



UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA
MOLECULAR

Ortologia e perfil de expressão dos genes da via de biossíntese de esteróis em
dendezeiro (*Eleais guineensis* Jacq.).

CLEYSON PANTOJA SERRÃO

BELEM-PA
AGOSTO/2021



UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA
MOLECULAR

Ortologia e perfil de expressões dos genes da via de biossíntese de esteróis
em dendezeiro (*Eleais guineensis* Jacq.).

CLEYSON PANTOJA SERRÃO

Dissertação submetida ao programa
de Pós-graduação em Genética e
Biologia molecular da Universidade
como requisito para obtenção do grau
de Mestre em Genética e Biologia
molecular.

Orientador: Prof Dr Rafael Azevedo Baraúna

BELÉM-PA

AGOSTO/2021

RESUMO

Os Esteróis vegetais são lipídeos responsáveis pela regulação da fluidez da membrana plasmática, sinalização celular, morfogênese da plantula, resposta a estresses ambientais e desenvolvimento sexual. A via de biossíntese dos esteróis vegetais é bem caracterizada em modelos como *Arabidopsis thaliana*, mas pouco se sabe sobre essa via em espécies da família *Arecaceae*. *Elaeis guineensis*, conhecido popularmente como dendezeiro, é uma palmeira oleaginosa com alto valor comercial e alto investimento em estudos genéticos. Esses estudos podem funcionar como fonte de informação biológica e genética para caracterização dos esteróis nessa planta. Nessa perspectiva, o objetivo desse trabalho é realizar a predição genômica, fazer a anotação funcional e de expressão gênica de genes ortólogos e parálogos da via de biossíntese dos esteróis em *Elaeis guineensis* usando *Arabidopsis thaliana* como modelo. Para predição da homologia entre as espécies foram usadas as ferramentas: BLASTp, PFAM, MEME motif, além da construção do Dendrograma usando MEGAX. A caracterização físico-química das proteínas foi feita utilizando o exPASy. A avaliação da expressão gênica foi feita pelo programa SALMON utilizando os transcriptomas: PRJNA474663, PRJNA700429, DRA001857 e PRJDB9517. Os resultados apontam que a via dos esteróis é altamente conservada entre as duas espécies, cada gene de *Arabidopsis* tem pelo menos dois homólogos com o dendezeiro. Os domínios funcionais são conservados e não apresentam nenhum indicativo de divergência funcional. Além disso, as proteínas têm propriedades físico-químicas, composição de aminoácidos e localização sub-celular semelhantes. Todos os genes têm pelo menos um ortólogo expresso em todos os tecidos avaliados e especialmente na flor na qual os genes estão hiperexpressos, mostrando forte relação com o desenvolvimento reprodutivo da planta.

Palavras-chave: fitoesteróis, dendê, proteínas, genes ortólogos, RNAseq.

ABSTRACT

Plant sterols are lipids that regulate plasma membrane fluidity, cell signaling, plantlet morphogenesis, response to environmental stresses and sexual development. The plant sterol biosynthesis pathway is well understood in models such as *Arabidopsis thaliana*, but remain unknown in species of the Arecaceae Familia. *Elaeis guineensis*, popularly known as dendezeiro, is an oil palm with high commercial value and high investment in genetic studies. These studies can serve as a source of biological and genetic information for characterization of sterols in this plant. In this perspective, the purpose of this study is perform genomic prediction and functional annotation of orthologous and paralogous genes of the sterol biosynthesis pathway in *Elaeis guineensis* using *Arabidopsis thaliana* as a model. To predict the homology between species, the following tools were used: BLASTp, PFAM, MEME motif, in addition to building the dendrogram using MEGAX. The physicochemical characterization of the proteins was performed using exPASy. Gene expression evaluation was performed by the SALMON program using transcriptomes: PRJNA474663, PRJNA700429, DRA001857 and PRJDB9517. The results show that the sterol pathway is highly conserved between the two species, the genes of *Arabidopsis* have at least two homologues with the palm tree. The functional domains are conserved and there is no functional divergence. Furthermore, proteins have similar physical-chemical properties, amino acid composition and sub-cellular location yes. All genes have at least one ortholog expressed in all adopted tissues and especially in the flower in which the genes are overexpressed, showing the strong relationship with the plant's reproductive development.

Keywords: phytosterols, oil palm, proteins, ortholog genes, RNAseq.

SUMÁRIO

1	INTRODUÇÃO.....	6
1.1	Esteróis: estrutura química, diversidade e função.....	6
1.2	Genes e proteínas envolvidos nas vias de biossíntese de esteróis. ...	11
1.3	<i>Elaeis guineensis</i> Jacq: o dendezeiro.	15
1.4	Transcriptômica e estudos de expressão gênica em <i>E. guineensis</i>	17
2	OBJETIVOS.....	21
2.2	Objetivos gerais.....	21
2.3	Objetivos específicos	21
3	MATERIAL E MÉTODOS	22
3.1	Busca de ortólogos: alinhamento local e domínios conservados	22
3.2	Caracterização das proteínas.....	23
3.3	Dendrograma, agrupamento e classificação dos genes.....	24
3.4	Análise da expressão gênica.....	24
4	RESULTADOS E DISCUSSÃO	27
4.1	Esqualeno Sintase (SQS1) e esqualeno epoxidase (SQE1)	27
4.2	Cicloartenol Sintase (CAS1) e lanosterol sintase (LAS1).	30
4.3	Esterol metil-transferases (SMT1/SMT2)	34
4.4	Complexo de desmetilação: SMO1, SMO2, β HSD, SKR e EGR28. ...	37
4.5	Ciclopropil-isomerase (CPI1) e Citocromo P51 (CYP51)	42
4.6	Genes que realizam as últimas reações da via de biossíntese de esteróis	45
4.7	Enzimas que fazem a conversão dos fito-esteróis às formas derivadas e ao campesterol.	50
4.8	Caracterização físico-química e localização das proteínas	55
4.9	Análise da expressão diferencial.....	61
5	CONCLUSÕES.....	63
6	REFERÊNCIAS BIBLIOGRÁFICAS.....	64

1 INTRODUÇÃO

Lipídeos são um grupo de moléculas bioquimicamente diverso, estimativas conservadoras sumarizam em 200 mil o número de lipídeos nas principais classes conhecidas. São definidos como lipídeos: moléculas altamente hidrofóbicas ou ligeiramente anfifílicas que podem ser solubilizadas em solventes orgânicos (ORESIC *et al.*, 2008; SMITH, 2000).

Várias classes de lipídeos desempenham papéis fundamentais na célula, os triacilgliceróis, conhecidos popularmente triglicerídeos, são importantes reservas energéticas. Os glicerofosfolipídeos e os esfingolipídeos fazem parte da constituição da membrana plasmática e regulam sua fluidez (LEHNINGER, 2014).

Os esteróis fazem parte da constituição da membrana plasmática ajudando na regulação da fluidez e da estabilidade, assim como os lipídeos citados anteriormente (ALBERTS, 2017), além disso, atuando como hormônios, os esteróis podem participar de uma gama de processos de sinalização celular: os hormônios esteroides regulam várias reações metabólicas em animais e plantas, são assim conhecidos por derivarem de lipídeos esteróis. Os hormônios sexuais de mamíferos são esteroides: os estrógenos, progesterona e testosterona, já as plantas têm os hormônios Brassinoesteróides como representantes esteroides (ZHENG *et al.*, 2018).

Nessa perspectiva, os esteróis constituem uma importante classe de lipídeos que podem coordenar diferentes processos biológicos, como desenvolvimento ontogenético, resposta a variações ambientais, sinalização celular, crescimento e maturação reprodutiva.

1.1 Esteróis: estrutura química, diversidade e função.

Esteróis são lipídeos isoprenos caracterizados pelo núcleo esteroide: quatro anéis carbônicos conjugados, três deles com seis carbonos e um com cinco. Um esterol típico é definido quimicamente como ciclopentano-peridrofenantreno (núcleo esteroide com quatro anéis rígidos) com um grupo hidroxila no carbono 3 e uma cadeia lateral de comprimento variável (8 a 10

carbonos) ligada ao carbono 17 (FERRER et al., 2017). A figura 1 ilustra a estrutura típica de um esterol.

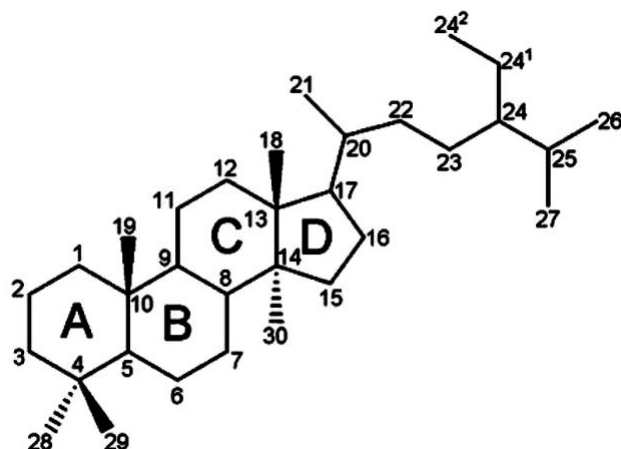


Figura 1: estrutura genérica de um esterol, os anéis A,B,C são fenatrenos hidratados e o anel D é um ciclopentano. Fonte: VALITOVA, 2016

Há uma grande diversidade de esteróis nos seres vivos, as variações mais comuns entre eles ocorrem na cadeia lateral C17 e no número de duplas ligações nos anéis do núcleo esteroide. Os esteróis mais comuns em animais e fungos são colesterol e ergosterol, respectivamente. Já em plantas β -sitosterol, estigmasterol, campesterol (GROSJEAN et al., 2015), e isofucosterol (SCHALLER, 2004) predominam.

O perfil de esteróis na planta modelo *Arabidopsis thaliana* corresponde a β -sitoesterol (64%), campesterol (11%), estigmasterol (6%), isofucosterol (3%), e brassicasterol (2%) (BENVENISTE, 2004). Em condições normais o perfil de esteróis não se altera muito para uma espécie em particular, no entanto, a proporção entre esteróis e os lipídeos de membrana pode variar muito entre tecidos, o que sugere uma diferença na expressão dos genes e na sua regulação entre os tecidos (VALITOVA, 2016).

Estruturalmente, esteróis vegetais variam no Carbono 22 e 24, β -sitoesterol tem um grupo etil ligado ao carbono 24, enquanto campesterol tem um grupo metil, já estigmasterol tem um grupo etil no carbono 24 e uma dupla ligação no carbono 22 (SCHALLER, 2004). A metilação do carbono 24 é realizada pelas enzimas C24 Esterol metil-transferases (SMT), é um dos passos

mais importantes da via de biossíntese dos esteróis (VALITOVA 2016). A figura 2 ilustra a estrutura química dos principais esteróis vegetais.

Esteróis de membrana podem se apresentar de maneira livre ou conjugada com ácidos graxos e carboidratos, formando esteróis ésteres e esteróis glicosídicos, respectivamente. Esteróis glicosídicos podem ainda ser conjugados com ácidos graxos e formarem esteróis acil-glicosídicos (FERRER et al., 2017).

As enzimas phospholipid-sterol acyltransferase (PSAT1) (BANAS et al., 2005) e sterol acyltransferase (ASAT1) (CHEN et al., 2007) favorecem a conjugação e formação de esteróis ésteres. Quando há superprodução de esteróis na célula, os esteróis são conjugados com ácidos graxos pelas enzimas PAST1 e ASAT1 sendo empacotados em *Lipid droplets* para regulação da homeostase lipídica da célula (SILVESTRO et., 2013).

A glicosilação de esteróis é favorecida pela enzima UDP-sterol glucosyltransferase, ocorre na hidroxila 3- β , essa enzima pode estar localizada na membrana plasmática, vesículas do complexo de golgi ou retículo endoplasmático (DEBOLT, 2009; MISHRA, 2015). Qualquer mudança na estrutura química de esteróis leva a mudança de suas propriedades fisiológicas, afetando sua função celular.

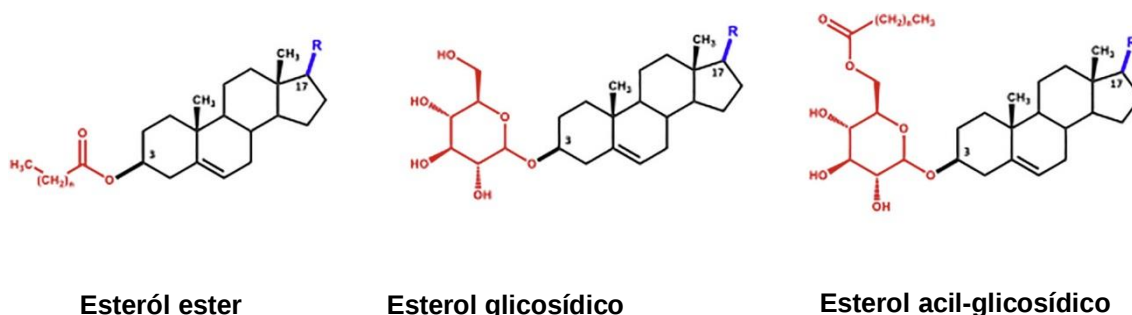


Figura 2: Estrutura química dos esteróis conjugados, a ligação dos ácidos graxos e glicídios acontece na hidroxila 3- β . Fonte: FERRER et al., 2017.

Esteróis estão diretamente envolvidos na constituição estrutural da membrana plasmática devido algumas particularidades químicas da molécula, como o grupo hidroxila 3- β , o esqueleto tetracíclico plano e uma cadeia lateral alifática de 8 a 10 carbonos. Todos os esteróis vegetais têm essas duas características e por isso interagem com fosfolipídios e esfingolipídios para

regular a fluidez da membrana limitando a sua mobilidade (GRONNIER et al., 2018). A importância dos esteróis na determinação das propriedades biofísicas da membrana também atribui aos níveis de esterol um papel proeminente nas respostas adaptativas das plantas a diferentes tipos de estresse abiótico e biótico, incluindo tolerância ao estresse térmico, hídrico e infecção de patógenos (KUCZYNSKA et al., 2019; ESTRADA et al., 2018).

Além disso, os fitoesteróis fazem a manutenção da heterogeneidade lipídica da membrana, sua capacidade de organizar a fluidez da membrana permite atuar na formação dos microdomínios lipídicos, conhecidos em inglês como *Lipids Rafts* (balsas lipídicas) (SIMON-PAS, 2011), os anéis do núcleo esteroide impõem uma conformação mais plana para as caudas alifáticas de esfingolipídios e a delta instauração no carbono 5 gera uma forte interação com os fosfolipídios, o que parece favorecer a constituição de uma superfície celular mais rígida para formação dos microdomínios.

Os *Lipids Rafts* são regiões espessas da membrana nas quais proteínas e sinalizadores celulares estão ancorados, são responsáveis por diversos processos de internalização e sinalização celular. Nos vegetais em particular estão envolvidos no crescimento do tubo polínico e dos pelos radiculares (BECK et al., 2007; ROCHE et al., 2008). Dois receptores de auxinas estão fortemente associados a *Lipids Rafts*, PIN1 e PIN2, quando ancorados aos *Lipids rafts* afetam a polarização da célula direcionando o crescimento celular, se a composição de esteróis da membrana é alterada por mutações nos genes da via de biossíntese, PIN1 e PIN2 não se distribuem da maneira correta e afetam o gravitropismo da raiz e o crescimento apropriado de diversos outros tecidos (MEN et al., 2008; YANG et al., 2013).

Técnicas genéticas de construção de modelos mutantes para genes das vias de biossíntese ajudam a entender as funções dos esteróis no crescimento e desenvolvimento vegetal. Mutações em genes da via de biossíntese no início do desenvolvimento da planta são letais, já outras apresentam fenótipos aberrantes como diminuição da estatura e nanismo (SCHALLER, 2004; SONG et al., 2019).

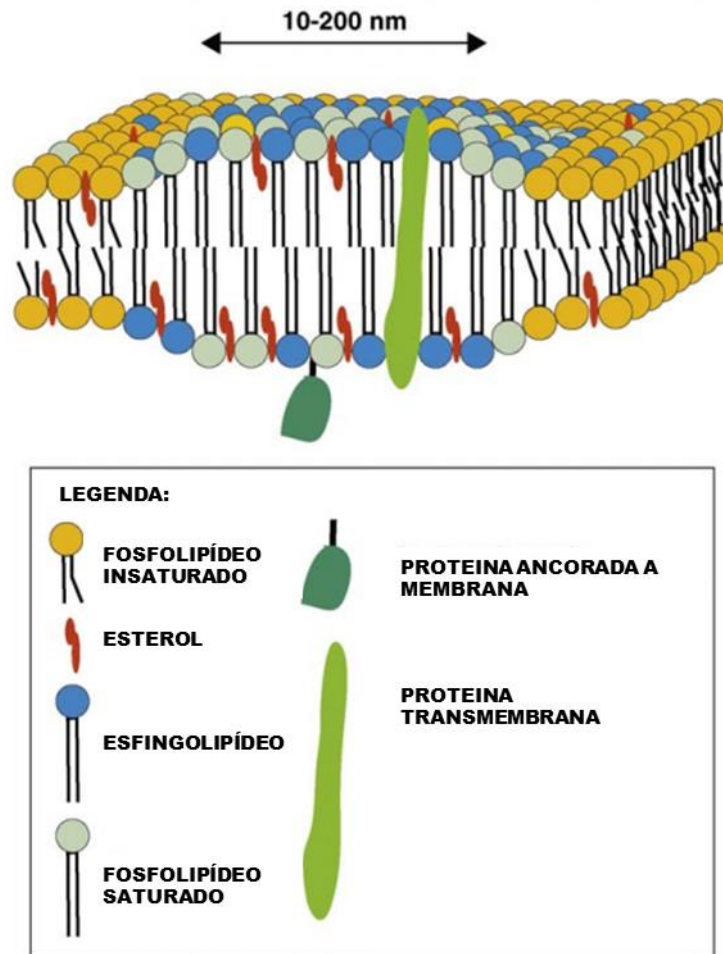


Figura 3: modelo de microdomínio lipídicos nas membranas de células vegetais, observar os esteróis centralizados entre duas moléculas de esfingolípídios e fosfolípídios. Fonte: (Boutté; Grebes, 2009)

Fitoesteróis também são responsáveis por processos de comunicação e sinalização celular, o campesterol é um dos precursores dos hormônios esteroides conhecidos como Brassinoesteróides (BR). Os Brassinoesteróides são uma classe de hormônios com mais de 60 compostos responsáveis pelos mais diferentes processos morfo-fisiológicos como: alongamento celular (caule, raiz), expansão foliar, fotomorfogênese, desenvolvimento floral, esterilidade masculina, desenvolvimento estomático e resistência ao estresse biótico e abiótico. Em *Arabidopsis* genes mutantes da via biossíntese de fitoesteróis geralmente apresentam fenótipos semelhantes aos deficientes em BR, incluindo nanismo e fertilidade masculina reduzida (NOLAN et al., 2020; ZHENG et al., 2018; VRIET et al., 2013).

Os esteróis vegetais têm impacto fisiológico em células animais pelo consumo na dieta. O esterol predominante em células animais é o colesterol. Mamíferos tem mecanismos de produção e aproveitamento do colesterol endógeno, mas a desregulação do metabolismo do colesterol causada pela dieta eleva a quantidade de lipoproteínas carreadoras de colesterol nos vasos sanguíneos e aumenta o risco de doença cardíaca, infarto do miocárdio e outros problemas cardíacos (MOREAU et al., 2010).

A relação entre o consumo de fitoesteróis e a redução do risco cardíaco está bem documentada (RAS et al., 2014), mas o consumo fitoesteróis também pode estar relacionado com aumento da atividade imunológica e no tratamento de inflamação no fígado. Os esteróis vegetais predominantes nas dietas humanas habituais são sitoesterol (66%), campesterol (22%), estigmasterol (8%) presentes em pão, cereais, vegetais, frutas, óleos vegetais e seus derivados (PLAT et al., 2019), (GENSER et al, 2012).

Esteróis são lipídios essenciais ao metabolismo celular vegetal, desempenham uma série de funções estruturais, como já visto, além de participarem da regulação do desenvolvimento e da resposta a estresses bióticos e abióticos, por isso, entender o processo de biossíntese e distribuição de esteróis nas plantas é entender como elas regulam seu metabolismo e respondem as variações do ambiente.

1.2 Genes e proteínas envolvidos nas vias de biossíntese de esteróis.

O processo de biossíntese de esteróis é complexo, envolve várias etapas, diversas enzimas e genes, é bem caracterizado no organismo modelo para plantas, *Arabidopsis thaliana*.

Todos os isoprenóides são derivados de um precursor comum, o difosfato de isopentenil (IPP) que pode ser sintetizado por duas vias diferentes: a via do mevalonato (MVA) no citosol e a via do metil eritritol fosfato (MEP) localizada no cloroplasto. A via do MVA direcionará o IPP para síntese dos fitoesteróis, já a via do MEP dará origem aos carotenoides (VRANOVA et al., 2013). A figura 4 representa as duas vias de síntese dos isoprenóides em plantas.

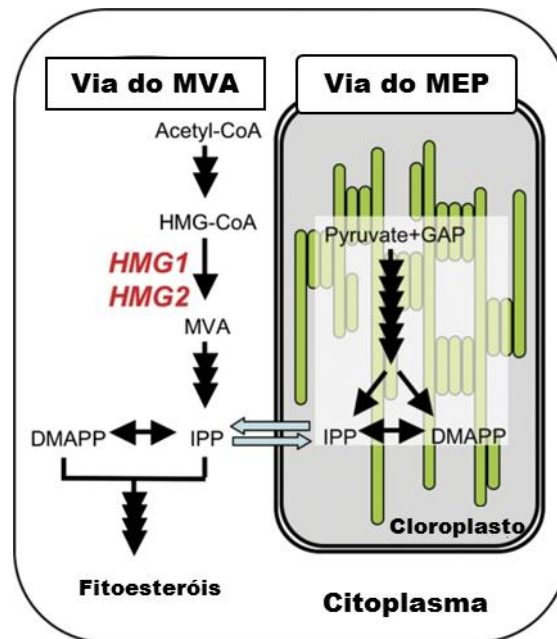


Figura 4: vias de biossíntese dos isoprenóides na célula vegetal, a via do mevalonato de localização citoplasmática e a via do metil eritrofosfato de localização plastídica. Fonte: SUZUKI et al., 2009

A via do mevalonato dá origem ao IPP por sete reações, a etapa que dá origem ao mevalonato como intermediário é a mais lenta da via, é catalisada pela enzima HMG-CoA redutase (SUZUKI et al., 2009). O IPP sintetizado na via do mevalonato será convertido em farnesil difosfato e este entrará na via de biossíntese dos esteróis propriamente dita.

O farnesil difosfato é convertido a esqualeno pela enzima Esqualeno Sintase, o esqualeno é o esqueleto de hidrocarboneto dos esteróis, sua oxigenação na extremidade da molécula é catalisada pela enzima Esqualeno Epoxidase. Essas são as duas primeiras reações da síntese dos fitoesteróis propriamente dita (LARANJEIRA et al., 2015; LIU & FU, 2018; SPANOVA & DAUM, 2011; SCHALLER, 2010).

A formação dos anéis do núcleo esteroide envolve uma reação ciclização do esqueleto de hidrocarbonetos, essa reação é favorecida por uma família de enzima chamada Oxidosqualeno ciclasas (OSC's). Em mamíferos e leveduras, a ciclização de 2,3 oxidosqualeno em lanosterol é favorecida pela enzima lanosterol sintase (LAS), já em plantas a ciclização acontece formando cicloartenol pela enzima cicloartenol sintase (CAS) (GAS-PASCUAL et al., 2014; THIMMAPPA et al., 2014).

Até pouco tempo atrás não havia conhecimento de que plantas produziam colesterol ou tinham atividade da enzima LAS, no entanto, já foi descoberto que algumas espécies apresentam baixas concentrações de colesterol e que sua produção pouco depende da atividade da enzima LAS (SONAWANE et al., 2016), também já está relatado que LAS também desempenha papel na biossíntese de fitoesteróis e de outros esteroides derivados (OHYAMA et al., 2009).

Duas reações importantes da via são as metilações no carbono 24 e as desmetilações no carbono quatro da molécula. As metilações são executadas pelas enzimas da família esterol metil-transferases, SMT1 e SMT2, essas duas reações caracterizam e diferenciam os esteróis vegetais dos esteróis de mamíferos, (CARLAND et al 2010; SCHALLER, 2010). As desmetilações no carbono quatro são feitas por um complexo enzimático que envolve de dois genes da família de esterol metil-oxidases dependente de ferro não-hemico (*SMO1*, *SMO2*), um gene da família de desidrogenases de cadeia curta(*β HSD*) e o gene da enzima esterol Ceto-redutase (*SKR*), (DARNET & RAHIER, 2019; SCHALLER, 2010; DARNET & RAHIER, 2004). As enzimas desse complexo enzimático ficam ligadas entre si e a parede do retículo endoplasmático com ajuda da enzima ERG28, que desempenha papel estrutural na via. Tanto a metilação do C-24 quanto a desmetilação do C-4 ocorrem duas vezes na via, a primeira após ciclização e a segunda após as modificações no núcleo esteroide, (SCHALLER, 2010).

Após a primeira metilação do C-25 e primeira desmetilação do C-4, ocorre uma série de reações de clivagem e modificações dos anéis do núcleo esteroide, ciclo-propil isomerase (CPI1) modifica o radical isopropano em uma das pontas da molécula. CYP51, uma enzima da super-família dos citocromos p450 retira um grupamento metil e forma uma dupla ligação no carbono 14, essa dupla ligação no C-14 é reduzida pela enzima Fackel (FK) no passo seguinte. Por fim, a enzima Hydra (HYD) faz a isomeração do anel B do núcleo esteróide (SHORT et al., 2018; SCHALLER, 2010; ROZHON et al 2013; KIM et al., 2010; SHIRICK et al., 2000; QIAN et al., 2013; RAHIER & KRAST, 2014).

Na etapa seguinte ocorre a bifurcação da via de biossíntese de esteróis. A enzima SMT2 catalisa uma nova metilação do carbono 24, moléculas com

grupo etil originarão β -Sitoesterol, no entanto, algumas moléculas não sofrem essa segunda metilação e seguirão o caminho para síntese do campesterol. Independente do caminho que seguirem, o complexo de desmetilação do C-4 a segunda desmetilação e direciona a molécula para o final da sua biossíntese. As reações finais de biossíntese consistem em uma serie de reduções e hidrogenações catalisadas pelas enzimas STE, DWF1 e DWF5 (TATON *et al.*, 2000; HUSSELSTEIN *et al.* 1999; ZHENG *et al.*, 2018; YOUN *et al.*, 2018).

Para finalizar a via, o estigmasterol é sintetizado a partir do β -sitoesterol o qual não tem dupla ligação no carbono 22. A enzima C-22 esterol desaturase (CYP710A) catalisa a formação dessa dupla ligação no carbono 22 do β - sitoesterol convertendo-a estigmasterol, essa a enzima também pertence à família dos citocromos p450 (SCHALLER, 2010).

Todas as reações da via de biossíntese foram caracterizadas em *Arabidopsis thaliana*, ou por ensaios complementação usando leveduras, ou inserindo mutações nos genes para observar os fenótipos dos mutantes. Mutações nos genes mais basilares da via são letais, ao passo que mutações nos genes que realizam as reações finais, como DWF1 e DWF5 geram fenótipos pleiotropicos, como nanismo, esterilidade, atraso na floração etc. Fenótipos característicos de deficiência na síntese de hormônios esteróis em plantas (SCHALLER, 2010; ZHENG *et al.*, 2018).

Considerando as funções e a importância dos esteróis vegetais para o desenvolvimento da planta como um todo, caracterizar esses genes em outras espécies é de suma importância, principalmente em espécies que apresentam evolução genômica rápida, a exemplo dos membros da família *Arecaceae* (JIAO *et al.*, 2014; LOW *et al.*, 2017), como veremos as espécies dessa família de plantas passaram por diversos eventos de duplicação do seu genoma e podem apresentar inovações genéticas na via: novos genes, novas funções e diferença na expressão dos genes e de suas isoformas nos tecidos das plantas. Várias espécies dessa família tem o genoma sequenciado, entre elas, *Cocos nucifera* (coqueiro), *Phoenix dactylifera* (tamareira) e *Eleais guineensis*(dendezeiro), essa última recebe especial destaque por ter papel destacado na produção de óleos e lipídeos.

1.3 *Elaeis guineensis* Jacq: o dendezeiro.

O dendezeiro, conhecido cientificamente como *Elaeis guineensis* Jacq., é uma espécie pertencente à família Arecaceae, esta família forma um grupo distinto de plantas entre as monocotiledôneas, popularmente conhecidas como Palmeiras. Outras espécies integram o gênero *Elaeis*, entre elas a espécie nativa das américas *Elaeis oleifera* (Kunth) Cortés, conhecida popularmente como Caiaué.

O dendezeiro é uma espécie nativa da região tropical úmida da África, marcadores genéticos fornecem evidências de que o centro original de distribuição do dendezeiro seja a Nigéria. Atualmente, o dendezeiro é cultivado em toda região tropical do globo, incluindo, África, Ásia, Américas do sul e central. No Brasil, o dendezeiro provavelmente foi introduzido durante o tráfico africanos negros no período colonial (VASCONCELOS, 2017; GODSWILL et al 2016).

Nos últimos 50 anos a expansão do cultivo do dendezeiro ocorreu pelos diversos modos de uso da palmeira, além da sua adaptabilidade a diversas condições ecológicas. O dendezeiro é interessante agronomicamente pela alta produção de óleo vegetal, atualmente, desponta como líder mundial no segmento, no período 2014 as 2015, foi responsável por 35% da produção de óleo vegetal entre as principais plantas oleaginosas. O valor bruto das commodities de dendê com agregação de valor das diferentes áreas pode chegar 150 bilhões de dólares (VASCONCELOS, 2017).

Levando em consideração a importância econômica global da palmeira, em 2013 foi lançado ao domínio público a primeira versão do sequenciamento do genoma da variedade *pisifera* do dendezeiro pelo Malaysian Palm Oil Board (MPOB) com uma sequência de ~ 1,5 Gb do genoma total de 1,8 Gb, com 50N de 1,05 Mb (SINGH et al., 2013), nesse projeto 30.772 genes foram identificados.

Já em 2016 o genoma da variedade *dura* de *E. guineensis* foi liberado com 1.7 Gb e N50 de 0,76 Mb, (JIN et al., 2016), cerca de 36.105 genes foram identificados nesse projeto de sequenciamento. Chan et al., (2017) publicou a revisão dos modelos gênicos resultando na identificação 26.059 genes codificantes com alta qualidade evidenciados por transcriptoma e Refseq.

Os dados do sequenciamento do genoma do dendezeiro mostram que o vegetal já passou por várias duplicações completas do genoma. Comparações entre o genoma do *E. guineensis* e o genoma de outras plantas como a tamareira, bananeira e *Arabidopsis*, mostram que a tamareira e o dendezeiro (ambas palmeiras) compartilham várias duplicações segmentares, isso acontece porque ao longo da história evolutiva da tamareira e do dendezeiro duas paleoduplicações aconteceram, uma compartilhada com todas as Monocotiledôneas, ou seja, no ancestral comum das monocot e a segunda ocorreu apenas na família Arecaceae, no ancestral comum das palmeiras (JIAO et al., 2014). Portanto, é possível encontrar vários genes duplicados nas palmeiras, frutos da duplicação do genoma todo. Genes originados da duplicação do genoma todo são chamados parálogos, já genes duplicados por processos de especiação são chamados ortólogos, os dois termos são usados como terminologia para relação evolutiva de homologia genética (KRISTENSEN et al., 2011; KAWASHIMA, 2019).

O sequenciamento do genoma do dendezeiro também mostrou que evolução genômica de *E. guineensis* e das Arecaceae parece ser mais lenta que a de outras monocotiledôneas. A eudicotiledônea *Nelumbo nucifera* tem mais blocos sintênicos ancestrais conservados com *E. guineensis* do que com *Oryza sativa* (Arroz), o que significa que os genes duplicados (parálogos) no arroz se modificam e se perdem mais rápido que no Dendê (JIAO et al., 2014), (LOW et al., 2017).

Estudos de evolução genômica são importantes, pois os pesquisadores de dendezeiro podem se beneficiar da miríade de descobertas de outras culturas com monocotiledôneas, usando informações ortológicas e regiões sintênicas de outras plantas para inferir regiões no genoma do dendê que podem controlar o mesmo traço agrônômico ou biológico de maneira geral (JIAO et al., 2014), (LOW et al., 2017).

Os dois projetos de sequenciamento identificaram genes essenciais associados ao desenvolvimento dos frutos e a produção de óleo neles, como *VIR* que controla a coloração do mesocarpo durante a maturação dos frutos e *SHELL* que controla a forma dos frutos, viabilizando a seleção de frutos como alto valor comercial. Estudos de expressão gênica identificaram que o gene

WRI1 é hiperexpresso em frutos que produzem altas taxas de óleo no mesocarpo e endosperma (LOW et al., 2017), (JIN et al., 2017).

A genômica do dendezeiro ajudou no entendimento do desenvolvimento do fruto, biossíntese de ácidos graxos, resistência a doenças e ao frio. No entanto, muito esforço ainda é necessário para melhorar a qualidade do plantio e como já dito anteriormente, os esteróis tem funções importantes na célula vegetal que influenciam seu desenvolvimento e fisiologia, a via de biossíntese de esteróis está bem caracterizadas em organismos modelos como *Arabidopsis thaliana* (SCHALLER, 2010) e algumas culturas como a de tomate (*Solanum lycopersicum*) (SONAWANE, 2016), mas ainda não é bem compreendida em organismos como dendezeiro.

1.4 Transcriptômica e estudos de expressão gênica em *E. guineensis*.

Após a conclusão do projeto do genoma humano, as atenções científicas começaram a se voltar a paisagem transcricional de todos os genes em um genoma, objetivo era investigar os mecanismos funcionais subjacentes às variações fenotípicas dos organismos. Nessa perspectiva, estudos biológicos sobre dados genéticos de alto rendimento vão do nível genômico ao nível transcriptômico (WAN; LI 2019).

O transcriptoma é uma coleção de todas as moléculas de RNA em uma célula ou em uma população de células, pode incluir informações de um conjunto de amostras em um certo estágio de desenvolvimento, condição ambiental ou tratamento experimental. Os dados de transcriptomas incluem bioinformação espaço-temporal afetada por diversos fatores, tipos de tecidos e ambiente interno / eventos externos. Portanto, os dados do transcriptoma são mais complexos e mais dinâmicos que os dados do genoma (SRIVASTAVA; GEORGE; KARUTURI, 2020), (WAN; LI 2019).

Em geral, as técnicas de transcriptoma envolvem dois princípios genéticos: hibridização e polimerização de cDNA (algumas técnicas mais recentes não utilizam mais amplificação de cDNA). Nas técnicas de hibridização o cDNA de interesse é marcado com fluorescência e incubado em uma matriz sólida de náilon, vidro ou sílica frequentemente chamada de chip gênico. O chip gênico contém um microarranjo de várias sondas de cDNAs as quais podem ser

customizadas ou comerciais, um scanner captura a fluorescência emitida com a hibridização entre as sondas e o cDNA de interesse indicando os níveis de expressão genica da amostra (SNUSTAD; SIMMONS, 2017; BUMGARNER, 2013).

As reações de sequenciamento envolvem a polimerização molécula de DNA seguida de um procedimento físico-químico para detectar qual base nitrogenada do DNA é adicionada a molécula sequencialmente. As primeiras técnicas de sequenciamento a se popularizarem seguiam o método de terminalização de cadeia com didesoxionucleotídeos (ddNTP's) do bioquímico britânico Frederick Sanger (SNUSTAD; SIMMONS, 2017), essa tecnologia foi substituída pelo chamado sequenciamento de nova geração (*Next generation sequencing* - NGS), também chamada de segunda geração de sequenciadores.

As tecnologias NGS permitem sequenciar, mapear e quantificar transcriptomas através de um procedimento de alto desempenho chamado RNAseq. No RNAseq o cDNA é fragmentado para produzir bibliotecas de sequenciamento que são ligadas a um adaptador, a metodologia de sequenciamento em si varia de acordo com a plataforma. As plataformas Ion Torrent (Life Technologies), 454 (Roche), Illumina utilizam amplificação, ao passo que SOLiD (Life Technologies) utiliza ligação de nucleotídeos para o sequenciamento (SHARMA et al., 2017; WANG et al., 2019).

As tecnologias mais atuais de sequenciamento não utilizam mais amplificação, nem bibliotecas de fragmentos gênicos, o sequenciamento acontece em tempo real, com uma única molécula, o processo todo foi simplificado. As plataformas NGS se valem de etapas repetidas de escaneamento e lavagem, que foram eliminadas nos sequenciadores de última geração. Conhecidas como terceira e quarta geração de sequenciadores, as plataformas de sequenciadores de última geração são *Single-Molecule Real-Time sequencing* da Pacific Biosciences (PacBio), *System True single molecule sequencing* da Helicos Genetic e *Nanopore Sequencing* da Oxford Nanopore technologies (AMEUR; KLOOSTERMAN; HESTAND, 2018), (SHARMA et al., 2017), (WANG et al., 2019).

As etapas finais do RNAseq são computacionais: alinhar e / ou montar as leituras de sequenciamento em um transcriptoma, quantificar leituras que se sobrepõem às transcrições, filtrar e normalizar entre amostras, por fim, modelagem estatística de mudanças significativas nos níveis de expressão de genes individuais e/ou transcritos entre grupos de amostra (STARK; GRZELAK; HADFIELD, 2019).

Normalmente, o resultado de um RNAseq após a etapa computacional é uma lista de genes diferencialmente expressos em condições fisiológicas e tratamentos diferentes. Um dos principais desafios da biologia computacional atual é fornecer métodos automáticos robustos para uma anotação funcional significativa das longas listas de genes ou proteínas derivados de tais estudos de alto rendimento. A análise de enriquecimento funcional é a metodologia mais popular disponível para derivar implicações funcionais de conjuntos de genes cooperantes, ela utiliza testes estatísticos para encontrar anotações significativas em grupos de genes (FONTANILLO, 2011).

A última versão do genoma do dendezeiro está disponível nos bancos de dados, além disso, mais de 25 projetos de transcriptoma de diversos cultivares também estão disponíveis. Diversos estudos já foram feitos utilizando os transcriptomas de *E.guineensis* depositados nos bancos de dados: Badai et al., (2019), utilizando dados de 28 transcriptomas de diversos tecidos da planta, identificaram 41 genes expressos apenas no mesocarpo dos frutos do dendezeiro, os autores elaboraram um heatmap que ilustra o perfil de expressão dos genes nos diversos tecidos da planta e no mesocarpo.

Além desses outros estudos já foram realizados utilizando unicamente dados depositados nos bancos de dados genômicos. As Diacilglicerol Aciltransferases (DGATs) foram caracterizadas em *E.guineensis* utilizando dados genômicos atuais depositados nos bancos de dados, essas enzimas realizam a última etapa de síntese dos trigliceróis, regulam a produção e o acúmulo desses lipídeos no citoplasma das células, portanto, são enzimas relacionadas a produção de óleo nos tecidos vegetais (ROSLI, et al 2018). Além disso, os autores também avaliaram a arquitetura dos genes e de várias isoformas de pré-mRNA para DGATs encontradas no dendezeiro, bem como a

expressão dos genes no mesocarpo, no endosperma e em vários tecidos vegetativos durante todo o desenvolvimento da planta (ROSLI, et al 2018).

Xiao et al., (2018) fizeram um levantamento de genes ortólogos relacionados ao metabolismo de ácidos graxos em *Cocos nucifera* (Coqueiro), *Phoenix dactylifera* (Tamarareira) e *Eleais guineensis* utilizando os genomas e transcriptomas das três espécies depositados nos bancos dados. A maioria dos genes relacionados ao metabolismo de ácidos graxos eram altamente conservados e retidos em segmentos homólogos nas três espécies. Os autores também realizaram uma análise da expressão dos genes no endosperma, mesocarpo e em tecidos vegetativos e mostraram que os genes envolvidos na conversão de piruvato em diversos tipos ácidos graxos tiveram uma expressão cinco a seis vezes maior no endosperma do coqueiro e no mesocarpo do dendezeiro do que nos tecidos da folha ou embrião, o que é esperado em palmeiras que tem alta produção de óleo, ao contrario da tamarareira (XIAO et al., 2018).

Como já dito anteriormente, a quinta versão do genoma do dendezeiro está disponível nos bancos de dados, portanto, é possível verificar a homologia da via de biossíntese dos esteróis no dendezeiro utilizando *Arabidopsis* como modelo, caracterizar os genes, seus domínios funcionais e localização genômica. Para além disso, utilizando dados de transcriptomas disponíveis na literatura e nos bancos de dados é possível construir o perfil de expressão gênica dos genes da via de biossíntese de esteróis em diferentes tecidos e cultivares de espécies do dendezeiro, semelhante aos estudos citados anteriormente. Estudos como esses ajudam explicitar o funcionamento celular em espécies não modelo, como *Eleais guineensis*, é um passo inicial para caracterizar novas vias e descobrir metabólitos.

2 OBJETIVOS

2.2 Objetivos gerais

- Realizar a predição genômica e analisar expressão dos genes ortólogos e parálogos da via de biossíntese dos esteróis em *Elaeis guineensis* usando *Arabidopsis thaliana* como modelo.

2.3 Objetivos específicos

- Alinhar os genes encontrados e avaliar o nível de similaridade entre os genes da via de biossíntese de esteróis em *E. guineensis* e *A. thaliana*.
- Destacar regiões e domínios conservados entre as sequências de proteínas de *E. guineensis* e *A. thaliana*.
- Nomear, agrupar e representar filogeneticamente os genes encontrados entre as duas espécies.
- Estabelecer blocos sintênicos de genes entre as duas espécies
- Realizar a anotação funcional de genes e os produtos gênicos
- Analisar a expressão dos genes, destacando genes diferencialmente expressos na raiz, flores e frutos do dendezeiro.

3 MATERIAL E MÉTODOS

3.1 Busca de ortólogos: alinhamento local e domínios conservados

As sequências de aminoácidos das enzimas da via de biossíntese de esteróis foram baixadas dos bancos de dados: Genbank, Uniprot e do banco de dados específico para *Arabidopsis thaliana* TAIR (The Arabidopsis Information Resource, <http://www.arabidopsis.org>).

As sequências foram baixadas em arquivos Multi-fasta e usadas para busca de ortólogos e parálogos na versão mais atual do genoma de *Elaeis guineensis*, EG5-GCA_000442705.1, utilizando a ferramenta de alinhamento local do Genbank para sequência de aminoácidos, o BLASTp (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>), os genes em *E.guineensis* que tinham E-value igual 0.0 ou próximo foram baixados e armazenados em arquivos multi-fasta, as referências e dados genômicos dos genes também foram anotados. As proteínas de *A. thaliana* usadas para o BLASTp estão grifadas em vermelho na tabela 1.

Tabela 1: lista de genes de *Arabidopsis thaliana* envolvidos na biossíntese de esteróis, os genes usados para busca contra o genoma de *Elaeis guineensis* estão grifados em vermelho, apenas os genes com evidencia funcional comprovada na literatura foram usados para busca.

Abreviação	Nome da enzima	Gene
SQS	squalene synthase	AT4G34640 (SQS1); AT4G34650 (SQS2);
SQE	squalene monooxygenase / squalene epoxidase	AT1G58440 (SQE1, XF1); AT2G22830 (SQE2); AT4G37760 (SQE3); AT5G24140 (SQE4, SQP2); AT5G24150 (SQE5, SQP1); AT5G24160 (SQE6);
CAS	2,3(S)-oxidosqualene-cycloartenol cyclase	AT2G07050 (CAS1);
LAS	lanosterol synthase	AT3G45130 (LAS1);
SMT1	cycloartenol-C24-methyltransferase	AT5G13710 (SMT1);
SMT2	24-methylene lophenol-methyltransferase	AT1G20330 (SMT2-1, SMT2); AT1G76090 (SMT2-2, SMT3);
SMO1	24-methylenecycloartanol 4 α -methyl oxidase	AT4G12110 (SMO1-1); AT4g22753 (SMO1-2); AT4g22756 (SMO1-3);
SMO2	4 α -methyl- Δ 7-sterol-4 α -methyl oxidase	AT1G07420 (SMO2-1); AT2G29390 (SMO2-2);

<i>βHSD</i>	3β-hydroxysteroid dehydrogenase/C4-decarboxylase	AT2G33630 (AT2G33630.1); AT2G43420 (); <i>AT1G47290</i> <i>(AT3BETAHSD/D1,CSD)</i> ; AT2G26260 (AT3BETAHSD/D2);
<i>SR</i>	sterone ketoreductase	At5g18210 (SKR);
<i>CPI</i>	cyclopropyl sterol isomerase	<i>AT5G50375 (CPI1)</i> ;
<i>C14DM</i>	obtusifoliol-14α-demethylase	<i>AT1G11680</i> <i>(OBT14DM1,CYP51G1)</i> ; AT2G17330 (OBT14DM1,CYP51G2); <i>AT1G20050 (HYD1)</i> ;
<i>7ISO</i>	Δ8-Δ7-sterol isomerase	<i>AT3G52940 (FK)</i> ;
<i>C14R</i>	Δ8,14-sterol-Δ14-reductase	<i>AT3G02580 (C5DES1,STE1)</i> ;
<i>C5D</i>	Δ7-sterol-C5-desaturase	AT3G02590 (C5DES2,STE2);
<i>C7R</i>	Δ5,7-sterol-Δ7-reductase	AT1G50430 (DWF5);
<i>C24R</i>	Δ5-sterol-Δ24-reductase (isomerase)	AT3G19820 (DWF1);
<i>C22D</i>	C-22 sterol desaturase	<i>AT2G28860 (CYP710A4)</i> ; AT2G28850 (CYP710A3); AT2G34500 (CYP710A1); AT2G34490 (CYP710A2);
<i>ERG28</i>	complex anchor	<i>AT1G10030 (ERG28)</i> ;
<i>USGT</i>	UDP-glucose: sterol glucosyltransferase	AT3G07020 (UDP1,SGT); <i>AT1G43620(UDP2,UGT80B1)</i>
<i>ASAT1</i>	acyl-CoA sterol acyl transferase 1	AT3G51970 (SAT1);
<i>PSAT</i>	Phospholipid--sterol O-acyltransferase	At1g04010 (LCAT2,PSAT1);

FONTE: Adaptado de Darnet & Rahier (2004), Schaller (2010)

Os domínios conservados dos genes homólogos do dendezeiro foram confirmados usando a ferramenta Pfam (<http://pfam.xfam.org/search/>). As informações dos domínios foram anotadas e comparadas, domínios que carregam posições semelhantes nas enzima, tem tamanhos semelhantes e a mesma família de motivos funcionais relacionados a síntese de esteróis foram considerados conservados entre as duas espécies, genes inicialmente anotados mas que não tinham domínios conservados entre dendê e *A. thaliana* foram descartados. O processo foi repetido para todas as enzimas da via de biossíntese de esteróis livres e conjugados.

3.2 Caracterização das proteínas

As proteínas foram caracterizadas quanto seu tamanho, peso molecular, ponto isoelétrico, composição de aminoácidos, índice de estabilidade, índice alifático, Hidropaticidade média (GRAVY), estruturas secundárias e regiões

transmembrana. O programa exPASy (<https://web.expasy.org/protparam/>) foi usado para avaliar as propriedades físico-químicas, tanto das proteínas de *A. thaliana* quando as de *E. guineensis*. O peso molecular proteico foi avaliado pelo programa Sequence Manipulation Suite (SMS) (https://www.bioinformatics.org/sms/prot_mw.html), já a localização sub-celular foi predita com a ferramenta online para biologia celular vegetal Plant-mPLOC (<http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/>).

3.3 Dendrograma, agrupamento e classificação dos genes.

Os alinhamentos globais múltiplos foram feitos com a ferramenta online Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>), os alinhamentos foram salvos em arquivos multi-fasta e sua visualização foi feita com a ferramenta SMS multialign show (http://www.bioinformatics.org/SMS/multi_align.html). A partir dos alinhamentos múltiplos foi possível construir a matriz de identidade e o Dendrograma dos genes, a matriz de identidade foi obtida dos dados do alinhamento no Clustal Omega e sua visualização foi feita pelo R. Já os Dendrogramas foram elaborados usando o programa MegaX (<https://www.megasoftware.net/>), usando neighbor joining como método de agrupamento e bootstrap com 1000 repetições. Os motivos conservados foram destacados usando a ferramenta online MEME (<http://meme-suite.org/tools/meme>). Proteínas com mínimo de 40% identidade foram classificadas na mesma família; sequências com identidade entre 55%-60% na mesma sub família (LEPESHEVA & WATERMAN, 2002); Isoformas ou alelos com identidade maior 97%.

A localização cromossômica dos genes será determinada utilizando a versão atual do genoma de *E. guineensis*, com visualização pela macro Mapdraw do pacote office.

3.4 Análise da expressão gênica

Para análise da expressão dos genes seis transcriptomas de diversos tecidos do dendezeiro foram utilizados: DRR286072 depositado por Wang et al (2020) no ncbi, PRJNA700429 depositado por Fooyontphanich et al (2021) também depositado no ncbi, PRJNA474663 depositado por Ooi et al (2019) no ncbi. Por fim, um transcriptoma do fruto dos três principais cultivares de

dendezeiro DRA001857 depositado por Jin et al (2017) banco de dados DNA do Japão. A tabela 2 resume as informações dos transcriptomas usados para análise da expressão dos genes da via de biossíntese.

Tabela 2: Lista de projetos utilizados para fazer a avaliação da expressão genica dos genes da via de biossíntese dos esteróis no dendezeiro.

ESPÉCIE	TECIDO	BIBLIOTECA	ANO	BIOPROJECT
E.GUINEENSIS TENERA	Raiz	Paired	2021	PRJDB9517
E.GUINEENSIS TENERA	Zona de abscisão do fruto	SINGLE	2021	PRJNA700429
E.GUINEENSIS PSIFERA	Carpelo Floral	Paired	2018	PRJNA474663
E.GUINEENSIS PSIFERA	Carpelo Floral Modificado	Paired	2018	PRJNA474663
E.GUINEENSIS PSIFERA	Estaminodio	Paired	2018	PRJNA474663
E.GUINEENSIS PSIFERA	Estaminodio Modificado	Paired	2018	PRJNA474663
E.GUINEENSIS DURA	Caroço	Paired	2017	DRA001857
E.GUINEENSIS PSIFERA	Caroço	Paired	2017	DRA001857
E.GUINEENSIS TENERA	Caroço	Paired	2017	DRA001857
E.GUINEENSIS DURA	Mesocarpo	Paired	2017	DRA001857
E.GUINEENSIS PSIFERA	Mesocarpo	Paired	2017	DRA001857
E.GUINEENSIS TENERA	Mesocarpo	Paired	2017	DRA001857

Os valores de qualidade das amostras foram avaliados usando o software FASTQC, leituras com qualidade menores que 20 foram retiradas usando o programa trimomatic (BOLGER et al., 2014). A quantificação da lista de genes resultado do alinhamento foi feita usando o programa Salmon (PATRO et al., 2017). Os valores de abundância da expressão retirados do Salmon foram importados para o programa de análises estatística R (R CORE TEAM, 2018) utilizando o pacote TXIMPORT (LOVE et al., 2014). A análise de expressão diferencial dos dados no programa R utilizando o pacote DESeq (SONESON et al., 2015; LOVE et al., 2014), leituras com P-value ajustado <0.05 e $\log_2\text{FoldChange} > 2$ foram consideradas diferencialmente expressas, a análise foi conduzida utilizando os transcriptomas dos tecidos da flor, fruto, raiz e zona de

abscisão do fruto. Os dados de expressão foram normalizados usando o Log2 e o pacote ComplexHeatmap foi utilizado para construção dos gráficos de expressão, para fins de comparação da expressão dos genes nos tecidos.

4 RESULTADOS E DISCUSSÃO

Em uma visão global, as 19 enzimas da via de biossíntese de esteróis foram alinhadas a 54 genes no genoma de *Elaeis guineensis*, cada gene alinhou para pelo menos um gene, indicando que a via é conservada nesse organismo. A análise de cada alinhamento, será feita separadamente.

4.1 Esqualeno Sintase (SQS1) e esqualeno epoxidase (SQE1)

Tanto esqualeno sintase quanto esqualeno epoxidase alinharam para dois genes no genoma de *Elaeis guineensis* cada uma. A tabela 3 resume os resultados do alinhamento.

Tabela 3: a tabela mostra os dois genes encontrados com alinhamento usando a ferramenta BLASTp e as informações genômicas dos gene.

Gene em A.thaliana	Gene em E.guineensis	Número de acesso do gene	CDS (bp)	Peptídeo (aa)
SQS1 AT4G34640	EgSQS1	NW_011561195	1640	408
	EgSQS2	NC_025993.1	1566	408
			1231	332
SQE1	EgSQE1	NC_025993.1	1992	520
AT1G58440	EgSQE2	NC_026005.1	2066	520

O gene *EgSQS2* tem duas isoformas, sendo que uma delas é menor, provavelmente produto de recomposição alternativa no pré-mRNA.

As três enzimas SQS1 tem o domínio *Squalene/phytoene synthase* (SQS_PSY) com assinatura funcional de 263 aminoácidos. Duas regiões ricas em aspartato caracterizam esse domínio: os motivos DTVEDD e DYLED são conservados em diferentes organismos, como as plantas *Arabidopsis thaliana*, *Selaginella moellendorffii*, *Populus trichocarpa*, o mamífero como *Rattus norvegicus* e até protozoários como *Trypanosoma cruzi* (ZHENG et al., 2018; LIU; FU, 2018).

Esses resíduos estão relacionados a ligação do farnesil difosfato ao sítio catalítico da enzima, portanto, mutações nessas regiões levam a alteração da atividade enzimática. Essas duas regiões estão conservadas nos três genes,

indicando assim a homologia funcional das enzimas esqualeno sintase em *E. guineensis*. A figura 5 mostra o alinhamento e os motivos conservados nas SQS.

Quanto a SQE1, as três enzimas apresentaram o domínio *Squalene epoxidase* (SE) com assinatura funcional de 275 aminoácidos, além desse, um domínio de ligação ao NADPH caracterizado pelos motivos: GXGXXG, DG, GD aparece ao longo das três proteínas. Outra característica das enzimas são os vários resíduos de tirosina conservados ao longo das sequências de SQE's nos mais diversos organismos. Em seres humanos a ligação de hidrogênio entre Y195 e Q168 é fundamental para atividade enzimática da proteína, há indicativos de que vários resíduos aromáticos como fenilalanina e tirosina parecem fazer parte do sítio ativo ao longo de esqualeno epoxidases. Todas essas características são conservadas em posições semelhantes nas duas enzimas de *E.guineensis* (PADYANA et al., 2019; ABE et al., 2007; RASBERY et al., 2007). A figura 6 apresenta os motivos conservados nas SQEs das duas espécies. A tabela 3 resume os resultados da busca de domínios no PFAM das enzimas SQS e SQE.

A análise do alinhamento (Figura 6), da matriz de identidade (figura 7) e do Dendrograma (Figura 37) mostram que os genes AtSQS1, EgSQS1 e EgSQS2 são da mesma sub-família, o grau de similaridade entre os três é de 71,32%. Outra observação importante é o índice de similaridade entre EgSQS1 e EgSQS2(100%) indicando que provavelmente são alelos ou isoformas.

Tabela 4: a tabela mostra os domínios dos respectivos genes encontrados na busca, é possível perceber que os genes do dendezeiro são idênticos

Gene	Pfam	Posição	Família	Gráfico
AtSQS1	PF00494.19	44-318	SQS_PSY	
EgSQS1	PF00494.19	44-316	SQS_PSY	
EgSQS2	PF00494.19	44-316	SQS_PSY	
AtSQE1	PF08491.10	211-484	SE	
EgSQE2	PF08491.10	206-479	SE	
EgSQE1	PF08491.10	206-479	SE	

Fonte: tabela autoral

Quanto a SQE, o percentual de identidade entre as enzimas de *E. guineensis* e *A. thaliana* é maior que 70%, portanto os três genes pertencem a mesma família. Entre as duas enzimas de *E. guineensis* o percentual de identidade é de 81% o que indica que as duas enzimas não são idênticas e há diferença na sequência de aminoácidos entre elas, portanto, não são alelos ou isoformas. As SQE de *E. guineensis* formam um agrupamento distinto com 90% de suporte no Dendrograma. A figura 7 mostra a matriz de identidade das enzimas SQE e SQS.

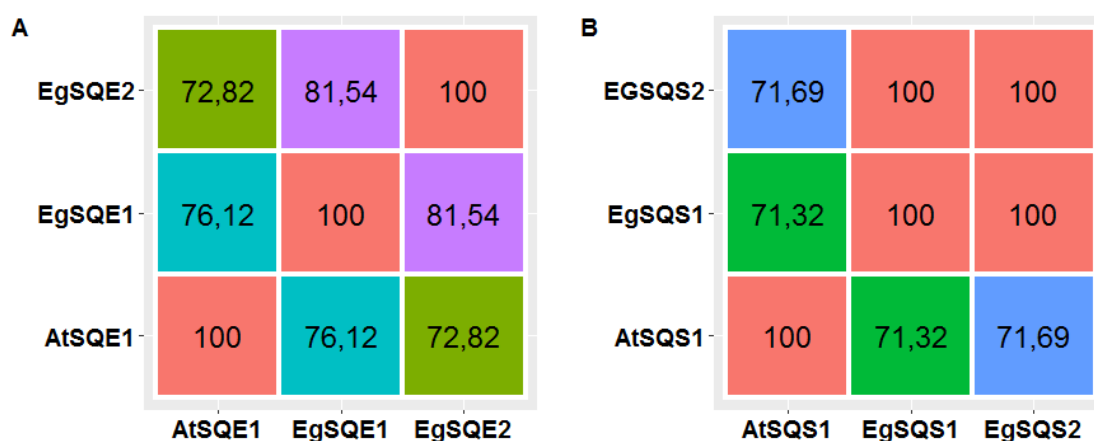


Figura 7: matriz de identidade dos genes envolvidos no metabolismo do esqualeno, em (A) esqualeno epoxidase e em (B) esqualeno sintase.

Esses resultados indicam que os genes *SQS* e *SQE* de *Arabidopsis* tem dois ortólogos cada em *E. guineensis*, sendo que apenas *SQE* acumulou diferença a ponto de formar um parálogo que ainda não divergiu funcionalmente dos seus ancestrais.

Quanto a expressão dos genes, o gene *EgSQE1* é expresso em todos tecidos e tem altos níveis de expressão no tecido do mesocarpo especificamente, o outro ortólogo para *SQE* também está expresso em todos os tecidos, assim como *EgSQS1*, esses três genes clusterizam juntos no heatmap da figura 8. As isoformas de *EgSQS* apresentam expressão baixa ou nenhuma expressão em alguns tecidos, mostrando que sua atividade não é essencial para o funcionamento celular ou que a transcrição dos outros genes no tecido é suficiente.

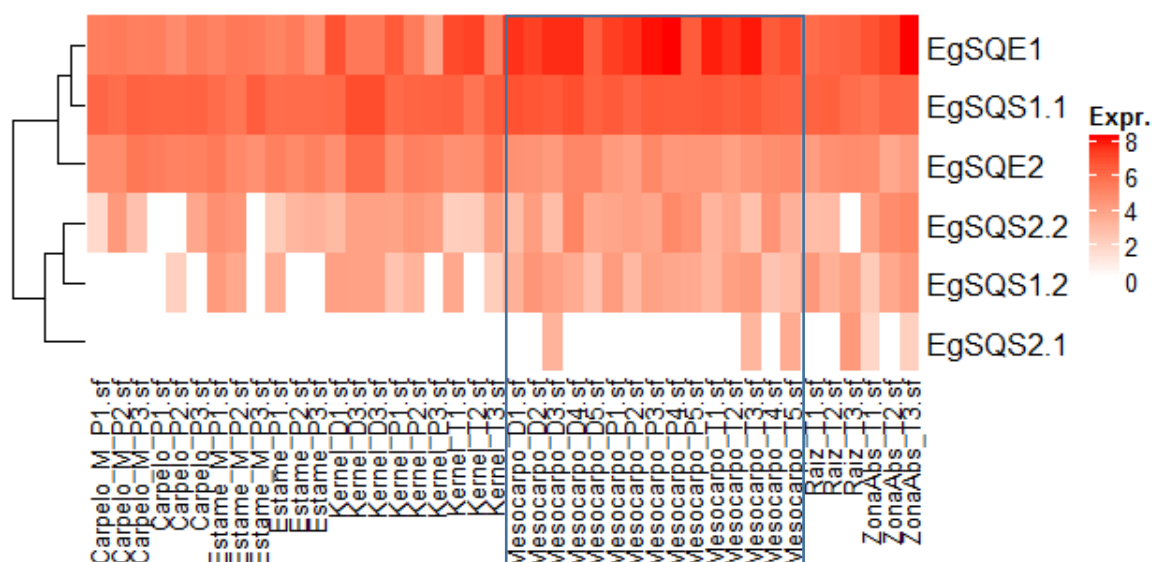


Figura 8: heatmap do valor de expressão dos genes normalizado por log2, os valores mais avermelhados indicam alta expressão, as isoformas dos genes são indicadas por números após o ponto.

4.2 Cicloartenol Sintase (CAS1) e lanosterol sintase (LAS1).

O BLASTp alinhou as duas enzimas, tanto cicloartenol sintase quanto lanosterol sintase, para os mesmos 5 genes em *E.guineensis*. LAS e CAS são proteínas semelhantes que atuam direcionando a via de biossíntese dos esteróis para a síntese de colesterol, no caso de LAS, ou para síntese dos fitoesteróis, no caso da CAS, ambas pertencem a mesma família de OSCs como dito na introdução deste trabalho, portanto são bastante similares(65%). A tabela 5 resume o resultado dos alinhamentos para as OSCs.

Tabela 5: resultado dos alinhamentos usando BLASTp e informações genômicas dos gene.

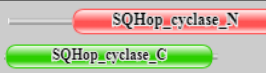
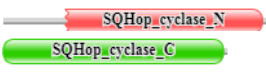
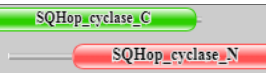
Gene em A.thaliana	Gene em E.guineensis	Número de acesso do gene	CDS (bp)	Peptídeo (aa)
CAS1	EgCAS4	NC_026002.1	2828	829
AT2G07050	EgCAS5	NW_011551001.1	3050	759
LAS	EgCAS1	NC_025999.1	3251	811
AT3G45130	EgCAS2	NC_025999.1	3197	759
	EgCAS3	NC_025999.1	3058	752
			3163	759

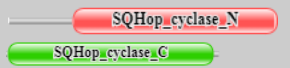
A predição de domínios no PFAM apontou que as cinco enzimas de *E.guineensis*, LAS1 e CAS1 de *A.thaliana* apresentam o domínio Squalene-hopene cyclase responsável pela ciclização do epoxiesqualeno em seus respectivos produtos, o domínio tem 319 e 291 aminoácidos em cada região terminal da proteína. Além desse domínio, as OSCs vegetais são caracterizadas pela sequência conservada DCTAE e pela assinatura de aminoácidos típica de terpeno-sintases DGSWyGsWAVcFtYG, esses motivos estão presentes em diversas plantas como *Nicotiana tabacum* (tabaco), *Nicotiana benthamiana*, *Solanum lycopersicon* (tomate) (GAS-PASCUA et al., 2014), e agora também foram encontrados nos cinco genes de *E.guineensis*. A figura 9 ilustra os domínios discutidos nesse parágrafo.

Os resíduos Tyr410, His477 and Ile481 são importantes para atividade catalítica da enzima CAS1, a mutação da His477 para Asn e de Ile481 para Val altera a atividade da enzima, convertendo a cicloartenol sintase em uma eficiente lanosterol sintase (THIMMAPPA et al., 2014). Os genes EgCAS3, EgCAS2, EgCAS5 não apresentam alterações consideráveis nesses resíduos citados, o que corrobora a atividade de cicloartenol sintase das enzimas e também fornece indicativo de homologia funcional, embora as proteínas sejam bem maiores que a proteína codificada pelo gene AT2G07050 de *A.thaliana*.

Tyr 410 é conservada em quase todos os genes, menos no gene EgCAS4. Já His477 não está presente em AT3G45130 (LAS1), EgCAS1 e EgCAS4. Em LAS1 His477 é substituída por Asn e nos dois outros genes de dendê há uma substituição por uma Gln nessa posição, essa substituição His477Gln está presente na lanosterol sintase de protozoários como *Trypanosoma*, *Leishmania* e em leveduras como *Saccharomyces* (SAWAI et al., 2006).

Tabela 6: resumo das buscas por domínios utilizando PFAM.

Gene	Pfam	Posição	Familia	Gráfico
AtCAS1	PF13249.6 PF13243.6	97-403 412-751	SQHOP-CYC- N/C	
AtLAS1	PF13249.6 PF13243.6	97-403 412-751	SQHOP-CYC- N/C	
EgCAS1	PF13249.6 PF13243.6	149-455 464-803	SQHOP-CYC- N/C	
EgCAS2	PF13249.6 PF13243.6	123-429 511-744	SQHOP-CYC- N/C	

EgCAS3	PF13249.6 PF13243.6	97-403 412-751	SQHOP-CYC- N/C	
EgCAS4	PF13249.6 PF13243.6	168-473 482-821	SQHOP-CYC- N/C	
EgCAS5	PF13249.6 PF13243.6	97-403 412-751	SQHOP-CYC- N/C	

Nos genes EgCAS1 e EgCAS4 há substituição da Ile481 por Val, valina na posição 481 é altamente conservada nas lanosteróis sintase nos mais diversos organismos, como *Trypanosoma*, *Leishmania*, *Saccharomyces*, *Mus musculus* e *Homo sapiens* (SAWAI et al., 2006). Esses resultados indicam que as enzimas codificadas nos genes EgCAS4 e EgCAS1 podem ter atividade de lanosterol sintase e que *E.guineensis* pode produzir lanosterol, um intermediário do colesterol e do ergosterol.

As proteínas de *E.guineensis* são da mesma subfamília das proteínas CAS1 e LAS1 de *A.thaliana*. Em relação a LAS1 todas tem o percentual de identidade variando entre o mínimo de 55,9% no gene EgCAS4 e o máximo 65,7% no gene EGCA55.

Já em relação a CAS as enzimas variam entre o mínimo de 61,2% no gene EgCAS4 e o máximo 77,6% no gene EgCAS5. A matriz de identidade indica que as enzimas são mais semelhantes a CAS1 do que a LAS1, esse resultado é corroborado pelo agrupamento dos genes no Dendrograma (figura 37), no qual CAS1 e os genes de *E.guineensis* agrupam juntos, justificando sua nomenclatura como cicloartenol sintases.

Os genes EgCAS1, EgCAS2, EgCAS3 e EgCAS5 são muito semelhantes entre si (inclusive três deles tem o mesmo número de acesso ao NCBI), carregam mais de 90% de identidade, como mostra a figura 10, estão localizados na mesma região cromossômica, provavelmente são frutos de duplicação do genoma todo. O gene EgCAS4 tem percentual de identidade de 65% com os outros genes de *E.guineensis* e está localizado no cromossomo 10.

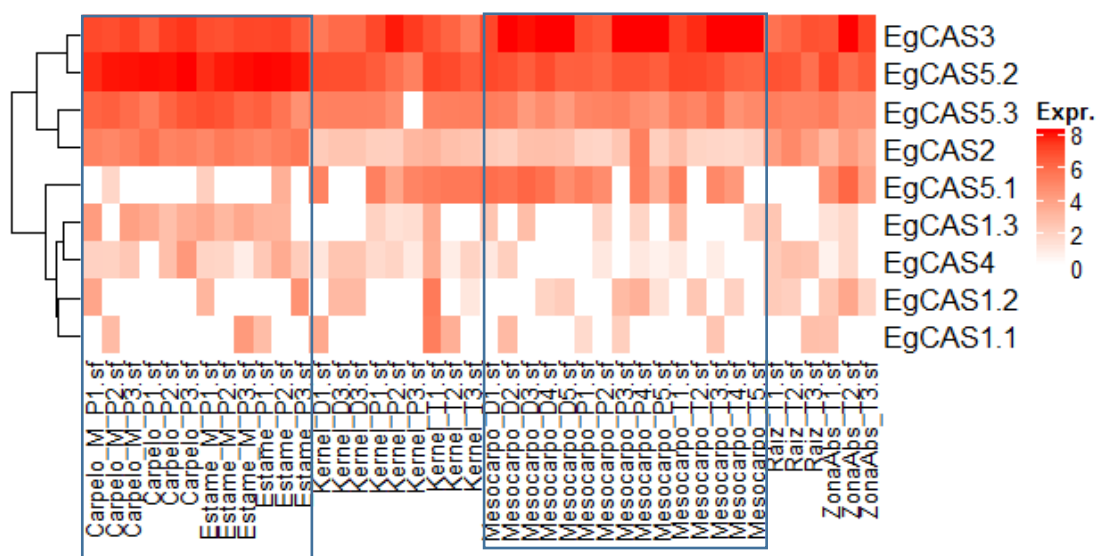
A análise dos domínios e semelhança na sequência de aminoácidos ilustrada pelo Dendrograma e matriz de identidade sugerem que os 5 genes do dendezeiro pertencem a família de OSC, no entanto, sua atividade funcional como lanosterol sintase ou clicoartenol sintase precisam ser confirmadas com experimentos mais específicos, principalmente quanto as enzimas codificadas pelos genes *EgCAS1* e *EgCAS4*.

EgCAS5	65.7	65.7	77.6	94.4	92.4	94.9	100
EgCAS3	64.3	65.09	77.4	96.4	93.9	100	94.9
EgCAS1	63.7	65.4	75.6	91.4	100	93.9	92.4
EgCAS2	64.8	65	77.6	100	91	96.4	94.4
AtCAS1	65.7	61.2	100	77.6	75.6	77.4	77
EgCAS4	55.9	100	61.2	65	65.4	65	65.7
AtLAS1	100	55.9	65.7	64.8	63.7	64.3	65.7
	AtLAS1	EgCAS4	AtCAS1	EgCAS2	EgCAS1	EgCAS3	EgCAS5

Figura 10: matriz de identidade das OSCs de *A.thaliana* e *E.guinensis*.

Quanto a expressão dos genes *EgCAS3* e as isoformas *EgCAS5.2* e *EgCAS5.3* apresentaram altos valores de expressão, principalmente na flor e no fruto, como mostra a figura 11, além disso, são transcritos em todos os tecidos avaliados com níveis relativamente estáveis de expressão.

Já os genes *EgCAS4* e *EgCAS1* e todas as suas 3 isoformas apresentam baixos níveis de expressão e não são transcritos em todos os tecidos, como dito anteriormente, esses dois genes podem apresentar atividade de lanosterol sintase, o que pode ser um indicativo ou uma explicação do porque sua expressão é reduzida em relação aos outros dois genes.



4.3 Esterol metil-transferases (SMT1/SMT2)

Tabela 7: resultado dos alinhamentos de SMTs usando o BLASTp e as informações genômicas deles.

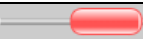
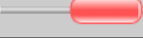
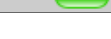
Tanto os genes de *A.thaliana* quanto de *E.guineensis* apresentam os domínios Sterol MTC terminal e Methyltransf 11 com assinatura de 65 e 94 aminoácidos respectivamente, com exceção do gene EgSMT1 1.

Esses domínios possuem quatro motivos conservados. Os motivos YE(WFY)GWGXSFHF e YS(AI)EATCHAP presentes tanto em SMT1 quanto em SMT2. Os outros domínios são **KPGQCFAAYEW** conservado apenas nas SMT1 ao passo que **KPG(AS)(LM)YVSYEW** é conservado apenas nas SMT2 (NEELEKADAN et al., 2009; BENVENISTE, 2004). Essas quatro regiões são responsáveis pelo sítio de ligação ao esterol substrato e são altamente conservadas em plantas, aparecem em *A.thaliana*, *Glycine max* (soja), *Nicotiana tabacum* e outros organismos como *Saccharomyces cerevisiae*, *Pneumocystis carinii* e *Trypanosoma brucei*, (NEELEKADAN et al., 2009; BENVENISTE, 2004), esses quatro motivos estão presentes nas SMTs de *E.guineensis*, com exceção de *EgSMT1_1*.

A maioria das metil-transferases alinhadas têm o sítio conservado LDVGCGXGGPXRXI rico em glicina, essa região é responsável pela ligação da S-Adenosilmetionina, a molécula que doa o agrupamento metil para reação de metilação. O alinhamento múltiplo e os motivos conservados nas esterol metil transferases podem ser visualizados na figura 12. A proteína do gene *EgSMT1_1* não tem nenhum dos motivos citados anteriormente, já na proteína do gene *EgSMT1_3* essas motifs apresentam várias mutações.

A matriz de identidade representada na figura 13 mostra os percentuais de identidade dos genes da família SMT e o Dendrograma na figura 37 mostra o agrupamento das enzimas baseado nesse percentual de identidade, há um nó divergindo entre SMT1 e SMT2.

Tabela 8: resumo da pesquisa de domínios utilizando PFAM das SMTs.

Gene	Pfam	Posição	Familia	Gráfico
AtSMT1	PF08241.12	99-197	Methyltransf_11	
	PF08498.10	271-336	Sterol_MT_C	
EgSMT1_3	PF08241.12	105-203	Methyltransf_11	
	PF08498.10	277-341	Sterol_MT_C	
EgSMT1_2	PF08241.12	105-203	Methyltransf_11	
	PF08498.10	277-341	Sterol_MT_C	
EgSMT1_1	PF08498.10	108-173	Sterol_MT_C	
AtSMT2	PF08241.12	128-226	Methyltransf_11	
	PF08498.10	290-355	Sterol_MT_C	
ATSMT2_1	PF08241.12	131-229	Methyltransf_11	
	PF08498.10	293-358	Sterol_MT_C	

ATSMT2_2	PF08241.12	131-229	Methyltransf_11	
	PF08498.10	293-358	Sterol_MT_C	

Entre SMT1, o grau de identidade dos genes de *E.guineensis* com o gene de *A.thaliana* está entre 50% e 80%, portanto, EgSMT1_3 e EgSMT1_2 pertencem a mesma subfamília que o gene AtSMT1 de *A.thaliana*. Esses resultados indicam que os genes EgSMT1_3 e EgSMT1_2 são parálogos entre si e ortólogos a AtSMT1. Já a proteína do gene EgSMT1_1 é muito reduzida em relação as outras, na predição de domínios do PFAM apresenta apenas o domínio *Sterol methyltransferase C-terminal* e não carrega motivos com traços funcionais conhecidos como as outras. Portanto, não é possível inferir homologia funcional desse gene em relação à família de genes SMTs, embora apresente mais de 57% de identidade com o gene de *Arabidopsis*.

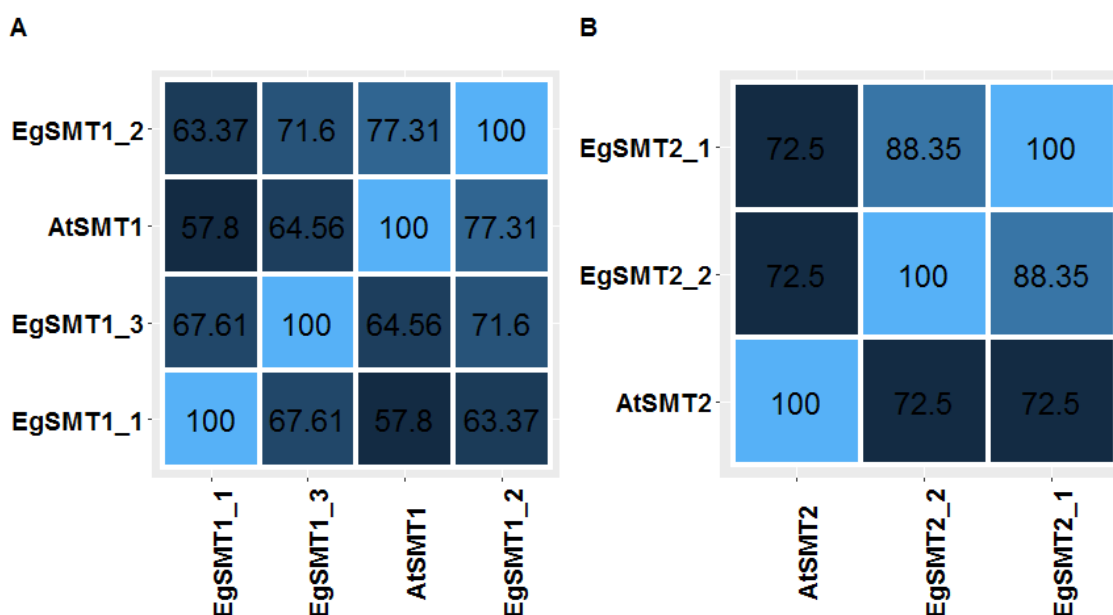


Figura 13: matriz de identidades dos dois da Família de esterol metil-transferases. Em (A) EgSMT1 e em (B) EgSMT2.

Quanto aos genes de SMT2 de *E.guineensis*, todos tem evidência de conservação funcional e carregam mais de 70% de identidade com o gene AtSMT2 de *A.thaliana* e menos de 90% de identidade entre si, indicando que pertencem a mesma subfamília. Apenas um gene tem localização cromossômica conhecida, o gene EgSMT2_1 está localizado no cromossomo 4. A conservação funcional, a proximidade filogenética e grau de identidade são indicativos que os dois genes são parálogos entre si, e ortólogos com AtSMT2.

Em relação a expressão dos genes, EgSMT1_2 e EgSMT2_1 são transcritos em todos os tecidos avaliados, SMT1 teve os mais altos níveis de transcrição no mesocarpo e nas flores, ao passo que SMT2 teve grande expressão no caroço (*kernel*) do fruto. O gene EgSMT1 e suas isoformas apresentam baixos níveis de expressão e não estão expressas em muitos tecidos, já o gene EgSMT2_2 não está expresso em nenhum tecido, por isso não aparece no heatmap.

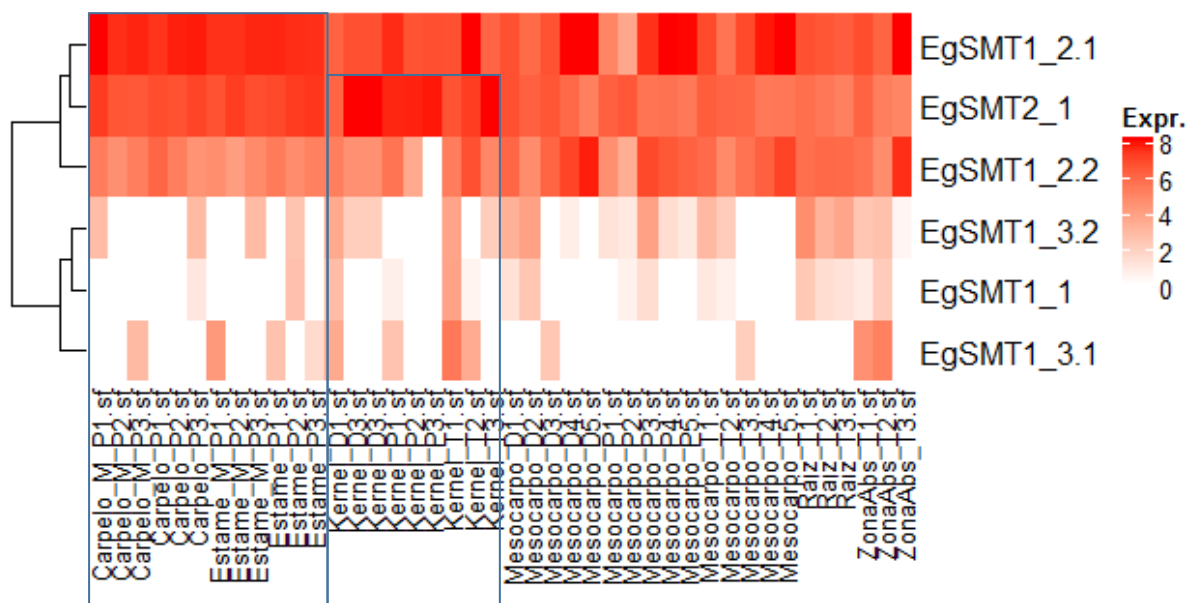


Figura 14: Heatmap que representa a expressão dos genes SMT em *Elaeis guineensis*, as células mais avermelhadas indicam maior nível de expressão.

4.4 Complexo de desmetilação: SMO1, SMO2, β HSD, SKR e EGR28.

Como citado anteriormente, o complexo enzimático formado por SMO1, SMO2, β HSD e SKR realiza as duas reações desmetilação do carbono 4 na síntese dos esteróis. Já a proteína EGR28 mantém o complexo enzimático unido e aderido à membrana do retículo endoplasmático.

As proteínas SMO1 e SMO2 pertencem à família de oxigenases de ferro não-hemicas, em *A.thaliana* os genes *AtSMO1* e *AtSMO2* dessa família carregam 40% de identidade entre si. *AtSMO1* alinhou para três genes e *AtSMO2* para quatro no genoma de *E.guineensis*.

O gene *At β HSD* de *A.thaliana* pertence à família de desidrogenase / redutase de cadeia curta (SDR), esse gene alinhou para dois em *E.guineensis*. A enzima esterol ceto-redutase (SKR) alinhou para 6 genes diferentes, enquanto

ERG28 alinhou para um apenas. A tabela 9 resume os resultados os alinhamentos realizados no BLASTp.

Tabela 9: resultado dos alinhamentos das enzimas que fazem a desmetilação do carbono quatro na via de biossíntese dos esteróis.

Gene em <i>A.thaliana</i>	Gene em <i>E.guineensis</i>	Número de acesso do gene	CDS (bp)	Peptídeo (aa)
AtSMO1	EgSMO1_1	NC_025996.1	1623	302
	EgSMO1_2	NW_011551042.1	1152	356
			1081	335
	EgSMO1_3	NW_011551727.1	1285	287
AtSMO2	EgSMO2_2	NC_026006.1	1415	261
	EgSMO2_3	NC_025997.1	1589	264
			1324	253
	EgSMO2_1	NC_025997.1	1239	267
	EgSMO2_4	NC_026002.1	1399	318
			1192	219
			1444	181
			1491	181
AtβHSD	EgβHSD1	NC_025994.1	2321	581
			2318	580
	EgBHSD2	NC_026005.1	2068	554
AtSKR	EgSKR1	NC_025995.1	1349	265
	EgSKR2	NW_011551471.1	1318	269
	EgSKR3	NC_026007.1	1111	264
	EgSKR4	NW_011551212.1	1033	257
	EgSKR5	NW_011551908.1	1083	260
	EgSKR6	NC_025993.1	1161	260
AtERG28	EgERG28	NC_025993.1	130	14.91

Apenas três genes de *E.guineensis* tem isoformas produtos de recomposição alternativa do pre-mRNA: o gene EgSMO1_2 tem duas, o gene EgMO2_4 tem quatro e EgβHSDs1 tem duas isoformas.

As SMO's de *A.thaliana* e *E.guineensis* apresentam o domínio Ácido graxo hydroxilase, em SMO's esse domínio possui três motivos funcionais ricos

em histidina HXXXH, HXXHH e HDYHH, constitutivos da pequena família de oxigenases de ferro não-hemicas que ficam ancoradas à membranas, estão geralmente envolvidas na oxidação de ácidos graxos e esteróis (DARNET et al., 2001). A figura 15 ilustra os alinhamentos e os domínios presentes nas SMO's.

Quanto as β HSDs, para predição no PFAM todas são caracterizadas pelo domínio 3Beta_HSD com assinatura 280 aminoácidos. As β HSDs de *A.thaliana* tem o motivo de ligação ao NADH rico em glicina na região N-terminal: TGGRGFAA, os resíduos GD e DG também aparecem ao longo da proteína para interagir com NADH. Além dessas regiões, os resíduos Asp-39, Asp-70, Thr-129, e Lys-163 são responsáveis pela interação com o agente redutor (RAHIER et al., 2009) e o motivo 159YXXXK163 tem importância para o sítio catalítico interagindo com o esterol substrato. Esses dois motivos não estão bem conservados em *E.guineensis*, poucas regiões de homologia entre as duas enzimas e várias mutações nos domínios ficam claras observando o alinhamento na figura 16. (RAHIER et al., 2009).

Para predição no PFAM as enzimas SKR apresentaram o domínio adh_short_C2 com assinatura de 234 aminoácidos, esse domínio é característico da família de enzimas Enoyl reductases que realiza a última etapa da via de biossíntese de ácidos graxos, como são NADH dependentes apresentam um sítio típico de ligação ao NADH de SDRs TGXXXGhG. Além desse domínio, apresentam as sequências conservadas YX(AS)(ST)K e GxhxhSSh. Por fim, apresentam o domínio h(KR)h(NS)xhxPGxxxT que carrega função estrutural e direciona a reação enzimática. Todas essas sequências de aminoácidos estão presentes nas enzimas das duas espécies como mostra a figura 17.

As duas enzimas ERG28 apresentam o domínio Erg28 do PFAM com assinatura 110 aminoácidos, não há motivos funcionais descritos na literatura para essa família, no entanto, as regiões RTFG, WTLL e HFL são altamente conservadas entre seres humanos, leveduras e vegetais, portanto podem estar envolvidos com a função da enzima (GACHOTTE et al., 2001). A figura 18 ilustra o alinhamento dos dois genes.

As SMO's de *A.thaliana* e *E.guineensis* são agrupadas em duas sub-famílias com base no percentual de identidade das matrizes e do agrupamento das enzimas no Dendrograma. Os percentuais de identidade entre as sequências de aminoácidos variam de 78% a 80% em SMO2, em SMO1 variam entre 64% e 85%, indicando que os respectivos genes estão relacionados às duas respectivas sub-famílias, a figura 19 resume os percentuais de identidade das SMO's e o Dendrograma corrobora esse resultado com 100 de suporte no bootstrap.

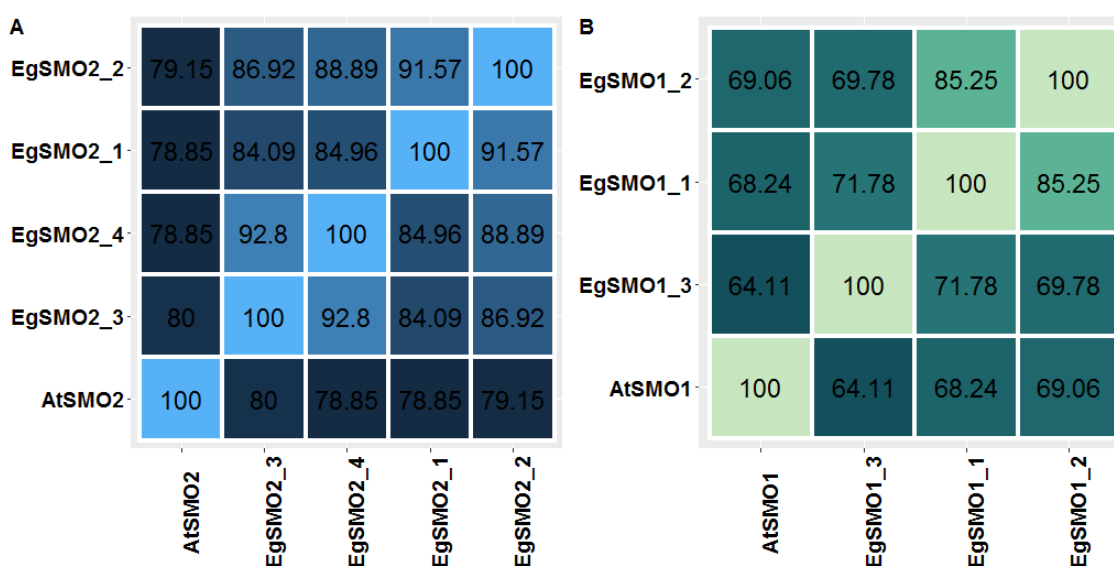


Figura 19: Matrizes de identidade das enzimas da família de esterol metil oxidases. Em (A) os valores de identidade mais altos entre SMO2 são indicados pelo azul mais claro, em (B) os valores de identidade mais altos entre SMO1 são indicados com o verde mais claro

Já está relatado na literatura que os genes da família SDR estão divididos em dois grupos, o grupo clássico e o grupo estendido. As β HSDs são normalmente incluídas no grupamento estendido, pela baixa similaridade com as outras SDR (RAHIER et al., 2009). Os genes *Eg β HSD1* e *Eg β HSD1* de *E.guineensis* apresentam 40,7% e 65,9% de identidade com a β HSDs de *A.thaliana*, respectivamente, como mostram a matriz de identidade na figura 20 e o Dendrograma na figura 37.

A similaridade das enzimas SKR de *E.guineensis* com *A.thaliana* está entre 52% na proteína EgSKR1 e 62% na proteína EgSKR6. A similaridade dos genes entre si é maior que 50% o que indica que todos são da mesma subfamília, como ilustra a figura 20. As enzimas também agrupam juntas no dendrograma,

com 100 suporte de bootstrap, o que indica a homologia entre as enzimas. A figura 37 ilustra o agrupamento dos genes SKR no Dendrograma.

ERG28 é bastante conservada entre as duas espécies, o único gene encontrado em *Eleais guineensis* tem 79% de similaridade com *Arabidopsis*.



Figura 20: Matrizes de identidade dos genes que realizam a desmetilação do carbono 4. Em (A) os genes para BHSD, em (B) os genes para SKR, em (C) os genes para ERG28. Maiores percentuais de identidade são indicados pelas células amarelas.

Esses resultados apontam que a etapa de desmetilação do carbono é conservada no dendezeiro, as enzimas que realizam esse ponto da via têm alta semelhança entre as duas espécies, portanto os genes do dendezeiro são ortólogos a via de *Arabidopsis*.

Quanto a expressão dos genes a maioria dos genes desse complexo enzimático está expressa nos tecidos estudados. O gene EgSMO2_4 e suas isoformas não estão expressos em diversos tecidos estudados, nos tecidos em que há expressão, os níveis são baixos, indicando que sua atividade não é

Heatmap showing gene expression data (Expr.) across various samples. The color scale ranges from 0 (white) to 8 (red). The samples are clustered hierarchically on the left. The genes are listed on the right, and the samples are listed at the bottom.

Genes (Y-axis):

- EgSMO1_1
- EgBHSD2
- EgSKR6
- EgSMO2_2
- EgSMO1_2.2
- EgSMO2_1
- EgSMO1_3
- EgSKR4
- EgSKR5
- EgBHSD1.2
- EgBHSD1.1
- EgSMO2_3.1
- EgERG28
- EgSKR1
- EgSKR3
- EgSMO2_3.2
- EgSKR2
- EgSMO2_4.2
- EgSMO1_2.1
- EgSMO2_4.1
- EgSMO2_4.4
- EgSMO2_4.3

Samples (X-axis):

- Carpe1
- Carpe2
- Carpe3
- Carpe4
- Carpe5
- Carpe6
- Carpe7
- Carpe8
- Carpe9
- Carpe10
- Carpe11
- Carpe12
- Carpe13
- Carpe14
- Carpe15
- Carpe16
- Carpe17
- Carpe18
- Carpe19
- Carpe20
- Carpe21
- Carpe22
- Carpe23
- Carpe24
- Carpe25
- Carpe26
- Carpe27
- Carpe28
- Carpe29
- Carpe30
- Carpe31
- Carpe32
- Carpe33
- Carpe34
- Carpe35
- Carpe36
- Carpe37
- Carpe38
- Carpe39
- Carpe40
- Carpe41
- Carpe42
- Carpe43
- Carpe44
- Carpe45
- Carpe46
- Carpe47
- Carpe48
- Carpe49
- Carpe50
- Carpe51
- Carpe52
- Carpe53
- Carpe54
- Carpe55
- Carpe56
- Carpe57
- Carpe58
- Carpe59
- Carpe60
- Carpe61
- Carpe62
- Carpe63
- Carpe64
- Carpe65
- Carpe66
- Carpe67
- Carpe68
- Carpe69
- Carpe70
- Carpe71
- Carpe72
- Carpe73
- Carpe74
- Carpe75
- Carpe76
- Carpe77
- Carpe78
- Carpe79
- Carpe80
- Carpe81
- Carpe82
- Carpe83
- Carpe84
- Carpe85
- Carpe86
- Carpe87
- Carpe88
- Carpe89
- Carpe90
- Carpe91
- Carpe92
- Carpe93
- Carpe94
- Carpe95
- Carpe96
- Carpe97
- Carpe98
- Carpe99
- Carpe100
- Carpe101
- Carpe102
- Carpe103
- Carpe104
- Carpe105
- Carpe106
- Carpe107
- Carpe108
- Carpe109
- Carpe110
- Carpe111
- Carpe112
- Carpe113
- Carpe114
- Carpe115
- Carpe116
- Carpe117
- Carpe118
- Carpe119
- Carpe120
- Carpe121
- Carpe122
- Carpe123
- Carpe124
- Carpe125
- Carpe126
- Carpe127
- Carpe128
- Carpe129
- Carpe130
- Carpe131
- Carpe132
- Carpe133
- Carpe134
- Carpe135
- Carpe136
- Carpe137
- Carpe138
- Carpe139
- Carpe140
- Carpe141
- Carpe142
- Carpe143
- Carpe144
- Carpe145
- Carpe146
- Carpe147
- Carpe148
- Carpe149
- Carpe150
- Carpe151
- Carpe152
- Carpe153
- Carpe154
- Carpe155
- Carpe156
- Carpe157
- Carpe158
- Carpe159
- Carpe160
- Carpe161
- Carpe162
- Carpe163
- Carpe164
- Carpe165
- Carpe166
- Carpe167
- Carpe168
- Carpe169
- Carpe170
- Carpe171
- Carpe172
- Carpe173
- Carpe174
- Carpe175
- Carpe176
- Carpe177
- Carpe178
- Carpe179
- Carpe180
- Carpe181
- Carpe182
- Carpe183
- Carpe184
- Carpe185
- Carpe186
- Carpe187
- Carpe188
- Carpe189
- Carpe190
- Carpe191
- Carpe192
- Carpe193
- Carpe194
- Carpe195
- Carpe196
- Carpe197
- Carpe198
- Carpe199
- Carpe200
- Carpe201
- Carpe202
- Carpe203
- Carpe204
- Carpe205
- Carpe206
- Carpe207
- Carpe208
- Carpe209
- Carpe210
- Carpe211
- Carpe212
- Carpe213
- Carpe214
- Carpe215
- Carpe216
- Carpe217
- Carpe218
- Carpe219
- Carpe220
- Carpe221
- Carpe222
- Carpe223
- Carpe224
- Carpe225
- Carpe226
- Carpe227
- Carpe228
- Carpe229
- Carpe230
- Carpe231
- Carpe232
- Carpe233
- Carpe234
- Carpe235
- Carpe236
- Carpe237
- Carpe238
- Carpe239
- Carpe240
- Carpe241
- Carpe242
- Carpe243
- Carpe244
- Carpe245
- Carpe246
- Carpe247
- Carpe248
- Carpe249
- Carpe250
- Carpe251
- Carpe252
- Carpe253
- Carpe254
- Carpe255
- Carpe256
- Carpe257
- Carpe258
- Carpe259
- Carpe260
- Carpe261
- Carpe262
- Carpe263
- Carpe264
- Carpe265
- Carpe266
- Carpe267
- Carpe268
- Carpe269
- Carpe270
- Carpe271
- Carpe272
- Carpe273
- Carpe274
- Carpe275
- Carpe276
- Carpe277
- Carpe278
- Carpe279
- Carpe280
- Carpe281
- Carpe282
- Carpe283
- Carpe284
- Carpe285
- Carpe286
- Carpe287
- Carpe288
- Carpe289
- Carpe290
- Carpe291
- Carpe292
- Carpe293
- Carpe294
- Carpe295
- Carpe296
- Carpe297
- Carpe298
- Carpe299
- Carpe300
- Carpe301
- Carpe302
- Carpe303
- Carpe304
- Carpe305
- Carpe306
- Carpe307
- Carpe308
- Carpe309
- Carpe310
- Carpe311
- Carpe312
- Carpe313
- Carpe314
- Carpe315
- Carpe316
- Carpe317
- Carpe318
- Carpe319
- Carpe320
- Carpe321
- Carpe322
- Carpe323
- Carpe324
- Carpe325
- Carpe326
- Carpe327
- Carpe328
- Carpe329
- Carpe330
- Carpe331
- Carpe332
- Carpe333
- Carpe334
- Carpe335
- Carpe336
- Carpe337
- Carpe338
- Carpe339
- Carpe340
- Carpe341
- Carpe342

4.5 Ciclopropil-isomerase (CPI1) e Citocromo P51 (CYP51)

Tabela 10: resultados do BLASTp para os genes CPI1 e CYP51

Gene em <i>A.thaliana</i>	Gene em <i>E.guineensis</i>	Número de acesso do gene	CDS (bp)	Peptídeo (aa)
CPI1	EgCPI3	NC_025996.1	1035	344
	EgCPI1	NC_026001.1	1390	351
	EgCPI2	NC_025994.1	1390	351
CYP51	EgCYP51G1	NW_011550908.1	2032	493
	EgCYP51G2	NC_026001.1	2236	490

42

domínios da conhecida superfamília de citocromos P450 com assinatura de 350 aminoácidos, como mostra a tabela 11.

Tabela 11: resultados da busca por domínios utilizando PFAM. CPI1 não apresentou nenhum domínio conhecido.

GENE NAME	PFAM	POSIÇÃO	FAMILIA	GRÁFICO
AtCYP51	PF00067.22	40-480	p450	
EGCYP51G1	PF00067.22	39-477	p450	
EGCYP51G2	PF00067.22	40-479	p450	

Os resíduos Gly28, Glu29 e Asp260 são importantes para atividade enzimática das enzimas CPI1 em *Arabidopsis*. O motivo G¹⁰⁸NYFWTHYFF¹¹⁷ constitui o bolso de anelamento do esterol, sendo que Trp112 e Thr113 estabilizam a ligação da enzima com esterol substrato, tanto os resíduos citados anteriormente quanto o motivo estão conservados em *Arabidopsis*, *Oryza sativa* (Arroz), *Solanum lycopersicum* (tomate) (RAHIER; KARST 2014) também em *E.guineensis*, como mostra a figura 21.

Embora nenhum domínio funcional tenha sido encontrado com o PFAM a presença desses motivos funcionais nos genes de *E.guineensis* é evidencia de conservação funcional entre as enzimas das duas espécies.

As CYP51G têm quatro regiões conservadas que servem de sítio de reconhecimento do substrato. As duas primeiras são duas alfa-helices com as seguintes assinaturas de aminoácidos 110YQFNVPTFGPGVVFDVD126 e 287FAGQHTSSIT296, as duas últimas são duas folhas-beta com assinaturas 352PPLIMLMRAS361 e 468NAMVVG VKG476. Os resíduos Val108, Tyr110, Asn113, Leu354 e Met470 estão localizados no bolso de anelamento dos substratos (QI et al., 2006). A tabela 11 representa os domínios funcionais conhecidos em CYP51G. Outros resíduos como Pro115, Phe117, Gly118, Gly189, His191, Ser193, Trp299, Arg359 são conservados nas CYP51 de mais de 30 organismos, entre fungos, plantas e animais, portanto podem estar relacionados funcionalidade da família de enzimas (QI et al., 2006). Essas regiões podem ser visualizadas no alinhamento múltiplo na figura 22.

As CPI agrupam em um mesmo ramo do Dendrograma com 100 de suporte no bootstrap, a análise da matriz de identidade mostra que os genes CPI

de *E.guineensis* tem percentual de identidade maior que 59% com o gene de *A.thaliana*. Esses resultados corroboram que os genes pertencem a mesma família, portanto são ortólogos entre as duas espécies.

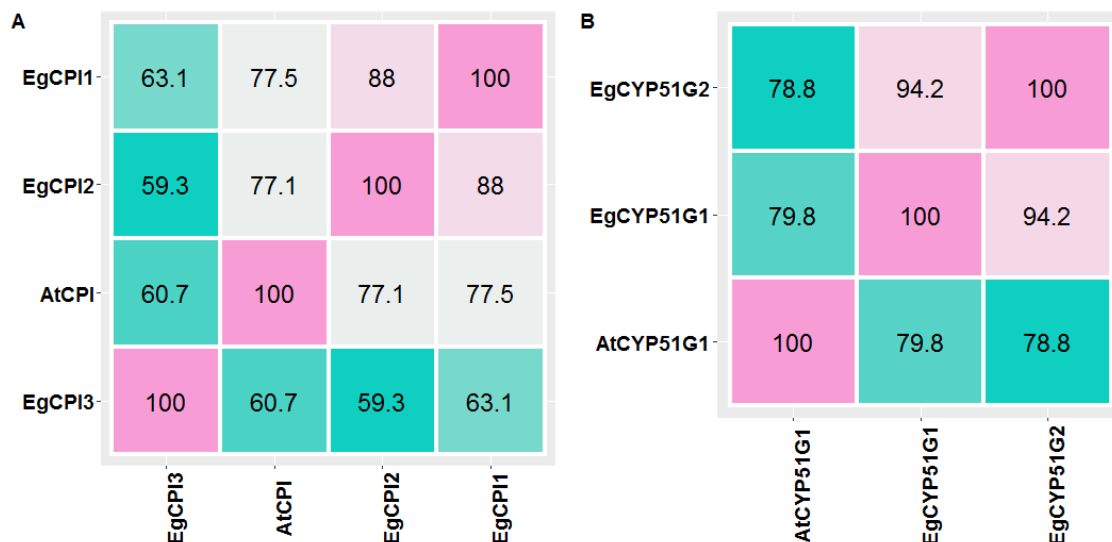


Figura 23: Percentual de identidade das enzimas CPI e CYP51G, células com valores mais arrosados apresentam maior percentual de identidade.

Como já dito, as CYP51G1 são enzimas da super-família dos citocromos P450, elas carregam similaridade com outras enzimas da via que também são citocromos, como CYP710A. No Dendrograma CYP51G1 e CYP710A agrupam-se em um único ramo com 100 de suporte do bootstrap, como mostra a figura 38, esse ramo se divide exatamente entre as duas famílias de enzimas, individualizando os ramos de cada uma das famílias.

Esses resultados são corroborados com os percentuais de identidade da matriz, os genes de *E.guineensis* apresentam quase 80% de identidade com os genes de *Arabidopsis*, indicando homologia entre as enzimas. Os genes do dendezeiro apresentam quase 95% de identidade entre si, possivelmente são frutos de duplicação, mas sem qualquer divergência funcional aparente.

Esses resultados indicam que o gene CPI1 de *Arabidopsis* tem três ortólogos com o dendezeiro e esses três genes são parálogos entre si. Já CYP51G1 tem dois ortólogos e dois parálogos entre si. Nenhum dos genes apresenta evolução ou inovação funcional de acordo com a análise dos domínios.

Tabela 12: resultados do BLASTp para os genes HYD1, FK, STE1, DWF5, DWF1.

Gene de <i>A.thaliana</i>	Gene em <i>E.guineensis</i>	Número de acesso do gene	CDS (bp)	Peptídeo (aa)
AtHYD1	EgHYD1	NW_011567723.1	1121	219
	EgHYD2	NC_026001.1	1131	219
AtFK	EGFK	NW_011552837.1	1429	369
			1261	313
			1386	308
			1213	297
AtSTE1	EgSTE1	NC_025995.1	1197	273
	EgSTE2	NC_025999.1	1182	277
AtDWF5	EgDWF5_1	NC_025999.1	1710	374
	EgDWF5_2	NC_026000.1	1621	434
AT3G19820 DWF1	EgDWF1_1	NC_026008.1	2272	560
	EgDWF1_2.	NC_026004.1	2109	560

Na predição realizada no PFAM, HYD1 é caracterizado pelo domínio EBP (proteína de ligação ao Emopamil), bem descrito em esterol isomerases de animais e fungos, também presente em esterol isomerases de plantas, com assinatura de 178 de aminoácidos. Nesse domínio, os aminoácidos Trp61, His69, Glu73, Asp102, Glu115, Asp114, Trp188 são altamente conservados entre eucariotos como *Zea mays* (milho), *Mus musculus*, *H.sapiens*, *S.cerevisiae*, *A.thaliana* (RAHIER et al., 2008). Além desses resíduos, Thr118 é responsável por estabilizar a ligação entre o esterol substrato e o sítio catalítico da enzima. (RAHIER et al., 2008). A figura 25 mostra que esses resíduos são conservados em regiões semelhantes na proteína do dendezeiro.

As FK's e DWF5 tem o domínio ERG4_ERG24 que é característico da família de proteínas esterol redutases encontradas principalmente em fungos e vertebrados, esse domínio tem assinatura de 432 aminoácidos. O motivo LLXSGWWGXXRH na região N-terminal da enzima caracteriza essa família de genes, embora não tenha função elucidada ainda, modificações na serina presente nesse motivo alteram a atividade da enzima, mostrando que essa região conservada é importante para o sítio catalítico das esterol redutases (QIAN et al., 2013), (SHIRICK et al 2000), as figuras 26 e 27 mostram a organização dos domínios funcionais no alinhamento das enzimas de *E.guineensis* e *A.thaliana*.

As enzimas STEs apresentam o domínio FA_hydroxylase (Fatty acid hidroxilase) assim como as SMO's, portanto apresentam os três domínios ricos em histidina HXXXH, HXXHH e HXXHH responsáveis pela ligação aos átomos de ferro que são co-fatores da reação, além desses domínios, apresentam uma região conservada de 20 aminoácidos PFA(S/G)(H/L)(A/S)FHP(V/I)DG típica Esterol C5-desasturases. A Treonina na posição 114 é responsável pela ligação ao esterol substrato, sua substituição acarreta perda da estabilidade da enzima com substrato (TATON et al., 2000; HUSSELSTEIN et al 1999). Essas regiões são conservadas em *Nicotiana Tabacum*, *A.thaliana* e também *E.guineensis* (HUSSELSTEIN et al 1999). A figura 28 ilustra a organização dos motivos funcionais das enzimas STE1.

DWF1 apresentam o domínio FAD_binding_4 com assinatura de 139 aminoácidos, esse domínio é caracterizado por dois motivos funcionais, um rico em glicina e outro rico em aminoácidos hidrofóbicos, as funções desses motivos ainda não são amplamente conhecidas. A figura 29 mostra os domínios presentes nas sequências de DWF1 no alinhamento múltiplo entre *A.thaliana* e *E.guineensis*.

Os domínios estão localizados em regiões semelhantes entre as proteínas de *E.guineensis* e *A.thaliana*, além disso. A tabela 13 resume as informações da busca de domínios utilizando PFAM.

Tabela 13: resultados da busca por domínios utilizando PFAM.

GENE NAME	PFAM	POSIÇÃO	FAMILIA	GRÁFICO
ATFK	PF01222.17	3-369	ERG4 ERG24	
EGFK	PF01222.17	1-308	ERG4 ERG24	
ATHYD1	PF05241.12	34-206	EBP	
EGHYD1	PF05241.12	35-209	EBP	
EGHYD2	PF05241.12	34-209	EBP	
ATSTE1	PF04116.13	133-262	FA_hydroxylase	
EGSTE1	PF04116.13	132-261	FA_hydroxylase	
EGSTE2	PF04116.13	136-265	FA_hydroxylase	
ATDWF1	PF01565.23	110-199	FAD_binding_4	
EGDWF1_2.	PF01565.23	109-198	FAD_binding_4	
EGDWF1_1	PF01565.23	108-198	FAD_binding_4	
ATDWF5	PF01222.17	62-432	ERG4 ERG24	
EGDWF5_1	PF01222.17	5-374	ERG4 ERG24	
EGDWF5_2	PF01222.17	49-434	ERG4 ERG24	

Como já dito anteriormente, FK e DWF5 pertencem à mesma família de esterol redutases, os dois grupos de enzimas tem 31% de identidade entre si e agrupam juntos no Dendrograma com 100 de suporte no bootstrap, a figura 30 mostra o percentual de identidade entre as enzimas, a imagem 38 ilustra o agrupamento entre os genes de esterol redutases. Os genes de *E.guineensis* são ate 85% similares ao gene AtDWF5 de *A.thaliana*. Já o único gene de FK em *E.guineensis* apresenta 72% de identidade com o gene de Arabidopsis.

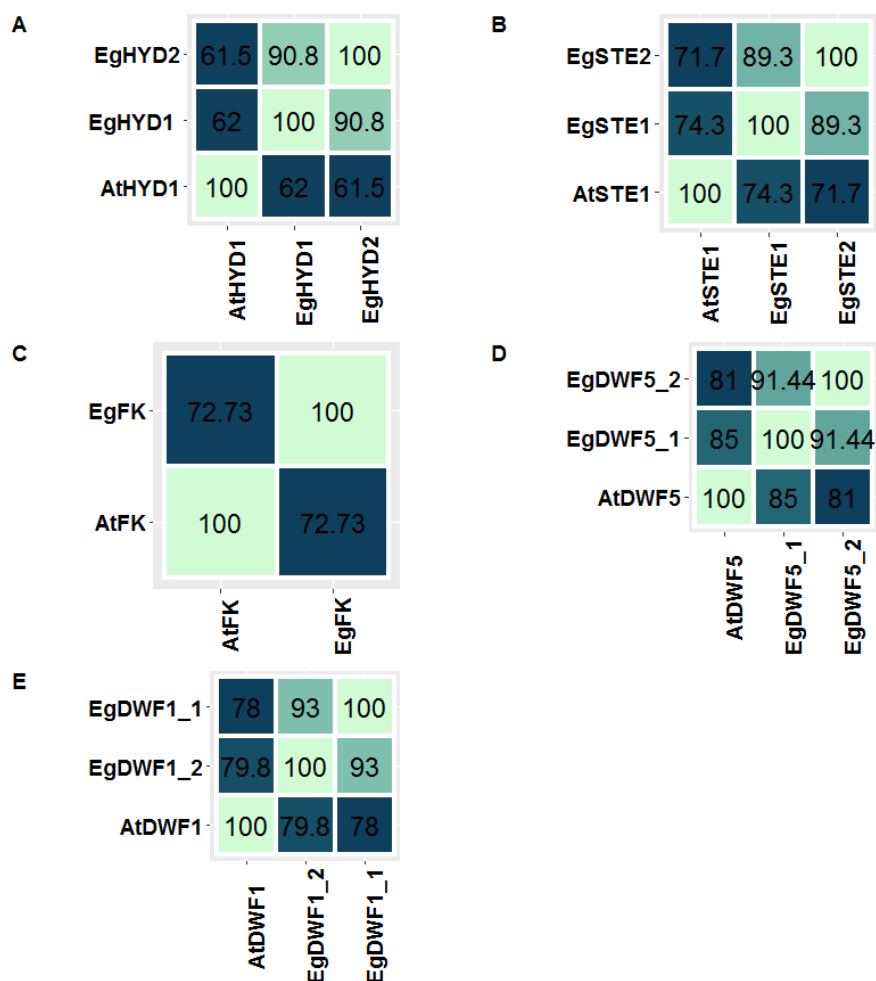


Figura 30: Matriz de identidade dos genes que realizam as reações finais da via de biossíntese dos esteróis, células em verde claro indicam alto percentual de identidade entre os genes. Em (A) HYD, em (B) STE, em (C) FK, em (D) em DWF5, em (E) DWF1

Esses resultados são evidência de que os genes de esterol redutases são ortólogos entre as duas espécies. Como no dendezeiro há dois genes para DFW5 e eles têm 91% de similaridade, é possível que esse gene tenha sido duplicado, portanto, os dois são parálogos entre si.

4.7 Enzimas que fazem a conversão dos fito-esteróis às formas derivadas e ao campesterol.

As CYP710 fazem a conversão de b-sitoesterol a estigmasterol. UDP2, SAT1, PSAT1 convertem os fitoesteróis em esteróis esterificados e glicosilados. Os genes *CYP710*, *UDP2* e *SAT1* de *Arabidopsis* alinham para dois genes em *E.guineensis* cada. *PSAT1* alinhou para apenas um gene. Os resultados dos alinhamentos utilizando blastp estão resumidos na tabela 1.

A família de proteínas CYP710A pertencem a super família de citocromos P450, esse domínio tem assinatura de 463 aminoácidos e ocupa boa parte da enzima. O motivo funcional característico de várias CYP710 carrega a sequência conservada FLFA(A/S)QDAS(T/S)S, duas alaninas conservadas nas posições 295-299 e uma cisteína conservada na região n-termianal da proteína, parecem estar ligados ao sítio ativo da proteína (ARNQVIST et al., 2008), (MORIKAWA et al., 2006). A figura 32 apresenta o alinhamento múltiplo e as sequências conservadas entre as enzimas CYP710.

As enzimas que fazem a glicosilação dos esteróis são as esterol glicosil transferases (UGTs). No PFAM os genes de UGT's de *Arabidopsis* e do dendezeiro apresentaram dois domínios funcionais da super-família de glicosil transferases, Glyco_transf_28 e UDPGT, com 139 e 499 aminoácidos como assinatura funcional, respectivamente. Quatro motivos funcionais conservados caracterizam as UGTs vegetais: a região AIIANPPAYGH é o sítio putativo de ligação do esterol as UGTs (WARNECKE et al., 1999).

Tabela 14 resultado dos alinhamentos feitos com o BLASTp.

Gene de A.thaliana	Gene em E.guineensis	Número de acesso do gene	CDS (bp)	Peptídeo (aa)
AT2G28860 CYP710A4	EgCYP710A1	NC_025999.1	3093	507
	EgCYP710A2	NC_025995.1	2796	505
AT1G43620 UDP2	EgUDP1.1	NC_026003.1	2058	613
			1860	553
			2018	580
	EgUDP2	NC_025994.1	2562	449
AT3G51970 SAT1	EgSAT1	NW_011551022.1	1278	345
	EgSAT2	NW_011551022.1	1159	339
AT1G04010	EgPSAT1	NC_026008.1	2685	648
			2662	640

PSAT1	2613	624
	2098	590

As regiões HIFFTMPWTPT e YIGFGS são conservadas entre as UGTs de diferentes espécies e também estão conservadas em *Arabidopsis* e *E.guineensis*. Por fim, a sequência HGGXGTTXAGLKAGCPTXXXPFPGDQ é assinatura da super família UDP-glycosiltransferase também está presente nas duas espécies em questão (WARNECKE et al., 1999; FERRER et al., 2017). A figura 33 mostra os domínios funcionais citados nesse e nos parágrafos anteriores nas UGTs

As PSAT1 têm o domínio LCAT (Lecithin:cholesterol acyltransferase), domínio pertencente e caracterizado em acyl transferases de mamíferos, a assinatura do domínio no PFAM tem 392 aminoácidos. Seis regiões conservadas estão presentes em esterol aciltransferases de diversos organismos, como *Mus musculus*, *Nicotiana tabacum*, *Homo sapiens* e *Arabidopsis thaliana* (BANÁS et al., 2005), além de estarem conservados em *E.guineensis*, a figura 34 mostra essas regiões no alinhamento múltiplo. Os domínios AVPYDYRLSP e GLPVS são responsáveis pela atividade da enzima, o primeiro se liga ao ácido graxo substrato e o segundo ao esterol substrato, respectivamente. Além desses, os outros quatro domínios não tem função definida, no entanto, mutações nos aminoácidos Ser195, Asp461 e His546 presentes nesses domínios levam a perda da função da enzima, o que mostra que eles são importantes para sítio ativo ou estrutura secundária da mesma (BANÁS et al., 2005).

As SAT1 são proteínas pertencentes a família de acil-transferases O ligadas a membrana (MBOAT), essa família é bem caracterizada em fungos e mamíferos, mas também está presente em vegetais, por isso no PFAM as enzimas apresentam o domínio MBOAT_2 com 83 aminoácidos de assinatura. Quatro motivos funcionais caracterizam a família MBOAT de fungos e mamíferos, no entanto, apenas dois estão conservados em *Arabidopsis* e no dendezeiro, os motivos WNLMV e MHEV. A figura 35 ilustra os domínios encontrados nas enzimas SAT1 (CHEN et al., 2007). A tabulação dos domínios funcionais dos genes citados nos últimos parágrafos está apresentada na tabela.

Tabela 19: resultado da busca por domínios no PFAM.

GENE NAME	PFAM	POSIÇÃO	FAMILIA	GRÁFICO
AtCYP710	PF00067.22	38-401	p450	
EGCYP710A1	PF00067.22	89-475	p450	
EGCYP710A2	PF00067.22	89-475	p450	
AtUDP2	PF03033.20	158-308	Glyco_transf_28	
	PF00201.18	411-556	UDPGT	
EGUDP1.1	PF03033.20	110-253	Glyco_transf_28	
	PF00201.18	354-498	UDPGT	
EGUDP2	PF03033.20	1-143	Glyco_transf_28	
	PF00201.18	240-385	UDPGT	
AtPSAT	PF02450.15	73-503	LCAT	
EGPSAT1	PF02450.15	82-502	LCAT	
AtSAT	PF13813.6	181-275	MBOAT_2	
EGSAT1	PF13813.6	181-265	MBOAT_2	
EGSAT2	PF13813.6	181-260	MBOAT_2	

Todas as sequências de aminoácidos com exceção das alinhadas a SAT1 tem mais de 50% de identidade entre *Arabidopsis* e *E.guineensis*. Os genes de CYP710 agrupam juntamente com CYP51 no dedrograma, ambos pertencem a super família de citocromos P450, o suporte para esse agrupamento da árvore é de 100% das repetições. Além disso, o grau de similaridade de AtCYP710 com as duas enzimas de *E. guineensis* é de 58%. Esses resultados são evidenciam que os dois genes de *E.guineensis* são ortólogos a AtCYP710. Os dois genes de CYP710 de *E.guineensis* tem 100% de similaridade entre si, o que indica que são alelos ou isoformas.

SAT1, UGT1 e PSAT1 também formam ramos únicos no Dendrograma, no entanto AtSAT1 tem baixa similaridade com os genes de *E. guineensis*, cerca de 39% de identidade, indicando que não pertencem à família MBOAT. AtUGT1 tem 54% e 76% de identidade com os genes de *E.guineensis*, portanto, os dois são da família de glycosil transferases. Por fim, os genes para PSAT tem 75% de similaridade e ambos agrupam juntos no dendrograma. A matriz de similaridade e o Dendrograma na figura 38 resumem o padrão de agrupamento e identidade dos genes CYP710, UGT1, SAT e PSAT1.

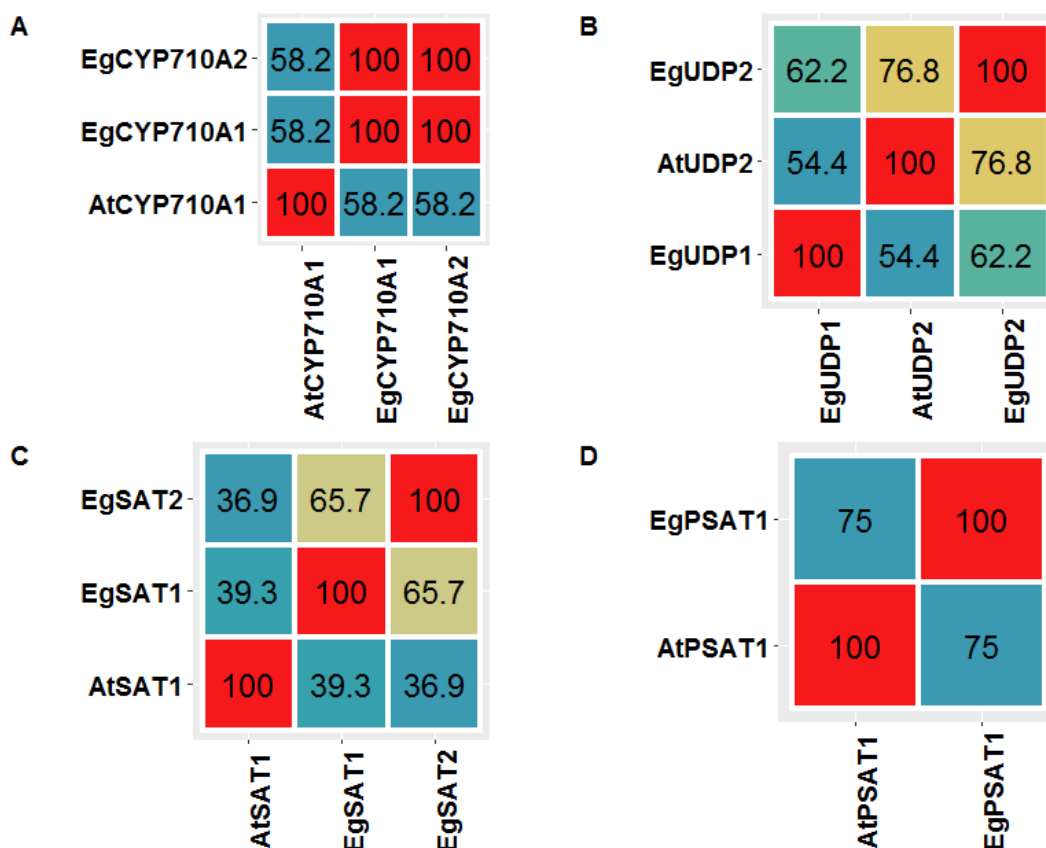


Figura 36: matriz de identidade entre as enzimas de *Eleais guineensis* e *Arabidopsis thaliana*, em (A) as proteínas da família CYP710A, em (B) proteínas da família UDP, em (C) as proteínas da família MBOAP, em (D) PSAT1.

Os níveis de expressão das proteínas que fazem a conversão dos esteróis as formas derivadas são medianos, apenas UDP1 apresenta os níveis máximos de expressão no caroço e apenas UDP2 está expresso em todos os tecidos. As EgCYP710A são expressas em todos os tecidos, indicando que há conversão de sitoesterol a estigmasterol nesses tecidos. A PSAT1 apresenta quatro isoformas, mas apenas uma está expressa em todos os tecidos, já entre os dois genes de SAT1 apenas um está expresso em todos os tecidos. Não é um gene fortemente expresso em nenhum tecido em específico, como nas outras enzimas.

4.8 Caracterização físico-química e localização das proteínas

As proteínas foram caracterizadas quanto ao seu Ponto isoelétrico, índice alifático, índice de estabilidade, composição de aminoácidos, índice de hidropaticidade e peso molecular hipotético, além de sua localização sub-celular. As propriedades físico-químicas das proteínas são determinadas pela sequência de aminoácidos, por isso tem efeito direto na função e na atividade das proteínas (ALBERTS, 2017). Portanto, é importante determinar se as proteínas de *Arabidopsis* e *E. guineensis* tem propriedades físico químicas semelhantes, se o comportamento e localização celular são semelhantes.

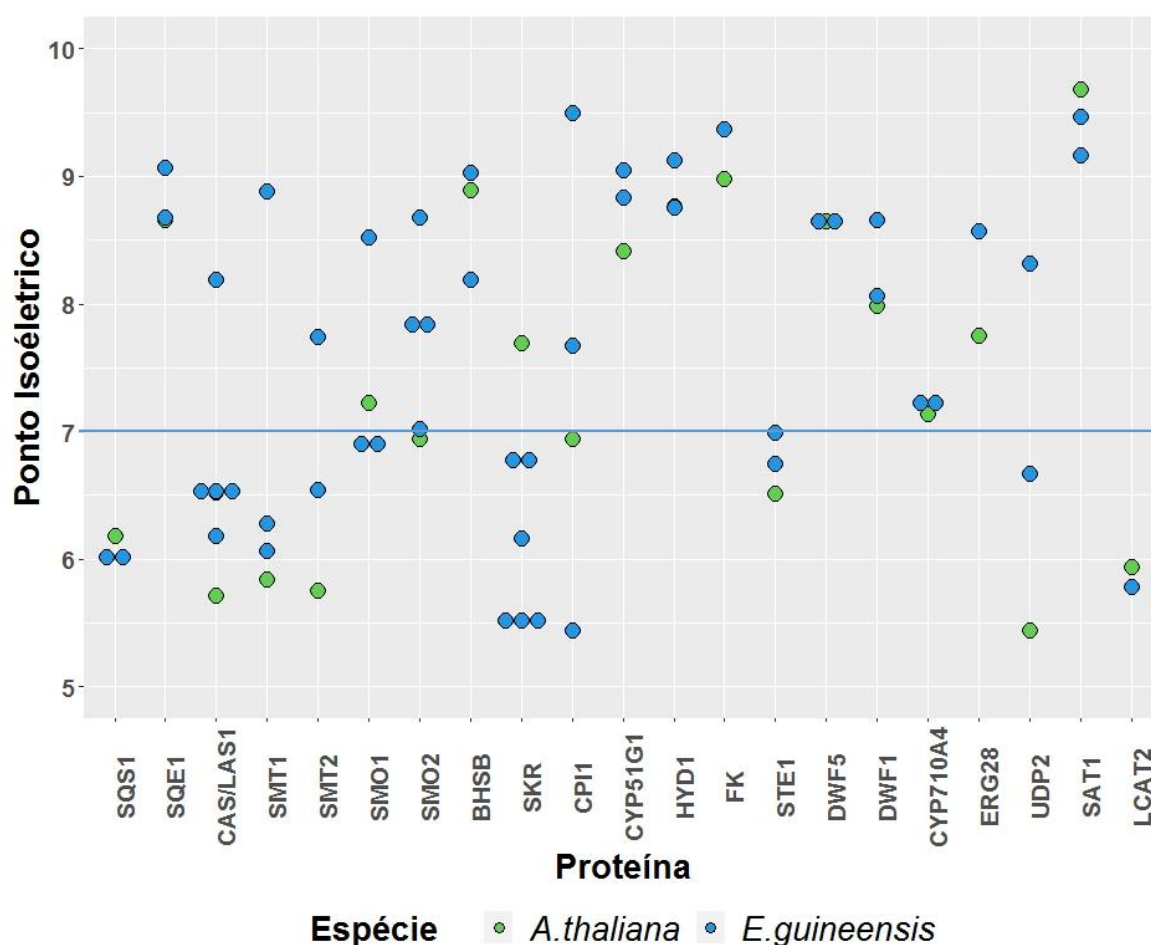


Figura 40: gráfico de ponto isoelétrico das enzimas da via biossíntese dos esteróis, a linha azul separa pontos isoelétricos que indicam acidez($pI < 7$) e basicidade($pI > 7$) das proteínas.

O ponto isoelétrico (pI) das proteínas envolvidas na síntese dos lipídeos esteróis de *E.guineensis* varia de 9,5 em EgCPI1-3 a 5,44 em EgSKR. Algumas proteínas têm altas variações de ponto isoelétrico entre as duas espécies, como CPI1, SMT2 e UDP2, o que torna difícil classificá-las quanto ao pH. pI é o valor de pH no qual as proteínas não têm carga ou a soma das cargas negativas e

positivas é igual a zero, ele fornece um indicativo da natureza da proteína em relação ao pontencial hidrogeniônico em condições naturais (ALBERTS, 2017). As enzimas SQS1, CAS1/LAS1, SMT1, STE1 e LCAT2 das duas espécies podem ser classificadas como levemente ácidas($pI < 7$), já SQE1, SMO2, β HSD, CYP51, DWF5, DWF1, HYD1, FK, ERG28 e SAT1 são classificadas como básicas($pI > 7$). Já SMO1 e CYP710 são classificadas como neutras tendo em vista o pI próximo a 7. A figura 40 é o gráfico dos pontos isoeltricos das enzimas da via.

O peso molecular é altamente conservado entre as espécies, ele varia de 14,9 kDa em EgERG28 a 94 kDa EgCAS1. Todas as enzimas têm peso molecular muito semelhantes entre as duas espécies, o que corrobora o percentual de identidade aferido pelo alinhamento. Kirag et al. (2007) reportou uma relação positiva entre o tamanho, comprimento e ponto isoelétrico de diversas proteínas fazendo uma metanálise usando proteomas das três divisões de seres vivos, no entanto, mais tarde essa correlação não foi corroborada para plantas (MOHANTA et al 2019).

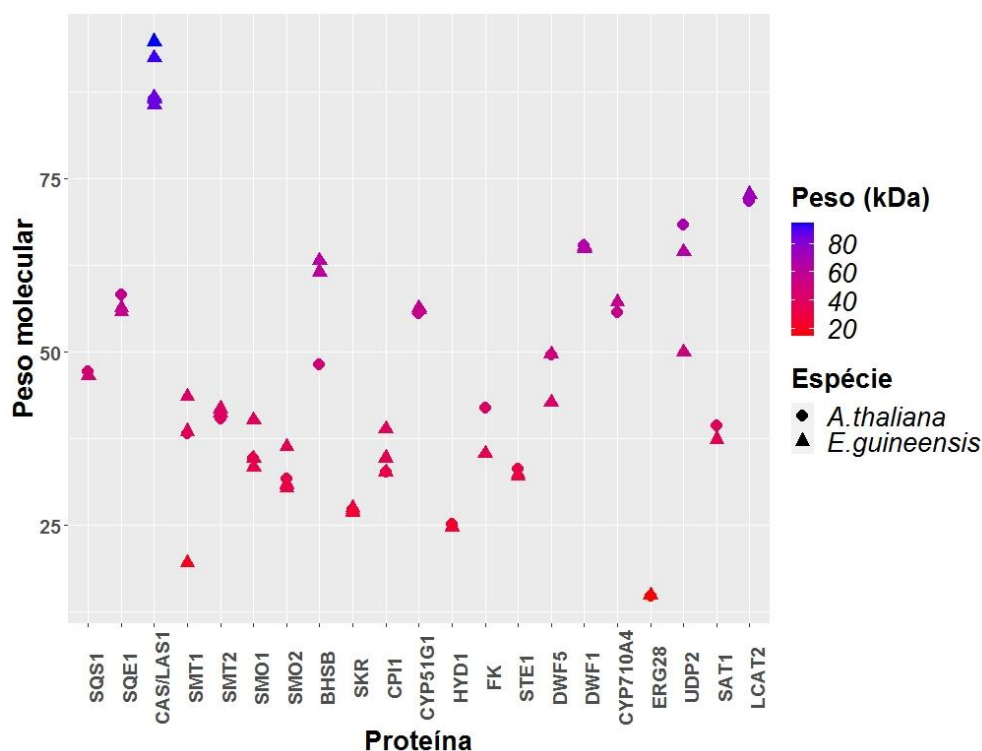


Figura 41: gráfico do peso molecular das Enzimas que realizam a via de biossíntese dos esteroides.

Essa relação também não foi observada entre os pesos moleculares, o tamanho e o ponto isoelétrico das enzimas deste trabalho, através da correlação de Pearson's (Peso~pI: $t = -0.84084$, $df = 71$, $p\text{-value} = 0.4033$, Tamanho~pI, $t = -0.84084$, $df = 71$, $p\text{-value} = 0.4033$, Tamanho~Peso: $t = 48.715$, $df = 71$, $p\text{-value} < 2.2e-16$), o figura 41 indica os peso moleculares das enzimas de *E.guineensis* e *Arabidopsis*.

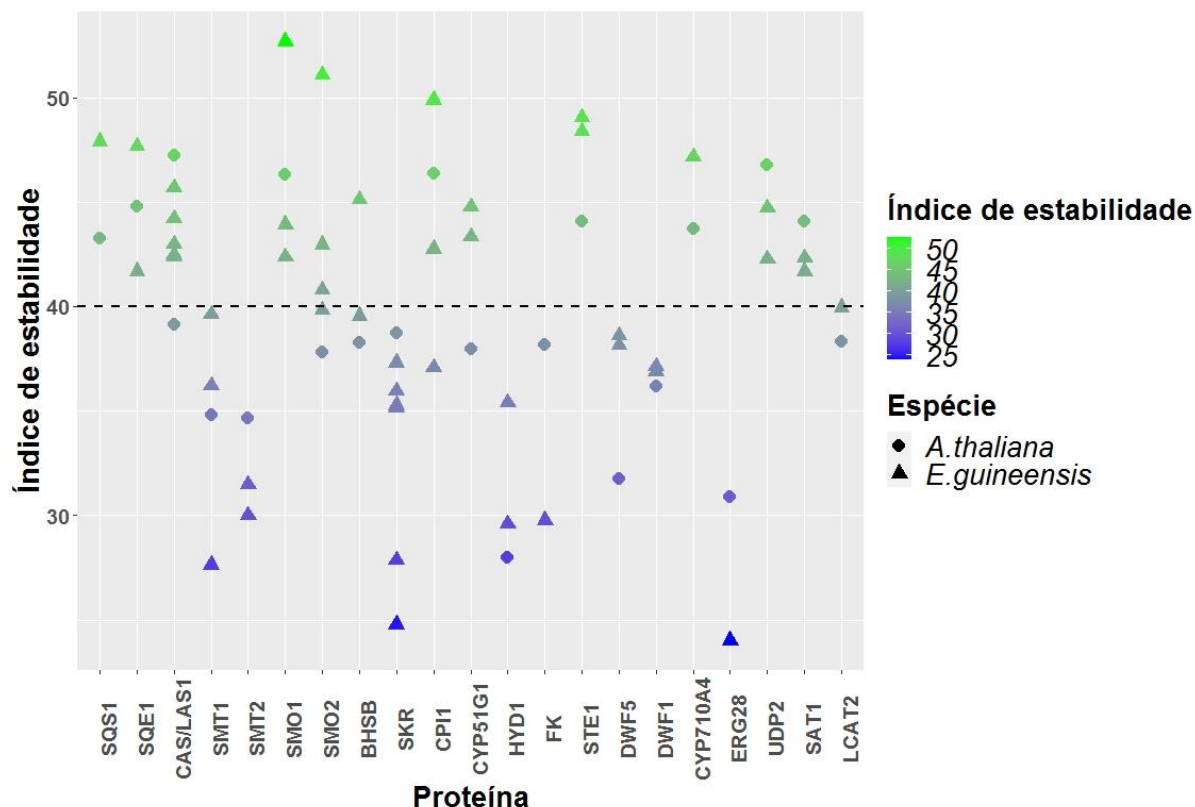


Figura 42: gráfico representando o índice da estabilidade das enzimas da via, a linha tracejada marca o ponto que separa proteínas estáveis e instáveis de acordo com (KAUR et al., 2020), (GURUPRASAD et al., 1990)

O índice de estabilidade dá um indicativo da estabilidade da enzima em condições citosólicas ou *in vitro*. Proteínas com índice de instabilidade < 40 são consideradas estáveis (meia-vida maior 16 horas) em relação àquelas com valor > 40 são referidas como instáveis (Meia-vida menor que 5 horas) (KAUR et al., 2020), (GURUPRASAD et al., 1990). No caso das enzimas da via de biossíntese dos esteróis, ERG28, DWF1, DWF5, FK, HYD1, SKR, SMT1 e SMT2 são estáveis, o restante das enzimas é instável em condições celulares normais. O menor valor de índice de estabilidade foi das enzimas ERG28 e SKR. ERG28 é uma proteína estrutural responsável por manter o complexo de desmetilação unido a membrana do retículo endoplasmático (GACHOTTE et al., 2001), já SKR

desempenha diversos papéis como proteína da super-família SDR que catalisa diversas reações de redução no citoplasma, não só na via dos esteróis (RAHIER et al., 2009), a figura 42 ilustra os pontos dos índices de estabilidade.

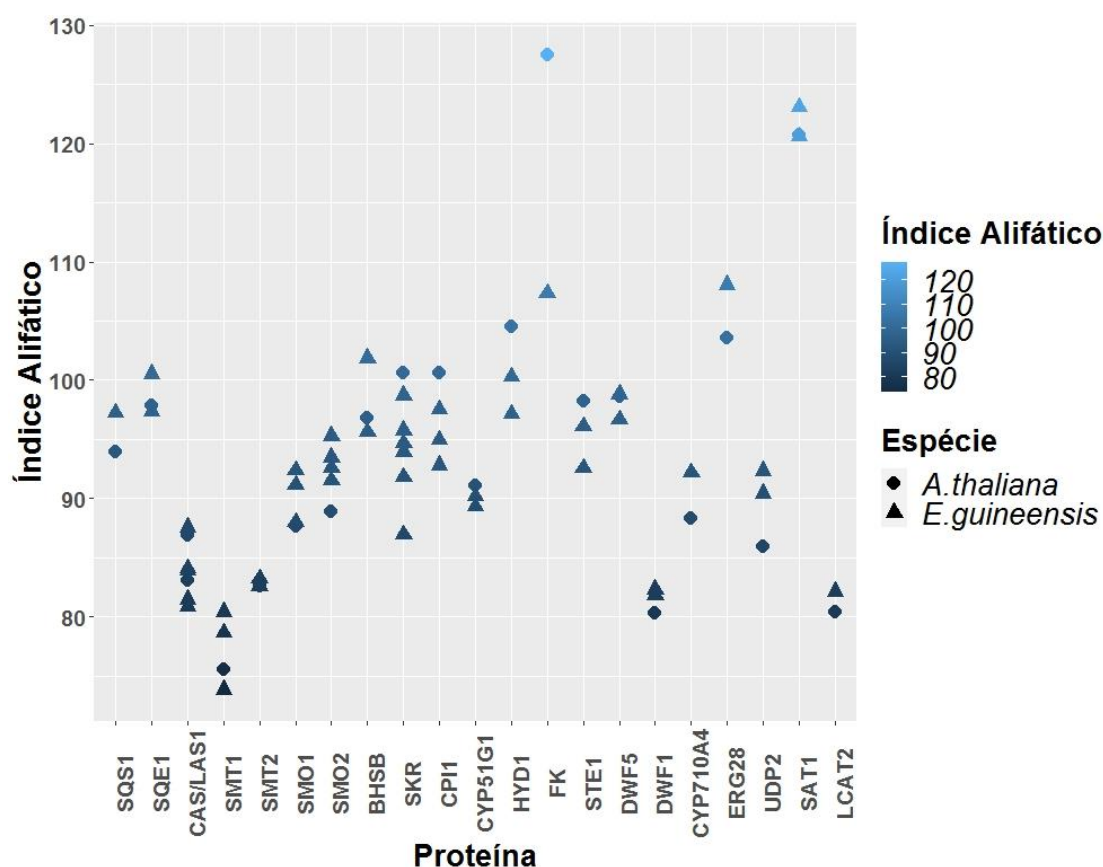


Figura 43: gráfico que representa o índice alifático das enzimas do dendezeiro e Arabidopsis.

O índice alifático é outro indicativo de estabilidade proteica, ele é definido como o volume relativo ocupado pelas cadeias laterais alifáticas de aminoácidos como A (alanina), V (valina), L (leucina) e I (isoleucina). Ikai (1980) estabeleceu uma relação entre o índice alifático e a termoestabilidade de proteínas, quanto mais alto o índice alifático mais termoestável e menos flexível uma proteína é (KAUR et al., 2020). Como mostra a figura 39, grande parte das proteínas deste trabalho tem são ricas em Leucina, alanina e valina, portanto, a maioria delas tem índice alifático alto, superior a 80, indicando alta termoestabilidade das proteínas da via. Uma hipótese que pode explicar essa característica das enzimas da via dos esteróis: é que eles estão envolvidas no metabolismo de lipídeos, estão acoradas à membranas lipídicas e interagem fortemente com

substâncias apolares que não são bons condutores de calor, portanto, é possível que compartilhem essa característica do seu ambiente celular proximal.

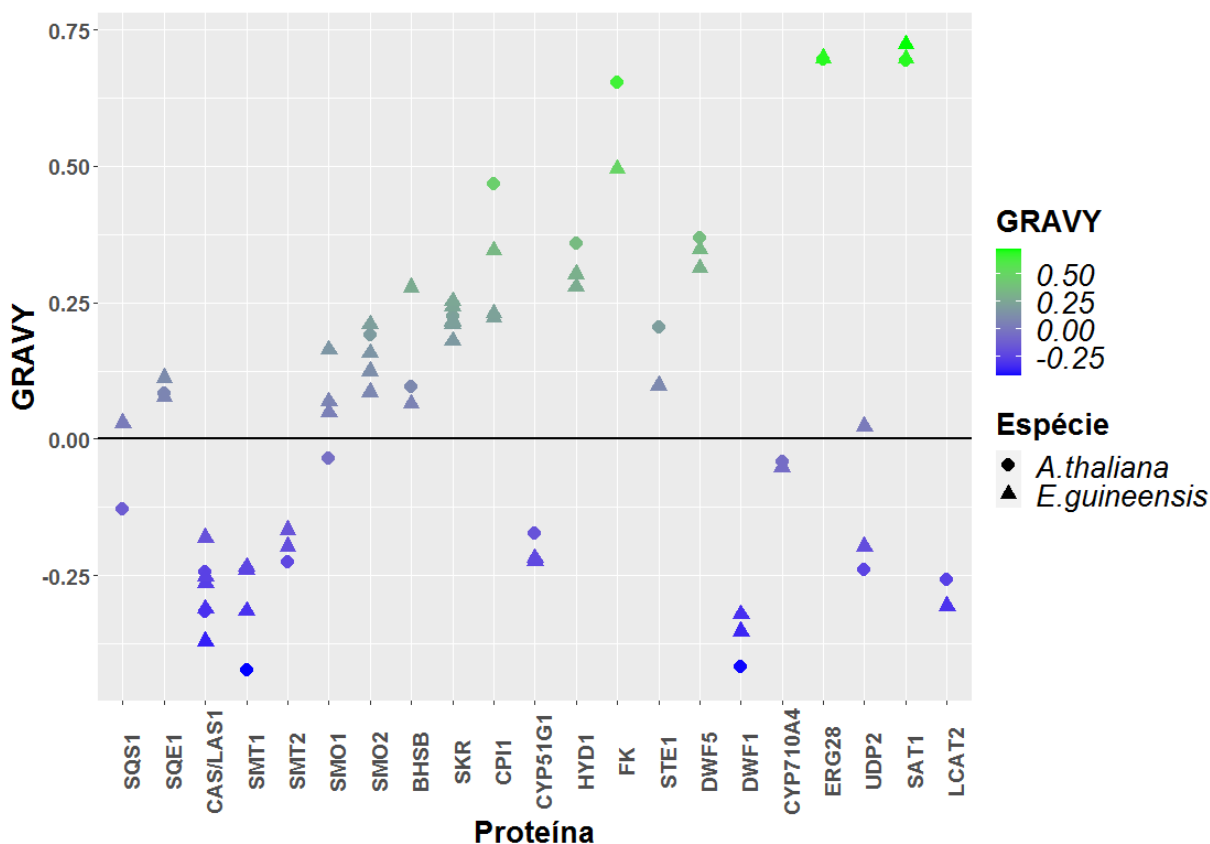


Figura 44: gráfico do índice de hidropaticidade das enzimas, valores maiores zero indicam que a enzima é hidrofóbica, valores menores que zero indicam que a enzima é hidrofílica.

O GRAVY (do inglês, grand average of hydropathicity) é um número que indica a hidropaticidade e solubilidade da proteína em água. O GRAVY para determinada proteína é calculado usando soma dos valores de hidropaticidade de todos os aminoácidos presentes na proteína, dividido pelo número de resíduos na mesma. Seu valor está entre -2 a +2, onde a pontuação negativa significa hidrofiliabilidade e pontuação positiva indica hidrofobicidade (KAUR et al., 2020). No caso das proteínas da via de biossíntese de esteróis 7 enzimas têm GRAVY negativo, no entanto, muito próximo a 0 indicando que são levemente hidrofílicas e pouco solúveis em água. O restante das enzimas (a maioria delas) tem GRAVY positivo, algumas próximo a 0,75, como ERG28 e SAT1. Novamente, essa característica pode ser explicada pela interação com componentes lipídicos das células, a figura 44 mostra o gráfico do índice de hidropaticidade das enzimas dos esteróis.

Parâmetros semelhantes para as características físico-químicas das enzimas da via foram obtidos por Zheng et al., (2018) com genes que realizam a biossíntese dos brasinoesteroides (derivados de esteróis) em macieiras e com esqualeno synthases de fabaceas (AMINFAR & TOHIDFAR 2018).

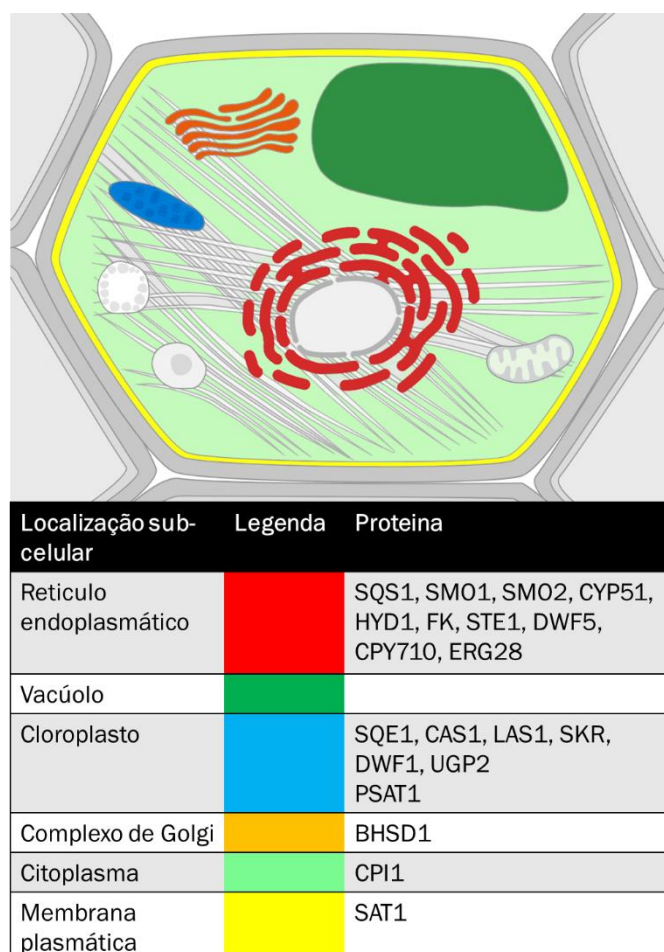


Figura 45: representação esquemática das localizações sub-celulares das enzimas da via de biossíntese dos esteróis.

Quanto a predição da localização sub-celular, das vinte enzimas da via dez foram localizadas no retículo endoplasmático, sete no cloroplasto, uma no complexo de golgi, uma na membrana plástica e uma no citoplasma. O esquema feito na imagem resume os resultados da predição da localização sub-celular das enzimas. Esses resultados corroboram parcialmente estudos anteriores que indicam que os genes da via dos esteróis estão localizados no reticulo endoplasmático, como indicam Benveniste (2004) e Schaller (2010), no entanto, o anabolismo de lipídeos isoprenóides ocorre em partes no cloroplasto é possível

as enzimas da via de biossíntese dos esteróides tenham peptídios sinais que as localizam no cloroplasto também (VRANOVA et al., 2013).

4.9 Análise da expressão diferencial

Para confirmar os resultados de expressão com dados estatísticos uma análise de expressão diferencial dos genes foi realizada em todos os tecidos. Como já dito anteriormente os tecidos das flores apresentam um perfil particular de expressão dos genes da via de biossíntese dos esteróides, é o tecido que apresenta mais genes diferencialmente expressos, pelo menos um gene ou isoforma de cada proteína da via. A figura 46 mostra os gráficos de genes diferencialmente expressos em diversos tecidos contrapostos ao tecido da flor.

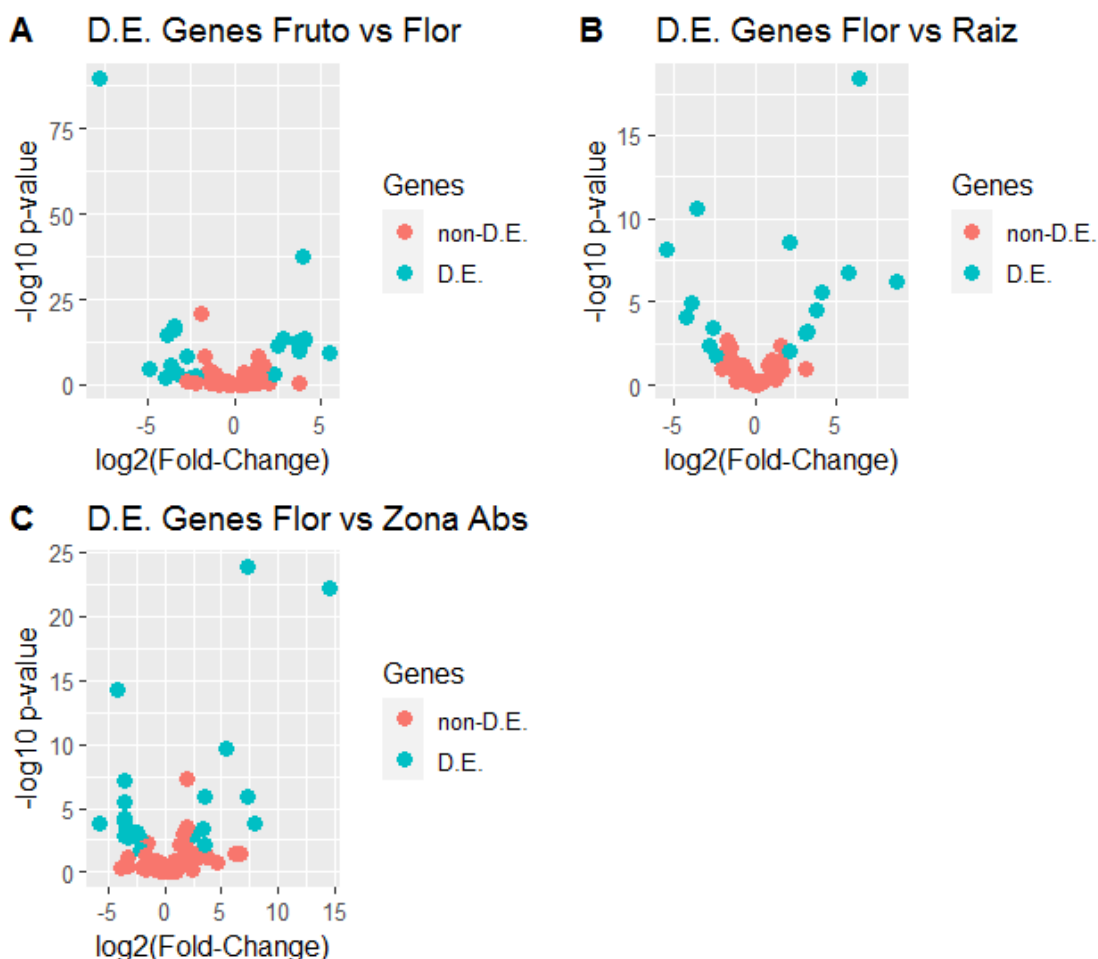


Figura 26: gráficos de expressão diferencial dos genes de *Elaeis guineensis*, os pontos azuis indicam genes que foram diferencialmente expressos.

O padrão de expressão dos genes no carpelo e no estame da flor mostra genes como EgDWF5, EgDWF1, EgCAS e EgCYP51A, EgSMT1 e EgSMT2 altamente expressos, ao passo que em outros tecidos a expressão desses é

mediana ou reduzida. Como já dito anteriormente mutações em genes da via biossíntese dos esteróis afetam o desenvolvimento e fertilidade das plantas (NOLAN et al., 2020; ZHENG et al., 2018; VRIET et al., 2013). O balanço entre sitoesterol e campesterol está diretamente ligado ao desenvolvimento das flores, plantas com altas taxas de campesterol e baixas taxas de sitoesterol tem flores aberrantes e fertilidade reduzida (SCHAEFFER et al, 2010), o controle da razão entre esses dois esteróis é feito pela segunda metilação que ocorre na via, catalisada pela enzima SMT2, as taxas elevadas de expressão desse gene nas flores podem estar relacionadas com essa atividade organogênica do gene.

Experimentos de mutagênese nos genes CYP51A de *Arabidopsis* também levam a esterilidade da planta por má formação das estruturas reprodutivas da flor (KIM et al., 2010). Choe et al., (2001) realizando experimentos encontraram altos níveis de expressão de DWF5 na flor e fertilidade reduzida em alelos mutados desse gene, também em *Arabidopsis*. Esses resultados dialogam com a função fisiológica dos esteróis como precursores de hormônios envolvidos no desenvolvimento sexual das plantas. Embora já existam dados que indiquem que a suplementação de brassinoesteroides não recupera alguns mutantes para genes da via de biossíntese dos esteróis, mostrando que nesses casos esses genes atuam na organogênese da flor por métodos ainda não conhecidos (SCHAEFFER et al, 2010; CARLAND et al., 2010).

O dendezeiro é uma palmeira produtora de óleo no mesocarpo e já está bem relatado na literatura que SAT1 e PSAT1 estão envolvidos no acúmulo de lipídeos no citoplasma das células. Esses dois genes sintetizam as formas conjugadas dos esteróis, estão localizadas em *lipid droplets* que armazenam lipídeos em excesso no citosol (LARA et al., 2018), esperava-se que essas enzimas estivessem super expressas nos tecidos do mesocarpo, não entanto, isso não ocorre, os níveis de expressão delas são relativamente baixos nesse tecido e não estão diferencialmente expressas em relação aos outros.

Outros resultados corroboram esse padrão de expressão de genes da via de biossíntese dos esteróis. Ines et al., (2019) avaliaram o perfil de expressão dos genes da via de biossíntese dos esteróis em *Olea europaea* (Oliveira) e encontraram resultados similares que mostram altos níveis de expressão de SMT2 e CYP51 durante antese floral, já durante a amadurecimento do fruto o

padrão de expressão não sofreu grandes alterações, embora a oliveira também seja uma planta produtora de óleos, como o azeite de oliva. No entanto, os autores destacam a importância da expressão dos genes no mesocarpo, indicando que esses lipídeos são importantes na manutenção da qualidade do azeite de oliva e dos seus benefícios à saúde humana.

É importante ressaltar que esses resultados precisam ser confirmados com experimentos *in vivo*, as análises da expressão dos genes foram feitas com transcriptomas de bibliotecas de espécimes diferentes, que foram sequenciadas por tecnologias diferentes, com extrações diferentes, isso pode afetar a comparação dos tecidos. Mas é importante levar em consideração que outros estudos já foram feitos com metodologias semelhantes e os resultados foram confirmados com experimentos *in vivo*, como RT-PCR dos tecidos do dendezeiro (ROSLI, et al 2018; XIAO *et al.*, 2018; BADAI *et al.*, 2018; KAUR *et al.*, 2020).

5 CONCLUSÕES

Os 22 genes de *Arabidopsis thaliana* foram alinhados para 51 genes de *Elaeis guineensis*, portanto um total de 51 genes ortólogos entre as duas espécies. A pesquisa de domínios não revelou nenhum do domínio novo para os 51 ortólogos, deste modo não houve divergência funcional, embora alguns genes claramente tenham se duplicado ao longo da evolução das palmeiras, a exemplo de CAS1 tem apenas um gene caracterizado em *Arabidopsis* e 5 genes com diversas isoformas no dendezeiro, no entanto, os cinco genes apresentam os mesmos domínios proteicos, características físico química semelhantes e tem a mesma localização subcelular. De todos os 51 genes apenas 2 genes e uma isoforma não foi expressa em tecido nenhum, pelo menos um dos genes de cada enzima da via está diferencialmente e fortemente nos tecidos reprodutivos das flores, corroborando assim, umas das funções fisiológicas dos esteróis no dendezeiro. Mais estudos são necessários para confirmar a homologia e o padrão de expressão dos genes da via biossíntese dos esteróis no dendezeiro

6 REFERÊNCIAS BIBLIOGRÁFICAS

- ABE, I. ABE, T. LOU, W. MASUOKA, T. NOGUCHI, H. Site-directed mutagenesis of conserved aromatic residues in rat squalene epoxidase. **Biochemical Biophysical Research Communication**, 352(1), 259-263. 2007.
- AMEUR, A. KLOOSTERMAN, W. P. HESTAND, M. S. Single-Molecule Sequencing: Towards Clinical Applications. **Trends Biotechnol**, 37(1), 72-85. 2019.
- ARNQVIST, L. PERSSON, M. JONSSON, L. DUTTA, P.C. SITBON, F. Overexpression of CYP710A1 and CYP710A4 in transgenic Arabidopsis plants increases the level of stigmasterol at the expense of sitosterol. **Planta**, 227(2), 309-317. 2008.
- ASHBURNER, M. BALL, C.A. BLAKE, J.A. BOTSTEIN, D. BUTLER, H. CHERRY, J. M. Sherlock, G.. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. **Nature Genetics**, 25(1), 25-29. 2000.
- ATTWOOD, T. K. MITCHELL, A. L. The Evolution of Protein Familia Databases. **Elsevier Reviews**. 34-45. 2019.
- BAGOWSKI, C.P. BRUINS, W. TE VELTHUIS, A.J. The nature of protein domain evolution: shaping the interaction network. **Current Genomics**, 11(5), 368-376. 2010.
- BANAS, A. CARLSSON, A.S. HUANG, B. LENMAN, M. BANAS, W. LEE, M. STYMNE, S. Cellular sterol ester synthesis in plants is performed by an enzyme (phospholipid:sterol acyltransferase) different from the yeast and mammalian acyl-CoA:sterol acyltransferases. **Journal Biological Chemistry**, 280(41), 34626-34634. 2005.
- BENVENISTE, P. Biosynthesis and accumulation of sterols. **Annual Reviews in Plant Biology**, 55, 429-457..141616. 2004.
- BECK, J. G. MATHIEU, D. LOUDET, C. BUCHOUX, S., & DUFOURC, E. J. Plant sterols in "rafts": a better way to regulate membrane thermal shocks. **FASEB Journal** , 21(8), 1714-1723. 2007.
- BOLGER, A. M., LOHSE, M., & USADEL, B. Trimmomatic: A flexible trimmer for Illumina Sequence Data. **Bioinformatics**, btu170. 2014.
- BOUTTE, Y. GREBE, M. Cellular processes relying on sterol function in plants. **Current Opinion Plant Biology**, 12(6), 705-713. 2009.
- BULJAN, M. BATEMAN, A. The evolution of protein domain families. **Biochemical Society transactions**, 37(Pt 4), 751-755. 2009.
- BUMGARNER, R. Overview of DNA microarrays: types, applications, and their future. **Current Protocols Molecular Biology**, Chapter 22, Unit 22 21. 2013.

- BUSQUETS, A. KEIM, V. CLOSA, M. DEL ARCO, A. BORONAT, A. ARRO, M. FERRER, A. *Arabidopsis thaliana* contains a single gene encoding squalene synthase. **Plant Molecular Biology**, 67(1-2), 25-36. 2008.
- CARLAND, F. FUJIOKA, S. NELSON, T. The sterol methyltransferases SMT1, SMT2, and SMT3 influence *Arabidopsis* development through nonbrassinosteroid products. **Plant Physiology**, 153(2), 741-756. 2010.
- CHAN, K.L. TATARINOVA, T.V. ROSLI, R. AMIRUDDIN, N. AZIZI, N. HALIM, M.A.A. LOW, E.T.L. Evidence-based gene models for structural and functional annotations of the oil palm genome. **Biology Direct**, 12(1). 2017
- CHEN, Q. STEINHAUER, L. HAMMERLINDL, J. KELLER, W. ZOU, J.. Biosynthesis of phytosterol esters: identification of a sterol o-acyltransferase in *Arabidopsis*. **Plant Physiology**, 145(3), 974-984. 2007.
- DARNET, S. RAHIER, A. Plant sterol biosynthesis: identification of two distinct families of sterol 4 α -methyl oxidases. **Biochemical Journal**, 378(Pt 3), 889-898. 2004.
- DARNET, S. SCHALLER, H. Metabolism and Biological Activities of 4-Methyl-Sterols. **Molecules**, 24(3). 2019.
- DAS, S. HIRANO, M. Comparative genomics and genome evolution. **Current Genomics**, 13(2), 85. 2012.
- DAWSON, N.L. DAS, S. LEES, J.G. ORENGO, C. Protein Structure Classification. **Methods in Molecular Biology**. 472-487. 2019.
- DEBOLT, S. SCHEIBLE, W. R. SCHRICK, K. AUER, M. BEISSON, F. BISCHOFF, V. SOMERVILLE, C. Mutations in UDP-Glucose:sterol glucosyltransferase in *Arabidopsis* cause transparent testa phenotype and suberization defect in seeds. **Plant Physiology**, 151(1), 78-87. 2009.
- DIENER, A. C. LI, H. ZHOU, W. WHORISKEY, W. J. NES, W. D. FINK, G. R. Sterol methyltransferase 1 controls the level of cholesterol in plants. **Plant Cell**, 12(6), 853-870. 2000.
- DINIZ, W.J. CANDURI, F. Bioinformatics: an overview and its applications. **Genetic and Molecular Research**, 16(1). 2017.
- DONG, Q. BRENDDEL, V. Computational identification of related proteins: BLAST, PSI-BLAST, and other tools. In J.M. Walker (ed.), **The Proteomics Protocols Handbook**, Humana Press, U.S.A., pp. 555-570. 2005.
- RAMIREZ-ESTRADA K, CASTILLO N, LARA JA, ARRÓ M, BORONAT A, FERRER A AND ALTABELLA T. Tomato UDP-Glucose Sterol Glycosyltransferases: A Familia of Developmental and Stress Regulated Genes that Encode Cytosolic and Membrane-Associated Forms of the Enzyme. **Front. Plant Sci**. 8:984. doi: 10.3389/fpls.2017.00984, 2017.
- FERRER, A., ALTABELLA, T., ARRO, M., & BORONAT, A. Emerging roles for conjugated sterols in plants. **Progress in Lipid Research**, 67, 27-37. 2017.

FONTANILLO, C. NOGALES-CADENAS, R. PASCUAL-MONTANO, A. DE LAS RIVAS, J.. Functional analysis beyond enrichment: non-redundant reciprocal linkage of genes and biological terms. **PLoS One**, 6(9), e24289. 2011.

GAS-PASCUAL, E. BERNA, A. BACH, T.J. SCHALLER, H. Plant oxidosqualene metabolism: cycloartenol synthase-dependent sterol biosynthesis in *Nicotiana benthamiana*. **PLoS One**, 9(10), e109156. 2014.

GENSER, B. SILBERNAGEL, G. DE BACKER, G. BRUCKERT, E. CARMENA, R. CHAPMAN, M.J. MARZ, W. Plant sterols and cardiovascular disease: a systematic review and meta-analysis. **European Heart Journal**, 33(4), 444-451. 2012.

GODSWILL, N.N. FRANK, N.-E. G. WALTER, A.N. EDSON, M.Y.J. KINGSLEY, T.M. ARONDEL, V. Emmanuel, Y. Oil Palm. 217-273. 2016.

GRONNIER, J. GERBEAU-PISSOT, P. GERMAIN, V. MONGRAND, S. SIMON-PLAS, F.. Divide and Rule: Plant Plasma Membrane Organization. **Trends in Plant Science**, 23(10), 899-917. 2018.

GROSJEAN, K. MONGRAND, S. BENEY, L. SIMON-PLAS, F. GERBEAU-PISSOT, P. Differential effect of plant lipids on membrane organization: specificities of phytosphingolipids and phytosterols. **Journal of Biological Chemistry**, 290(9), 5810-5825. 2015.

HUSSELSTEIN, T. SCHALLER, H. GACHOTTE, D. BENVENISTE, P. Delta7-sterol-C5-desaturase: molecular characterization and functional expression of wild-type and mutant alleles. **Plant Molecular Biology**, 39(5), 891-906. 1999.

JIN, J. LEE, M. BAI, B. SUN, Y. QU, J., RAHMADSYAH, YUE, G.H. Draft genome sequence of an elite Dura palm and whole-genome patterns of DNA variation in oil palm. **DNA Research**, 23(6), 527-533. 2016.

KANEHISA, M. FURUMICHI, M. TANABE, M. SATO, Y. MORISHIMA, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. **Nucleic Acids Research**, 45(D1), D353-D361. 2017.

KAWASHIMA, T. Comparative and Evolutionary Genomics. **Elsevier Reviews**, 257-267. 2019.

KIM, B. KIM, G., FUJIOKA, S. TAKATSUTO, S. CHOE, S. Overexpression of 3beta-hydroxysteroid dehydrogenases/C-4 decarboxylases causes growth defects possibly due to abnormal auxin transport in *Arabidopsis*. **Molecules and Cells**, 34(1), 77-84. 2012.

KIM, H.B. LEE, H. OH, C.J. LEE, H.Y. EUM, H.L. KIM, H.S. CHOI, S. B. Postembryonic seedling lethality in the sterol-deficient *Arabidopsis* cyp51A2 mutant is partially mediated by the composite action of ethylene and reactive oxygen species. **Plant Physiology**, 152(1), 192-205. 2010.

KRISTENSEN, D. M. WOLF, Y. I. MUSHEGIAN, A. R. KOONIN, E. V. Computational methods for Gene Orthology inference. **Briefings in Bioinformatics**, 12(5), 379-391. 2011.

KUCZYNSKA, A. CARDENIA, V. OGRODOWICZ, P. KEMPA, M. RODRIGUEZ-ESTRADA, M.T. MIKOLAJCZAK, K. Effects of multiple abiotic stresses on lipids

and sterols profile in barley leaves (*Hordeum vulgare* L.). **Plant Physiology and Biochemistry**, 141, 215-224. 2019.

KUMAR, S., STECHER, G. TAMURA, K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, 33(7), 1870-1874.

LARANJEIRA, S. AMORIM-SILVA, V. ESTEBAN, A. ARRO, M. FERRER, A.TAVARES, R.M. AZEVEDO, H. Arabidopsis Squalene Epoxidase 3 (SQE3) Complements SQE1 and Is Important for Embryo Development and Bulk Squalene Epoxidase Activity. **Molecular Plant**, 8(7), 1090-1102. 2015.

LIU, G. FU, J. Squalene synthase cloning and functional identification in wintersweet plant (*Chimonanthus zhejiangensis*). **Botanical Studies**, 59(1), 30. 2018.

LESLIE LOW, E.T. The Oil Palm Genome Revolution. **Journal of Oil Palm Research**, 29(4), 456-468. 2018.

LEPESHEVA G.I., WATERMAN M.R. Structural basis for conservation in the CYP51 Familia. *Biochim. Biophys. Acta Proteins Proteom*. 2011;1814:88–93. doi: 10.1016/j.bbapap.2010.06.006.

LOVE MI, HUBER W, ANDERS S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.” **Genome Biology**, 15, 550. doi: 10.1186/s13059-014-0550-8. 2014.

MEN, S., BOUTTE, Y. IKEDA, Y. LI, X., PALME, K. STIERHOF, Y. D. GREBE, M. Sterol-dependent endocytosis mediates post-cytokinetic acquisition of PIN2 auxin efflux carrier polarity. **Nature Cell Biology**, 10(2), 237-244. 2008.

MISHRA, M. K. SINGH, G., TIWARI, S., SINGH, R., KUMARI, N., & MISRA, P. Characterization of Arabidopsis sterol glycosyltransferase TTG15/UGT80B1 role during freeze and heat stress. **Plant Signal Behavior**, 10(12), e1075682. 2015.

MOREAU, R.A. WHITAKER, B.D. HICKS, K.B. Phytosterols, phytostanols, and their conjugates in foods: structural diversity, quantitative analysis, and health-promoting uses. **Progress in Lipid Research**, 41(6), 457-500. 2002.

MORIKAWA, T. MIZUTANI, M. AOKI, N. WATANABE, B. SAGA, H. SAITO, S. OHTA, D. Cytochrome P450 CYP710A encodes the sterol C-22 desaturase in Arabidopsis and tomato. **Plant Cell**, 18(4), 1008-1022. 2006.

NEELAKANDAN, A. K. SONG, Z. WANG, J. RICHARDS, M. H. WU, X. VALLIYODAN, B. NES, W. D.. Cloning, functional expression and phylogenetic analysis of plant sterol 24C-methyltransferases involved in sitosterol biosynthesis. **Phytochemistry**, 70(17-18), 1982-1998. 2009.

NOLAN, T. M. VUKASINOVIC, N. LIU, D. RUSSINOVA, E. YIN, Y.. Brassinosteroids: Multidimensional Regulators of Plant Growth, Development, and Stress Responses. **Plant Cell**, 32(2), 295-318. 2020.

OHYAMA, K. SUZUKI, M., KIKUCHI, J. SAITO, K. MURANAKA, T. Dual biosynthetic pathways to phytosterol via cycloartenol and lanosterol in Arabidopsis. **Proceedings of the National Academy of Sciences of USA**, 106(3), 725-730. 2009.

ORESIC, M., HANNINEN, V. A., & VIDAL-PUIG, A. LIPIDOMICS: a new window to biomedical frontiers. **Trends in Biotechnology**, 26(12), 647-652. 2008.

P. BAGOWSKI, C. BRUINS, W.J.W. TE VELTHUIS, A. The Nature of Protein Domain Evolution: Shaping the Interaction Network. **Current Genomics**, 11(5), 368-376. 2010.

PADYANA, A.K. GROSS, S. JIN, L. CIANCHETTA, G. NARAYANASWAMY, R. WANG, F. SMOLEN, G.A. Structure and inhibition mechanism of the catalytic domain of human squalene epoxidase. **Nature Communications**, 10(1), 97. 2019.

PATERSON, A.H. WANG, X. TANG, H. LEE, T.H. Synteny and Genomic Rearrangements in: WENDEL F.J. GREILHUBER, J. DOLEZEL, J. **Plant genome diversity**. New York. Springer 195-207. 2012.

PATRO, R., DUGGAL, G., LOVE, M. I., IRIZARRY, R. A., & KINGSFORD, C. Salmon provides fast and bias-aware quantification of transcript expression. **Nature Methods**. 2017.

PEARL, F.M.G. SILLITOE, I. ORENGO, C.A. Protein Structure Classification. 1-10. 2015.

PEREZ-WOHLFEIL, E. DIAZ-DEL-PINO, S. TRELLES, O. Ultra-fast genome comparison for large-scale genomic experiments. **Scientific Reports**, 9(1), 10274.

PERSSON, B. KALLBERG, Y. OPPERMAN, U. JÖRNVALL, H.. Coenzyme-based functional assignments of short-chain dehydrogenases/reductases (SDRs). *Chemico-Biological Interactions*, 143-144, 271-278. 2003.

PLAT, J. BAUMGARTNER, S. VANMIERLO, T. LUTJOHANN, D. CALKINS, K. L. BURRIN, D.G. MENSINK, R.P. Plant-based sterols and stanols in health & disease: "Consequences of human development in a plant-based environment?". **Progress in Lipid Research**, 74, 87-102. 2019.

QIAN, P. HAN, B. FORESTIER, E. HU, Z. GAO, N. LU, W. HOU, S. Sterols are required for cell-fate commitment and maintenance of the stomatal lineage in Arabidopsis. **Plant Journal**, 74(6), 1029-1044. 2013.

RAHIER, A. BERGDOLL, M. GENOT, G. BOUVIER, F. CAMARA, B. Homology modeling and site-directed mutagenesis reveal catalytic key amino acids of 3beta-hydroxysteroid-dehydrogenase/C4-decarboxylase from Arabidopsis. **Plant Physiology**, 149(4), 1872-1886. 2009.

RAHIER, A. DARNET, S. BOUVIER, F. CAMARA, B. BARD, M. Molecular and enzymatic characterizations of novel bifunctional 3beta-hydroxysteroid dehydrogenases/C-4 decarboxylases from Arabidopsis thaliana. *Journal of Biological Chemistry*, 281(37), 27264-27277. 2006.

RAHIER, A. KARST, F. Plant cyclopropylsterol-cycloisomerase: key amino acids affecting activity and substrate specificity. **The Biochemical Journal**, 459(2), 289-299. 2014.

RAHIER, A. PIERRE, S. RIVEILL, G. KARST, F. Identification of essential amino acid residues in a sterol 8,7-isomerase from Zea mays reveals functional

homology and diversity with the isomerases of animal and fungal origin. **The Biochemical Journal**, 414(2), 247-259. 2008.

RASBERRY, J. M. SHAN, H. LECLAIR, R. J. NORMAN, M. MATSUDA, S. P. BARTEL, B. *Arabidopsis thaliana* squalene epoxidase 1 is essential for root and seed development. **Journal Biological Chemistry**, 282(23), 17002-17013. 2007.

ROCHE, Y. GERBEAU-PISSOT, P. BUHOT, B., THOMAS, D. BONNEAU, L. GRETI, J. SIMON-PLAS, F.. Depletion of phytosterols from the plant plasma membrane provides evidence for disruption of lipid rafts. **FASEB Journal**, 22(11), 3980-3991. 2008.

ROZHON, W. HUSAR, S. KALAIVANAN, F. KHAN, M. IDLHAMMER, M. SHUMILINA, D. POPPENBERGER, B. Genetic variation in plant CYP51s confers resistance against voriconazole, a novel inhibitor of brassinosteroid-dependent sterol biosynthesis. **PLoS One**, 8(1), e53650. 2013.

SAWAI, S. AKASHI, T. SAKURAI, N. SUZUKI, H. SHIBATA, D. AYABE, S. AOKI, T. Plant lanosterol synthase: divergence of the sterol and triterpene biosynthetic pathways in eukaryotes. **Plant Cell Physiology**, 47(5), 673-677. 2006.

SCHALLER, H. New aspects of sterol biosynthesis in growth and development of higher plants. **Plant Physiology and Biochemistry**, 42(6), 465-476. 2004.

SCHALLER, H. sterol and steroid biosynthesis and metabolism in plants and microorganisms. 755-787. 2010.

SHARMA, T. R. DEVANNA, B. N. KIRAN, K. SINGH, P. K. ARORA, K. JAIN, P. MONDAL, T. K. Status and Prospects of Next Generation Sequencing Technologies in Crop Plants. **Current Issues in Molecular Biology**, 27, 1-36. 2018.

SCHAEFFER, A., BRONNER, R., BENVENISTE, P. AND SCHALLER, H. The ratio of campesterol to sitosterol that modulates growth in *Arabidopsis* is controlled by STEROL METHYLTRANSFERASE 2;1. **Plant J.** 25, 605– 615. 2001.

SHORT, E. LEIGHTON, M. IMRIZ, G. LIU, D. COPE-SELBY, N. HETHERINGTON, F. LINDSEY, K. Epidermal expression of a sterol biosynthesis gene regulates root growth by a non-cell-autonomous mechanism in *Arabidopsis*. **Development**, 145(10). 2018.

SILVESTRO, D. ANDERSEN, T. G. SCHALLER, H. JENSEN, P. E. Plant sterol metabolism. Delta(7)-Sterol-C5-desaturase (STE1/DWARF7), Delta(5,7)-sterol-Delta(7)-reductase (DWARF5) and Delta(24)-sterol-Delta(24)-reductase (DIMINUTO/DWARF1) show multiple subcellular localizations in *Arabidopsis thaliana* (Heynh) L. **PLoS One**, 8(2), e56429. 2013.

SIMON-PLAS, F. PERRAKI, A. BAYER, E. GERBEAU-PISSOT, P. MONGRAND, S.. An update on plant membrane rafts. **Current Opinion Plant Biology**, 14(6), 642-649. 2011.

SINGH, R. LOW, E. T. OOI, L. C. ONG-ABDULLAH, M. TING, N. C. NAGAPPAN, J. MARTIENSSEN, R. A. The oil palm SHELL gene controls oil yield and encodes a homologue of SEEDSTICK. **Nature**, 500(7462), 340-344. 2013.

SONAWANE, P. D. POLLIER, J. PANDA, S. SZYMANSKI, J. MASSALHA, H. YONA, M. AHARONI, A. Plant cholesterol biosynthetic pathway overlaps with phytosterol metabolism. **Nature Plants**, 3, 16205. 2016.

SONESON C, LOVE MI, ROBINSON MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. **F1000Research**, 4. doi: 10.12688/f1000research.7563.1. 2015.

SONG, J., SUN, S., REN, H., GRISON, M., BOUTTE, Y., BAI, W., & MEN, S. The SMO1 Familia of Sterol 4alpha-Methyl Oxidases Is Essential for Auxin- and Cytokinin-Regulated Embryogenesis. **Plant Physiology**, 181(2), 578-594. 2019.

SPANOVA, M. DAUM, G. Squalene - biochemistry, molecular biology, process biotechnology, and applications. **European Journal of Lipid Science and Technology**, 113(11), 1299-1320. 2011.

SRIVASTAVA, G. SANDEEP, GARG, A. MISRA, R. C. CHANOTIYA, C. S. GHOSH, S.. Transcriptome analysis and functional characterization of oxidosqualene cyclases of the arjuna triterpene saponin pathway. **Plant Science**, 292, 110382. 2020.

SRIVASTAVA, A. GEORGE, J. KARUTURI, R.K.M. Transcriptome Analysis. **Encyclopedia of Bioinformatics and Computational Biology**, 792-805. 2019.

STARK, R. GRZELAK, M. HADFIELD, J. RNA sequencing: the teenage years. **Nature Reviews Genetics**, 20(11), 631-656. 2019.

SUZUKI, M. NAKAGAWA, S. KAMIDE, Y. KOBAYASHI, K. OHYAMA, K. HASHINOKUCHI, H. NAGATA, N. Complete blockage of the mevalonate pathway results in male gametophyte lethality. **Journal Experimental Botany**, 60(7), 2055-2064. 2009.

TATON, M. HUSSELSTEIN, T. BENVENISTE, P. RAHIER, A. Role of highly conserved residues in the reaction catalyzed by recombinant Delta7-sterol-C5(6)-desaturase studied by site-directed mutagenesis. **Biochemistry**, 39(4), 701-71. 2000.

THE GENE ONTOLOGY, C. Expansion of the Gene Ontology knowledgebase and resources. **Nucleic Acids Research**, 45(D1), D331-D338. 2017.

THIMMAPPA, R. GEISLER, K. LOUVEAU, T. O'MAILLE, P. OSBOURN, A. Triterpene biosynthesis in plants. **Annual Reviews in Plant Biology**, 65, 225-257. 2014.

VALITOVA, J. N. SULKARNAYEVA, A. G. MINIBAYEVA, F. V. Plant Sterols: Diversity, Biosynthesis, and Physiological Functions. **Biochemistry (Mosc)**, 81(8), 819-834. 2016.

VRANOVA, E. COMAN, D. GRUISSEM, W. Structure and dynamics of the isoprenoid pathway network. **Molecular Plant**, 5(2), 318-333. 2012.

- VRANOVA, E. COMAN, D. GRUISSEM, W. Network analysis of the MVA and MEP pathways for isoprenoid synthesis. **Annual Reviews Plant Biology**, 64, 665-700. 2013.
- VRIET, C. RUSSINOVA, E. REUZEAU, C. From squalene to brassinolide: the steroid metabolic and signaling pathways across the plant kingdom. **Molecular Plant**, 6(6), 1738-1757. 2013.
- WAN, X. LI, Z. Plant Comparative Transcriptomics Reveals Functional Mechanisms and Gene Regulatory Networks Involved in Anther Development and Male Sterility. 2019.
- WANG, B. KUMAR, V. OLSON, A. WARE, D. Reviving the Transcriptome Studies: An Insight Into the Emergence of Single-Molecule Transcriptome Sequencing. *Frontiers in Genetics*, 10, 384. 2019.
- YANG, H. RICHTER G. L., WANG, X. MLODZINSKA, E. CARRARO, N. MA, G. MURPHY, A. S. Sterols and sphingolipids differentially function in trafficking of the Arabidopsis ABCB19 auxin transporter. **Plant Journal**, 74(1), 37-47. 2013.
- YOUN, J. H. KIM, T.W. JOO, S.H. SON, S.H. ROH, J. KIM, S. KIM, S.K. Function and molecular regulation of DWARF1 as a C-24 reductase in brassinosteroid biosynthesis in Arabidopsis. **Journal of Experimental Botany**, 69(8), 1873-1886. 2018.
- ZHENG, L. ZHAO, C. MAO, J. SONG, C. MA, J. ZHANG, D. AN, N. Genome-wide identification and expression analysis of brassinosteroid biosynthesis and metabolism genes regulating apple tree shoot and lateral root growth. **Journal of Plant Physiology**, 231, 68-85. 2018.