros genes dessa via, estas enzimas são codificadas por genes de cópia única em *Arabidopsis thaliana* (HEMMERLIN et al., 2012), mostrando que o gene não precisa de mais outras cópias pra realizar suas funções nas vias de terpenos.

A atividade da *MVK* está envolvida na regeneração e crescimento das plantas, e também está relacionada ao rebrote e formação de raízes em *Pinus strobes* (HEMMERLIN et al., 2012). A *DXR* também é uma enzima importante envolvida na biossíntese de terpenóides pela via plastídica do MEP, ela catalisa a segunda etapa da biossíntese dos terpenóides no cloroplasto (THOLL, 2015). A inativação de *DXR* resulta na parada do desenvolvimento e plantas albinas em *Arabidopsis* (XING et al., 2010).

As algas, plantas basais e superiores são capazes de sintetizar simultaneamente terpenoides pelas vias MEP e MVA em paralelo. A manutenção da presença dos genes de terpenoides em plantas indica que estes compostos provavelmente proporcionaram vantagens evolutivas para as espécies vegetais mais antigas e foram mantidas ao longo da evolução (KUMARI et al., 2013). Desta forma, o estudo sobre a evolução torna-se interessante para essas famílias gênicas (RODRÍGUEZ-CONCEPCIÓN; BORONAT, 2015). A duplicação de genes, mutação e seleção natural são processos essenciais que aumentam a diversidade genética das plantas e podem facilitar a sua adaptação entre a planta e o ambiente (PANCHY et al., 2016; QIAO et al., 2019).

A análise evolutiva de genes específicos destas vias, ainda é incipiente e necessita de estudos mais aprofundados, especialmente em cafeeiros (PICHERSKY; RAGUSO, 2018). Apesar de ter sido estudados em plantas modelos, dado o grande número de terpenoides especializados produzidos por várias espécies vegetais, as principais funções da grande maioria dos genes envolvidos nestas vias ainda não foram bem determinadas para o gênero *Coffea* spp. Os estudos filogenéticos são importantes para uma compreensão dos processos evolutivos que moldaram o metabolismo de isoprenoides em cafeeiro.

1.3. Os isoprenoides no gênero *Coffea* spp

O gênero *Coffea* (Rubiaceae) possui grande importância alimentar e agrônômica (FAOSTAT, 2017). Embora existam 124 espécies descritas para esse gênero (DAVIS et al., 2011), apenas duas possuem maior interesse econômico: *C. canephora* e *C. arabica*. A espécie *C. canephora* é diploide e apresenta 22 cromossomos ($2n = 2x = 22$), enquanto que a *C. arabica* é alotetraploide e possui 44 cromossomos ($2n = 4x = 44$). *C. arabica* provavelmente originou-se de uma hibridização natural que ocorreu entre *C. canephora* e *C. eugenioides* no continente africano (LASHERMES et al., 1999; LASHERMES et al., 2016). Nos últimos anos, a sequência dos genomas de *C. canephora* (DENOEUDE et al., 2014), *C. arabica* (NCBI SRA *accession number* PRJNA497895) e *C. eugenioides* (NCBI SRA *accession number* PRJNA497891) passou a ser disponível em bancos de dados públicos, trazendo uma oportunidade única de estudos bioquímicos, evolutivos e genômicos.

Por exemplo, os estudos bioquímicos relacionados aos lipídios (triterpeno formado a partir da via MVA) são de grande interesse para a qualidade do grão em

cafeeiro (PEREIRA; IVAMOTO, 2015). Os lipídios representam aproximadamente 10-20% de os compostos químicos totais que são encontrados em *Coffea arabica* (SPEER; KÖLLING-SPEER, 2006; PEREIRA; IVAMOTO, 2015). Os lipídios desempenham um papel importante na qualidade da bebida quanto ao sabor e aroma, principalmente por causa da hidrólise de triglicerídeos e o liberação de ácidos graxos que são oxidados e inferem sabores característicos para bebidas de café (CZECH, et al., 2016; MAHMUD; SHELLIE; KEAST, 2020).

Os estudos moleculares sobre a via de isoprenoides no gênero *Coffea* ainda é incipiente. Um estudo anterior avaliou o gene *CaHMGR1* em *C. arabica* e observou atividade transcricional em folha, botão floral, polpa e no início do desenvolvimento do perisperma (TISKI et al., 2011). Genes relacionados a biossíntese de monoterpenos em café já foram funcionalmente validados para a via MVA (DEL TERRA et al., 2013). Os níveis transcricionais dos genes das vias dos isoprenoides (MVA e MEP) já foram estudados em cafeeiros utilizando dados de sequenciamento em larga escala de RNAs mensageiros (RNA-seq), sem genoma de referência, e foi observado que esses genes apresentavam diferentes padrões de expressão de acordo com o órgão e tecido de *C. arabica* analisado (PEREIRA; IVAMOTO, 2015). Quantificações bioquímicas e de atividade transcricional de genes relacionados com a biossíntese de diterpenos em *C. arabica*, em diferentes órgãos e tecidos de cafeeiro, demonstraram que existe uma modulação genética envolvida na produção desses compostos (KITZBERGER et al., 2013; IVAMOTO et al., 2017). Porém, ainda é necessário que mais estudos sejam realizados para definir um número mais preciso de cópias existentes para cada um dos genes das vias MVA e MEP, agora que são disponíveis dados do genoma completo. A realização de análises evolutivas e a determinação do perfil transcricional desses genes nas espécies que deram origem a espécie *C. arabica* (*C. canephora* e *C. eugenioides*) irão aumentar o conhecimento sobre os processos evolutivos que possivelmente moldaram o metabolismo de isoprenoides em cafeeiros.

1.4. A utilização de elicitores para a indução da biossíntese de compostos especializados

Atualmente, sabemos que elicitores podem induzir a produção de terpenoides (ARANEGABOU et al., 2014; MANIVANNAN et al., 2016). Os elicitores são compostos orgânicos produzidos pelas plantas ou animais que atuam como um sistema de alerta da planta (ex: fitormônios) para induzir compostos relacionados a defesa contra

predadores ou patógenos (FINITI et al., 2014). Os elicitores também podem ser sintetizados artificialmente e aplicado na planta, e agem como sinalizadores celulares (ARANEGABOU et al., 2014). O ácido hexanoico (C₆H₁₂O₂), derivado do hexano, é reconhecido como elicitador (SCALSCHI et al., 2013). Estudos realizados em *Arabidopsis* e tomateiros mostram que o ácido hexanoico (Hx) pode proteger as duas espécies vegetais contra o fungo *Botrytis cinerea* (VICEDO et al., 2009; KRAVCHUK et al., 2011). Além disso, já foi observado que o Hx pode atuar na indução de vias da produção de terpenoides, contribuindo com o aumento da expressão de genes de terpenoides em órgãos de *Citrus* e cafeeiro. (LLORENS et al., 2016; SILVA et al., 2020).

1.5. Referencial bibliográfico

ARANEGA-BOU, Paz et al. Priming of plant resistance by natural compounds. Hexanoic acid as a model. **Frontiers in plant science**, v. 5, p. 488, 2014.

CARRETERO-PAULET, Lorenzo et al. Functional and evolutionary analysis of DXL1, a non-essential gene encoding a 1-deoxy-D-xylulose 5-phosphate synthase like protein in *Arabidopsis thaliana*. **Gene**, v. 524, n. 1, p. 40-53, 2013.

CZECH, Hendryk et al. Resolving coffee roasting-degree phases based on the analysis of volatile compounds in the roasting off-gas by photoionization time-of-flight mass spectrometry (PI-TOFMS) and statistical data analysis: toward a PI-TOFMS roasting model. **Journal of agricultural and food chemistry**, v. 64, n. 25, p. 5223-5231, 2016.

DAVIS, Aaron P. et al. Growing coffee: *Psilanthus* (Rubiaceae) subsumed on the basis of molecular and morphological data; implications for the size, morphology, distribution and evolutionary history of *Coffea*. **Botanical Journal of the Linnean Society**, v. 167, n. 4, p. 357-377, 2011.

DENOEUD, France et al. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. **Science**, v. 345, n.6201, p.1181-1184, 2014.

DEL TERRA, Lorenzo et al. Functional characterization of three *Coffea arabica* L. monoterpene synthases: Insights into the enzymatic machinery of coffee aroma. **Phytochemistry**, v. 89, p. 6-14, 2013.

FAOSTAT. Exports/Commodities by Country. Available online: <http://www.fao.org/faostat/en/?#data/TP>

FINITI, Ivan et al. Hexanoic acid protects tomato plants against *Botrytis cinerea* by priming defence responses and reducing oxidative stress. **Molecular Plant Pathology**, v. 15, n. 6, p. 550-562, 2014.

GALATA, Mariana; MAHMOUD, Soheil. Bioactive plant isoprenoids: Biosynthetic and biotechnological approaches. *Studies in Natural Products Chemistry*. Elsevier, v.37, p. 135-171, 2012.

HEMMERLIN, Andréa; HARWOOD, John L.; BACH, Thomas J. A raison d'être for two distinct pathways in the early steps of plant isoprenoid biosynthesis? **Progress in Lipid Research**, v. 51, n. 2, p. 95-148, 2012.

IVAMOTO, Suzana Tiemi et al. Transcriptome analysis of leaves, flowers and fruits perisperm of *Coffea arabica* L. reveals the differential expression of genes involved in raffinose biosynthesis. **PloS One**, v. 12, n. 1, 2017.

KITZBERGER, Cíntia Sorane Good et al. Diterpenes in green and roasted coffee of *Coffea arabica* cultivars growing in the same edapho-climatic conditions. *Journal of Food Composition and Analysis*, v. 30, n. 1, p. 52-57, 2013.

KRAVCHUK, Zhana et al. Priming for JA-dependent defenses using hexanoic acid is an effective mechanism to protect Arabidopsis against *B. cinerea*. **Journal of Plant Physiology**, v. 168, n. 4, p. 359-366, 2011.

KUMARI, Sangita et al. Structural and biochemical perspectives in plant isoprenoid biosynthesis. **Phytochemistry Reviews**, v. 12, n. 2, p. 255-291, 2013.

LASHERMES, Philippe et al. Inter-genomic DNA exchanges and homeologous gene silencing shaped the nascent allopolyploid coffee genome (*Coffea arabica* L.). **G3: Genes, Genomes, Genetics**, v. 6, n. 9, p. 2937-2948, 2016.

LASHERMES, Philippe et al. Molecular characterisation and origin of the *Coffea arabica* L. genome. **Molecular and General Genetics**, v. 261, n. 2, p. 259-266, 1999.

LI, Wei et al. Species-specific expansion and molecular evolution of the 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGR) gene family in plants. **PloS One**, v. 9, n. 4, 2014.

LLORENS, Eugenio et al. Priming by hexanoic acid induce activation of mevalonic and linolenic pathways and promotes the emission of plant volatiles. **Frontiers in Plant Science**, v. 7, p. 495, 2016.

MANIVANNAN, Abinaya et al. Chemical elicitor-induced modulation of antioxidant metabolism and enhancement of secondary metabolite accumulation in cell suspension cultures of *Scrophularia kakudensis* Franch. **International Journal of Molecular Sciences**, v. 17, n. 3, p. 399, 2016.

MAHMUD, MM Chayan; SHELLIE, Robert A.; KEAST, Russell. Unravelling the relationship between aroma compounds and consumer acceptance: Coffee as an example. **Comprehensive Reviews in Food Science and Food Safety**, v. 19, n. 5, p. 2380-2420, 2020.

OLOFSSON, Linda et al. Relative expression of genes of terpene metabolism in different tissues of *Artemisia annual*L. **BMC Plant Biology**, v. 11, n. 1, p. 45, 2011.

PANCHY, Nicholas; LEHTI-SHIU, Melissa; SHIU, Shin-Han. Evolution of gene duplication in plants. **Plant Physiology**, v. 171, n. 4, p. 2294-2316, 2016.

PAN, Xuhao et al. Bioinformatics study of 1-deoxy-d-xylulose-5-phosphate synthase (DXS) genes in Solanaceae. **Molecular Biology Reports**, v. 46, p. 5175-5184, 2019.

PEREIRA, Luiz Filipe Protasio; IVAMOTO, Suzana Tiemi. **Characterization of Coffee Genes Involved in Isoprenoid and Diterpene Metabolic Pathways**. In: Coffee in health and disease prevention. Academic Press, 2015. p. 45-51.

PICHERSKY, Eran; RAGUSO, Robert A. Why do plants produce so many terpenoid compounds? **New Phytologist**, v. 220, n. 3, p. 692-702, 2018.

QU, Yang et al. A lettuce (*Lactuca sativa*) homolog of human Nogo-B receptor interacts with cis-prenyltransferase and is necessary for natural rubber biosynthesis. **Journal of Biological Chemistry**, v. 290, p. 1898-1914, 2015.

RODRÍGUEZ-CONCEPCIÓN, Manuel; BORONAT, Albert. Breaking new ground in the regulation of the early steps of plant isoprenoid biosynthesis. **Current Opinion in Plant biology**, v. 25, p. 17-22, 2015.

SCALSCHI, Loredana et al. Hexanoic acid is a resistance inducer that protects tomato plants against *Pseudomonas syringae* by priming the jasmonic acid and salicylic acid pathways. **Molecular Plant Pathology**, v. 14, n. 4, p. 342-355, 2013.

SILVA, Natacha et al. Low-Copy Genes in Terpenoid Metabolism: The Evolution and Expression of *MVK* and *DXR* Genes in Angiosperms. **Plants**, v. 9, n. 4, p. 525, 2020.

SPEER, Karl; KÖLLING-SPEER, Isabelle. The lipid fraction of the coffee bean. **Brazilian Journal of Plant Physiology**, v. 18, n. 1, p. 201-216, 2006.

TAIZ, Lincoln et al. **Fisiologia e desenvolvimento vegetal**. Artmed Editora, 2017.

TISKI, Iris et al. Characterization and expression of two cDNA encoding 3-Hydroxy-3-methylglutaryl coenzyme A reductase isoforms in coffee (*Coffea arabica* L.). **Omics: a journal of integrative biology**, v. 15, n. 10, p. 719-727, 2011.

THOLL, Dorothea. **Biosynthesis and biological functions of terpenoids in plants**. In: Biotechnology of isoprenoids. Springer International Publishing Switzerland, v. 148, p.63-106, 2015.

THOLL, Dorothea; LEE, Sungbeom. **Terpene specialized metabolism in *Arabidopsis thaliana***. In: The Arabidopsis Book. American Society of Plant Biologists, v. 9, 2011.

VICEDO, Begonya et al. Hexanoic acid-induced resistance against *Botrytis cinerea* in tomato plants. **Molecular Plant-Microbe Interactions**, v. 22, n. 11, p. 1455-1465, 2009.

VRANOVÁ, Eva; COMAN, Diana; GRUISSEM, Wilhelm. Network analysis of the MVA and MEP pathways for isoprenoid synthesis. **Annual Review of Plant Biology**, v. 64, p. 665-700, 2013.

VRIET, Cécile; RUSSINOVA, Eugenia; REUZEAU, Christophe. From squalene to brassinolide: the steroid metabolic and signaling pathways across the plant kingdom. **Molecular Plant**, v. 6, p. 1738-1757, 2013.

XING, Shufan et al. Disruption of the 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR) gene results in albino, dwarf and defects in trichome initiation and stomata closure in Arabidopsis. **Cell Research**, v. 20, n. 6, p. 688-700, 2010.

ZHOU, Fei; PICHERSKY, Eran. More is better: the diversity of terpene metabolism in plants. **Current Opinion in Plant Biology**, v. 55, p. 1-10, 2020.

ZI, Jiachen; MAFU, Sibongile; PETERS, Reuben J. To gibberellins and beyond! Surveying the evolution of (di) terpenoid metabolism. **Annual Review of Plant Biology**, v. 65, p. 259-286, 2014.

YOU, M. K. et al. The organ-specific differential roles of rice DXS and DXR, the first two enzymes of the MEP pathway, in carotenoid metabolism in *Oryza sativa* leaves and seeds. **BMC Plant Biology**, v. 20, p. 1-16, 2020.

2. OBJETIVO GERAL

O objetivo do trabalho foi compreender aspectos de diversidade genômica, evolutiva e transcricional dos genes da via de síntese de isoprenoides em angiospermas, com ênfase no gênero *Coffea*.

2.1. Objetivos específicos

- Identificar os genes iniciais da via de síntese de isoprenoides em *Coffea*, avaliar se o ácido hexanoico é capaz de modular a transcrição destes genes em *C. arabica*, bem como analisar em termos evolutivos os genes chave da via MVA (*HMGR*) e MEP (*DXS*).
- Analisar a evolução dos genes *MVK* (via MVA) e *DXR* (via MEP) em espécies de angiospermas e avaliar se o ácido hexanoico é capaz de modular o perfil transcricional desses genes em plantas de *C. arabica*.

3. CAPÍTULO 1: ANÁLISE GENÔMICA DAS VIAS DE BIOSÍNTESE DE ISOPRENOIDES EM *Coffea* spp.

3.1. Introdução

Os isoprenoides ou terpenoides são derivados de duas moléculas precursoras: o isopentenil difosfato (IPP) e seu isômero dimetilalil difosfato (DMAPP) (PICHESKY; RAGUSO, 2018). Existem duas vias metabólicas capazes de produzir o IPP e o DMAPP em plantas superiores: a via citosólica do mevalonato ou ácido mevalônico (MVA) e a via plastidial do metileritritol fosfato (MEP) (PICHESKY; RAGUSO, 2018).

A via MVA inicia-se pelas reações químicas ligadas as enzimas acetoacetil-CoA tialose (AACT), 3-hidroxi-3-metilglutaril-Coa sintase (HMGS), HMG-CoA redutase (HMGR), mevalonato quinase (MVK), fosfomevalonato quinase (PMK) e mevalonato difosfato descarboxilase (MVD). E na via MEP, pelas reações enzimáticas de 1-Desoxi-D-xilulose 5-fosfato sintase (DXS), 1-Desoxi-D-xilulose 5-fosfato redutoisomerase (DXR), 2-C-metil-D-eritritol 4-fosfato citidiltransferase (MCT), 4-difosfocitidil-2-C-metil-D-eritritol quinase (CMK), 2-C-metil-d-eritritol-2,4-ciclodifosfato sintase (MDS), 4-hidroxi-3-metilbut-2-enil difosfato sintase (HDS) e 4-hidroxi-3-metilbut-2-enil difosfato redutase (HDR) (RODRÍGUEZ-CONCEPCIÓN; BORONAT, 2015).

O gene *HMGR* da via MVA, é importante pois codifica a enzima catalisadora responsável pela conversão de HMG-CoA em ácido mevalônico (LI et al., 2014). O gene *DXS* também é importante por codificar a enzima responsável pela condensação de gliceraldeído 3-fosfato (GAP) mais piruvato, que dá início ao primeiro passo da via MEP. As enzimas codificadas por estes genes são consideradas as etapas limitantes nas duas vias de síntese de terpenoides (LI et al., 2014; PAN et al., 2019).

Uma cópia do gene *HMGR* (*CaHMGR1*) apresentou atividade transcricional em folha, botão floral, polpa e início do desenvolvimento do perisperma em *C. arabica* (TISKI et al., 2011). Observou-se também que a super expressão de *HMGR* em folha de tabaco influenciou o aumento da concentração de sesquiterpenos (β -cariofileno e 5-epi-aristolocheno) e monoterpenos (β -ocimeno e linalol) nas plantas de *Nicotiana tabacum* (HENRY et al., 2018). Além das altas taxas de expressão do gene, análises evolutivas sugeriram que a função e evolução da família gênica de *HMGR* foram conservadas em angiospermas (LI et al., 2014).

O gene *DXS* também está relacionado a etapas regulatórias para a biossíntese de isoprenoides da via MEP nos cloroplastos (XU et al., 2019). A enzima 1-Desoxi-D-xilulose 5-fosfato sintase (DXS) é responsável pela condensação de gliceraldeído 3-fosfato (GAP) e piruvato, para produzir uma molécula 1-Desoxi-D-xilulose 5-fosfato

(DXP), um passo importante para o início da via (CARRETERO-PAULET et al., 2013; YOU et al., 2020). Os níveis transcricionais de *DXS* foram altos em folhas de arroz (*Oryza sativa*), sugerindo um papel relevante desse gene nos tecidos fotossintéticos (YOU et al., 2020). Uma alta expressão de *DXS* em tricomas de tomate (*Solanum lycopersicum*) evidenciou uma relação diretamente proporcional com o conteúdo de sesquiterpenos (HOFBERGER et al., 2015). O contrário, o silenciamento de *DXS*, foi observado em estudos com *Arabidopsis* em que a planta apresentou folhas de tamanhos menores, falta de pigmentação e alteração nos plastídios e até mesmo fenótipo albino (ESTÉVEZ et al., 2000; ESTÉVEZ et al., 2001; CROWELL et al., 2003; CARRETERO-PAULET et al., 2006).

Além dos estudos relacionados ao perfil transcricional dos genes das vias de terpenos, são de grande importância a compreensão da evolução desses genes, para maior embasamento sobre a biossíntese de terpenoides. Estudos filogenéticos podem permitir a formulação de hipóteses sobre a evolução da estrutura e função dos genes (LI et al., 2014). Como por exemplo, *HMGR* e *DXS*, genes da síntese de terpenoides, são altamente conservados em plantas superiores - monocotiledôneas e eudicotiledôneas (LI et al., 2014; PAN et al., 2019)

Para aumentar ainda mais o conhecimento e embasamento sobre a produção de terpenoides, alguns estudos utilizaram a aplicação de ácido hexanoico (Hx) como um elicitador de compostos de metabolismo especializado e observaram que ocorreu a modulação da atividade transcricional dos genes envolvidos na via de biossíntese de terpenoides (NIINEMETS, 2010; SCALSCHI et al., 2013; LLORENS et al., 2016). Os elicitadores são compostos orgânicos produzidos pelas plantas ou animais que atuam como um sistema de alerta da planta (ex: fitormônios) para induzir a produção de compostos que promovam a defesa da planta contra o ataque de pragas e/ou patógenos (PARÉ et al., 2005). Além disso, os elicitadores também podem ser sintetizados artificialmente, disponibilizados junto com a solução nutritiva e podem atuar na indução da produção de moléculas sinalizadoras de estresse ambiental para que a planta possa se proteger (ARANEGABOU et al., 2014).

Em *Citrus sinensis* foi observado que o Hx induz a resistência efetiva da planta contra o fungo *Alternaria alternata*. Além disso, foi observado alta expressão dos genes *MVD* e *HMGR* da via MVA em Citros com a presença do Hx (LLORENS et al., 2016).

Estudos realizados em *Arabidopsis* e tomateiros mostram que o Hx pode ativar compostos de proteção dessas espécies vegetais contra o fungo *Botrytis cinerea* (KRAVCHUK et al., 2011; FINITI et al., 2014). Hx também pode ativar compostos de proteção em batatas contra larvas do besouro *Leptinotarsa decemlineata* (LÓPEZ-GALIANO et al., 2019). O Hx, como elicitor, induz a ativação de diversos mecanismos de respostas de defesa da planta (ARANEGABOU et al., 2014), e já foram observadas evidências de indução transcricional dos genes das vias MVA em outros sistemas vegetais (LLORENS et al., 2016).

Este trabalho teve como objetivo: i) identificar o número de cópias dos genes envolvidos nos passos iniciais da via de biossíntese de isoprenoides em *C. arabica*; ii) avaliar se a aplicação exógena de Hx em solução nutritiva é capaz de modular os níveis transcricionais de genes dessa via de folhas e raízes laterais a transcrição dos genes *HMGR* e *DXS* em cafeeiros; iii) compreender a filogenia dos genes *HMGR* e *DXS* do gênero *Coffea* e espécies de angiospermas.

3.2 Material e métodos

3.2.1. Material vegetal, tratamento com ácido hexanoico e extração de RNA total

A espécie utilizada no estudo com ácido hexanoico foi *Coffea arabica* cultivares Catuaí Vermelho e Obatã. A espécie já vem sendo pesquisada, principalmente os tecidos folhas e raízes, pelo grupo de pesquisa do laboratório GeTransPlant – Unesp Rio Claro, coordenada pelo orientador do atual trabalho, Prof. Dr. Douglas Silva Domingues. Logo, esta pesquisa de doutorado dá continuidade aos trabalhos de pesquisas do grupo. A espécie *Coffea arabica* cultivar Catuaí Vermelho é importante devido à grande adaptabilidade ao ambiente de cultivo. A Catuaí é originada do cruzamento entre as outras cultivares de arabica que são Mundo Novo e Caturra Amarelo. Ela apresenta como características a alta produtividade e rusticidade da Mundo Novo e pequeno porte da Caturra. Podem apresentar frutos com coloração avermelhada ou amarelada e alta adaptação em condições de seca (Disponível em: <http://www.consorciopesquisacafe.com.br/>). A cultivar Obatã também é de grande importância para as lavouras de café por ser de alta resistência a doenças, como por exemplo a ferrugem. É derivada do cruzamento das cultivares de arabica Villa Sarchi e Híbrido de Timor e possível hibridização natural com Catuaí Vermelho. Ela apresenta porte baixo, exigência em nutrição, maturação tardia com frutos grandes vermelhos e

alta produtividade podendo superar Catuaí Vermelho em determinadas regiões (Disponível em: <http://www.consorciopesquisacafe.com.br/>).

Assim, as plantas com cinco meses de idade (3 a 4 pares de folhas) de *C. arabica* cv. Catuaí Vermelho IAC 144 e Obatã IAC 1669-20 foram selecionadas por uniformidade de desenvolvimento e transferidas para vasos (3 L) contendo solução nutritiva aerada (SNA), adaptada de Clark (1975) por de Carvalho et al. (2013). A solução nutricional foi composta por K_2SO_4 (1068 mM), $MgSO_4 \cdot 7H_2O$ (332,5 mM), KH_2PO_4 (266 mM), $CaCl_2 \cdot 2H_2O$ (66 mM) e NH_4NO_3 (5333 mM). Ferro e micronutrientes foram fornecidos usando uma mistura de sais quelada comercial (ConMicros Standard, Conplant) nas seguintes concentrações ($0,2 \text{ mg} \cdot L^{-1}$): Fe (363), Cu (91), Zn (37), Mn (91), B (91), Mo (18) e Ni (17). Os experimentos foram realizados em sala de crescimento de plantas, com a temperatura ($23^\circ C \pm 2^\circ C$), fotoperíodo (12 h) e pH (5,5 ~ 5,6) controlados diariamente. As plantas foram mantidas sob iluminação de painéis de LED ($400 \mu\text{mol} \cdot m^{-2} \cdot s^{-1}$).

Após um período de 96 h de aclimação na sala de crescimento, os seguintes tratamentos foram aplicados: a) SNA (controle) e b) SNA + 0,55 mM de ácido hexanóico (Merck, Darmstadt, Alemanha) por 48 h. Cada tratamento foi composto por seis vasos, cada um contendo três plantas. Cada grupo de vasos (composto de 18 plantas) foi considerado uma replicata biológica. As folhas maduras do terço médio e as raízes laterais das plantas foram coletadas três horas após o início do fotoperíodo, em seguida foram maceradas em nitrogênio líquido e armazenadas em freezer a $-80^\circ C$ até a extração do RNA. O RNA total dos tecidos foi extraído usando o kit RNeasy Plant, e purificado utilizando o kit RNeasy MinElute Cleanup Kit (Qiagen, Hilden, Renânia do Norte-Vestfália, Alemanha). A integridade das amostras de RNA foi verificada por eletroforese em gel de agarose a 1% e o RNA foi quantificado utilizando o fluorômetro Qubit (Thermo Fisher).

3.2.2. Análise do perfil transcricional dos genes das vias MVA e MEP em *C. arabica*.

Os RNAs totais de folhas maduras e raízes laterais foram utilizados para o preparo de bibliotecas de sequenciamento de alto rendimento, em um total de 24 bibliotecas (2 tecidos, 2 condições - controle e tratamento com Hx, 2 cultivares e 3 replicatas biológicas). O preparo de bibliotecas poli-A e sequenciamento foram realizados pela empresa LC Sciences (Houston, Texas, Estados Unidos da América). As bibliotecas foram sequenciadas na plataforma Illumina NovaSeq, com leituras de

150 pares de base (pb), *paired-end*. A qualidade das sequências nucleotídicas foi verificada com o programa FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) e apenas as que possuíam $Q > 20$ foram mantidas. As sequências foram mapeadas no genoma de referência público de *C. arabica* disponível na plataforma NCBI (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Coffea_arabica/100/), com o programa HISAT2 (KIM et al., 2015). A contagem de sequências do RNA-seq mapeadas no genoma de referência foi realizada com o programa StringTie (PERTEA et al., 2015). Os resultados de quantificação de atividade transcricional *in silico* (*digital gene expression* - DGE) foram obtidos em valores de FPKMs (*Fragments Per Kilobase per Million*).

As análises estatísticas foram realizadas no programa XLSTAT v.2014. Os valores de expressão (FPKM) de cada uma das amostras em triplicatas, de folhas e raízes controles e tratadas com Hx, foram transformadas em Log_{10} para a conversão dos dados em distribuição normal. Em seguida foram feitos os Testes *T* de Student - two sample ($p < 0,05$) nas triplicatas de cada um dos genes das vias MVA e MEP, nos dois órgãos (folha e raiz) separadamente. Ou seja, os valores de expressões de cada um do gene em folha controle (Folha C) foram comparados com folha tratada com Hx (Folha Hx). E os valores de expressões de cada gene em raiz controle (Raiz C) foram comparados com raiz tratada com Hx (Raiz Hx). As figuras com os valores de expressão gênica *in silico* (*Digital Gene Expression* - DGE) foram desenvolvidas com a ferramenta online *BAR HeatMapper Plus Tool* (<http://bar.utoronto.ca/welcome.htm>) e editadas no programa CoreIDRAW 2019 (64-Bit).

3.2.3. Identificação dos genes das vias MVA e MEP nos genomas de *Coffea* spp

Foram utilizados como *queries* as sequências de proteína (aminoácidos) dos genes *AACT* (*At5g48230*), *HMGS* (*At4g11820*), *HMGR* (*At1g76490*), *MVK* (*At5g27450*), *PMK* (*At1g31910*), *MVD* (*At2g38700*), *DXS* (*At4g15560*), *DXR* (*At5g62790*), *MCT* (*At2g02500*), *CMK* (*At2g26930*), *MDS* (*At1g63970*), *HDS* (*At5g60600*) e *HDR* (*At4g34350*) de *Arabidopsis thaliana*. Estes genes foram comparados contra os genomas das espécies *C. canephora*, *C. eugeniodies* e *C. arabica* através da ferramenta Blast. Foi utilizado o genoma de *C. canephora* disponível na plataforma PLAZA 4.0 (VAN BEL et al., 2017) e os dados do *GenBank* do *National Center for Biotechnology Information* (SAYERS et al., 2019) para *C.*

arabica e *C. eugenoides*. Apenas os genes que apresentaram valores de *score* ≥ 200 e *E-value* $\leq e-50$ foram selecionados para análises posteriores.

3.2.4. Análise filogenética dos genes *HMGR* e *DXS*

A seleção das espécies para as análises evolutivas foram a partir das espécies que se tem genoma completo e disponível no banco de dados genômicos públicos. Além disso, foram baseadas ao que se tem na literatura para análises filogenéticas dos genes *HMGR* e *DXS* (Li et al 2014, Pan et al., 2019, You et al. 2020).

As sequências codificantes (CDSs) de *HMGR* e *DXS* de cinco espécies de plantas angiospermas (*Sorghum bicolor*, *Oryza sativa*, *Zea mays*, *Carica papaya*, *Vitis vinifera*) foram obtidas por BLAST na plataforma PLAZA 4.0 (VAN BEL et al., 2017) utilizando sequências de *A. thaliana* como *queries*, com os mesmos procedimentos descritos no item 3.2.3. As sequências nucleotídicas codificadoras foram traduzidas em aminoácidos e alinhadas com o software MUSCLE (EDGAR, 2004), usando a ferramenta TranslatorX (ABASCAL et al., 2010). As árvores filogenéticas dos genes *HMGR* e *DXS* foram inferidas pelo método *Neighbor-Joining*, baseado no modelo LG (LE; GASCUEL, 2008). O *bootstrap* da árvore foi inferido a partir de 1000 réplicas e foi utilizado para representar a história evolutiva dos genes de interesse. A análise evolutiva e as árvores finais foram obtidas utilizando o programa MEGAX (KUMAR et al., 2018). Os nomes e as chaves dos Grupos I, II, III e IV de *HMGR* (LI et al., 2014) e dos Tipos I, II e III de *DXS* (YOU et al., 2020) foram editadas no programa CoreIDRAW 2019 (64-Bit).

3.2.5. Determinação das estruturas gênicas de *HMGR* e *DXS* em plantas

Foi realizada a análise de estrutura gênica de *HMGR* e *DXS* nas seis espécies de plantas angiospermas citadas no item 3.2.4 (*A. thaliana*, *C. papaya*, *O. sativa*, *S. bicolor*, *V. vinifera*, *Z. mays*), incluindo as três espécies de *Coffea*.

As sequências de nucleotídeos foram recuperadas na plataforma PLAZA 4.0 (VAN BEL et al., 2017) para as espécies de angiospermas e *C. canephora*, e na base de dados *GenBank* do *National Center for Biotechnology Information* (SAYERS et al., 2019) para *C. arabica* e *C. eugenoides*. Para analisar a estrutura éxon-íntron dos genes *HMGR* e *DXS*, utilizamos o programa *Gene Structure Display Server v2.0* (HU et al., 2014). O CDS no formato FASTA correspondentes aos genes das espécies

foram inseridos, alinhadas por Est2genome, para gerar estruturas gênicas. Em seguida, cada estrutura éxon-íntron foram adicionadas aos ramos de cada uma das espécies, juntos nas árvores filogenéticas geradas no MEGAX, e realizados as figuras finais (árvore mais estruturas éxon-íntron) no programa CoreIDRAW 2019 (64-Bit).

3.3 Resultados e discussão

3.3.1 Identificação do número de cópias dos genes das vias MVA e MEP em *Coffea* spp

Em *C. canephora* observou-se que a maior parte dos genes da via MVA (*CcAACT*, *CcHMGS*, *CcMVK*, *CcPMK*, *CcMVD*) são codificados por uma única cópia gênica, com exceção de *HMGR*, que possui dois genes codificantes (Tabela 1). *C. eugenoides* apresentou genes de cópia única (*CeHMGS*, *CeMVK*), duas cópias (*CeHMGR*, *CeMVD*), e quatro cópias (*CeAACT*, *CePMK*) (Tabela 1). Na espécie alotetraplóide *C. arabica*, foi observado um número maior para quase todos os genes: uma cópia para *CaAACT*, duas cópias para *CaHMGS* e *CaMVK*, três cópias para *CaHMGR*, *CaPMK* e quatro cópias para *CaMVD*. Além disso, o ganho de um gene para *CaMVD* em *C. arabica* sugere uma duplicação recente desse gene (Tabela 1).

As três espécies do genero *Coffea* spp apresentaram mais de uma cópia do gene *HMGR*. O número de cópias de *HMGR* identificados em *C. canephora* (2) e *C. eugenoides* (2), foi diferente da observada em *C. arabica* (3) (Tabela 1).

C. canephora e *C. eugenoides* são codificadas por duas cópias de *HMGR*, em cada uma das duas espécies, o mesmo número de cópias de *HMGR* foram vistos em *Cucumis melo*, *Citrus sinensis* e *Arabidopsis thaliana* e (Lange e Ghassemian, 2003 (esta citação está em Tiski 2011); Li et al., 2014). Assim, *HMGR* ser codificado por duas cópias sugere que as duas cópias tenha um papel importante para a biossíntese de isoprenóides na via MVA. Porém é válido ressaltar que número de cópias de alguns genes das vias MVA e MEP podem variar de espécie pra espécie (PICHERSKY; RAGUSO, 2018).

Como por exemplo, em estudos anteriores observaram uma cópia do gene *HMGR* para a codificação desta proteína em *C. arabica* (TISKI et al., 2011; PEREIRA; IVAMOTO, 2015). Porém, os nosso estudos mostraram que existem três cópias que codificam *HMGR* para *C. arabica*.

Tabela 1. Genes codificadores das enzimas da via MVA em *C. canephora*, *C. eugenioides* e *C. arabica*

Via MVA	<i>C. canephora</i>	<i>C. eugenioides</i>	<i>C. arabica</i>
AACT	Cc05_g01610	XM_027314663	XM_027208355
		XM_027314664	
		XM_027314665	
		XM_027314666	
HMGS	Cc08_g13850	XM_027325230	XM_027225456
			XM_027229425
HMGR	Cc10_g00230	XM_027293193	XM_027218637
	Cc07_g12220	XM_027322981	XM_027222532
MVK	Cc00_g15600	XM_027304863	XM_027235546
			XM_027251860
PMK	Cc08_g03790	XM_027327059	XM_027212763
		XM_027327060	XM_027226201
		XM_027327061	XM_027226202
		XM_027327062	
MVD	Cc02_g09750	XM_027306037	XM_027248959
		XM_027306039	XM_027248960
			XM_027255523
			XM_027255524

Assim como na via MVA, a via MEP também apresentou diferenças no número de cópias gênicas para as três espécies de *Coffea* spp analisadas. *C. canephora* apresentou o menor número de cópias para quase todas as enzimas da via MEP. Seis genes foram de cópia única (*CcDXR*, *CcMCT*, *CcCMK*, *CcMDS*, *CcHDS*, *CcHDR*) e um gene com três cópias (*CcDXS*) (Tabela 2). Em *C. eugenioides*, cinco genes apresentam cópia única (*CeDXR*, *CeMCT*, *CeCMK*, *CeMDS*, *CeHDS*), um gene com duas cópias (*CeHDR*) e um gene com quatro cópias (*CeDXS*). Novamente, *C. arabica* apresentou um número maior de cópias para a maioria dos genes, no qual dois genes são de cópia única (*CaMCT*, *CaCMK*), quatro genes com duas cópias (*CaDXR*, *CaMDS*, *CaHDS*, *CaHDR*) e um gene com seis cópias (*CaDXS*) (Tabela 2).

DXS foi o gene que apresentou o maior número de cópias na via MEP, nas três espécies de *Coffea* spp. O número de cópias gênicas de *DXS* em *C. arabica* (6) foi diferente do somatório das espécies *C. canephora* (3) e *C. eugenioides* (4). Um estudo anterior identificou quatro cópias de *DXS* em *C. arabica* utilizando dados de RNA-seq (PEREIRA; IVAMOTO, 2015), porém estes dados não englobaram todos os órgãos da planta, apenas folhas, flores e frutos.

Tabela 2. Genes codificadores das enzimas da via MEP em *C. canephora*, *C. eugenioides* e *C. arabica*

Via MEP	<i>C. canephora</i>	<i>C. eugenioides</i>	<i>C. arabica</i>
DXS	<i>Cc06_g13680</i>	<i>XM_027309974</i>	<i>XM_027211297</i>
	<i>Cc07_g13880</i>	<i>XM_027318006</i>	<i>XM_027215027</i>
	<i>Cc03_g02540</i>	<i>XM_027318004</i>	<i>XM_027260306</i>
		<i>XM_027322583</i>	<i>XM_027265354</i>
			<i>XM_027222632</i>
DXR	<i>Cc04_g14010</i>	<i>XM_027311920</i>	<i>XM_027267332</i>
			<i>XM_027270403</i>
MCT	<i>Cc02_g23630</i>	<i>XM_027305187</i>	<i>XM_027257511</i>
CMK	<i>Cc01_g07140</i>	<i>XM_027305705</i>	<i>XM_027226336</i>
			<i>XM_027247923</i>
MDS	<i>Cc08_g15230</i>	<i>XM_027324840</i>	<i>XM_027229283</i>
			<i>XM_027224378</i>
HDS	<i>Cc01_g11160</i>	<i>XM_027324546</i>	<i>XM_027230473</i>
			<i>XM_027252084</i>
HDR	<i>Cc02_g17960</i>	<i>XM_027303576</i>	<i>XM_027256537</i>
		<i>XM_027303260</i>	<i>XM_027250035</i>

3.3.2 Perfis transcricionais dos genes das vias MVA e MEP

Foi realizada uma análise do perfil transcricional *in silico* dos genes das vias MVA e MEP, em folhas e raízes de cafeeiros arábica das variedades Catuaí Vermelho (CaCV; Figuras 1A e 2A) e Obatã (CaOB; Figuras 1B e 2B), observando-se o efeito da aplicação de Hx.

Dentre os 15 genes da via MVA analisados em CaCV, 13 deles apresentaram indução em sua atividade transcricional em folhas quando expostas ao elicitador Hx: *AACT* (*XM_027208355*), *HMGS* (*XM_027225456*), *HMGR* (*XM_027218637*, *XM_027222532*), *MVK* (*XM_027251860*, *XM_027258791*), *PMK* (*XM_027212763*, *XM_027226201*, *XM_027226202*) e *MVD* (*XM_027248959*, *XM_027255523*, *XM_027255524*, *XM_027248960*). E sete genes de CaCV apresentaram uma redução em seus níveis de expressão: *AACT* (*XM_027208355*), *HGMS* (*XM_027229425*), *HMGR* (*XM_027218637*, *XM_027222532*), *MVK* (*XM_027251860*), *PMK* (*XM_027226201*, *XM_027226202*) em raízes tratadas com Hx em comparação as plantas controle.

Os dados sugerem que o Hx tem potencial de aumentar a expressão da maioria dos genes na via MVA e em folhas. Em estudos da espécie *Citrus sinensis* foram

vistos maior expressão dos genes *HMGR* e *MVD* da via MVA em folhas tratadas com Hx (LLORENS et al., 2016).

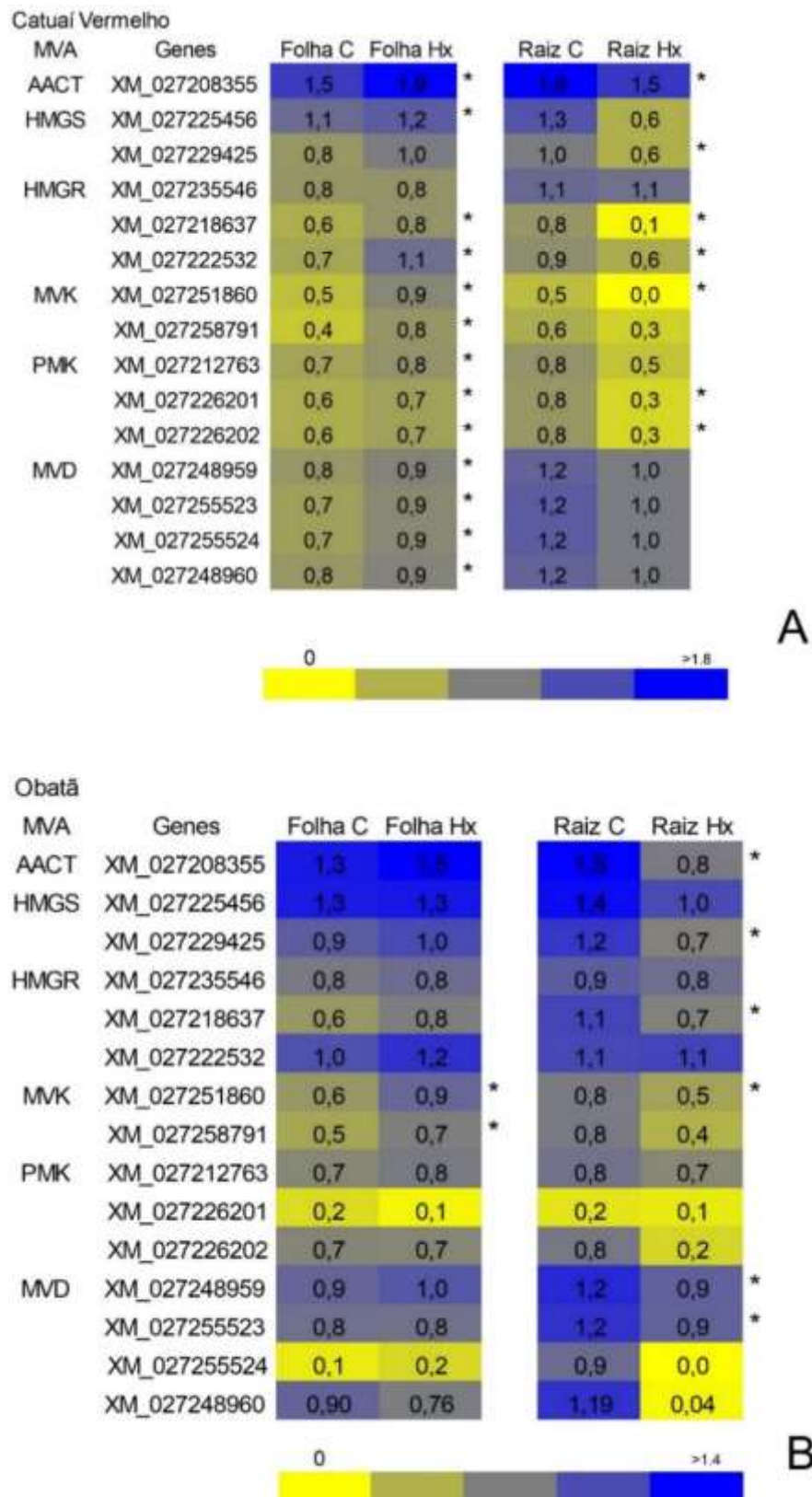
Em folhas de CaOB, apenas os dois genes *MVK* (*XM_027251860*, *XM_027258791*) da via MVA apresentaram diferenças significativas na expressão e foram induzidos na presença de Hx (Figura 1B). Nas raízes de CaOB, os genes *AACT* (*XM_027208355*), *HMGS* (*XM_027229425*), *HMGR* (*XM_027218637*), *MVK* (*XM_027251860*) e *MVD* (*XM_027248959*, *XM_027255523*) da via MVA foram reprimidos na presença de Hx (Figura 1B).

A expressão dos genes envolvidos na via MEP também foram modulados por Hx, isto ocorreu de maneira oposta em folhas e raízes, onde foi observada uma predominância de indução e repressão, respectivamente. Em folhas de CaCV, sete genes dos 17 genes analisados foram estatisticamente mais expressos na presença de Hx (Figura 2A): *DXS* (*XM_027211297*, *XM_027215027*, *XM_027260306*), *DXR* (*XM_027267332*), *CMK* (*XM_027247923*, *XM_027226336*) e *MDS* (*XM_027224378*). Nas raízes, o Hx reduziu significativamente a expressão dos genes *DXS* (*XM_027222632*, *XM_027220323*) e *DXR* (*XM_02726732*, *XM_027270403*).

Com relação os níveis transcricionais dos genes da via MEP em CaOB, o Hx aumentou significativamente a expressão de 11 genes em folhas : *DXS* (*XM_027211297*, *XM_027215027*, *XM_027222632*, *XM_027220323*), *DXR* (*XM_027267332*, *XM_027270403*), *CMK* (*XM_027247923*, *XM_027226336*), *MDS* (*XM_027229283*) e *HDS* (*XM_027230473*, *XM_027252084*). Nas raízes dessa cultivar, o Hx reprimiu significativamente a expressão dos genes *DXS* (*XM_027215027*, *XM_027222632*) e *DXR* (*XM_027267332*, *XM_027270403*) (Figura 2B). A análise geral dos genes da via MEP também mostrou que uma das cópias gênicas de *HDR* (*XM_027256537*) foi bastante expresso nos dois tecidos analisados e pouco alterado por Hx nas duas variedades (Figura 2).

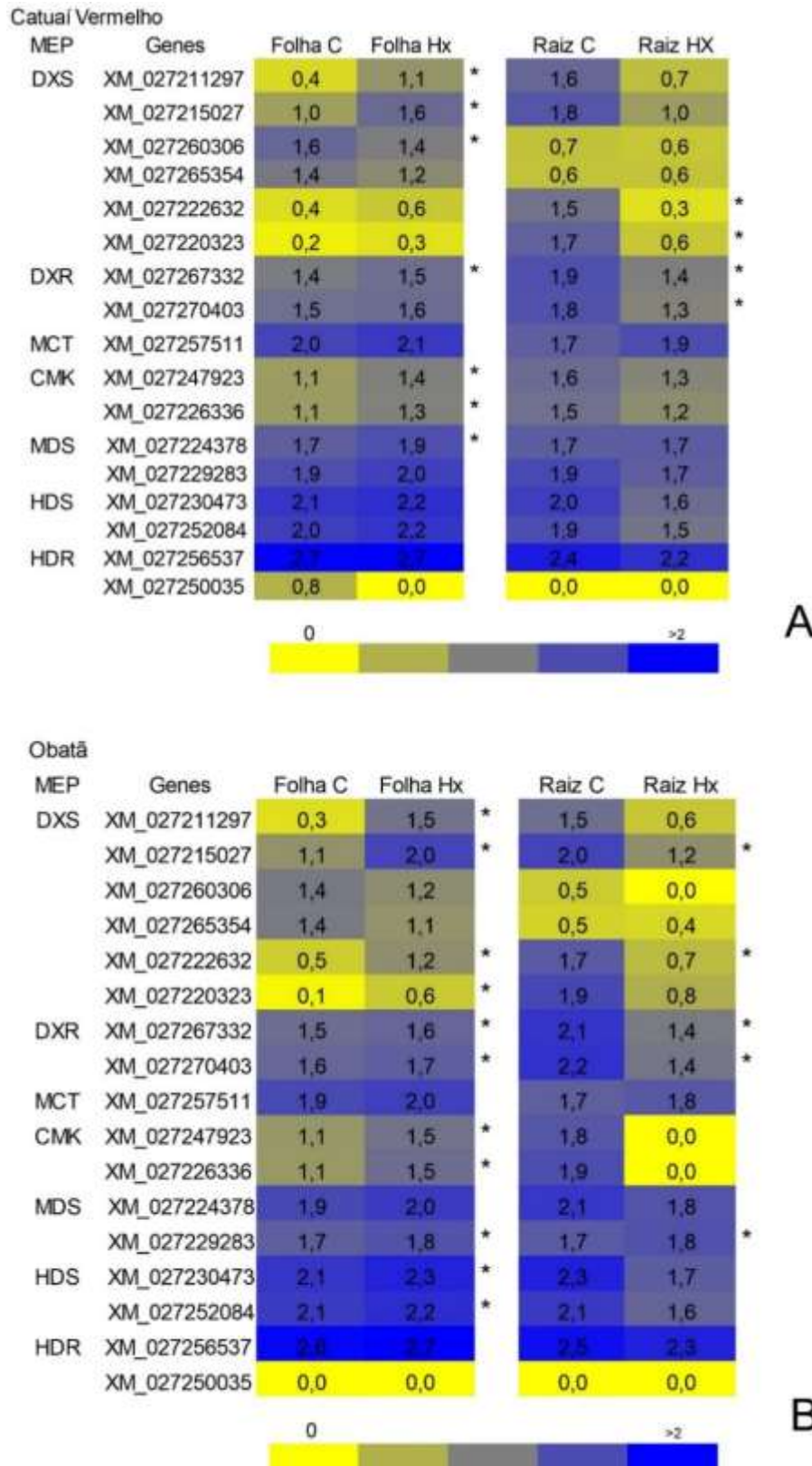
Assim, observou-se que o padrão geral de resposta ao tratamento com Hx para os níveis de expressão das duas vias foi indução nas folhas e repressão nas raízes, tanto em CaCV quanto em CaOB. Provavelmente, o Hx atua mais, quanto aos melhores níveis de expressão dos genes, nos tecidos de folha.

Figura 1. Perfil transcricional dos genes da via MVA



C. arabica cv. Catuaí Vermelho (A) e *C. arabica* cv. Obatã (B), expressão dos genes em FPKM em folha e raiz controle (C) e tratamento com ácido hexanoico (Hx). Os valores de FPKM estão representados em escala de Log₁₀. Foram realizados Testes *t* de Student ($p < 0,05$) para os valores amostrais e os asteriscos (*) indicam diferenças significativas entre os tratamentos (controle e Hx) e tecidos (folhas e raízes), separadamente. Fonte: próprio autor.

Figura 2. Perfil transcricional dos genes da via MEP



C. arabica cv. Catuaí Vermelho (A) e *C. arabica* cv. Obatã (B), em folha e raiz controle (C) e tratamento com ácido hexanoico (Hx). Os valores de FPKM estão representados em escala de Log₁₀. Foram realizados Testes *t* de Student ($p < 0,05$) para os valores amostrais e os asteriscos (*) indicam diferenças significativas entre os tratamentos (controle e Hx) e tecidos (folhas e raízes), separadamente. Fonte: próprio autor.

3.3.3 Número de cópias e análise filogenética dos genes *HMGR* e *DXS*

Foram identificados os números de cópias dos genes de *HMGR* e *DXS* em espécies monocotiledôneas e eudicotiledôneas selecionadas (Tabela 3). Duas cópias de *HMGR* foram identificados em *C. canephora*, *C. eugenioides*, *A. thaliana*, *V. vinifera* e *O. sativa*. *C. arabica*, *C. papaya* e *S. bicolor* apresentaram três cópias de *HMGR* e *Z. mays* foi a espécie com maior número de genes *HMGR*, com cinco cópias (Tabela 3).

DXS apresentou três cópias para as espécies *C. canephora*, *A. thaliana*, *S. bicolor* e *Z. mays*, enquanto que em *C. eugenioides* foi observado quatro cópias. *C. arabica* e *V. vinifera* foram as espécies com o maior número de *DXS* (6). *O. sativa* apresentou dois genes *DXS* e a espécie com o menor número foi *C. papaya* com apenas um gene (Tabela 3).

Tabela 3. Identificação dos genes *HMGR* e *DXS* em angiospermas

Linhagem	Espécie	Gene ID		
		<i>HMGR</i>	<i>DXS</i>	
Eudicotiledôneas	<i>Coffea canephora</i>	<i>Cc10_g00230</i>	<i>Cc06_g13680</i>	
		<i>Cc07_g12220</i>	<i>Cc07_g13880</i>	
	<i>Coffea eugenioides</i>	<i>XM_027293193</i>	<i>Cc03_g02540</i>	<i>XM_027309974</i>
		<i>XM_027322981</i>		<i>XM_027318006</i>
				<i>XM_027318004</i>
				<i>XM_027322583</i>
	<i>Coffea arabica</i>	<i>XP_027091347</i>		<i>XP_027121155</i>
		<i>XP_027074438</i>		<i>XP_027116107</i>
		<i>XP_027078333</i>		<i>XP_027067098</i>
				<i>XP_027070828</i>
				<i>XP_027078433</i>
				<i>XP_027076124</i>
	<i>Arabidopsis thaliana</i>	<i>AT1G76490</i>		<i>AT4G15560</i>
		<i>AT2G17370</i>		<i>AT3G21500</i>
			<i>AT5G11380</i>	
<i>Carica papaya</i>	<i>Cpa.g.sc232.9</i>		<i>Cpa.g.sc16.90</i>	
	<i>Cpa.g.sc6.348</i>			
	<i>Cpa.g.sc26.7</i>			
<i>Vitis vinifera</i>	<i>GSVIVG01026444001</i>		<i>GSVIVG01029109001</i>	
	<i>GSVIVG01013435001</i>		<i>GSVIVG01017832001</i>	
			<i>GSVIVG01029112001</i>	
			<i>GSVIVG01003955001</i>	
			<i>GSVIVG01029169001</i>	
			<i>GSVIVG01035799001</i>	
Monocotiledôneas	<i>Sorghum bicolor</i>	<i>Sobic.002G250400</i>	<i>Sobic.002G064500</i>	
		<i>Sobic.004G262100</i>	<i>Sobic.009G135500</i>	
		<i>Sobic.007G207600</i>	<i>Sobic.010G032900</i>	
	<i>Oryza sativa</i>	<i>Os09g31970</i>		<i>Os05g33840</i>
		<i>Os08g40180</i>		<i>Os07g09190</i>
	<i>Zea mays</i>	<i>Zm00001d006040</i>		<i>Zm00001d045383</i>
		<i>Zm00001d017826</i>		<i>Zm00001d038170</i>
		<i>Zm00001d030595</i>		<i>Zm00001d019060</i>
		<i>Zm00001d051626</i>		
		<i>Zm00001d020963</i>		

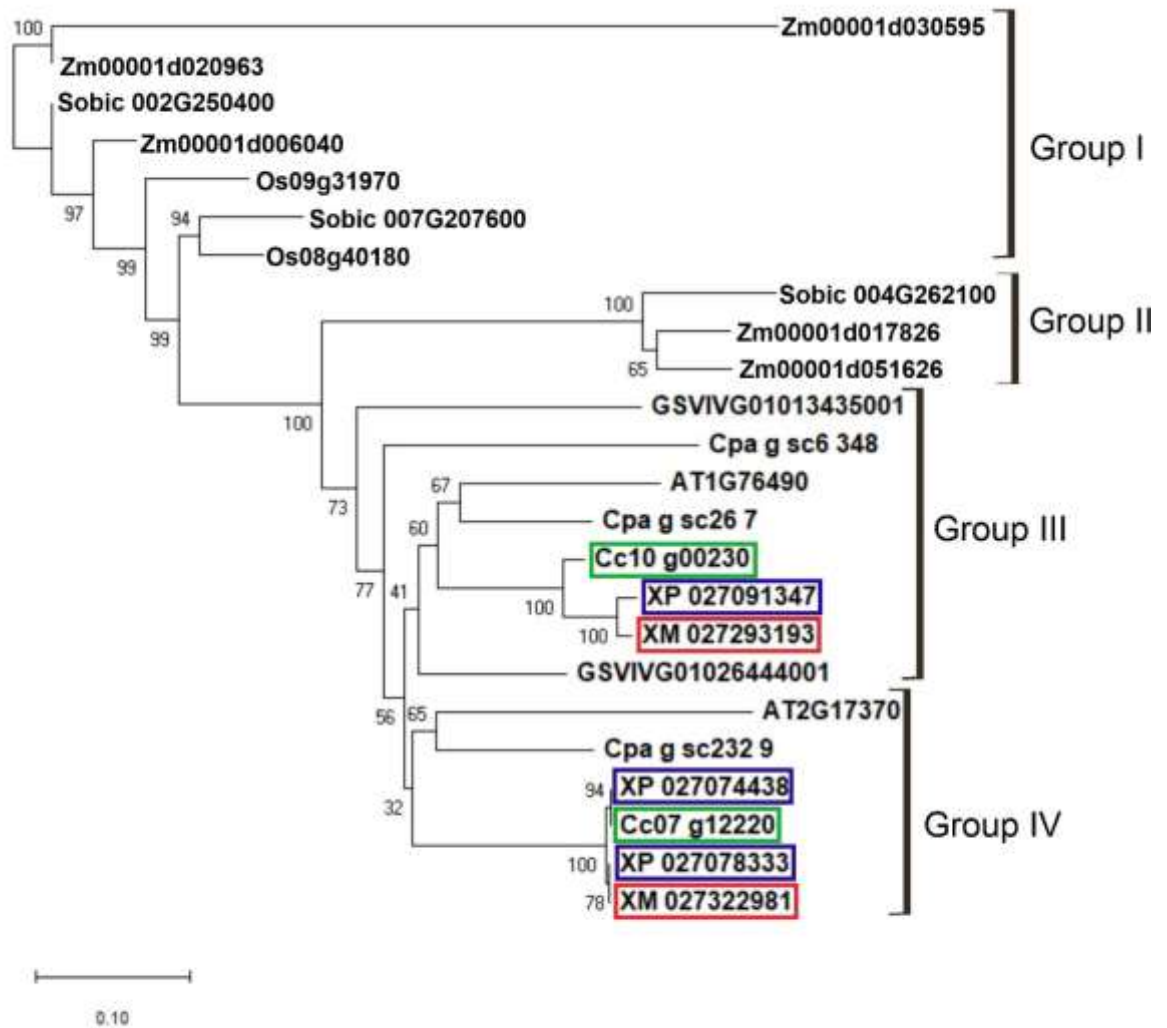
A análise filogenética de *HMGR* resultou na formação de quatro grupos gênicos (Figura 3). O grupo I e II são compostos pelas espécies monocotiledôneas e os grupos III e IV pelas espécies eudicotiledôneas (Figura 3). As espécies *C. canephora* e *C. eugenioides* apresentaram um gene no grupo III e um gene no grupo IV, respectivamente. Em *C. arabica* observou-se um gene no grupo III e dois no grupo IV (Figura 3).

Duas cópias de *HMGR* de *S. bicolor* foram alocadas no grupo I (*Sobic002G250400*, *Sobic007G207600*) e uma cópia no grupo II (*Sobic004G262100*). Três cópias de *HMGR* de *Z. mays* no grupo I (*Zm00001d006040*, *Zm00001d020963*, *Zm00001d030595*) e duas das cópias no grupo II (*Zm00001d017826*, *Zm00001d051626*). A espécie *A. thaliana* apresentou um gene no grupo III (*AT1G76490*) e um gene no grupo IV (*AT2G17370*). O padrão filogenético dos quatro grupos de *HMGR* e a divisão dos genes de *A. thaliana*, *S. bicolor* e *Z. mays* corroboram com os observados no trabalho de Li et al. (2014).

A árvore filogenética de *DXS* dividiu-se em três tipos (Figura 4), o mesmo padrão foi observado em um estudo anterior com espécies da família Solanaceae (PAN et al., 2019) e *O. sativa* (YOU et al., 2020). Os três grupos filogenéticos (Type I, II e III) formados não separaram as espécies eudicotiledôneas das monocotiledôneas.

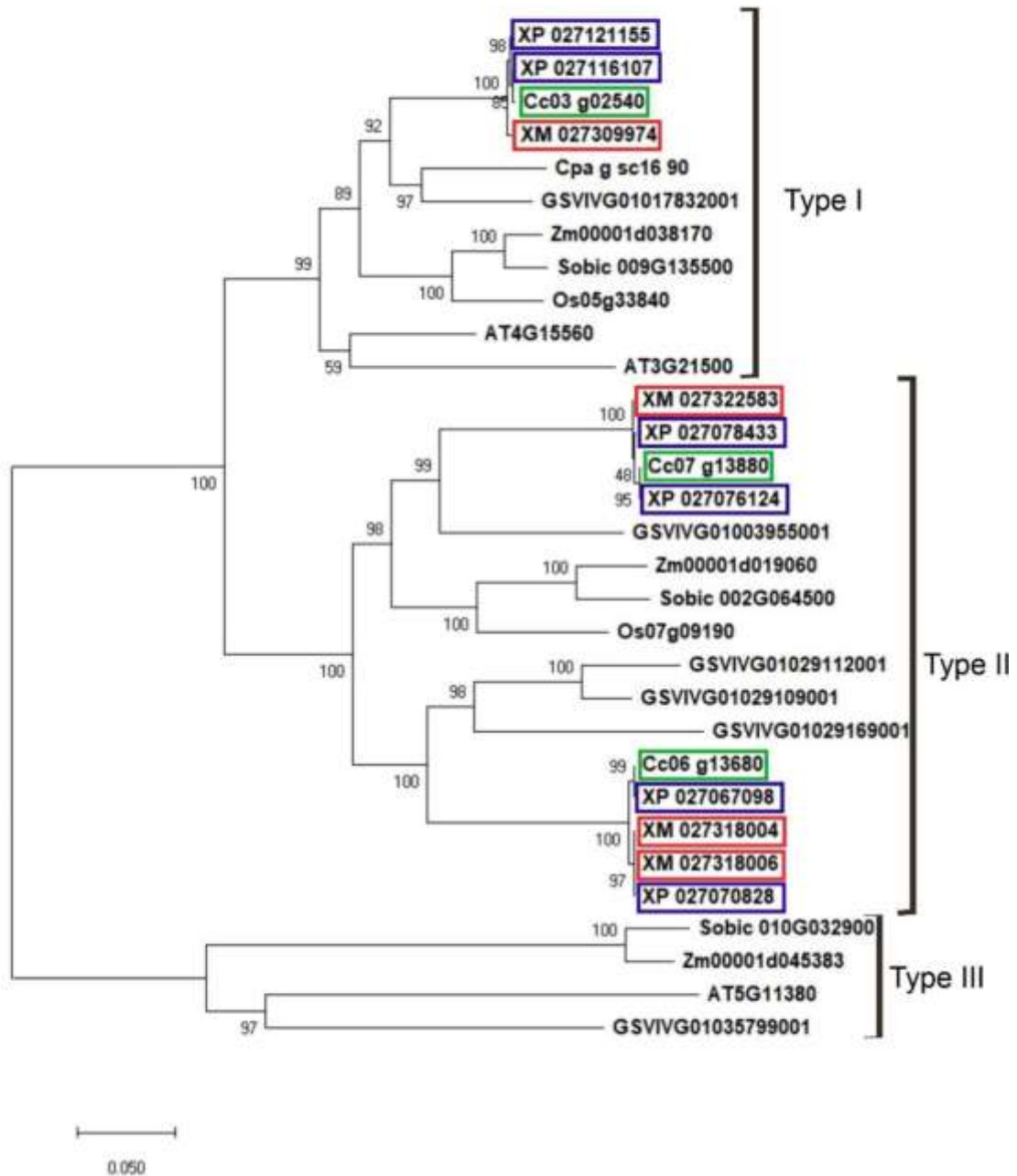
Os genes de *Coffea* spp apresentaram distribuição nos tipos I e II. *C. canephora* apresentou um gene no tipo I (*Cc03g02540*) e dois genes no tipo II (*Cc07g13880*, *Cc06g13680*). *C. eugenioides* apresentou um gene no tipo I (*XM_027309974*) e três genes no tipo II (*XM_027322583*, *XM_027318004*, *XM_027318006*). *C. arabica* apresentou dois genes no tipo I (*XP_027121155*, *XP_027116107*) e quatro genes no tipo II (*XP_027078433*, *XP_027076124*, *XP_027067098*, *XP_027070828*). *A. thaliana* apresentou dois genes no tipo I (*AT4G15560*, *AT3G21500*) e um gene no tipo III (*AT5G11380*). Observou-se ainda que os genes de milho (*Z. mays*) estavam presentes nos tipos I (*Zm00001d038170*), II (*Zm00001d019060*) e III (*Zm00001d045383*). Essa distribuição é similar ao observado nos estudos de You et al. (2020) e Pan et al. (2019).

Figura 3. Árvore filogenética de proteínas HMGR de nove espécies angiospermas



Os genes das espécies *C. canephora*, *C. eugenioides* e *C. arabica* estão representados dentro de quadros verdes, vermelhos e azuis, respectivamente. A história evolutiva foi inferida usando 23 sequências de aminoácidos HMGR. A árvore foi feita usando o método de Neighbor-Joining, com 1000 bootstraps (mostradas como porcentagens). A análise evolutiva e o desenho da árvore foram realizados no software MEGAX. Fonte: próprio autor. *Zea mays* (Zm), *Oryza sativa* (Os), *Sorghum bicolor* (Sobic), *Vitis vinífera* (GS), *Carica papaya* (Cpa), *Arabidopsis thaliana* (AT).

Figura 4. Árvore filogenética de proteínas DXS de nove espécies de angiospermas



Os genes das espécies *C. canephora*, *C. eugenioides* e *C. arabica* estão representados dentro de quadros verdes, vermelhos e azuis, respectivamente. A história evolutiva foi inferida usando 29 sequências de aminoácidos DXS. A árvore foi realizada usando o método de Neighbor-Joining, com 1000 bootstraps (mostradas como porcentagens). A análise evolutiva e o desenho da árvore foram realizados no software MEGAX. Fonte: próprio autor. *Zea mays* (Zm), *Oryza sativa* (Os), *Sorghum bicolor* (Sobic), *Vitis vinifera* (GS), *Carica papaya* (Cpa), *Arabidopsis thaliana* (AT).

3.3.4. Estrutura éxon-íntron dos genes *HMGR* e *DXS*

As estruturas éxon-íntron foram analisadas para aprimorar nossa compreensão sobre a diversidade estrutural dos genes *HMGR* (Figura 5) e *DXS* (Figura 6) em espécies de angiospermas. A análise da estrutura e organização dos genes revelou que o número de éxons e íntrons nos genes que codificam *HMGR* e *DXS* são distintos entre as espécies estudadas.

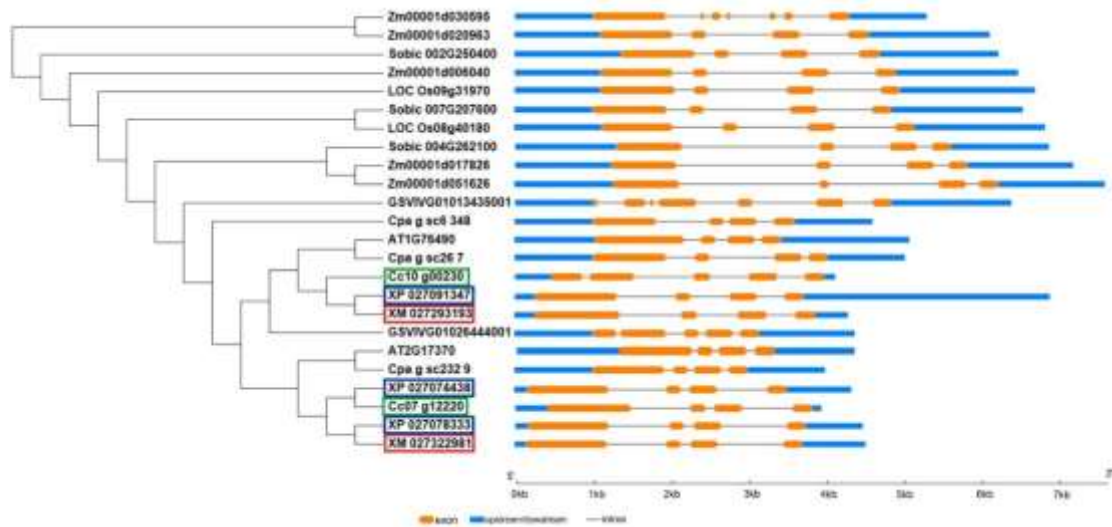
O número de éxons de *HMGR* variou entre quatro e sete (Figura 5). Os genes da espécie *Z. mays* possuem quatro éxons em sua composição (*Zm00001d006040*, *Zm00001d017826*, *Zm00001d051626*, *Zm00001d020963*) e um gene possui sete éxons (*Zm00001d030595*). As espécies *S. bicolor*, *O. sativa*, *C. papaya* e *A. thaliana* apresentaram quatro éxons na estrutura de todos os genes codificadores de *HMGR* (Figura 5). *V. vinifera* apresentou cinco e sete éxons para os genes *GSVIVG01026444001* e *GSVIVG01013435001*, respectivamente. Esses mesmos padrões, de quatro a sete de éxons na estrutura de *HMGR*, foram visto anteriormente nas mesmas espécies *Z. mays*, *S. bicolor*, *O. sativa*, *C. papaya*, *A. thaliana* e *V. vinifera* (LI et al, 2014; LENG et al, 2017). *C. arabica* (*XP_027091347*, *XP_027074438*, *XP_027078333*) *C. eugeniodes* (*XM_027293193*, *XM_027322981*) e *C. canephora* (*Cc07_g12220*) apresentaram quatro éxons na estrutura dos genes de *HMGR*, exceto em um gene de *C. canephora* que apresentou cinco éxons (*Cc10g00230*) (Figura 5).

O gene *DXS* também apresentou números variáveis de éxons (três à dez) entre as espécies analisadas (Figura 6). A espécie *Z. mays* apresentou dez éxons no gene *Zm00001d045383*, nove em *Zm00001d038170*, três em *Zm00001d019060*. *O. sativa* apresentou nove e quatro éxons nos genes *Os05g33840* e *Os07g09190*, respectivamente. Em *S. bicolor* observou-se que os genes *Sobic.002G064500*, *Sobic.009G135500* e *Sobic.010G032900* apresentaram três, sete e dez éxons, respectivamente. *A. thaliana* e *V. vinifera* apresentaram uma menor variação em número de éxons na estrutura da maioria dos seus genes *DXS*, onde foi observado sete (*GSVIVG01035799001*), nove (*AT4G15560*, *AT3G21500*, *GSVIVG01029109001*, *GSVIVG01029169001*) e 10 (*AT5G11380*, *GSVIVG01017832001*, *GSVIVG01029112001* e *GSVIVG01003955001*) éxons.

Os genes das três espécies de *Coffea* spp. (*Cc06_g13680*, *Cc07_g13880*, *Cc03_g02540*, *XM_027309974*, *XM_027318006*, *XM_027318004*, *XM_027322583*, *XP_027121155*, *XP_027116107*, *XP_027067098*, *XP_027070828*, *XP_027078433*, *XP_027076124*) e *C. papaya* (*Cpa.g.sc16.90*) possuem dez éxons que compõem

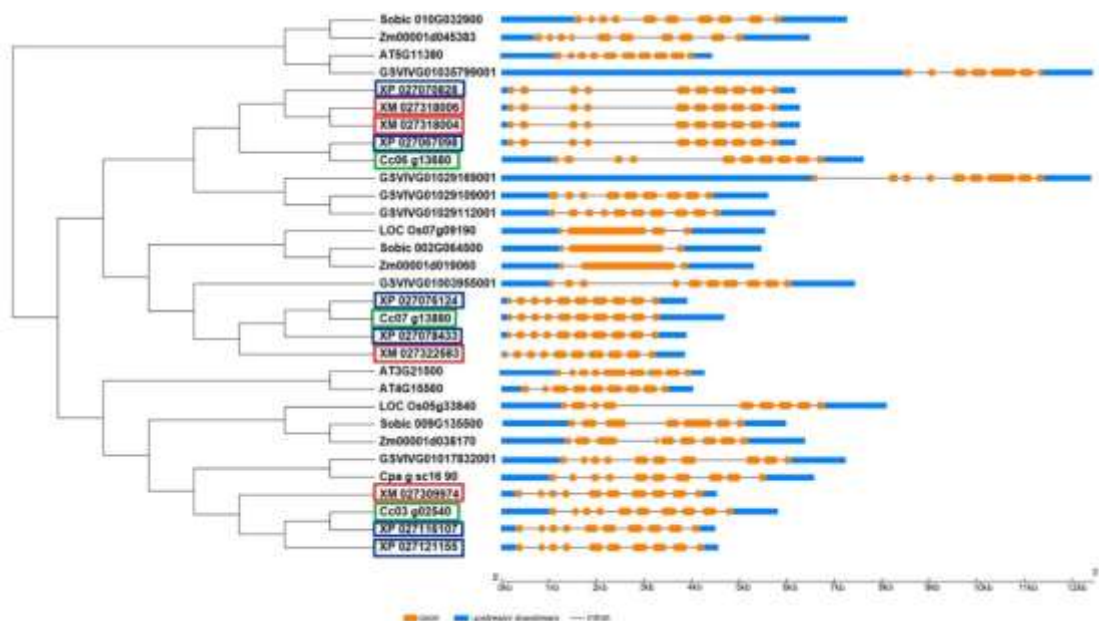
todos os seus genes que codificam *DXS* (Figura 6). O padrão de nove ou dez éxons na estrutura de *DXS* corroboram com estudos anteriores das espécies *V. vinifera* (LENG et al. 2017) e *Solanum lycopersicum* (PAETZOLD et al. 2010).

Figura 5. Estrutura éxon-íntron dos genes *HMGR*



Os genes das espécies *C. canephora*, *C. eugenioides* e *C. arabica* estão representados dentro de quadros verdes, vermelhos e azuis, respectivamente. Fonte: próprio autor.

Figura 5. Estrutura éxon-íntron dos genes *DXS*



Os genes das espécies *C. canephora*, *C. eugenioides* e *C. arabica* estão representados dentro de quadros verdes, vermelhos e azuis, respectivamente. Fonte: próprio autor.

3.4. Conclusão

Genes codificantes das enzimas dos passos iniciais da via MVA e MEP foram identificadas no alotetraploide *C. arabica* e comparadas com as possíveis espécies diploides parentais. A maioria dos passos metabólicos é codificado por mais de um gene. Em termos evolutivos, os genes-chave deste processo HMGR e DXS são duplicados na maioria das espécies analisadas, reforçando a importância da redundância gênica para manutenção destas funções enzimáticas. O Hx modulou o padrão transcricional dos genes das vias MVA e MEP. Os transcritos das vias foram mais expressos em folha tratada com Hx. Potencialmente, o Hx pode ser utilizado como elicitador para induzir genes das vias terpenicas em folhas.

3.5 Referencial bibliográfico

ARANEGA-BOU, Paz et al. Priming of plant resistance by natural compounds. Hexanoic acid as a model. **Frontiers in Plant Science**, v. 5, p. 488, 2014.

CARRETERO-PAULET, Lorenzo et al. Enhanced flux through the methylerythritol 4-phosphate pathway in Arabidopsis plants overexpressing deoxyxylulose 5-phosphate reductoisomerase. **Plant molecular biology**, v. 62, n. 4, p. 683-695, 2006.

CARRETERO-PAULET, Lorenzo et al. Functional and evolutionary analysis of *DXL1*, a non-essential gene encoding a 1-deoxy-D-xylulose 5-phosphate synthase like protein in *Arabidopsis thaliana*. **Gene**, v. 524, n. 1, p. 40-53, 2013.

CLARK, R. B. Differential magnesium efficiency in corn inbreds: I. Dry-matter yields and mineral element composition. **Soil Science Society of America Journal**, v. 39, n. 3, p. 488-491, 1975.

CROWELL, Dring N. et al. Identification of an allele of *CLA1* associated with variegation in *Arabidopsis thaliana*. **Physiologia Plantarum**, v. 118, n. 1, p. 29-37, 2003.

DE CARVALHO, Kenia et al. Nitrogen starvation, salt and heat stress in coffee (*Coffea arabica* L.): identification and validation of new genes for qPCR normalization. **Molecular Biotechnology**, v. 53, n. 3, p. 315-325, 2013.

EDGAR, Robert C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. **BMC Bioinformatics**, v. 5, n. 1, p. 113, 2004.

ESTÉVEZ, Juan M. et al. Analysis of the expression of *CLA1*, a gene that encodes the 1-deoxyxylulose 5-phosphate synthase of the 2-C-methyl-D-erythritol-4-phosphate pathway in *Arabidopsis*. **Plant physiology**, v. 124, n. 1, p. 95-104, 2000.

ESTÉVEZ, Juan M. et al. 1-Deoxy-D-xylulose-5-phosphate synthase, a limiting enzyme for plastidic isoprenoid biosynthesis in plants. **Journal of Biological Chemistry**, v. 276, n. 25, p. 22901-22909, 2001.

FINITI, Ivan et al. Hexanoic acid protects tomato plants against *Botrytis cinerea* by priming defense responses and reducing oxidative stress. **Molecular Plant Pathology**, v. 15, n. 6, p. 550-562, 2014.

HANSEN, Kasper D.; BRENNER, Steven E.; DUDOIT, Sandrine. Biases in Illumina transcriptome sequencing caused by random hexamer priming. **Nucleic Acids Research**, v. 38, n. 12, p. e131-e131, 2010.

HENRY, Laura K. et al. Contribution of isopentenyl phosphate to plant terpenoid metabolism. **Nature Plants**, v. 4, n. 9, p. 721-729, 2018.

HOFBERGER, Johannes A. et al. Large-scale evolutionary analysis of genes and supergene clusters from terpenoid modular pathways provides insights into metabolic diversification in flowering plants. **PLoS One**, v. 10, n. 6, p. e0128808, 2015.

HU, Bo et al. GSDS 2.0: an upgraded gene feature visualization server. **Bioinformatics**, v. 31, n. 8, p. 1296-1297, 2015.

KIM, Daehwan; LANGMEAD, Ben; SALZBERG, Steven L. HISAT: a fast spliced aligner with low memory requirements. **Nature Methods**, v. 12, n. 4, p. 357-360, 2015.

KRAVCHUK, Zhana et al. Priming for JA-dependent defenses using hexanoic acid is an effective mechanism to protect *Arabidopsis* against *B. cinerea*. **Journal of Plant Physiology**, v. 168, n. 4, p. 359-366, 2011.

KUMAR, Sudhir; STECHER, Glen; TAMURA, Koichiro. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. **Molecular Biology and Evolution**, v. 33, n. 7, p. 1870-1874, 2016.

LE, Si Quang; LARTILLOT, Nicolas; GASCUEL, Olivier. Phylogenetic mixture models for proteins. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 363, n. 1512, p. 3965-3976, 2008.

LENG, Xiangpeng et al. Genome-wide identification and characterization of genes involved in carotenoid metabolic in three stages of grapevine fruit development. **Scientific Reports**, v. 7, n. 1, p. 1-13, 2017.

LI, Wei et al. Species-specific expansion and molecular evolution of the 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGR) gene family in plants. **PloS One**, v. 9, n. 4, 2014.

LLORENS, Eugenio et al. Priming by hexanoic acid induce activation of mevalonic and linolenic pathways and promotes the emission of plant volatiles. **Frontiers in Plant Science**, v. 7, p. 495, 2016.

LÓPEZ-GALIANO, María José et al. Identification of Stress Associated microRNAs in *Solanum lycopersicum* by High-Throughput Sequencing. **Genes**, v. 10, n. 6, p. 475, 2019.

NIINEMETS, Ülo. Responses of forest trees to single and multiple environmental stresses from seedlings to mature plants: past stress history, stress interactions, tolerance and acclimation. **Forest Ecology and Management**, v. 260, n. 10, p. 1623-1639, 2010.

PAETZOLD, Heike et al. The isogene 1-deoxy-D-xylulose 5-phosphate synthase 2 controls isoprenoid profiles, precursor pathway allocation, and density of tomato trichomes. **Molecular Plant**, v. 3, n. 5, p. 904-916, 2010.

PAN, Xuhao et al. Bioinformatics study of 1-deoxy-d-xylulose-5-phosphate synthase (DXS) genes in Solanaceae. **Molecular Biology Reports**, v. 46, n. 5, p. 5175-5184, 2019.

PARE, Paul W. et al. Elicitors and priming agents initiate plant defense responses. **Photosynthesis Research**, v. 85, n. 2, p. 149-159, 2005.

PEREIRA, Luiz Filipe Protasio; IVAMOTO, Suzana Tiemi. **Characterization of Coffee Genes Involved in Isoprenoid and Diterpene Metabolic Pathways**. Coffee in health and disease prevention, p. 45-51, 2015.

PERTEA, Mihaela et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. **Nature Biotechnology**, v. 33, n. 3, p. 290-295, 2015.

PICHERSKY, Eran; RAGUSO, Robert A. Why do plants produce so many terpenoid compounds? **New Phytologist**, v. 220, n. 3, p. 692-702, 2018.

RAMAK, Parvin et al. Inhibition of the mevalonate pathway enhances carvacrol biosynthesis and *DXR* gene expression in shoot cultures of *Satureja khuzistanica* Jamzad. **Journal of Plant Physiology**, v. 170, n. 13, p. 1187-1193, 2013.

RAO, Shen et al. Characterization and functional analysis of two novel 3-hydroxy-3-methylglutaryl-coenzyme A reductase genes (GbHMGR2 and GbHMGR3) from *Ginkgo biloba*. **Scientific Reports**, v. 9, n. 1, p. 1-13, 2019.

ROBINSON, Mark D.; MCCARTHY, Davis J.; SMYTH, Gordon K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. **Bioinformatics**, v. 26, n. 1, p. 139-140, 2010.

RODRÍGUEZ-CONCEPCIÓN, Manuel; BORONAT, Albert. Breaking new ground in the regulation of the early steps of plant isoprenoid biosynthesis. **Current Opinion in Plant Biology**, v. 25, p. 17-22, 2015.

SAYERS, Eric W. et al. GenBank. **Nucleic Acids Research**, v. 48, n. D1, p. D84-D86, 2020.

SCALSCHI, Loredana et al. Hexanoic acid is a resistance inducer that protects tomato plants against *Pseudomonas syringae* by priming the jasmonic acid and salicylic acid pathways. **Molecular Plant Pathology**, v. 14, n. 4, p. 342-355, 2013.

SILVA, Natacha et al. Low-Copy Genes in Terpenoid Metabolism: The Evolution and Expression of MVK and DXR Genes in Angiosperms. **Plants**, v. 9, n. 4, p. 525, 2020.

SONG, Qiling et al. Transcriptome-Guided Gene Isolation, Characterization and Expression Analysis of a Phosphomevalonate Kinase Gene (GbPMK) from *Ginkgo biloba*. **International Journal of Agriculture and Biology**, v. 20, n. 5, p. 1080-1088, 2018.

TAO, Tingting et al. Cloning and characterisation of the gene encoding acetyl-coa c-acetyltransferase in *Matricaria chamomilla*. **Journal of Pharmaceutical, Chemical and Biological**, v. 4, n. 3, p. 386-393, 2016.

TONG, Yuru et al. Molecular cloning and characterization of DXS and DXR genes in the terpenoid biosynthetic pathway of *Tripterygium wilfordii*. **International Journal of Molecular Sciences**, v. 16, n. 10, p. 25516-25535, 2015.

VAN BEL, Michiel et al. PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. **Nucleic Acids Research**, v. 46, n. D1, p. D1190-D1196, 2018.

VISHWAKARMA, Rishi K. et al. Molecular cloning, biochemical characterization, and differential expression of an acetyl-CoA C-acetyltransferase gene (AACT) of Brahmi (*Bacopa monniera*). **Plant Molecular Biology Reporter**, v. 31, n. 3, p. 547-557, 2013.

WRIGHT, Louwrence P. et al. Deoxyxylulose 5-phosphate synthase controls flux through the methylerythritol 4-phosphate pathway in Arabidopsis. **Plant Physiology**, v. 165, n. 4, p. 1488-1504, 2014.

XU, Chen et al. Evaluation, characterization, expression profiling, and functional analysis of DXS and DXR genes of *Populus trichocarpa*. **Plant Physiology and Biochemistry**, v. 142, p. 94-105, 2019.

YOU, M. K. et al. The organ-specific differential roles of rice DXS and DXR, the first two enzymes of the MEP pathway, in carotenoid metabolism in *Oryza sativa* leaves and seeds. **BMC Plant Biology**, v. 20, p. 1-16, 2020.

4. CAPÍTULO 2:

Article

Low-Copy Genes in Terpenoid Metabolism: The Evolution and Expression of *MVK* and *DXR* Genes in Angiosperms

Natacha Silva ^{1†}, Suzana Tiemi Ivamoto-Suzuki ^{1†}, Paula Oliveira Camargo ¹, Raíssa Scalzoni Rosa ¹, Luiz Filipe Protasio Pereira ² and Douglas Silva Domingues ^{1*}

¹ Departamento de Biodiversidade, Instituto de Biociências, Universidade Estadual Paulista, UNESP, Rio Claro-SP, Brazil; natacha.s7.ns@gmail.com (N.S.); suzanatiemi@yahoo.com.br (S.T.I.S.); paula_camargoo@outlook.com (P.O.C.); raissa.scalzoni@gmail.com (R.S.R.)

² Laboratório de Biotecnologia Vegetal, Empresa Brasileira de Pesquisa Agropecuária (Embrapa-Café), Londrina-PR, Brazil; filipe.pereira@embrapa.br

* Correspondence: doug@rc.unesp.br, +55 (19) 3526-4207

† These authors contributed equally to this work

Plants 2020, 9(4), 525; <https://doi.org/10.3390/plants9040525>

Received: 7 December 2019 / Revised: 10 April 2020 / Accepted: 15 April 2020 / Published: 19 April 2020

Abstract: Terpenoids are a diverse class of metabolites that impact plant metabolism in response to environmental cues. They are synthesized either via a predominantly cytosolic (MVA) pathway or a plastidic pathway (MEP). In *Arabidopsis*, several enzymes from the MVA and MEP pathways are encoded by gene families, excluding *MVK* and *DXR*, which are single-copy genes. In this study, we assess the diversity, evolution and expression of *DXR* and *MVK* genes in selected angiosperms and *Coffea arabica* in particular. Evolutionary analysis revealed that *DXR* and *MVK* underwent purifying selection, but the selection effect for *DXR* was stronger than it was for *MVK*. Digital gene expression (DGE) profile analysis of six species revealed that expression levels of *MVK* in flowers and roots were high, whereas for *DXR* peak values were observed in leaves. In *C. arabica*, both genes were highly expressed in flowers, and *CaDXR* was upregulated in response to methyl jasmonate. *C. arabica* DGE data were validated by assessing gene expression in selected organs, and by plants treated with hexanoic acid (Hx) using RT-qPCR. *MVK* expression was upregulated in roots treated with Hx. *CaDXR* was downregulated in leaves by Hx treatment in a genotype-specific manner, indicating a differential response to priming.

Keywords: *Coffea*; MVA and MEP pathways; purifying selection; RT-qPCR.

4.1. Introduction

Terpenoids are a large and diverse class of metabolites that include compounds essential for cellular functions and environment interactions [1]. They are the largest and most diverse class of metabolites, containing over 40,000 substances. The molecules are industrially relevant and are used as flavors, pigments, polymers and drugs [1,2].

Terpenoids are produced in all living organisms, but they are most abundant and possess a greater degree of diversity in the plant kingdom [1]. Their biological functions affect plant membrane structure (sterols), respiration (ubiquinone), photosynthesis (chlorophylls, carotenoids, prenylquinones) and the

regulation of plant development (cytokines, brassinosteroids, gibberellins, abscisic acid, strigolactones) [3].

They are derived from two precursor molecules: isopentenyl diphosphate (IPP) and its isomer, dimethylallyl diphosphate (DMAPP) [4]. There are two pathways used to produce IPP and DMAPP in higher plants: the mevalonate pathway (MVA) and methylerythritol phosphate (MEP) pathways. The MVA pathway is predominantly in the cytosol but is also distributed between the endoplasmic reticulum and peroxisomes [5,6], and the MEP pathway is located in plastids [7]. The key rate-limiting enzymes of the MEP pathway and the MVA pathway are, respectively, the extensively studied 1-deoxy-D-xylulose-5-phosphate synthase (DXS) and hydroxymethylglutaryl-CoA reductase (HMGR) [7–10]. However, there are enzymes in the MVA and MEP pathways that are still poorly addressed.

The *mevalonate kinase* (*MVK*) gene encodes one important biosynthetic enzyme in the MVA pathway [3]. It catalyzes the phosphorylation of mevalonate to produce 5-phosphate mevalonate [3]. *MVK* activity is involved in plant regeneration and growth, and it is also related to the frequency of shoot and root formation in white pine (reviewed in [7]).

1-deoxy-d-xylulose 5-phosphate reductoisomerase (*DXR*) is also an important enzyme involved in terpenoid biosynthesis via the MEP plastid pathway, and is responsible for catalyzing the second step of terpenoid biosynthesis in the chloroplast [1]. Inactivation of the *1-deoxy-D-xylulose-5-phosphate reductoisomerase* (*DXR*) gene results in strong developmental arrest and albino plants in *Arabidopsis* [11].

Only plants are capable of simultaneously synthesizing terpenoids via MEP and MVA pathways in parallel, suggesting that genes involved in terpenoid biosynthesis in plants have adapted evolutionarily in basal to terrestrial plant lineages [12,13]. In this sense, it is important to understand the evolutionary relationships and duplication events that have occurred within these gene families [3]. Gene duplication, mutation and natural selection are essential processes that expand plant genetic diversity and facilitate biological adaptation [14,15]. Phylogenetic studies can provide important information needed to enhance our knowledge regarding the mechanism by which *MVK* and *DXR* genes influence terpenoid biosynthesis in higher plants. These genes are both single copy genes in the model plant *Arabidopsis thaliana*. Although previous studies report the phylogenetic analyses of *MVK* and *DXR* [16–18], these studies were focused in species-specific gene characterization and did not seek to further investigate the evolutionary aspects of *MVK* and *DXR* in plants.

The *Coffea* genus (Rubiaceae) is comprised of 124 species [19]. *Coffea arabica* and *C. canephora* are of the species of greatest economic importance and have garnered worldwide interest [20]. Few molecular studies have addressed the terpenoid biosynthesis pathways in coffee plants. Previously, researchers investigated the roles of the *3-hydroxy-3-methylglutaryl-CoA reductase* (*HMGR*) gene, which is responsible for the third step of the MVA pathway, and their respective expression profiles in several organs [21]. Another study validated the function of three *C. arabica* monoterpene synthases (limonene, linalool and β -myrcene synthases), and studied the evolution of the genes, which affect the composition of coffee aroma [22]. Thus far, there have been no studies addressing the molecular responses of hub genes of these pathways.

Resistance inducers are known to modulate the expression pattern of genes involved in terpenoid biosynthesis [23]. Hexanoic acid (Hx) is a short, naturally occurring monocarboxylic acid that induces resistance to pathogens in several plant systems [24]. In *Citrus* plants, Hx has been shown to induce genes involved in the MVA pathway [25]. This prompted us to hypothesize that terpenoid pathway genes can be modulated by the application of Hx in coffee plants.

The main goal of this study was to enhance our understanding of the evolution of both *DXR* and *MVK* genes in angiosperm species, and to characterize the transcriptional profiles of these genes in *C. arabica* plants. Our results provide important knowledge regarding the evolutionary dynamics of the *DXR* and *MVK* gene families, and evaluate differences in expression via assessment of gene transcription. In addition, these genes can be further analyzed to verify whether they are involved in plant defense mechanisms aimed at enhancing the biosynthesis of terpenoids.

4.2. Results

4.2.1. Copy Number of MVK and DXR Genes in Plants

We used a diverse array of angiosperm species to perform genome-wide analysis of MVK and DXR genes and three species as outgroups: one basal angiosperm, one basal Viridiplantae and one Chlorophyta (Figure 1). Thirty-one MVK and 31 DXR sequences were identified in 24 species. We identified 22 MVK and 21 DXR genes in 16 eudicotyledonous species, while in monocotyledons we identified 7 MVK and 6 DXR sequences in five species. In the three outgroup species assessed, we identified 2 MVK and 4 DXR genes, except in the unicellular green algae, *Chlamydomonas reinhardtii*, which does not have a gene encoding MVK (Table 1). *Chenopodium quinoa*, *Glycine max*, *Gossypium raimondii*, *Marchantia polymorpha*, *Populus trichocarpa*, *Ricinus communis*, *Musa acuminata* and *Setaria italica* had between two and three genes encoding MVK and DXR (Table 1). As expected, predictions of the subcellular localization of MVK and DXR proteins for all 24 species revealed that MVK is likely located in the cytosol and DXR is likely located in chloroplasts.

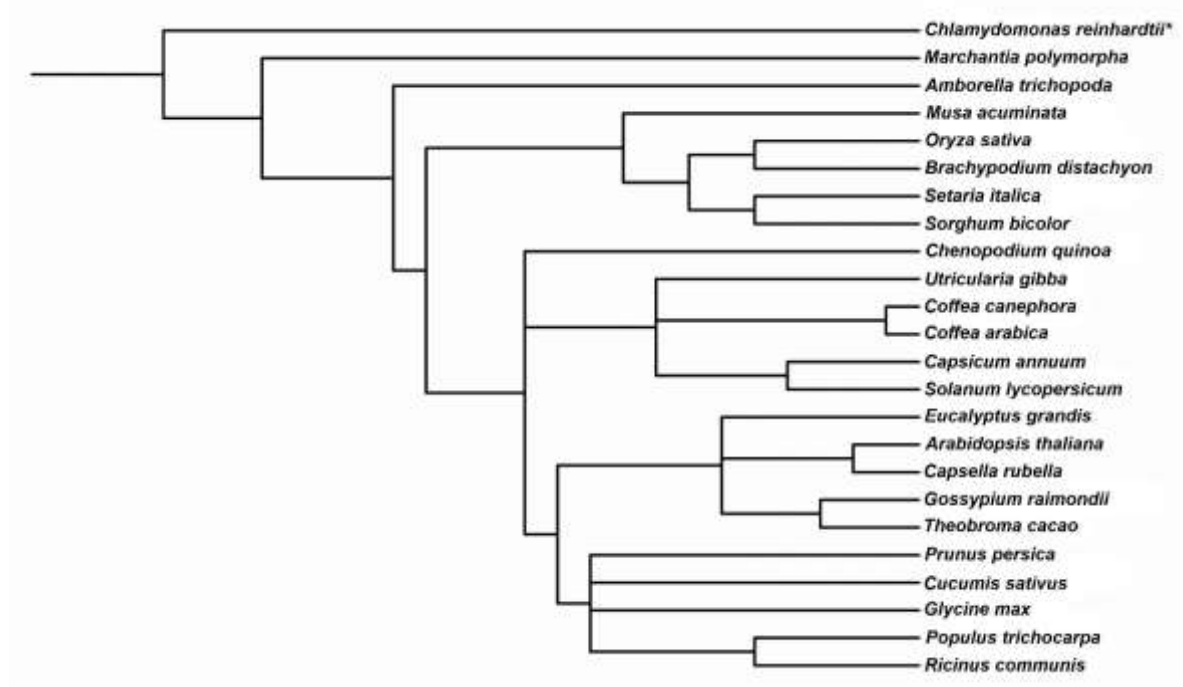


Figure 1. Phylogenetic tree of the analyzed species based on NCBI species taxonomy. This tree was constructed with PhyloT [26] and edited using iTOL [27]. An asterisk (*) indicates that the species lacks an MVK gene.

Table 1. Identification of MVK and DXR genes in 24 plant species.

Lineage	Species	Gene ID	
		MVK	DXR
Outgroups	<i>Amborella trichopoda</i>	ATR0561G047	ATR0730G147
	<i>Chlamydomonas reinhardtii</i>		Cre12.g546050
	<i>Marchantia polymorpha</i>	Mapoly0214s0007	Mapoly0064s0074 Mapoly0034s0078
Eudicotyledons	<i>Arabidopsis thaliana</i>	AT5G27450	AT5G62790
	<i>Capsella rubella</i>	Carubv10001182m.g	Carubv10026322m.g
	<i>Capsicum annuum</i>	CAN.G1077.21	CAN.G781.11
	<i>Chenopodium quinoa</i>	AUR62005132	AUR62030965
		AUR62000841	AUR62008212
	<i>Coffea arabica</i>	CaMVK_comp31934	CaDXR_comp33448
	<i>Coffea canephora</i>	Cc00_g15600	Cc04_g14010
	<i>Cucumis sativus</i>	Cucsa.091600	Cucsa.196910
	<i>Eucalyptus grandis</i>	Eucgr.B01993	Eucgr.B00316
	<i>Glycine max</i>	Glyma.03G239000	Glyma.16G08900
		Glyma.19G236200	Glyma.17G089600
			Glyma.05G037500
	<i>Gossypium raimondii</i>	Gorai.007G296000	Gorai.007G001700
		Gorai.007G342800	Gorai.006G189200
	<i>Populus trichocarpa</i>	Potri.013G024000	Potri.012G080900
		Potri.005G034600	Potri.015G076200
		Potri.005G035300	
	<i>Prunus persica</i>	Prupe.6G327600	Prupe.5G174000
	<i>Ricinus communis</i>	RCO.g.30128.000221	RCO.g.30147.000220
		RCO.g.29656.000023	
<i>Solanum lycopersicum</i>	Solyc01g098840.2	Solyc03g114340.2	
<i>Theobroma cacao</i>	TCA.TCM_026574	TCA.TCM_01200	
<i>Utricularia gibba</i>	UGL.Sc00073.6483	UGL.Sc00226.11960	
Monocotyledons	<i>Brachypodium distachyon</i>	Bradi2g13360	Bradi2g00650
		MAC03G2585	MAC12G2064
		MAC04G2409	MAC07G0719
	<i>Oryza sativa</i>	Os10g18220	Os01g01710
	<i>Setaria italica</i>	Seita.3G273700	Seita.5G071800
		Seita.2G354500	
	<i>Sorghum bicolor</i>	Sobic.004G012400	Sobic.003G103300

4.2.2. Phylogenetic Analyses

We focused our phylogenetic analysis on plant species whose complete genomes were available within the PLAZA 4.0 database (Figures 2 and 3). The MVK gene family was divided into three groups (Figure 2). The first group was composed of outgroup species (red). The second group (blue) contained monocotyledon species and the last group (black) contained eudicotyledon species (Figure 2). The phylogenetic tree of DXR revealed a more complex division than MVK (Figure 3). We observed one group containing the outgroup species *C. reinhardtii* and *M. polymorpha* (red), while the monocotyledon genes were grouped into one major group (blue) and dicotyledons were divided into two groups (black), one of which was near the outgroup species *A. trichopoda*. In *C. arabica*, MVK (comp31934) and DXR (comp33448) sequences were retrieved from a transcriptome set previously developed by our group [28]. In this case, we demonstrate that these sequences are closely related to *C. canephora* genes (Figure 4).

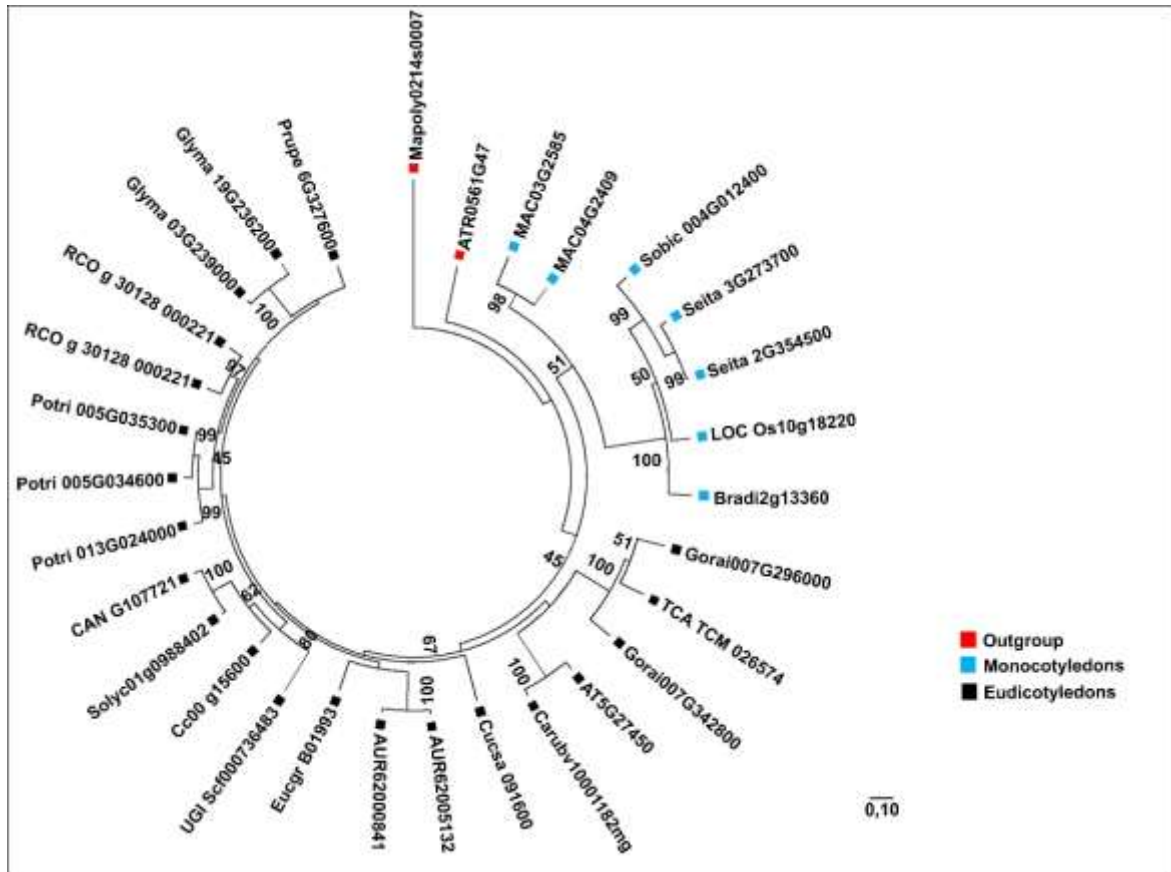


Figure 2. Phylogenetic tree of MVK genes from 23 plant species analyzed in this study. Outgroup species genes are indicated in red, while monocotyledon and eudicotyledon genes are indicated in blue and black, respectively. Evolutionary history was inferred using 30 MVK amino acid sequences. The tree was made using the maximum likelihood method, based on the LG model, with 1000 bootstraps (shown as percentages). All positions containing missing data and gaps were deleted using Gblocks. There were a total of 360 positions in the final dataset. The evolutionary analysis and tree design were performed using MEGAX software.

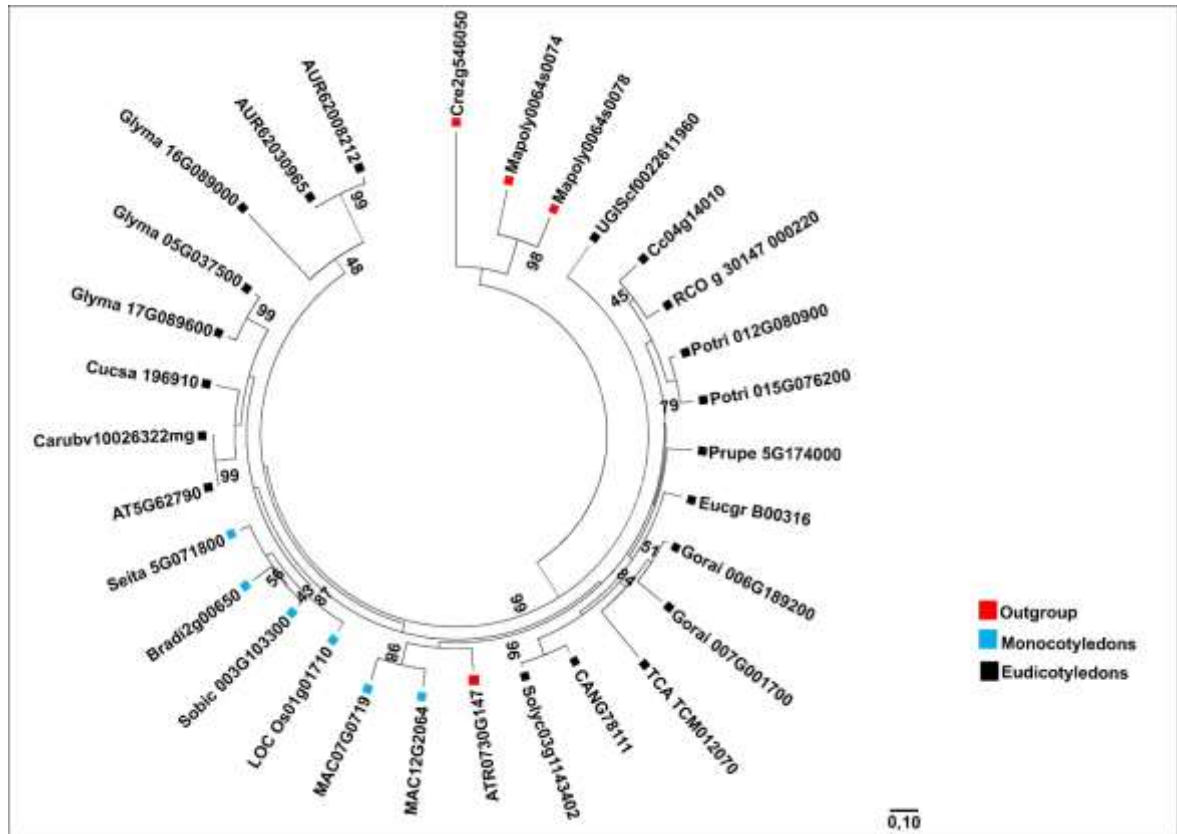


Figure 3. Phylogenetic tree of *DXR* genes from 24 plant species considered in this study. Outgroup species genes, monocotyledon genes and eudicotyledon genes are indicated using red, blue and black boxes, respectively. Evolutionary history was inferred using 30 *DXR* amino acid sequences. The tree was drawn using the maximum likelihood method, based on the LG model, with 1000 bootstraps (shown as a percentage). All positions containing missing data and gaps were deleted using Gblocks. There were a total of 374 positions in the final dataset. The evolutionary analysis and tree design were performed using MEGAX software.

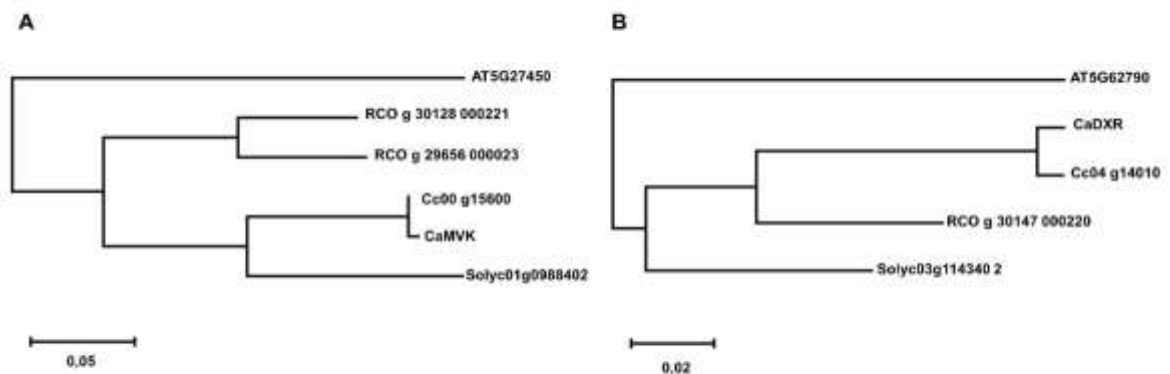


Figure 4. Phylogenetic analysis of *MVK* and *DXR C. arabica* transcripts using five selected species (*A. thaliana*, *C. arabica*, *C. canephora*, *R. communis* and *S. lycopersicum*). The tree was drawn using the maximum likelihood method, which was based on the LG model.

4.2.3. Exon–Intron Structure of MVK and DXR Genes

Exon–intron structures were analyzed to enhance our understanding of the structural diversity of the MVK and DXR genes of angiosperm species. To perform the analyses, publicly available genomic sequences from the PLAZA database (Figure 5A,B) were used. According to these analyses, the MVK gene from all species have five exons, except the *LOC_Os10g18220* gene from *O. sativa*, which has four (Figure 5A). Further, most DXR genes have between 11 and 13 exons, but for the *CAN.G1077.21* gene of *C. annuum* we observed only four exons (Figure 5B).

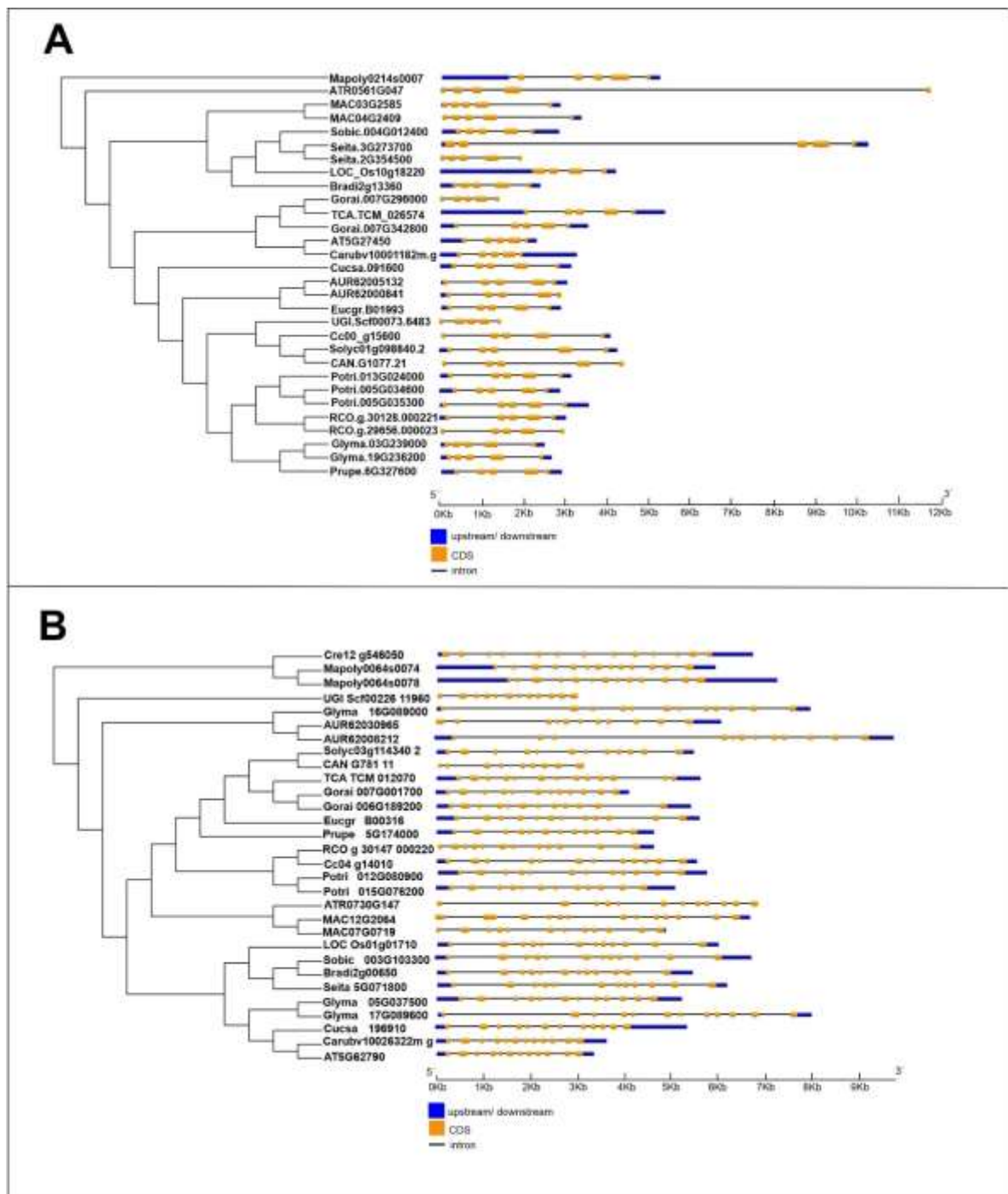


Figure 5. Exon–intron structure of MVK (A) and DXR (B) genes.

The shortest *MVK* and *DXR* full length gene occur in *U. gibba* (\approx 2Kb), while the longest ~12 kb *MVK* full length gene is from *S. italica* (*Seita.3G273700*) and the longest *DXR* full length gene (~10 kb) is in *C. quinoa* (*AUR62008212*) (Figures 5A,B).

4.2.4. Nucleotide Substitution and Selection in *MVK* and *DXR* Genes

To verify the possibility of distinct evolutionary pressures, we analyzed the individual selection profile of each amino acid, as well as the non-synonymous (dN) and synonymous (dS) substitutions in eudicotyledonous and monocotyledonous *MVK* and *DXR* genes. Our results showed that the dN/dS ratios of sequences among the *MVK* and *DXR* genes were distinct (Table 2). In addition, all estimated dN/dS values were substantially less than 1, suggesting that all *MVK* and *DXR* sequences within each group assessed are under strong purifying selection pressure. We observed lower dN/dS ratios in *DXR* than *MVK* (Table 2). Both genes had lower dN/dS ratios in eudicotyledons than in monocotyledons, except for *DXR* in model M5 (Table 2).

Table 2. Average non-synonymous (dN) and synonymous (dS) substitution ratios for *MVK* and *DXR* genes in eudicotyledonous and monocotyledonous plants.

Gene		dN/dS		Positive sites*		Selection	
<i>MVK</i>	Models	Eudycots	Monocots	Eudycots	Monocots	Eudycots	Monocots
	M8(beta+w>=1)	0.19831	0.29343	112, 206, 213	0	Purifying	Purifying
	M7(beta)	0.22751	0.35528	0	0	Purifying	Purifying
	M5(gamma)	0.18986	0.26649	0	0	Purifying	Purifying
<i>DXR</i>	M8(beta+w>=1)	0.11136	0.14043	0	0	Purifying	Purifying
	M7(beta)	0.11877	0.14589	0	0	Purifying	Purifying
	M5(gamma)	0.16233	0.13052	0	0	Purifying	Purifying

*Calculations were performed using a Bayesian inference approach in Selecton Server ($p < 0.05$) using three distinct selection models.

We found three positively selected sites in eudicots within the *MVK* gene using a Bayesian inference approach to calculate site-specific positive and purifying selection. We also evaluated the identification of sites of positive selection using three methods: SLAC (single likelihood ancestor counting), FEL (fixed-effects likelihood) and MEME (Mixed Effects Model of Evolution). FEL and SLAC software did not detect any positively selected codon sites. MEME analyses, however, did positively identify the selected sites (Table 3). In *MVK*, the MEME model detected three positively selected sites in eudicotyledons and four sites in monocotyledons. Site 42 was identified as a positively selected site in both angiosperm groups. With regard to the analysis of *DXR* using MEME, five positively selected sites in eudicotyledons and two in monocotyledons were identified.

Table 3. Predicted number and location of codons under positive selection in *MVK* and *DXR*.

Gene		Positive selection sites*		Selection	
<i>MVK</i>	Methods	Eudicot.	Monocot.	Eudycot.	Monocot.
	FEL	0	0	Purifying	Purifying
	SLAC	0	0	Purifying	Purifying
	MEME	42, 129, 330	42, 115, 122, 317	Purifying	Purifying
<i>DXR</i>	FEL	0	0	Purifying	Purifying
	SLAC	0	0	Purifying	Purifying
	MEME	12, 48, 52, 332, 409	142, 145	Purifying	Purifying

*Calculations were performed using the Datamonkey platform ($p < 0.05$).

4.2.5. Estimated Time of Duplication of *MVK* and *DXR* Genes

The duplication rates of the *MVK* and *DXR* genes were estimated for species with more than one gene copy (*M. polymorpha*, *G. raimondii*, *P. trichocarpa*, *R. communis*, *M. acuminata*, *S. italica*, *C. quinoa* and

G. max). We used substitution data and inferred nucleotide substitution rates (Tables 4 and 5). In order to clarify cases of tandem duplication, segmental duplication and whole-genome duplication, we also retrieved a synteny analysis from Plaza database 4.0 (Figure S1).

The *MVK* gene has been duplicated in seven angiosperm species (Table 4). Five species are eudicotyledons (*C. quinoa*, *G. max*, *G. raimondii*, *P. trichocarpa* and *R. communis*) and two species are monocotyledons (*M. acuminata* and *S. italica*). In *MVK*, the lowest dN/dS ratio and the most recent duplication event was identified in *P. trichocarpa* (Table 4), which corresponds to a tandem duplication (Figure S1) between two copies of *MVK*—*Potri.005G034600* and *Potri.005G035300*—which occurred less than 1 million years ago (Figure S1 and Table 5). The oldest event was determined to occur in *M. acuminata* (Table 4). Among the species considered, only *R. communis* and *S. italica* had more than one copy of *MVK* and no recent genomic duplication events were identified [14], suggesting that a species-specific event generated these gene duplications. We identified that *G. raimondii* and *P. trichocarpa* that had undergone a recent whole-genome duplication also have duplicated *MVK* coding genes (Figure S1 and Table 4).

Table 4. Duplication time estimated by dN/dS analysis in *MVK* genes of seven plant species.

Species	Duplicated <i>MVK</i> genes		dN	dS	dN/dS	Time (mya)
<i>Populus trichocarpa</i>	<i>Potri.005G034600</i>	<i>Potri.005G035300</i>	0.02	0.02	1.00	0.67
	<i>Potri.005G034600</i>	<i>Potri.013G024000</i>	0.05	0.20	0.25	6.67
	<i>Potri.005G035300</i>	<i>Potri.013G024000</i>	0.04	0.19	0.21	6.33
<i>Gossypium raimondii</i>	<i>Gorai.007G296000</i>	<i>Gorai.007G342800</i>	0.11	0.46	0.24	15.33
<i>Ricinus communis</i>	<i>RCO.g.30128.000221</i>	<i>RCO.g.29656.000023</i>	0.06	0.27	0.22	9.00
<i>Musa acuminata</i>	<i>MAC03G2585</i>	<i>MAC04G2409</i>	0.10	0.46	0.22	35.38
<i>Setaria italica</i>	<i>Seita.3G273700</i>	<i>Seita.2G354500</i>	0.02	0.06	0.33	4.62
<i>Chenopodium quinoa</i>	<i>AUR62005132</i>	<i>AUR62000841</i>	0.02	0.08	0.25	2.67
<i>Glycine max</i>	<i>Glyma.03G239000</i>	<i>Glyma.19G236200</i>	0.03	0.17	0.18	5.67

Abbreviation: mya, million years ago.

Table 5. Duplication time estimated by dN/dS analysis for *DXR* genes in six plant species.

Species	Duplicated <i>DXR</i> genes		dN	dS	dN/dS	Time (mya)
<i>Gossypium raimondii</i>	<i>Gorai.007G001700</i>	<i>Gorai.006G189200</i>	0.06	0.34	0.18	11.33
<i>Marcanthia polymorpha</i>	<i>Mapoly0064s0074</i>	<i>Mapoly0034s0078</i>	0.15	n/c	-	-
<i>Populus trichocarpa</i>	<i>Potri.012G080900</i>	<i>Potri.015G076200</i>	0.02	0.16	0.13	5.33
<i>Musa acuminata</i>	<i>MAC12G2064</i>	<i>MAC07G0719</i>	0.05	0.46	0.11	15.33
<i>Chenopodium quinoa</i>	<i>AUR62030965</i>	<i>AUR62008212</i>	0.04	0.09	0.44	3.00
<i>Glycine max</i>	<i>Glyma.16G08900</i>	<i>Glyma.17G089600</i>	0.14	0.6	0.23	20.00
	<i>Glyma.16G08900</i>	<i>Glyma.05G037500</i>	0.15	0.58	0.26	19.33
	<i>Glyma.17G089600</i>	<i>Glyma.05G037500</i>	0.02	0.07	0.29	2.33

Abbreviations: mya, million years ago; n/c, not calculated.

All species with duplicated *MVK* genes also have duplicated *DXR*, except for *S. italica* and *R. communis*. Thus, *DXR* duplication appears to be more genome duplication-dependent than *MVK* duplication. *P. trichocarpa* and *G. max* have an unusual profile of the genes in which each have three *DXR* genes and three *MVK* genes, respectively (Table 5).

The synteny analysis (Figure S1) corroborate that most observed duplications correspond to WGD with two cases of local gene duplication. It was not possible to calculate distance for *DXR* in *M. polymorpha* (Table 5).

4.2.6. Recombination Events of *MVK* and *DXR* Genes

We identified intragenic recombination events in *MVK* and *DXR* using Geneconv, RDP and MaxChi methods. Five *MVK* genes—in two eudicot species, two monocot species and one outgroup—experienced recombination events (Table 6). A total of eight *DXR* genes—in five eudicot species, two monocots and one outgroup—also had recombination events (Table 7). In the analysis of the *MVK* genes,

Seita.3G273700 from *S. italica* had the highest number of recombinations (5). The lowest number (1) was observed in *M. acuminata*, in the gene MAC04G2409 (Table 7). For *DXR*, the highest number of recombinations (6) was observed in *M. polymorpha*, in the Mapoly0064s0074 gene, and the lowest number of recombination events (1) in *T. cacao*, in the TCA.TCM_01200 gene (Table 7).

Table 6. Recombination events predicted for the *MVK* gene family using the RDP, GENECONV and MAXCHI recombination methods.

Species	RDP	GENECONV	MAXCHI	Gene ID
<i>E. grandis</i>	1	0	4	<i>Eucgr.B01993</i>
<i>C. rubella</i>	0	0	2	<i>Carubv10001182m.g</i>
<i>M. polymorpha</i>	0	0	3	<i>Mapoly0214s0007</i>
<i>M. acuminata</i>	0	0	1	<i>MAC04G2409</i>
<i>S. italica</i>	0	0	5	<i>Seita.3G273700</i>

Table 7. Recombination events predicted for the *DXR* gene family using the RDP, GENECONV and MAXCHI recombination methods.

Species	RDP	GENECONV	MAXCHI	Gene ID
<i>C. sativus</i>	2	0	0	<i>Cucsa.196910</i>
<i>T. cacao</i>	0	1	1	<i>TCA.TCM_01200</i>
<i>G. raimondii</i>	0	3	0	<i>Gorai.007G001700</i>
<i>R. communis</i>	0	0	2	<i>RC30147G06220</i>
<i>C. annuum</i>	0	0	5	<i>CANG781.11</i>
<i>S. bicolor</i>	0	0	3	<i>Sobic.003G103300</i>
<i>M. acuminata</i>	0	0	4	<i>MAC07G0719</i>
<i>M. polymorpha</i>	0	0	6	<i>Mapoly0064s0074</i>

4.2.7. RNA-seq-Based Expression Profiles

To determine the *MVK* and *DXR* gene expression profiles among angiosperms, we used public RNA-seq data from seven plant species (*A. thaliana*, *C. canephora*, *P. trichocarpa*, *O. sativa*, *S. bicolor*, *B. distachyon* and *S. italica*) and four tissues (leaves, roots, flowers and fruits) (Figure 6). *MVK* transcripts were expressed most highly in *A. thaliana* roots with 56 RPKM (*At5g27450*), *P. trichocarpa* roots with 60 RPKM (*Potri.013g024000*), *C. canephora* flowers (*Cc00_g15600*) with 70 RPKM and *S. bicolor* flowers with 42 RPKM (*Sobic.004g012400*) (Figure 6). *DXR*, on the other hand, was most highly expressed in *A. thaliana* leaves (*At5g62790*) with 181 RPKM *C. canephora* roots with 88 RPKM (*Cc04_g14010*), *O. sativa* leaves with 308 RPKM (*Os01g01710*), *S. bicolor* leaves with 215 RPKM (*Sobic.003g103300*) and *S. lycopersicum* flowers with 608 RPKM (*Solyc03g114340.2*). For both genes, the lowest levels of expression were observed in *P. trichocarpa* in most organs analyzed (Figure 6). Our results indicated that *MVK* is mostly expressed in roots and flowers, and *DXR* in leaves.

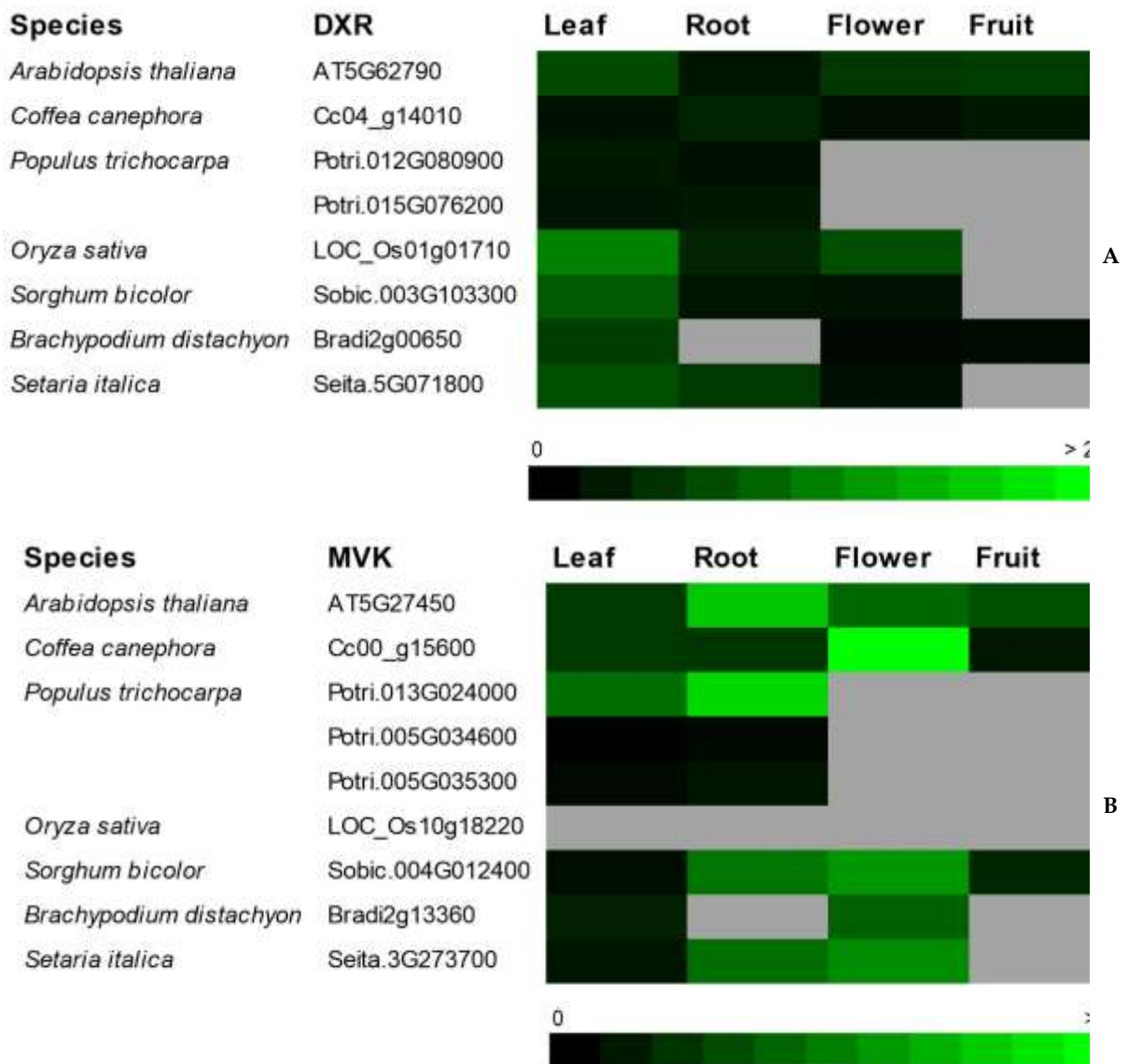


Figure 6. Digital expression profile of *MVK* (A) and *DXR* (B) genes in organs of *A. thaliana* (AT5G27450 and AT5G62790), *C. canephora* (Cc00_g15600 and Cc04_g14010), *P. trichocarpa* (Potri.013g024000, Potri.005g034600, Potri.005g035300, Potri.012g080900 and Potri.015g076200), *O. sativa* (Os10g18220 and Os01g01710), *S. bicolor* (Sobic.004G012400 and Sobic.003G103300) and *S. italica* (Seita.3G273700 and Seita.5G071800). The black color indicates that there were no transcripts detected, and the gray color indicates that no data were available. The intensity of the green color is proportional to the number of transcripts observed in reads per kilo base per million mapped reads (RPKM) values.

Digital gene expression (DGE) profile analysis of *CaMVK* (Figure 7A,B) and *CaDXR* (Figure 7C,D) from several organs of *C. arabica* was also analyzed. The *CaDXR* (Figure 7C) gene had higher transcript levels than *CaMVK* (Figure 7A) in leaves, flowers and fruits (30 to 150 DAF). However, while *CaMVK* steadily decreases its expression along fruit development, *CaDXR* is decreased only in the late stages of fruit development. In addition, we observed that application of methyl jasmonate upregulated the expression of *CaMVK* and *CaDXR* in fruits (Figure 7B,D).

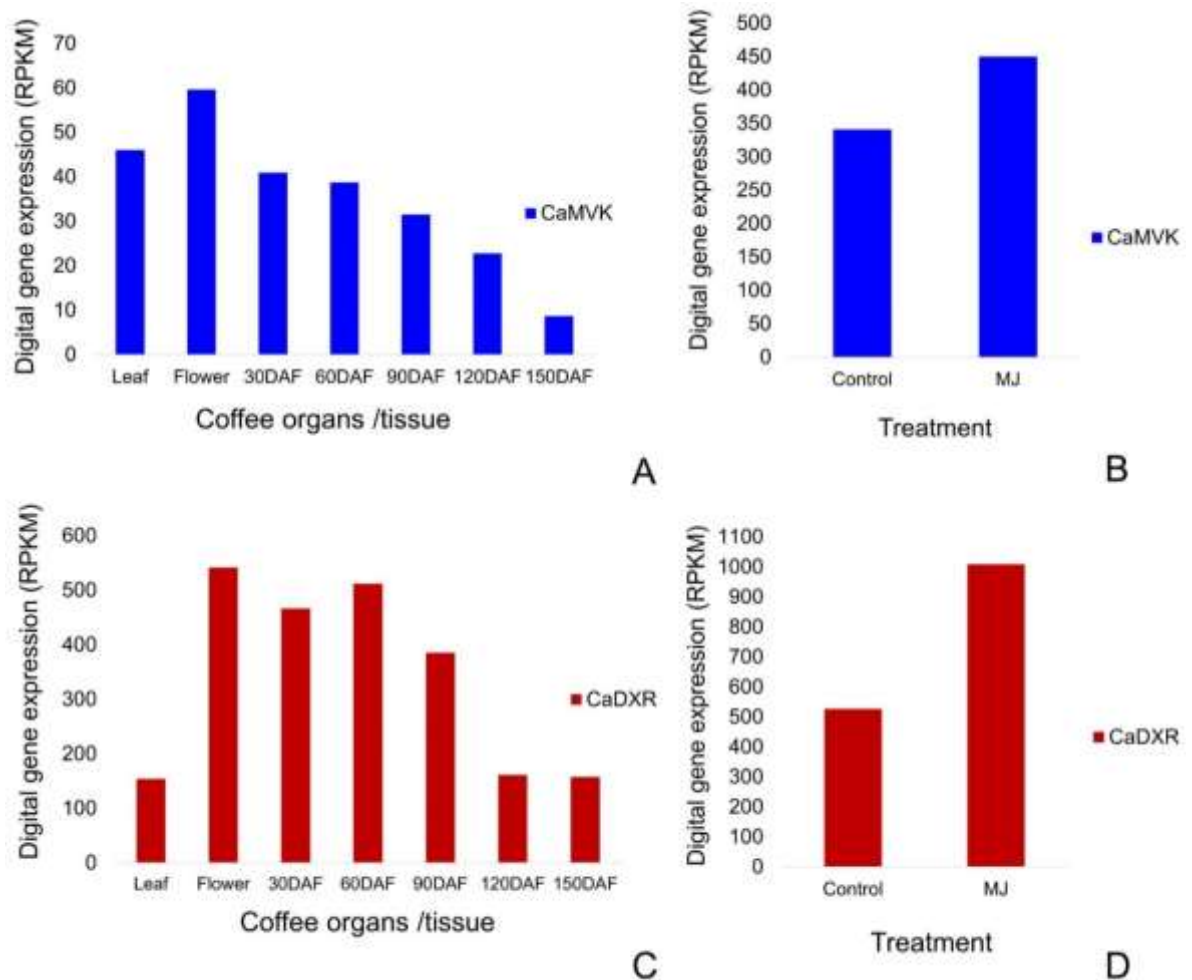


Figure 7. Digital expression profiles of *CaMVK* (blue) and *CaDXR* (red) genes in leaves, flowers and fruits (30 to 150 days after flowering, DAF) in *C. arabica* (A and C) and under the application of 2mM methyl jasmonate (MJ) (B and D). Transcript levels are expressed in reads per kilo base per million mapped reads (RPKM) values.

In order to provide an overall picture of transcriptional responses of MVA and MEP pathways in coffee, we provide DGE analysis of 23 genes of these pathways after MJ application (Figure S2) and in several organs (Figure S3), including *CaMVK* and *CaDXR*. Eleven genes (52.3%) were upregulated similarly to *CaMVK* and *CaDXR* (*CaDXS2*, *CaDXS4*, *CaMCT*, *CaCMK*, *CaHDS1*, *CaHDS2*, *CaIDI*, *CaGGPPS1*, *CaGGPPS2*, *CaAACT2*, *CaHMGS*, *CaMPDC* and *CaIDI*), while six were downregulated (*CaDXS1*, *CaDXS3*, *CaMECPS*, *CaGGPPS3*, *CaAACT1*, *CaPMK*) and two genes (*CaIDS* and *CaHMGR*) did not show significant differences under MJ application (Figure S2). We observed in the fruit development DGE analysis (Figure S3) that 18 genes showed similar expression pattern with *CaMVK* and *CaDXR*—high transcript levels in flowers and low in leaves (*CaDXS2*, *CaDXS4*, *CaCMK*, *CaMECPS*, *CaHDS1*, *CaHDS2*, *CaIDI*, *CaIDS*, *CaGGPPS1*, *CaAACCT1*, *CaAACCT2*, *CaHMGS*, *CaHMGR*, *CaPMK*, *CaMPDC* and *CaIDI*).

4.2.8. RT-qPCR-Based Transcriptional Profile of MVK and DXR Genes in *C. arabica*

Levels of *CaMVK* and *CaDXR* gene expression in flowers, endosperm and the perisperm of fruits 120 days after flowering (DAF) were evaluated using RT-qPCR. Both genes were more active in flowers, followed by perisperm and endosperm (Figure 8A,B).

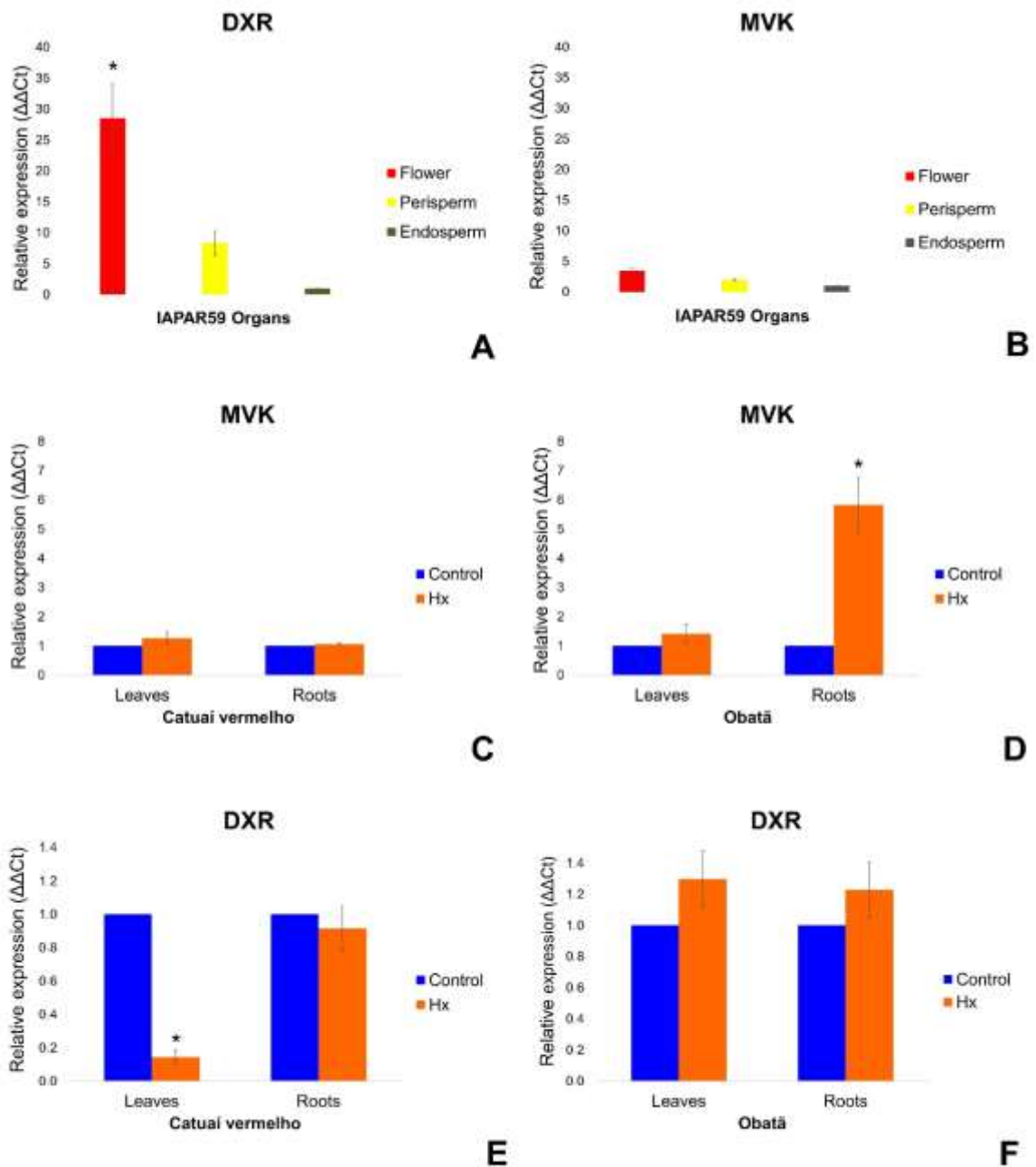


Figure 8. Transcriptional profile of *CaMVK* and *CaDXR* genes in *C. arabica* cv. Iapar 59 (A and B), *C. arabica* cv. Catuaí Vermelho (C and E) and *C. arabica* cv. Obatã (D and F). Experiments were performed using three biological and technical replicates each. Bars indicate standard deviation, and significant differences in which $p < 0.05$ (ANOVA followed by Tukey test) have been indicated with an asterisk (*). Abbreviations: Hexanoic acid application (Hx). Controls represent samples without Hx application.

We also verified that exogenous application of hexanoic acid, a well-known natural priming substance, was capable of modulating *CaMVK* and *CaDXR* transcriptional profiles in leaves and roots. *CaMVK* was upregulated in the roots of Obatã cultivar (Figure 8D) post-Hx treatment. *CaDXR* was downregulated in leaves of Catuaí Vermelho (Figure 8E) treated with Hx.

4.3. Discussion

MVK and *DXR* are key genes involved in MVA and MEP terpenoid biosynthesis pathways. They produce several terpenoid precursors, however, little is known about their molecular evolution [3]. These genes are present in several plant kingdom lineages, such as algae, mosses, monocotyledons and eudicotyledons [9,10,29]. This study revealed that *MVK* and *DXR* are low copy number genes in all 24 species analyzed. In contrast to *Arabidopsis*, which only has one copy of each gene, several species have more than one copy of *MVK* and *DXR*.

We observed that the outgroup species, *C. reinhardtii*, does not have the *MVK* gene. A previous study has shown that another MVA pathway gene, *HMGR*, was also absent in the species [10]. These data suggest that throughout evolution the MVA terpene pathway may have been lost in Chlorophyta. Unusual patterns of MVA pathway has already been reported in a previous study in bacteria [29].

When analyzing the evolutionary relationships of *MVK* and *DXR* genes among angiosperm species, we observed that the phylogenetic trees displayed distinct clades, which divided outgroup species, monocotyledon and eudicotyledon lineages.

The *MVK* gene tree, clearly divided between monocots and eudicots, is highly similar to *HMGR* [10], suggesting a similar pattern of gene evolution among the MVA pathway. However, *DXR* in eudicotyledons have a reticulate pattern, which was not observed for *MVK*, suggesting differential evolutionary forces acting on the two genes. A reticulated pattern of gene evolution was also observed in *DXS*, the rate-limiting enzyme of the MEP pathway [9], suggesting that this pattern is common for genes related to the MEP pathway.

Previous studies have shown that *DXR* has a greater impact on carotenoid and steroid synthesis than *MVK* [30] and the drastic effect of *DXR* loss on plant development is well known [11]. In this sense, we can also speculate that these observations might be associated with a greater impact of *DXR* on terpenoid metabolism.

Most duplications of these genes were associated with recent WGD (whole-genome duplication) events detected in plant species [15], or to polyploid characteristics of species analyzed. In *C. quinoa*, duplication could be explained by the polyploid character of its genome [31]. In *G. max*, *G. raimondii*, *P. trichocarpa* and *M. acuminata* the identification of more than one copy of *MVK* can be also associated to recent genome duplication events [15]. Interestingly, *R. communis* (Table 4) and *S. italica* (Table 5) possessed more than one copy of *MVK* and did not experience any recent genome duplication events, but increased the number of genes nonetheless. They underwent a possible local gene species-specific duplication of *MVK*, which occurred less than 10 and 5 million years ago, respectively—an unusual occurrence in angiosperms.

Gene structure and organization analyses revealed that the number of introns and exons within these genes are similar among species. However, the *DXR* gene has a more complex exon/intron organization than *MVK*, which may also explain its distinct phylogenetic pattern. This observation is in accordance with our findings, indicating that a greater number of recombination events occurred within the *DXR* gene than the *MVK* gene (Tables 6 and 7). Gene structure and organization analyses revealed that the number of introns and exons within these genes are similar among species.

We observed that the *MVK* and *DXR* gene families in angiosperms experienced a strong purifying pressure ($dN/dS < 1$). This fact highlights the functional importance of these genes during their evolutionary process [32]. Some positive sites indicate a possible diversification between strains of both genes in plants [32]. Thus, in sites under positive selection, duplication accelerated, which facilitated functional divergence and resulted in the formation of gene subgroups.

Throughout evolution, gene duplication events may be exposed to divergent selection pressure, which leads to non-functionalization (loss of original functions), neofunctionalization (acquisition of new functions) or sub-functionalization (partition of original functions) [9,10,14]. In this study, dN/dS ratios for duplicated *MVK* and *DXR* genes were less than 1, suggesting that genes were predominantly under strong purifying selection pressure.

To better understand selective pressure, we estimated the age of *MVK* and *DXR* gene duplications using dN/dS data. This confirmed the recent tandem duplication of *MVK* in *P. trichocarpa* and WGD duplications in *G. max*.

We also observed a direct relation between copy number among *MVK* and *DXR* with the respective rate-limiting genes from these pathways—*HMGR* and *DXS*. As examples of gene duplication in *MVK*, we can highlight *G. raimondii*, *G. max* and *P. trichocarpa*. These species were also the ones with higher number of copies of *HMGR* [10]. In *DXR*, *Arabidopsis thaliana* and *Eucalyptus grandis* are species with a single copy of *DXR*; both species have three copies of *DXS* [9]. Conversely, *G. max*, the species with the highest number of *DXR* copies, also harbored the highest number of *DXS* copies in a similar analysis [9].

Using MEME analysis, we identified amino acid sites that experienced episodic positive selection. We observed an opposite pattern of positively selected sites in *MVK* and *DXR* genes: while more positively selected sites were identified in monocot *MVK* genes, for *DXR* most positively selected sites were in dicot genes (Table 3). The amount of positively selected sites helped us to highlight natural selection effects on individual sites that are usually hindered by the low averaged dN/dS values. The analysis of dN/dS among individual amino acid residues in *DXR* (Figure S4 and S5) also indicated that higher values were in the N-terminal, where the protein has its targeting signal to organelles [33].

The transcriptional profile of genes involved in the initial steps of terpenoid biosynthesis in the *Coffea* genus has, thus far, been poorly investigated. However, the *MVK* and *DXR* transcriptional profiles in model plants were already extensively studied [8,34]. *MVK* was highly expressed in roots and flowers, while the transcript levels of *DXR* was more abundant in photosynthetic tissues [7,8,34].

We verified the expression levels of *CaMVK* and *CaDXR* genes in *C. arabica* organs using RNA-seq data and RT-qPCR. We observed higher transcript levels of these genes in flowers for both analyses. The high levels of expression observed in flowers can be explained by the fact that several terpenes are produced in this organ [34,35], most of which are plant volatile compounds, such as limonene, one of the main monoterpene compounds responsible for the aroma of coffee [22]. In addition, we observed in the RNA-seq data that *CaMVK* and *CaDXR* are upregulated in coffee fruits under MJ treatment compared to mock-control plants, including most of the genes involved in the MVA and MEP pathways. Several studies already described methyl jasmonate function as a potent elicitor of secondary metabolite production [36,37] and hexanoic acid as a plant resistance inductor by producing natural products [24,25].

Our RT-qPCR results showed that hexanoic acid is able to upregulate expression of *CaMVK* in *C. arabica* cv. Obatã roots, a result that is in agreement with previous reports investigating gene expression in *Citrus* plants [25]. In the same Hx experiment, the *DXR* gene was downregulated in the leaves of *C. arabica* cv. Catuaí Vermelho, suggesting that modulation of expression might be genotype-specific. The upregulation of transcription of the *MVK* gene in Hx-treated roots was detected only in the Obatã cultivar. Upregulation of the terpenoid pathway can provide protection against stresses, as previously observed in studies investigating tobacco [38], tomato [39] and maize [40]. In addition, it is important to highlight that Obatã and Catuaí Vermelho cultivars have distinct breeding stories, and they are classified in distinct genetic groups [41]. In this study, we provide insights to show that distinct genetic groups generated by coffee breeding might result in dissimilar transcriptional responses.

4.4. Materials and Methods

4.4.1. Identification and Annotation of *MVK* and *DXR* Gene Families in Plants

The coding sequences (CDSs) of *MVK* and *DXR* were retrieved for 24 plant species with sequenced genomes within the PLAZA 4.0 platform [42]. The coding sequences were translated into amino acids using TranslatorX [43]. Gene characterizations were based on characteristics observed in *A. thaliana* *MVK* and *DXR* genes (AT5G27450 and AT5G62790). Sequences on the PLAZA platform that contained the InterPro domains IPR006205 (Mevalonate kinase) and IPR013750 (GHMP kinase, C-terminal domain) were further selected as *MVK* genes. To identify *DXR*, we retrieved genes that simultaneously

contained the InterPro domains IPR003821 (DXP_reductoisomerase), IPR013512 (DXP_reductoisomerase_N), and IPR026877 (DXP C-terminal domain reductoisomerase) for further analysis.

4.4.2. Multiple Sequence Alignments and Phylogenetic Analysis

Coding nucleotide sequences were aligned with MUSCLE [44] and TranslatorX software [43]. In TranslatorX, we were able to eliminate positions containing gaps and missing data with the Gblocks tool [45]. The evolutionary history of genes was inferred using the maximum likelihood method based on the LG model [46]. The consensus tree bootstrap inferred from 1000 replicates was used to represent the evolutionary history of genes analyzed. The percentages of the replicated trees in which the associated rates were grouped in the bootstrap test (1000 repetitions) were shown near the branches. Evolutionary analysis and final trees were performed using MEGAX [47]. To predict the cellular localization of proteins produced by MVK and DXR genes, we used the TargetP 1.1 Server [48], CELLO v.2.5 [49] and Plant-mPLOC [50].

4.4.3. Determination of Gene Structures

To analyze the exon–intron structure of MVK and DXR genes, we used the Gene Structure Display Server v2.0 program with its default parameters [51]. The genomic and CDS in FASTA format corresponding to genes of 24 plant species were inserted to generate gene structures.

4.4.4. Selection Pressure and Evolutionary Analyses

Non-synonymous (dN) and synonymous (dS) nucleotide substitutions within DXR and MVK genomic sequences were used to calculate dN/dS ratios. The indices $dN/dS = 1$, $dN/dS < 1$ and $dN/dS > 1$ represent neutral evolution, purifying selection or positive Darwinian selection, respectively. Individual dN/dS indices for each amino acid of predicted proteins for each gene were also determined using the statistical test suite available on the Selecton Server platform (<http://selecton.tau.ac.il/>) [52], using the following models: M8 ($\omega \geq 1$), M7 (beta) and M5 (gamma). In addition, FEL, SLAC and MEME methods with default settings incorporated into the Datamonkey web interface [53] were also used to identify selection type.

4.4.5. Gene Duplication Analysis

The alignment of duplicate gene pairs of MVK and DXR in *P. trichocarpa*, *G. raimondii*, *R. communis*, *M. acuminata*, *S. italica*, *C. quinoa* and *G. max* species was performed using MUSCLE program implemented in MEGAX [47]. Then, pairwise synonymous (dS) and non-synonymous (dN) numbers of substitutions were calculated using MEGAX. Based on synonymous substitutions per year (λ) equal to 6.5×10^{-9} for monocotyledons [54] and 1.5×10^{-8} for dicotyledons [55], by substituting the calculated dS values, the approximate age of the duplicated events of the duplicate DXR or MVK gene pairs was estimated using the following equation (1):

$$T = Ks/2\lambda \times 10^{-6} \text{ mya.} \quad (1)$$

Synteny information was retrieved from the PLAZA 4.0 platform [42].

4.4.6. Gene Recombination Analysis

Recombination events between divergent nucleotide sequences were investigated to detect recombination signals using RDP.v4 software with default parameters (<http://web.cbio.uct.ac.za/~darren/rdp.html>). In this study, we used the following methods: RDP [56], Geneconv [57] and MaxChi [58] and the *p*-value was set to 0.05.

4.4.7. RNA-seq-Based Expression Profile Analysis of MVK and DXR Genes

We analyzed the digital expression profile of *A. thaliana*, *C. canephora*, *P. trichocarpa*, *O. stiva*, *S. bicolor* and *S. lycopersicum* genes in leaf, root, fruit and flower tissues from the Gene Expression Atlas platform (<http://www.ebi.ac.uk/gxa>) [59]. *C. arabica* digital expression data from leaves, flowers, fruits and fruits treated with methyl jasmonate (2 mM) were obtained from previously published work [28]. Figures depicting expression maps were developed using Genesis software [60].

4.4.8. Transcriptional Profiles of MVK and DXR Genes in *C. arabica* by RT-qPCR

4.4.8.1. Plant Material

Samples were obtained from 20-year-old individual *C. arabica* cv. IAPAR59 plants grown at the Agronomic Institute of Paraná (Londrina—Brazil) under full-sun field conditions with standard irrigation and fertilization practices. We collected open flowers and fruits 120 DAF in January of 2013. Fruit tissues were separated into pulp, perisperm and endosperm, and only perisperm and endosperm tissues were selected for RT-qPCR analysis. All samples were collected between 9 and 11 a.m., transferred immediately to liquid nitrogen and stored at -80°C until RNA was extracted.

4.4.8.2. Hexanoic Acid Treatment: Experimental Procedures

Five-month-old plants of *C. arabica* cv. Catuaí Vermelho IAC 144 and Obatã IAC 1669-20 (4–5 leaf pairs) were used to assess treatment with hexanoic acid. Plants were selected based on size uniformity, and were transferred to dark pots containing 3 L of aerated nutrient solution (ANS), adapted from Clark (1975) by de Carvalho et al. (2013) [61,62]. The nutrition solution was composed of K₂SO₄ (1068 mM), MgSO₄·7H₂O (332.5 mM), KH₂PO₄ (266 mM), CaCl₂·2H₂O (66 mM) and NH₄NO₃ (5333 mM). Iron and micronutrients were supplied by using a commercial chelated salt mixture (ConMicros Standard, Conplant) at the following concentrations (in µg.L⁻¹): Fe (363), Cu (91), Zn (37), Mn (91), B (91), Mo (18) and Ni (17). Experiments were carried out in a plant growth room, with the temperature set to 23 °C (ranging between 21 °C and 25°C). Plants were maintained using a 12 h day/night cycle. LED panels with a photosynthetically active photon flux density of approximately 400 µmol.m⁻².s⁻¹ provided artificial lighting. The pH of the nutrient solution was adjusted and maintained between 5.5 and 5.6 daily using chloridric acid or sodium hydroxide. Plants were acclimatized approximately 96 h. After the acclimatization period, nutrient solutions were replaced. The following treatments were assessed: a) ANS (control); b) ANS + hexanoic acid (Merck, final concentration 0.55 mM) for 48 h. Plants were grown in 3 to 6 plastic pots in which three plants received each treatment. The potted plants were grouped in “pools” (made of 9 to 18 plants), which were considered a replicate. The experiments were repeated 3 times to obtain biological replicates. The mature leaves of the middle third and lateral roots of the plants were collected within the 3rd hour of the light period, macerated in liquid nitrogen and stored in a freezer at -80°C until RNA extraction was performed.

4.4.8.3. Transcriptional Profiles of MVK and DXR Genes using RT-qPCR and Statistical Analyses

Total RNA from *C. arabica* mature leaves, roots, flowers and fruits were extracted using the RNeasy Plant kit (Qiagen, Hilden, North Rhine-Westphalia, Germany). Total RNA samples were purified using the RNeasy Minielute Cleanup kit (Qiagen, Hilden, North Rhine-Westphalia, Germany). RNA integrities were verified via 1% agarose gel electrophoresis. The purity of RNA was determined using a NanoDrop ND-100 spectrophotometer (Thermo Scientific, San Jose, USA) and concentrations were obtained using Qubit fluorimeter (Thermo Fisher Scientific, Wilmington, DE, USA). Complementary DNAs (cDNAs) were synthesized using the High Capacity cDNA Reverse Transcriptase kit (Thermo Fisher Scientific, Wilmington, DE, USA) according to the manufacturer’s instructions. A final volume of 20 µL with 5 µg of total RNA was used. Primers were designed with Primer 3 software to amplify products ranging from 95 to 127 bp (Table S1). Primer efficiencies were calculated using LinRegPCR software v. 11.0 [63]. RT-qPCR reactions were performed using a QuantiStudio3 system (Applied Biosystems, Carlsbad, CA, EUA) in accordance with basic procedures reported in a previous publication in coffee plants [28]. The

reaction mixture contained 7.5 μL of GoTaq[®] Hot Start Polymerase (Promega, Madison, Wisconsin, USA), 0.3 μL of each primer (3 μM), 1 μL of cDNA (10 ng/ μL) and 5.5 μL of Milli-Q water in a final volume of 15 μL . The RT-qPCR conditions began with an initiation step at 95 °C for 2 min; and were followed by 40 cycles at 94 °C for 30 s and 60 °C for 60 s. Melting curves were assessed to verify the presence of a single product and included a negative control. All reactions were performed with three biological and technical replicates, and we followed the MIQE guidelines for RT-qPCR experiments [64].

Relative expression was calculated using the following formula (Equation (2)):

$$(1 + E)^{-\Delta\Delta\text{Ct}} \quad (2)$$

where E = primer efficiency, $\Delta\text{Ct} = \text{Ct}_{\text{target gene}} - \text{Ct}_{\text{reference gene}}$ and $\Delta\Delta\text{Ct} = \Delta\text{Ct}_{\text{target}} - \Delta\text{Ct}_{\text{internal calibrator}}$, as previously described [62]. In all cases, GAPDH was the normalizer, as recommended for coffee plants [62,65]. Internal calibrators were endosperm for IAPAR59 and control samples (without application of hexanoic acid) for the hexanoic acid experiment. Data were analyzed via two-way ANOVA and the post-hoc Tukey test ($p < 0.05$) using XLSTAT.2014 software [66].

4.5. Conclusions

No *MVK* genes were identified in *C. reinhardtii* in the present study, suggesting a very specific evolution pattern for terpenoid biosynthesis in Chlorophyta. Seven species have more than one copy of the *MVK* gene, and six species had more than one copy of *DXR*. Most duplications are due to WGD and also happened with other genes of the pathway, except for *R. communis* and *S. italica*. The *DXR* gene was subjected to a stronger purifying selection pressure, producing lower dN/dS values compared to *MVK*. Our analyses still raise the need for further studies characterizing the impact that *MVK* knockdown could have on the phenotype and terpenoid production of model plants.

Furthermore, the *DXR* gene was shown to undergo a greater number of recombinations. We observed that *DXR* is more transcriptionally active than *MVK* in leaves, while *MVK* transcripts are more abundant in angiosperm roots and flowers. We finally observed that Hx is able to modulate the expression of these genes in coffee leaves and roots in a genotype-specific manner.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1.

Figure S1 – *MVK* and *DXR* Synteny plot.

Figure S2 - RPKM values of *C. arabica* terpenoid pathway genes under methyl jasmonate application.

Figure S3 - RPKM values of *C. arabica* terpenoid pathway genes in flowers, leaves and along fruit development.

Figure S4 - Distribution of dN/dS over individual sites in *MVK*.

Figure S5 - Distribution of dN/dS over individual sites in *DXR*.

Table S1 - Primer sequences used for RT-qPCR analyses.

Conflict of Interest: The authors declare that they have no competing political, personal, religious, ideological, academic, intellectual, and commercial or any other competing interests.

4.6. References

1. Tholl, D. Biosynthesis and biological functions of terpenoids in plants. In *Biotechnology of Isoprenoids*; Schrader, J., Bohlmann, J., Eds.; Springer: Cham, Switzerland, 2015; Volume 148, pp. 63–106, doi:10.1007/10_2014_295.
2. Nogueira, M.; Enfissi, E.M.; Almeida, J.; Fraser, P.D. Creating plant molecular factories for industrial and nutritional isoprenoid production. *Curr. Opin. Biotechnol.* **2018**, *49*, 80–87, doi:10.1016/j.copbio.2017.08.002.
3. Rodríguez-Concepción, M.; Boronat, A. Breaking new ground in the regulation of the early steps of plant isoprenoid biosynthesis. *Curr. Opin. Plant Biol.* **2015**, *25*, 17–22, doi:10.1016/j.pbi.2015.04.001.
4. Böttger, A.; Böttger, A.; Vothknecht, U.; Bolle, C.; Wolf, A. Terpenes and Terpenoids. In *Lessons on Caffeine, Cannabis & Co*; Böttger, A., Vothknecht, U., Bolle, C., Wolf, A., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 153–170, doi:10.1007/978-3-319-99546-5.

5. Simkin, A.J.; Guirimand, G.; Papon, N.; Courdavault, V.; Thabet, I.; Ginis, O.; Bouzid, S.; Giglioli-Guivarc'h, N.; Clastre, M. Peroxisomal localisation of the final steps of the mevalonic acid pathway in planta. *Planta* **2011**, *234*, 903–914, doi:10.1007/s00425-011-1444-6.
6. Pulido, P.; Toledo-Ortiz, G.; Phillips, M.A.; Wright, L.P.; Rodríguez-Concepción, M. *Arabidopsis* J20 delivers the first enzyme of the plastidial isoprenoid pathway to protein quality control. *Plant Cell* **2013**, *25*, 4183–4194, doi:10.1105/tpc.113.113001.
7. Hemmerlin, A.; Harwood, J.L.; Bach, T.J. A raison d'être for two distinct pathways in the early steps of plant isoprenoid biosynthesis? *Prog. Lipid Res.* **2012**, *51*, 95–148, doi:10.1016/j.plipres.2011.12.001.
8. Pichersky, E.; Raguso, R.A. Why do plants produce so many terpenoid compounds? *New Phytol.* **2016**, *220*, 692–702, doi:10.1111/nph.14178.
9. Hofberger, J.A.; Ramirez, A.M.; Van den Bergh, E.; Zhu, X.; Bouwmeester, H.J.; Schuurink, R.C.; Schranz, M.E. Large-scale evolutionary analysis of genes and supergene clusters from terpenoid modular pathways provides insights into metabolic diversification in flowering plants. *PLoS ONE* **2015**, *10*, e0128808.
10. Li, W.; Liu, W.; Wei, H.; He, Q.; Chen, J.; Zhang, B.; Zhu, S. Species-specific expansion and molecular evolution of the 3-hydroxy-3-methylglutaryl coenzyme A reductase (*HMGR*) gene family in plants. *PLoS ONE* **2014**, *9*, e94172, doi:10.1371/journal.pone.0094172.
11. Xing, S.; Miao, J.; Li, S. Disruption of the 1-deoxy-D-xylulose-5-phosphate reductoisomerase (*DXR*) gene results in albino, dwarf and defects in trichome initiation and stomata closure in *Arabidopsis*. *Cell Res.* **2010**, *20*, 688–700, doi:10.1038/cr.2010.54.
12. Kumari, S.; Priya, P.; Misra, G.; Yadav, G. Structural and biochemical perspectives in plant isoprenoid biosynthesis. *Phytochem. Rev.* **2013**, *12*, 255–291, doi:10.1007/s11101-013-9284-6.
13. Priya, P.; Kumari, S.; Yadav, G. Quantification of the plant terpenome: Predicted versus actual emission potentials. *Ind. J. Plant Physiol.* **2016**, *21*, 569–575, doi:10.1007/s40502-016-0256-x.
14. Panchy, N.; Melissa, L.; Shin-Han, S. Evolution of gene duplication in plants. *Plant Physiol.* **2016**, *171*, 2294–2316, doi:10.1104/pp.16.00523.
15. Qiao, X.; Li, Q.; Yin, H.; Qi, K.; Li, L.; Wang, R.; Zhang, S.; Paterson, A.H. Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. *Genome Biol.* **2019**, *20*, 38, doi:10.1186/s13059-019-1650-2.
16. Sando, T.; Takaoka, C.; Mukai, Y.; Yamashita, A.; Hattori, M.; Ogasawara, N.; Fukusaki, E.; Kobayashi, A. Cloning and characterization of mevalonate pathway genes in a natural rubber producing plant, *Hevea brasiliensis*. *Biosci. Biotechnol. Biochem.* **2008**, doi:10.1271/bbb.80165.
17. Singh, S.; Pal, S.; Shanker, K.; Chanotiya, C.S.; Gupta, M.M.; Dwivedi, U.N.; Shasany, A.K. Sterol partitioning by *HMGR* and *DXR* for routing intermediates toward withanolide biosynthesis. *Physiol. Plant.* **2014**, *152*, 617–633, doi:10.1111/ppl.12213.
18. Kumari, U.; Vishwakarma, R.K.; Sonawane, P.; Abbassi, S.; Khan, B.M. Biochemical characterization of recombinant mevalonate kinase from *Bacopa monniera*. *Int. J. Biol. Macromol.* **2015**, *72*, 776–783.
19. Davis, A.P.; Tosh, J.; Ruch, N.; Fay, M.F. Growing coffee: *Psilanthus* (Rubiaceae) subsumed on the basis of molecular and morphological data; implications for the size, morphology, distribution and evolutionary supphistory of *Coffea*. *Bot. J. Linn. Soc.* **2011**, *167*, 357–377, doi:10.1111/j.1095-8339.2011.01177.x.
20. FAOSTAT. Exports/Commodities by Country. Available online: <http://www.fao.org/faostat/en/#data/TP> (accessed on 17 November 2019).
21. Tiski, I.; Marraccini, P.; Pot, D.; Vieira, L.G.E.; Pereira, L.F.P. Characterization and expression of two cDNA encoding 3-Hydroxy-3-methylglutaryl coenzyme A reductase isoforms in coffee (*Coffea arabica* L.). *Omics A J. Integr. Biol.* **2011**, *15*, 719–727, doi:10.1089/omi.2010.0140.
22. Del Terra, L.; Lonzarich, V.; Asquini, E.; Navarini, L.; Graziosi, G.; Suggi Liverani, F.; Pallavicini, A. Functional characterization of three *Coffea arabica* L. monoterpene synthases: Insights into the enzymatic machinery of coffee aroma. *Phytochemistry* **2013**, *89*, 6–14, doi:10.1016/j.phytochem.2013.01.005.
23. Fitza, K.N.; Payn, K.G.; Steenkamp, E.T.; Myburg, A.A.; Naidoo, S. Chitosan application improves resistance to *Fusarium circinatum* in *Pinus patula*. *S. Afr. J. Bot.* **2013**, *85*, 70–78, doi:10.1016/j.sajb.2012.12.006.
24. Aranega-Bou, P.; de la O Leyva, M.; Finiti, I.; García-Agustín, P.; González-Bosch, C. Priming of plant resistance by natural compounds. Hexanoic acid as a model. *Front. Plant Sci.* **2014**, *5*, 488, doi:10.3389/fpls.2014.00488.

25. Llorens, E.; Camañes, G.; Lapeña, L.; García-Agustín, P. Priming by hexanoic acid induce activation of mevalonic and linolenic pathways and promotes the emission of plant volatiles. *Front. Plant Sci.* **2016**, *7*, 495, doi:10.3389/fpls.2016.00495.
26. Letunic, I.; Bork, P. Interactive tree of life v2: Online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.* **2011**, *39*, 475–478, doi:10.1093/nar/gkr201.
27. Letunic, I.; Bork, P. Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **2016**, *44*, 242–245, doi:10.1093/nar/gkw290.
28. Ivamoto, S.T.; Júnior, O.R.; Domingues, D.S.; Dos Santos, T.B.; De Oliveira, F.F.; Pot, D.; Leroy, T.; Vieira, L.G.E.; Carazzolle, M.F.; Pereira, G.A.G.; et al. Transcriptome analysis of leaves, flowers and fruits perisperm of *Coffea arabica* L. reveals the differential expression of genes involved in raffinose biosynthesis. *PLoS ONE* **2017**, *12*, e0169595, doi:10.1371/journal.pone.0169595.
29. Lombard, J.; Moreira, D. Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life. *Mol. Biol. Evol.* **2010**, *28*, 87–99, doi:10.1093/molbev/msq177.
30. Lange, I.; Poirier, B.C.; Herron, B.K.; Lange, B.M. Comprehensive assessment of transcriptional regulation facilitates metabolic engineering of isoprenoid accumulation in *Arabidopsis*. *Plant Physiol.* **2015**, *169*, 1595–1606, doi:10.1104/pp.15.00573.
31. Walsh, B.M.; Adhikary, D.; Maughan, P.J.; Emshwiller, E.; Jellen, E.N. *Chenopodium* polyploidy inferences from Salt Overly Sensitive 1 (SOS1) data. *Am. J. Bot.* **2015**, *102*, 533–543, doi:10.3732/ajb.1400344.
32. Benderoth, M.; Textor, S.; Windsor, A.J.; Mitchell-Olds, T.; Gershenzon, J.; Kroymann, J. Positive selection driving diversification in plant secondary metabolism. *Proc. Natl. Acad. Sci.* **2006**, *103*, 9118–9123.
33. Bruce, B.D. The paradox of plastid transit peptides: Conservation of function despite divergence in primary structure. *Biochim. Biophys. Acta* **2001**, *1541*, 2–21, doi:10.1016/S0167-4889(01)00149-5.
34. Vranová, E.; Coman, D.; & Gruißem, W. Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Annu. Rev. Plant Biol.* **2013**, *64*, 665–700, doi:10.1146/annurev-arplant-050312-120116.
35. Dudareva, N.; Klempien, A.; Muhlemann, J.K.; Kaplan, I. Biosynthesis, function and metabolic engineering of plant volatile organic compounds. *New Phytol.* **2013**, *198*, 16–32, doi:10.1111/nph.12145.
36. Singh, A.; Dwivedi, P. Methyl-jasmonate and salicylic acid as potent elicitors for secondary metabolite production in medicinal plants: A review. *J Pharm. Phytochem.* **2018**, *7*, 750–757.
37. Pan, L.; Zhao, X.; Chen, M.; Fu, Y.; Xiang, M.; & Chen, J. Effect of exogenous methyl jasmonate treatment on disease resistance of postharvest kiwifruit. *Food Chem.* **2020**, *305*, 125483, doi:10.1016/j.foodchem.2019.125483.
38. Steppuhn, A.; Gaquerel, E.; Baldwin, I.T. The two alpha-dox genes of *Nicotiana attenuata*: Overlapping but distinct functions in development and stress responses. *BMC Plant Biol.* **2010**, *10*, 171, doi:10.1186/1471-2229-10-171.
39. Finiti, I.; de la O Leyva, M.; Vicedo, B.; Gómez-Pastor, R.; López-Cruz, J.; García-Agustín, P.; Real, M.D.; González-Bosch, C. Hexanoic acid protects tomato plants against *Botrytis cinerea* by priming defence responses and reducing oxidative stress. *Mol. Plant Pathol.* **2014**, *15*, 550–562, doi:10.1111/mpp.12112.
40. Block, A.K.; Vaughan, M.M.; Schmelz, E.A.; Christensen, S.A. Biosynthesis and function of terpenoid defense compounds in maize (*Zea mays*). *Planta* **2019**, *249*, 21–30, doi:10.1007/s00425-018-2999-2.
41. Maluf, M.P.; Sivestrini, M.; de Campos Ruggiero, L.M.; Filho, O.G.; Colombo, C.A. Genetic diversity of cultivated *Coffea arabica* inbred lines assessed by RAPD, AFLP and SSR marker systems. *Sci. Agric.* **2005**, *62*, 366–373, doi:10.1590/S0103-90162005000400010.
42. Van Bel, M.; Diels, T.; Vancaester, E.; Kreft, L.; Botzki, A.; Van de Peer, Y.; Coppens, F.; Vandepoele, K. PLAZA 4.0: An integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Res.* **2017**, *46*, 1190–1196, doi:10.1093/nar/gkx1002.
43. Abascal, F.; Zardoya, R.; Telford, M.J. TranslatorX: Multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* **2010**, *38*, W7–W13, doi:10.1093/nar/gkq291.
44. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797, doi:10.1093/nar/gkh340.
45. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **2000**, *17*, 540–552, doi:10.1093/oxfordjournals.molbev.a026334.
46. Le, S.Q.; Gascuel, O. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **2008**, *25*, 1307–1320, doi:10.1093/molbev/msn067.

47. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549, doi:10.1093/molbev/msy096.
48. Emanuelsson, O.; Nielsen, H.; Brunak, S.; Von Heijne, G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **2000**, *300*, 1005–1016, doi:10.1006/jmbi.2000.3903.
49. Yu, C.S.; Chen, Y.C.; Lu, C.H.; Hwang, J.K. Prediction of protein subcellular localization. *Proteins Struct. Funct. Bioinform.* **2006**, *64*, 643–651, doi:10.1002/prot.21018.
50. Chou, K.C.; Shen, H.B. Plant-mPLoc: A top-down strategy to augment the power for predicting plant protein subcellular localization. *PLoS ONE* **2010**, *5*, 11335, doi:10.1371/journal.pone.0011335.
51. Hu, B.; Jin, J.; Guo, A.Y.; Zhang, H.; Luo, J.; Gao, G. GSDS 2.0: An upgraded gene feature visualization server. *Bioinformatics* **2014**, *31*, 1296–1297, doi:10.1093/bioinformatics/btu817.
52. Stern, A.; Doron-Faigenboim, A.; Erez, E.; Martz, E.; Bacharach, E.; Pupko, T. Selecton 2007: Advanced models for detecting positive and purifying selection using a Bayesian inference approach. *Nucleic Acids Res.* **2007**, *35*, W506–W511, doi:10.1093/nar/gkm382.
53. Weaver, S.; Shank, S.D.; Spielman, S.J.; Li, M.; Muse, S.V.; Kosakovsky Pond, S.L. Datamonkey 2.0: A modern web application for characterizing selective and other evolutionary processes. *Mol. Biol. Evol.* **2018**, *35*, 773–777, doi:10.1093/molbev/msx335.
54. Gaut, B.S.; Morton, B.R.; McCaig, B.C.; Clegg, M.T. Substitution rate comparisons between grasses and palms: Synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcl*. *PNAS* **1996**, *93*, 10274–10279, doi:10.1073/pnas.93.19.10274.
55. Koch, M.A.; Haubold, B.; Mitchell-Olds, T. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). *Mol. Biol. Evol.* **2000**, *17*, 1483–1498, doi:10.1093/oxfordjournals.molbev.a026248.
56. Martin, D.; Rybicki, E. RDP: Detection of recombination amongst aligned sequences. *Bioinformatics* **2000**, *16*, 562–563, doi:10.1093/bioinformatics/16.6.562.
57. Padidam, M.; Sawyer, S.; Fauquet, C.M. Possible emergence of new geminiviruses by frequent recombination. *Virology* **1999**, *265*, 218–225, doi:10.1006/viro.1999.0056.
58. Smith, J.M. Analyzing the mosaic structure of genes. *J. Mol. Evol.* **1992**, *34*, 126–129, doi:10.1007/bf00182389.
59. Kapushesky, M.; Emam, I.; Holloway, E.; Kurnosov, P.; Zorin, A.; Malone, J.; Rustici, G.; Williams, E.; Parkinson, H.; Brazma, A. Gene expression atlas at the European Bioinformatics Institute. *Nucleic Acids Res.* **2010**, *38*, 690–698, doi:10.1093/nar/gkp936.
60. Sturn, A.; Quackenbush, J.; Trajanoski, Z. Genesis: Cluster analysis of microarray data. *Bioinformatics* **2002**, *18*, 207–208, doi:10.1093/bioinformatics/18.1.207.
61. Clark, R.B. Differential Magnesium Efficiency in Corn Inbreds. I. Dry-Matter Yields and Mineral Element Composition 1. *Soil Sci. Soc. Am. J.* **1975**, *39*, 488–491.
62. de Carvalho, K.; Bepalhok Filho, J.C.; Dos Santos, T.B.; de Souza, S.G.H.; Vieira, L.G.E.; Pereira, L.F.P.; Domingues, D.S. Nitrogen starvation, salt and heat stress in coffee (*Coffea arabica* L.): Identification and validation of new genes for qPCR normalization. *Mol. Biotechnol.* **2013**, *53*, 315–325, doi:10.1007/s12033-012-9529-4.
63. Ramakers, C.; Ruijter, J.M.; Deprez, R.H.; Moorman, A.F. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci. Lett.* **2003**, *339*, 62–66, doi:10.1016/S0304-3940(02)01423-4.
64. Taylor, S.; Wakem, M.; Dijkman, G.; Alsarraj, M.; Nguyen, M. A practical approach to RT-qPCR—Publishing data that conform to the MIQE guidelines. *Methods* **2010**, *50*, S1–S5, doi:10.1016/j.ymeth.2010.01.005.
65. Cruz, F.; Kalaoun, S.; Nobile, P.; Colombo, C.; Almeida, J.; Barros, L.; Romano, E.; Grossi-de-sa, M.F.; Vaslin, M.F.S.; Alves-Ferreira, M. Evaluation of coffee reference genes for relative expression studies by quantitative real-time RT-PCR. *Mol. Breed.* **2009**, *23*, 607–616, doi:10.1007/s11032-009-9259-x.
66. XLSTAT Version 2014.5.03 Copyright Addinsoft 1995–2014, 2014 XLSTAT and Addinsoft are Registered Trademarks of Addinsoft. Available online: <https://www.xlstat.com>.



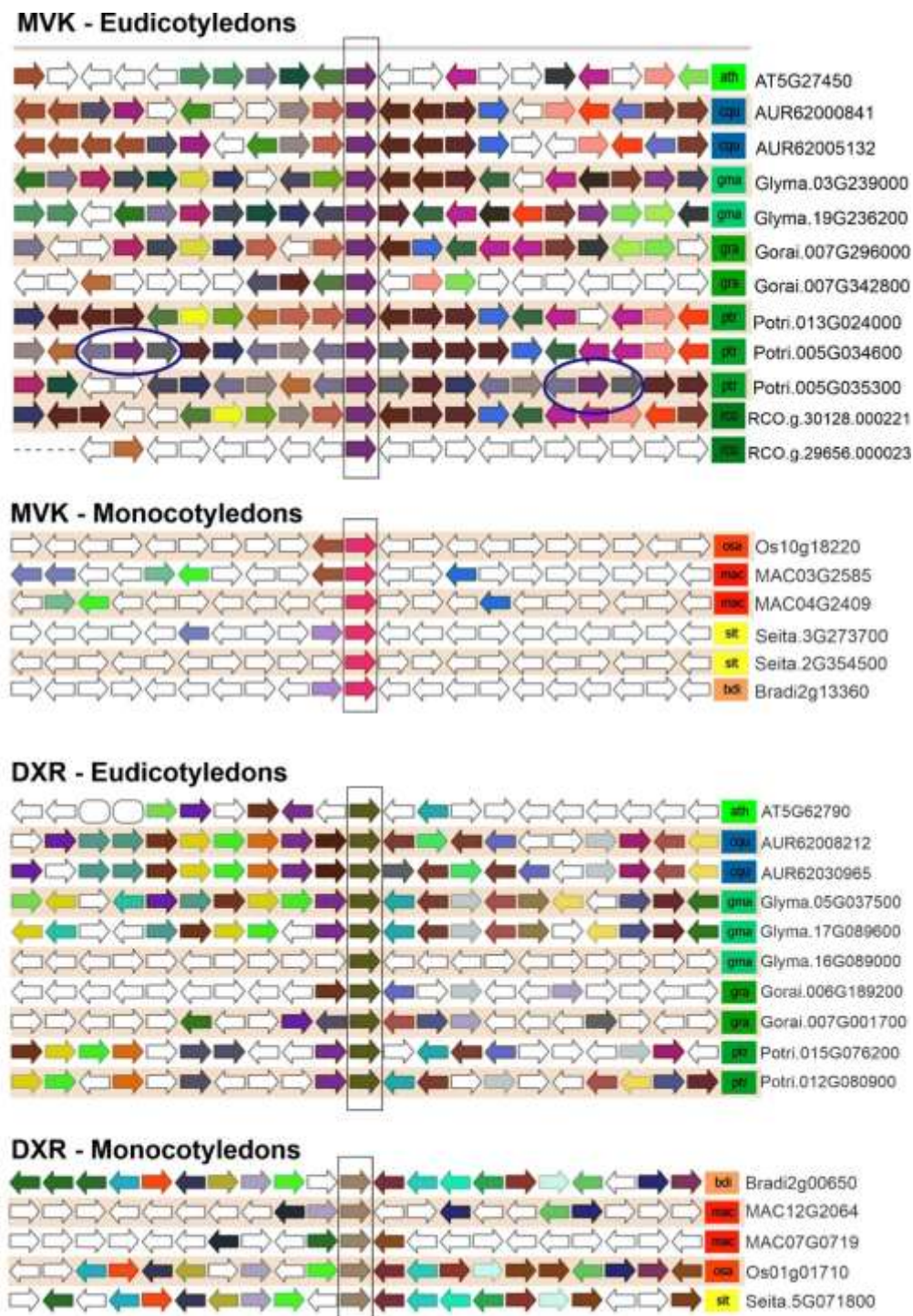
© 2020 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Supplementary materials

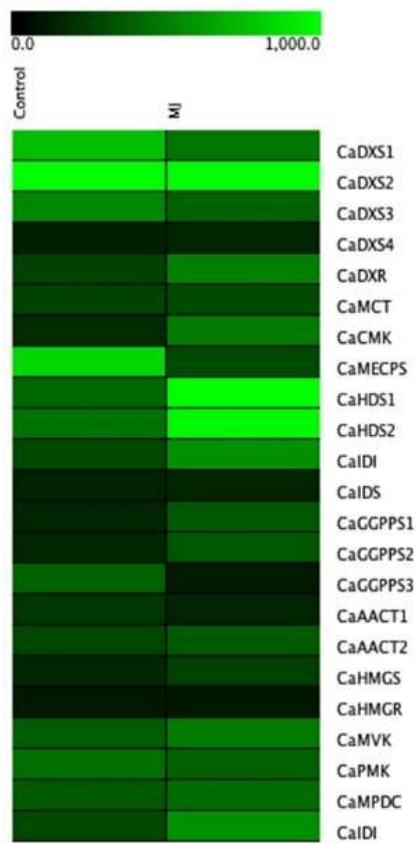
Supplementary Table 1. Primer sequences used for RT-qPCR analyses

Gene ID	Primer sequence	Size (bases)	T _m (°C)	Amplicon (bp)	Efficiency
<i>CaMVK</i>	F: ATGTTCCCTCCGCTTC	18	59,3	95	97%
	R: TCCAACCGCCAAGAAAAT	19	59,1		
<i>CaDXR</i>	F: CCTGTTCTCCCCTACGACT	19	59,3	127	92%
	R: GCCTCCATTGTCCTTCCT	18	59,1		
<i>CaGAPDH</i>	F: AGGCTGTTGGAAAGTTCTTC	21	59,5	70	98%
	R: ACTGTTGGAACCTCGGAATGC	20	58,4		

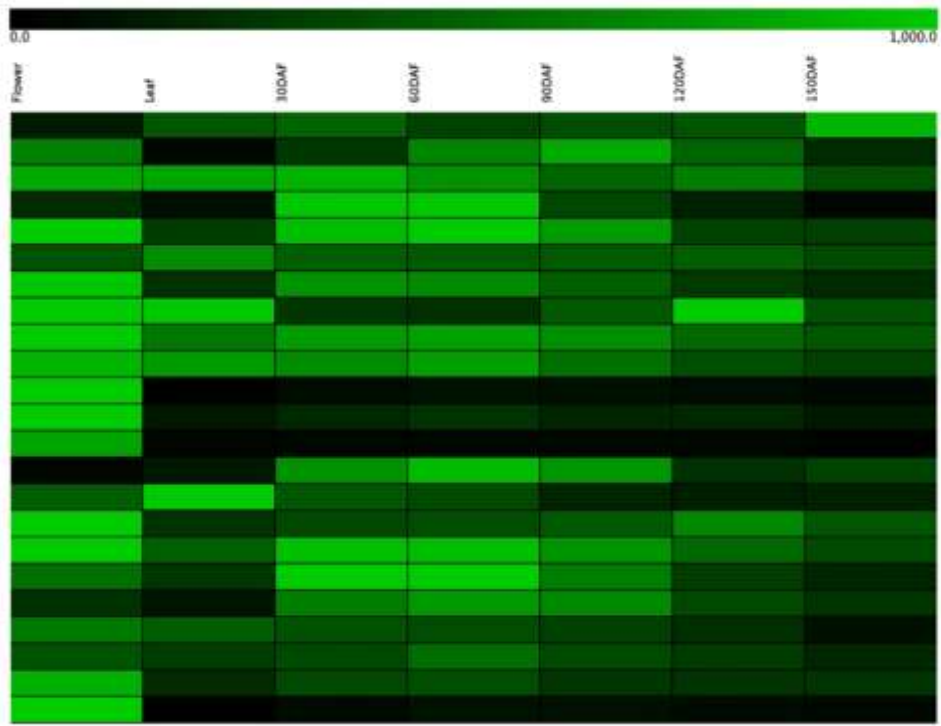
Supplementary figures



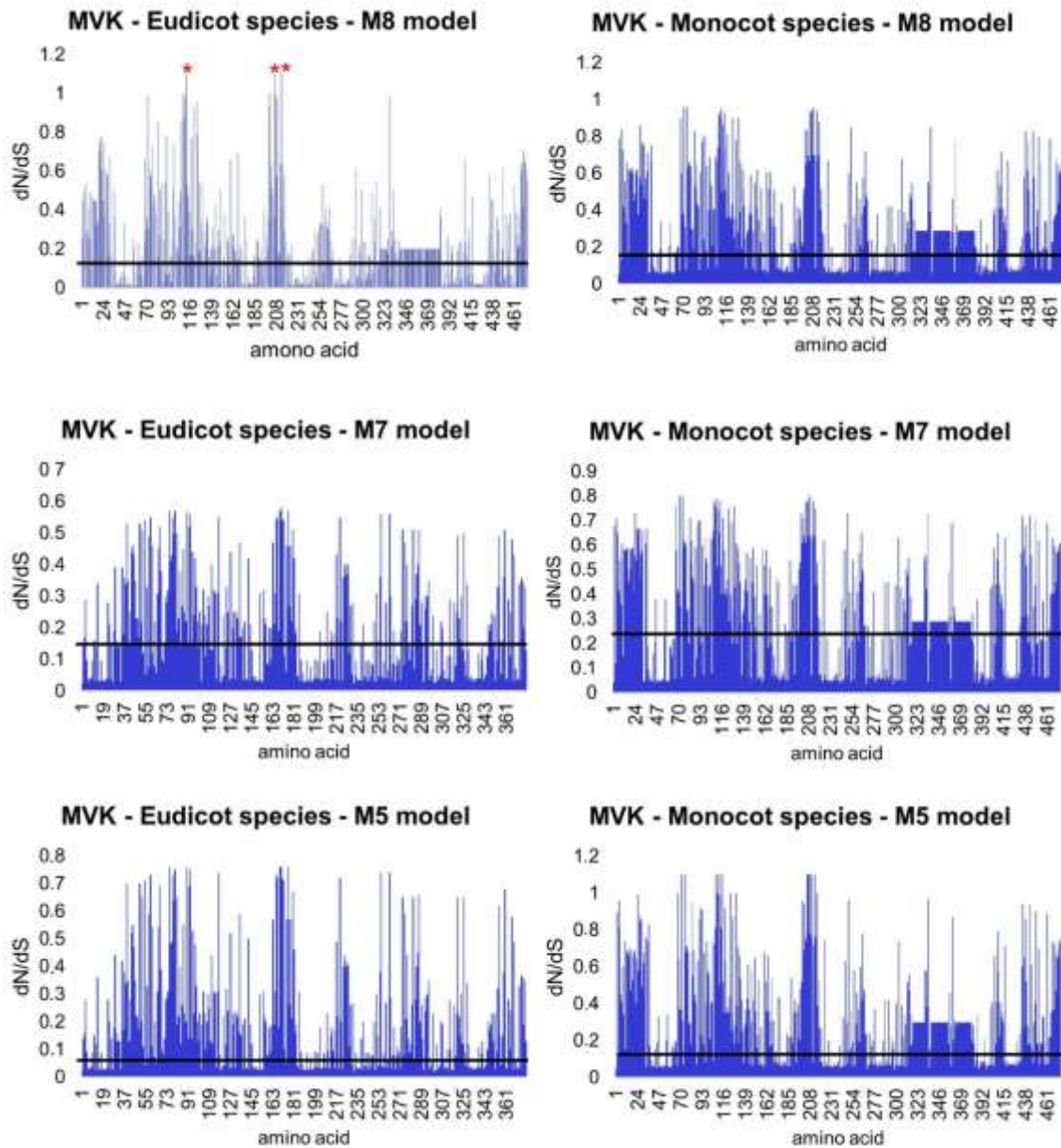
Supplementary Figure 1. Schematic representation of genomic regions containing *MVK* or *DXR* genes. We highlighted in a rectangle *DXR* and *MVK*. The figure was based on the “Synteny plot” tool of Plaza 4.0. Genes with the same color have similar function. Eudicot *MVK* genes share similar genomic neighborhood. We circled the *MVK* tandem duplication in *P. trichocarpa*. In monocots, genomic neighborhood among *MVK* genes has less collinear genes. In *DXR*, we also observe high similar genomic neighborhood, suggesting WGD events in soybean (*Glyma*) and *P. trichocarpa* (*Potri*). Only one gene in soybean (*Glyma.16G089000*) and one in cotton (*Gorai.006G189200*) do not share a significant genomic neighborhood, suggesting that these genes arose from segmental duplication.



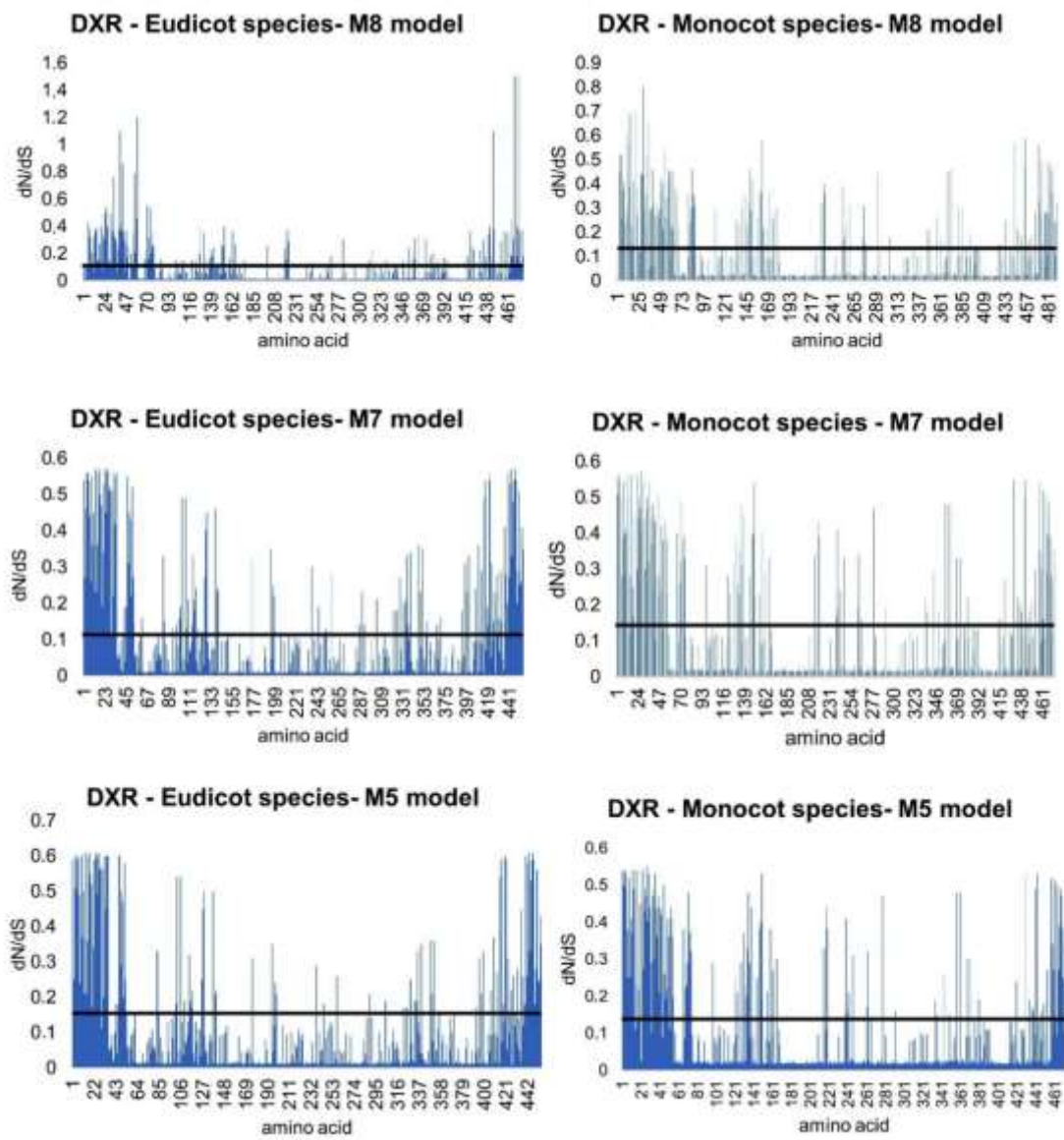
Supplementary Figure 2. RPKM values of *C. arabica* terpenoid pathway genes under methyl jasmonate application. Legend: C – control; MJ – methyl jasmonate application.



Supplementary Figure 3. RPKM values of *C. arabica* terpenoid pathway genes in flowers, leaves and along fruit development. Legend: DAF – perisperm in days after flowering.



Supplementary Figure 4. Distribution of dN/dS over individual sites in MVK, based on Selecton statistics. The horizontal line represents the average dN/dS. Positively selected sites (*), based on Bayesian analysis ($p < 0,05$).



Supplementary Figure 5. Distribution of dN/dS over individual sites in *DXR*, based on Selecton statistics. The horizontal line represents the average dN/dS. Positively selected sites (*), based on Bayesian analysis ($p < 0,05$).

5. CONSIDERAÇÕES FINAIS

O atual estudo apresentou informações e dados fundamentais para as futuras pesquisas, que ainda são incipientes em termos da genômica evolutiva e a biologia molecular das vias de isoprenoides em *Coffea*.

Ao identificar os genes e as cópias codificantes das enzimas dos passos iniciais da via MVA e MEP em *C. arabica* mostrou-se um maior embasamento sobre os genes responsáveis pela produção e a diversidade de terpenoides. Os levantamentos genicos poderão ser utilizados como dados comparativos para futuras pesquisas sobre terpenoides em angiospermas. Porém, nossas análises ainda levantam a necessidade de mais estudos sobre silenciamento dos quatros genes caracterizando o impacto que possivelmente poderia ter sobre o fenótipo e a produção de terpenóides em cafeeiro ou em plantas modelos.

Os estudos evolutivos dos genes-chave das vias de isoprenoides *HMGR*, *DXS*, *MVK* e *DXR* reforçam que estes genes são importantes na manutenção das funções enzimáticas de plantas.

O elicitor Hx poderá ser utilizado em futuros trabalhos, pois apresenta um potencial modulador do padrão transcricional dos genes iniciais das vias MVA e MEP, principalmente em folhas e raízes.

Por fim, o estudo genômico evolutivo e a biologia molecular de *Coffea* gera maior compreensão sobre o metabolismo de isoprenoides e a relação que integra a diversidade de terpenos e suas funções com o meio ambiente.