



**MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL RURAL DA AMAZÔNIA
INSTITUTO SÓCIOAMBIENTAL E DOS RECURSOS HÍDRICOS – ISARH
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA APLICADA À
AGROPECUÁRIA - PPGBAA**

WASHINGTON OLEGÁRIO VIEIRA

**IDENTIFICAÇÃO, ANÁLISE *IN SILICO* DE GENES E PROTEÍNAS E
ELABORAÇÃO DE ESTRATÉGIAS DE ISOLAMENTO DE GENES
POTENCIALMENTE ENVOLVIDOS NA BIOSÍNTESE E INATIVAÇÃO DE 20-
HIDROXIECDISONA (20E) EM PLANTAS**

**BELÉM
2021**

WASHINGTON OLEGÁRIO VIEIRA

**IDENTIFICAÇÃO, ANÁLISE *IN SÍLICO* DE GENES E PROTEÍNAS E
ELABORAÇÃO DE ESTRATÉGIAS DE ISOLAMENTO DE GENES
POTENCIALMENTE ENVOLVIDOS NA BIOSÍNTESE E INATIVAÇÃO DE 20-
HIDROXIECDISONA (20E) EM PLANTAS**

Dissertação apresentada ao Curso de Pós-Graduação em Biotecnologia Aplicada à Agropecuária, como requisito parcial para a obtenção do título de Mestre.

Orientador: Prof. Dr. Reginaldo Alves Festucci Buselli

BELÉM

2021

Dados Internacionais de Catalogação na Publicação (CIP)
Bibliotecas da Universidade Federal Rural da Amazônia
Gerada automaticamente mediante os dados fornecidos pelo(a) autor(a)

V657i Vieira, Washington Olegário

Identificação, análise in silico de genes e proteínas e elaboração de estratégias de isolamento de genes potencialmente envolvidos na biossíntese e inativação de 20-hidroxicdisona (20E) em plantas / Washington Olegário Vieira. - 2021.

102 f. : il. color.

Dissertação (Mestrado) - Programa de PÓS-GRADUAÇÃO em Biotecnologia Aplicada à Agropecuária (PPGBAA), Campus Universitário de Belém, Universidade Federal Rural Da Amazônia, Belém, 2021.

Orientador: Prof. Dr. Reginaldo Alves Festucci-Buselli

1. 20-hidroxicdisona. 2. Ferramentas BLAST. 3. Bioinformática. 4. Sequências de aminoácidos. 5. Genes. I. Festucci-Buselli, Reginaldo Alves, *orient.* II. Título

CDD 630.72

WASHINGTON OLEGÁRIO VIEIRA

**IDENTIFICAÇÃO, ANÁLISE *IN SÍLICO* DE GENES E PROTEÍNAS E
ELABORAÇÃO DE ESTRATÉGIAS DE ISOLAMENTO DE GENES
POTENCIALMENTE ENVOLVIDOS NA BIOSÍNTESE E INATIVAÇÃO DE 20-
HIDROXIECDISONA (20E) EM PLANTAS**

Dissertação apresentada ao Curso de Pós-Graduação em Biotecnologia Aplicada à Agropecuária, como requisito parcial para a obtenção do título de Mestre.

Aprovado em 25/08/2021

BANCA EXAMINADORA



Prof. Dr. Reginaldo Alves Festucci Buselli – Presidente/Orientador
UNIVERSIDADE FEDERAL RURAL DA AMAZÔNIA



Dr. Roberto Lisboa Cunha – 1ª Examinador
EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA/AMAZÔNIA ORIENTAL



Dra. Elisa Ferreira Moura Cunha – 2ª Examinadora
EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA/AMAZÔNIA ORIENTAL



Dr. Igor Guerreiro Hamoy – 3º Examinador
UNIVERSIDADE FEDERAL RURAL DA AMAZÔNIA

Dra. Simone de Miranda Rodrigues – Membro Suplente
EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA/AMAZÔNIA ORIENTAL

DEDICATÓRIA

Dedico esse trabalho a minha mãe, Maria do Perpétuo Socorro Olegário (*in memoriam*), com todo o meu amor e gratidão, pois ela será a eterna força e fonte de inspiração na minha vida.

AGRADECIMENTOS

À minha mãe, que embora não esteja mais ao meu lado, agradeço pelo amor incondicional, pelos incentivos escolares e acadêmicos, pelas preocupações, pelos momentos de alegria e ajuda. Amarei-te eternamente.

À minha família, em especial à minha madrastra, Odete Vieira, que me ensinou que não precisa compartilhar do mesmo sangue para ser uma família de verdade, pois desde quando a conheci ela fez parte de todos os momentos felizes e tristes, sempre dando força e carinho e desejando tudo de bom, se tornando uma das pessoas com o coração mais bondoso que já conheci.

À Universidade Federal Rural da Amazônia, instituição a qual faço parte desde a minha graduação.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) peça concessão de bolsa e apoio financeiro.

Ao Programa de Pós-Graduação em Biotecnologia Aplicada à Agropecuária pela oportunidade na realização desse mestrado.

Ao Igor Guerreiro Hamoy, coordenador do Programa de Pós-Graduação em Biotecnologia Aplicada à Agropecuária por seu empenho, dedicação e esforço.

Ao meu Orientador, Professor Dr. Reginaldo Alves Festucci Buselli, pela oportunidade, ensinamentos, conselhos, paciência, amizade, compreensão e competente orientação.

Às minhas amigas da graduação e da pós-graduação, Carol Maciel, Ana Catarina Furtado (some, mas depois lembra e aparece), Bruna Santos, Samara Sousa e Patrícia Navegantes e também às minhas companheiras que a UFRA me deu, Beatriz e Cilene, muito obrigado pelos almoços, pelas conversas, pelas risadas.

Ao Victor Carréra, por estar sempre ao meu lado nos momentos que mais precisei, alegrando minha vida e me ajudando a crescer cada vez mais, com amor, companheirismo e amizade.

À minha querida Gabrielle amiga advogada, sempre disposta a tomar um sorvete nos nossos encontros de casais.

À Lene Moreira e Beatriz Figueiredo, pelas risadas e almoços compartilhados.

Às queridas Izana Magno e Luma Pontes, salinha onde a risada é sempre garantida, principalmente nos aniversários.

Agradeço a todos que, direta ou indiretamente, fizeram parte da minha vida pessoal e acadêmica e proporcionaram o crescimento profissional e ético que tenho hoje.

RESUMO

A produção de metabólitos secundários é uma das principais formas de defesa das plantas contra insetos herbívoros. A molécula 20-hidroxicdisona (20E) é um hormônio produzido por insetos, cuja função está relacionada com diversos processos durante o desenvolvimento do inseto. Além destes organismos, algumas plantas também são capazes de sintetizar 20E, utilizando-o possivelmente como parte do mecanismo de defesa durante a herbivoria. A caracterização e isolamento de genes e proteínas envolvidos em mecanismos de defesas de plantas é essencial para se elaborar estratégias que visam o controle de insetos que atacam espécies de interesse econômico, entretanto a rota de biossíntese de 20E ainda é parcialmente conhecida em insetos e desconhecida em plantas. A molécula 20E também apresenta aplicabilidade farmacológica, medicinal e biotecnológica. Neste contexto, o objetivo deste trabalho foi realizar a identificação e análise *in silico* de genes e proteínas e elaborar estratégias de isolamento de genes potencialmente envolvidos na biossíntese de 20E em plantas. Para isto, foi realizada uma busca por proteínas similares em banco de dados por meio da ferramenta BLAST/NCBI utilizando como base as sequências de aminoácidos de proteínas que atuam a rota parcialmente conhecida de biossíntese (proteínas Neverland, Shade, Shadow, Disembodied e Phantom) e inativação (proteínas Ecdysone oxidase, 3-Dehydroecdysone 3beta-reductase e 3-Dehydroecdysone 3alfa-reductase) de 20E em insetos. Foram identificadas e analisadas *in silico* 239 proteínas potencialmente homólogas que apresentaram valores de similaridade entre 25 e 45%, altos valores de cobertura e valor-E de pelo menos e^{-5} . Foram realizados alinhamentos em pares e múltiplos, predição de valores físico-químicos (peso molecular, ponto isoelétrico teórico, número de resíduos negativos e positivos, coeficiente de extinção, índice de instabilidade, índice alifático e índice de hidrofobicidade), a localização subcelular e filogenia. Foram detectadas em maior quantidade proteínas similares a Shade, Shadow e 3DE 3beta-reductase e em várias espécies de plantas, enquanto as proteínas similares a Disembodied e Neverland foram detectadas, respectivamente, em uma e duas espécies. Foram propostas por meio do uso de oligonucleotídeos, construídos a partir de regiões conservadas, obtidas por meio de alinhamentos múltiplos de sequência de nucleotídeos dos genes, estratégias que visam o isolamento de genes potencialmente envolvidos com a biossíntese de 20E em plantas. Várias proteínas encontradas em vegetais apresentam função ainda desconhecidas, podendo estar envolvidas na biossíntese de 20E em plantas; entretanto, seu envolvimento e real função ainda precisam ser analisadas.

Palavras-chave: 20-hidroxicdisona. Ferramentas BLAST. Bioinformática. Sequências de aminoácidos. Genes.

ABSTRACT

The production of secondary metabolites is one of the main defenses of plants against herbivorous insect attacks. The 20-hydroxyecdysone (20E) molecule is a hormone produced by insects, whose function is related to several processes during insect development. Besides the production in insects, some plants are also capable of synthesizing 20E, possibly using it as part of the defense mechanism during herbivory. The characterization and isolation of genes and proteins involved in plant defense mechanisms is essential to develop strategies aimed at controlling insects that attack species of economic interest, however the 20E biosynthesis route is partially known in insects and unknown in plants. The 20E molecule also has pharmacological, medicinal and biotechnological applicability. In this context, the objective of this work was to carry out the identification and *in silico* analysis of genes and proteins and to elaborate strategies for the isolation of genes potentially involved in the biosynthesis of 20E in plants. For this, a search for similar proteins in a database was performed using the BLAST/NCBI tool using as a basis the amino acid sequence of proteins that act in the partially known biosynthesis route (Neverland, Shade, Shadow, Disembodied and Phantom proteins) and inactivation (Ecdysone oxidase, 3-Dehydroecdysone 3beta-reductase and 3-Dehydroecdysone 3alpha-reductase proteins) of 20E in insects. 239 potentially homologous proteins were identified and analyzed *in silico*, with similarity values between 25 and 45%, high coverage values and an E-value of at least e^{-5} . Pair and multiple alignments, prediction of physicochemical values (molecular weight, theoretical isoelectric point, number of negative and positive residues, extinction coefficient, instability index, aliphatic index and hydrophobicity index), subcellular location and phylogeny were performed. Shade, Shadow, and 3DE 3beta-reductase-like proteins were detected in higher amounts and in several plant species, while Disembodied and Neverland-like proteins were detected in one and two species, respectively. Strategies aimed at the isolation of genes potentially involved with 20E biosynthesis in plants have been proposed through the use of oligonucleotides, constructed from conserved regions, obtained through multiple nucleotide sequence alignments of genes. Several proteins found in vegetables have a function that is still unknown and may be involved in the biosynthesis of 20E in plants; however, its involvement and actual role still need to be analyzed.

Keywords: 20-hydroxyecdysone. BLAST tools. Bioinformatics. Amino acid sequences.

Genes.

LISTA DE FIGURAS

Figura 1 – Interação inseto planta. Mecanismos centrais e co-influência de outros organismos exibindo uma extensa rede multitrófica.	15
Figura 2 – Fluxograma da linha de defesa das plantas.	16
Figura 3 – Estrutura em comum dos ecdisteroides.	19
Figura 4 – Alteração nas concentrações do hormônio 20-hidroxicdisona durante o desenvolvimento da mosca dos frutos <i>Drosophilla melanogaster</i>	21
Figura 5 – A rota de biossíntese da molécula 20-Hidroxicdisona e as enzimas ecdiesteroidogênicas atuantes na rota em insetos. A caixa-preta (Black Box) representa os passos ainda desconhecidos na rota.	21
Figura 6 – Rota de biossíntese de 20-hidroxicdisona a partir do Mevalonato (MVA) em espinafre (<i>Spinacia oleracea</i> L.).	23
Figura 7 – Função dos citocromos P450 em plantas.	26
Figura 8 – Ciclo catalítico do citocromo P450.	28
Figura 9 – Estrutura geral do gene de um organismo eucarioto. Os genes, após a transcrição, apresentam grandes regiões não traduzidas (sigla UTR, em inglês) e uma extensa cauda poli-A, proveniente de eventos pós-transcricionais.	31
Figura 10 – Quantidade de proteínas similares e diversidade de espécies de plantas provenientes da busca em banco de dados pela ferramenta Blast/NCBI a partir de proteínas atuantes na rota de biossíntese de 20-hidroxicdisona em <i>Drosophila melanogaster</i> e <i>Spodoptera littoralis</i> . .	50
Figura 11 – Quantidade de proteínas similares e diversidade de espécies de plantas provenientes da busca em banco de dados pela ferramenta BLAST/NCBI a partir de proteínas atuantes na rota de biossíntese de 20-hidroxicdisona em <i>Drosophila melanogaster</i> e <i>Spodoptera littoralis</i>	51
Figura 12 – Árvores filogenéticas construídas com proteínas vegetais similares às proteínas de <i>Drosophila melanogaster</i> , a partir do alinhamento múltiplo ClustalW utilizando o software MEGA ver. 10.8.1. A árvore foi construída pelo método de Máxima Verossimilhança e validada por análise Bootstrap (1000 repetições). As diferentes cores aplicadas representam o grupo formado por proteínas. A) Proteína Neverland; B) Proteína Shade; C) Proteína Shadow; D) Proteína Phantom.	80

Figura 13 – Árvores filogenéticas construídas com proteínas vegetais similares às proteínas de *Drosophila melanogaster*, a partir do alinhamento múltiplo ClustalW utilizando o *software* MEGA ver. 10.8.1. A árvore foi construída pelo método de Máxima Verossimilhança e validada por análise Bootstrap (1000 repetições). As diferentes cores aplicadas representam o grupo formado por proteínas. A) Proteína Ecdysone oxidase; B) Proteína 3-dehydroecdysone 3alpha-reductase; C) Proteína 3-dehydroecdysone 3alpha-reductase. 81

LISTA DE TABELAS

Tabela 1 – Proteínas isoladas e caracterizadas da rota parcial da biossíntese de 20-hidroxiecdisona em <i>D. melanogaster</i> M. (1830) e <i>S. littoralis</i> Boisduval (1833).	46
Tabela 2 – Diversidade de espécies de plantas e número de proteínas por espécie provenientes da busca em banco de dados pela ferramenta Blast/NCBI a partir de proteínas atuantes na rota de biossíntese e neutralização de 20-hidroxiecdisona em <i>Drosophila melanogaster</i> e <i>Spodoptera littoralis</i> . Total de proteínas encontradas = 239.....	48
Tabela 3 – Proteínas vegetais que apresentam similaridade à proteína Neverland de <i>Drosophila melanogaster</i> a partir da busca em banco de dados pela ferramenta BLAST/NCBI, apresentando a porcentagem de Cobertura (Cob.), de Similaridade (Sim.), o Valor-E, o acesso cadastrado, a presença ou ausência de 20E (+ ou -, respectivamente) e os bancos de dados de origem.	52
Tabela 4 – Proteínas vegetais que apresentam similaridade à proteína Shade de <i>Drosophila melanogaster</i> a partir da busca em banco de dados pela ferramenta BLAST/NCBI, apresentando a porcentagem de Cobertura (Cob.), de Similaridade (Sim.), o Valor-E, o acesso cadastrado, a presença ou ausência de 20E (+ ou -, respectivamente) (DS = Desconhecido) e o banco de dados de origem.	53
Tabela 5 – Proteínas vegetais que apresentam similaridade à proteína Shadow de <i>Drosophila melanogaster</i> a partir da busca em banco de dados pela ferramenta BLAST/NCBI, apresentando a porcentagem de Cobertura (Cob.), de Similaridade (Sim.), o Valor-E, o acesso cadastrado, a presença ou ausência de 20E (+ ou -, respectivamente) (DS = desconhecido) e o banco de dados de origem.	55
Tabela 6 – Proteína vegetal que apresenta similaridade à proteína Disembodied de <i>Drosophila melanogaster</i> a partir da busca em banco de dados pela ferramenta BLAST/NCBI, apresentando a porcentagem de Cobertura (Cob.), de Similaridade (Sim.), o Valor-E, o acesso cadastrado, a presença ou ausência de 20E (+ ou -, respectivamente) (DS = desconhecido) e o banco de dados de origem.	57
Tabela 7 – Proteínas vegetais que apresentam similaridade à proteína Phantom de <i>Drosophila melanogaster</i> a partir da busca em banco de dados pela ferramenta BLAST/NCBI, apresentando a porcentagem de Cobertura (Cob.), de Similaridade (Sim.), o Valor-E, o acesso cadastrado, a presença ou ausência de 20E (+ ou -, respectivamente) (DS = desconhecido) e o banco de dados de origem.	58

Tabela 8 – Proteínas vegetais que apresentam similaridade à proteína Ecdysone Oxidase de *Drosophila melanogaster* a partir da busca em banco de dados pela ferramenta BLAST/NCBI, apresentando a porcentagem de Cobertura (Cob.), de Similaridade (Sim.), o Valor-E, o acesso cadastrado, a presença ou ausência de 20E (+ ou -, respectivamente) (DS = desconhecido) e o banco de dados de origem. 59

Tabela 9 – Proteínas vegetais que apresentam similaridade à proteína 3-dehydroecdysone 3beta-reductase de *Spodoptera littoralis* a partir da busca em banco de dados pela ferramenta BLAST/NCBI, apresentando a porcentagem de Cobertura (Cob.), de Similaridade (Sim.), o Valor-E, o acesso cadastrado, a presença ou ausência de 20E (+ ou -, respectivamente) (DS = desconhecido) e o banco de dados de origem..... 61

Tabela 10 – Proteínas vegetais que apresentam similaridade à proteína 3-dehydroecdysone 3alfa-reductase de *Spodoptera littoralis* a partir da busca em banco de dados pela ferramenta BLAST/NCBI, apresentando a porcentagem de Cobertura (Cob.), de Similaridade (Sim.), o Valor-E, o acesso cadastrado, a presença ou ausência de 20E (+ ou -, respectivamente) (DS = desconhecido) e o banco de dados de origem..... 62

Tabela 11 – Parâmetros físico-químicos computados das proteínas vegetais que apresentam similaridade à proteína Neverland de *Drosophila melanogaster*, por meio da ferramenta Expsy ProtParam, apresentando o tamanho das sequências de aminoácidos, o peso molecular, o Ponto Isoelétrico teórico (PI), número de resíduos negativos (-) e positivos (+), Coeficiente de Extinção (CE), Índice de Instabilidade (II), Índice Alifático (IA), Grande média de hidropatia (*Grand Average of Hydropathy* – GRAVY). A localização subcelular (LocSC) foi prevista pela ferramenta Plant-mPloc. 64

Tabela 12 – Parâmetros físico-químicos computados das proteínas vegetais que apresentam similaridade à proteína Shade de *Drosophila melanogaster*, por meio da ferramenta Expsy ProtParam, apresentando o tamanho das sequências de aminoácidos, o peso molecular, o Ponto Isoelétrico teórico (PI), número de resíduos negativos (-) e positivos (+), Coeficiente de Extinção (CE), Índice de Instabilidade (II), Grande média de hidropatia (*Grand Average of Hydropathy* – GRAVY). A localização subcelular (LocSC) foi prevista pela ferramenta Plant-mPloc..... 65

Tabela 13 – Parâmetros físico-químicos computados das proteínas vegetais que apresentam similaridade à proteína Shadow de *Drosophila melanogaster*, por meio da ferramenta Expsy ProtParam, apresentando o tamanho das sequências de aminoácidos, o peso molecular, o Ponto

Isoelétrico teórico (PI), número de resíduos negativos (-) e positivos (+), Coeficiente de Extinção (CE), Índice de Instabilidade (II), Grande média de hidropatia (<i>Grand Average of Hydropathy</i> – GRAVY). A localização subcelular (LocSC) foi prevista pela ferramenta Plant-mPloc.....	68
Tabela 14 – Parâmetros físico-químicos computados das proteínas vegetais que apresentam similaridade à proteína Disembodied de <i>Drosophila melanogaster</i> , por meio da ferramenta Expsy ProtParam, apresentando o tamanho das sequências de aminoácidos, o peso molecular, o Ponto Isoelétrico teórico (PI), número de resíduos negativos (-) e positivos (+), Coeficiente de Extinção (CE), Índice de Instabilidade (II), Grande média de hidropatia (<i>Grand Average of Hydropathy</i> – GRAVY). A localização subcelular (LocSC) foi prevista pela ferramenta Plant-mPloc.....	70
Tabela 15 – Parâmetros físico-químicos computados das proteínas vegetais que apresentam similaridade à proteína Phantom de <i>Drosophila melanogaster</i> , por meio da ferramenta Expsy ProtParam, apresentando o tamanho das sequências de aminoácidos, o peso molecular, o Ponto Isoelétrico teórico (PI), número de resíduos negativos (-) e positivos (+), Coeficiente de Extinção (CE), Índice de Instabilidade (II), Grande média de hidropatia (<i>Grand Average of Hydropathy</i> – GRAVY). A localização subcelular (LocSC) foi prevista pela ferramenta Plant-mPloc.....	71
Tabela 16 – Parâmetros físico-químicos computados das proteínas vegetais que apresentam similaridade à proteína Ecdysone Oxidase de <i>Drosophila melanogaster</i> , por meio da ferramenta Expsy ProtParam, apresentando o tamanho das sequências de aminoácidos, o peso molecular, o Ponto Isoelétrico teórico (PI), número de resíduos negativos (-) e positivos (+), Coeficiente de Extinção (CE), Índice de Instabilidade (II), Grande média de hidropatia (<i>Grand Average of Hydropathy</i> – GRAVY). A localização subcelular (LocSC) foi prevista pela ferramenta Plant-mPloc.....	73
Tabela 17 – Parâmetros físico-químicos computados das proteínas vegetais que apresentam similaridade à proteína 3-dehydroecdysone 3beta-reductase de <i>Spodoptera littoralis</i> , por meio da ferramenta Expsy ProtParam, apresentando o tamanho das sequências de aminoácidos, o peso molecular, o Ponto Isoelétrico teórico (PI), número de resíduos negativos (-) e positivos (+), Coeficiente de Extinção (CE), Índice de Instabilidade (II), Grande média de hidropatia (<i>Grand Average of Hydropathy</i> – GRAVY). A localização subcelular (LocSC) foi prevista pela ferramenta Plant-mPloc.	75

Tabela 18 – Parâmetros físico-químicos computados das proteínas vegetais que apresentam similaridade à proteína 3-dehydroecdysone 3alpha-reductase de <i>Spodoptera littoralis</i> , por meio da ferramenta Expsy ProtParam, apresentando o tamanho das sequências de aminoácidos, o peso molecular, o Ponto Isoelétrico teórico (PI), número de resíduos negativos (-) e positivos (+), Coeficiente de Extinção (CE), Índice de Instabilidade (II), Grande média de hidropatia (Grand Average of Hydropathy – GRAVY). A localização subcelular (LocSC) foi prevista pela ferramenta Plant-mPloc.	76
Tabela 19 – Oligonucleotídeos <i>Forwards</i> e <i>Reverse</i> s propostos com base na identidade de sequências com o gene <i>Neverland</i> de <i>Drosophila melanogaster</i> . Estatística dos oligonucleotídeos realizada por meio da ferramenta PCR Primer Stats (Disponível em: https://www.bioinformatics.org/sms2/pcr_primer_stats.html).	82
Tabela 20 – Oligonucleotídeos <i>Forwards</i> e <i>Reverse</i> s propostos com base na identidade de sequências com o gene <i>Shade</i> de <i>Drosophila melanogaster</i> . Estatística dos oligonucleotídeos realizada por meio da ferramenta PCR Primer Stats (Disponível em: https://www.bioinformatics.org/sms2/pcr_primer_stats.html).	83
Tabela 21 – Oligonucleotídeos <i>Forwards</i> e <i>Reverse</i> s propostos com base na identidade de sequências com o gene <i>Shadow</i> de <i>Drosophila melanogaster</i> . Estatística dos oligonucleotídeos realizada por meio da ferramenta PCR Primer Stats (Disponível em: https://www.bioinformatics.org/sms2/pcr_primer_stats.html).	84
Tabela 22 – Oligonucleotídeos <i>Forwards</i> e <i>Reverse</i> s propostos com base na identidade de sequências com o gene <i>Disembodied</i> de <i>Drosophila melanogaster</i> . Estatística dos oligonucleotídeos realizada por meio da ferramenta PCR Primer Stats (Disponível em: https://www.bioinformatics.org/sms2/pcr_primer_stats.html).	85
Tabela 23 – Oligonucleotídeos <i>Forwards</i> e <i>Reverse</i> s propostos com base na identidade de sequências com o gene <i>Phantom</i> de <i>Drosophila melanogaster</i> . Estatística dos oligonucleotídeos realizada por meio da ferramenta PCR Primer Stats (Disponível em: https://www.bioinformatics.org/sms2/pcr_primer_stats.html).	85
Tabela 24 – Oligonucleotídeos <i>Forwards</i> e <i>Reverse</i> s propostos com base na identidade de sequências com o gene <i>Ecdysone Oxidase</i> de <i>Drosophila melanogaster</i> . Estatística dos oligonucleotídeos realizada por meio da ferramenta PCR Primer Stats (Disponível em: https://www.bioinformatics.org/sms2/pcr_primer_stats.html).	86

Tabela 25 – Oligonucleotídeos *Forwards* e *Reverse*s propostos com base na identidade de sequências com o gene *3-dehydroecdysone 3beta-reductase* de *Spodoptera littoralis*. Estatística dos oligonucleotídeos realizada por meio da ferramenta PCR Primer Stats (Disponível em: https://www.bioinformatics.org/sms2/pcr_primer_stats.html)..... 87

Tabela 26 – Oligonucleotídeos *Forwards* e *Reverse*s propostos com base na identidade de sequências com o gene *3-dehydroecdysone 3alpha-reductase* de *Spodoptera littoralis*. Estatística dos oligonucleotídeos realizada por meio da ferramenta PCR Primer Stats (Disponível em: https://www.bioinformatics.org/sms2/pcr_primer_stats.html)..... 88

SUMÁRIO

1 CONTEXTUALIZAÇÃO	11
REFERÊNCIAS	13
2 REVISÃO DE LITERATURA	15
2.1 Interação Planta X Inseto	15
2.2 Caracterização e biossíntese de 20-Hidroxicdisona (20E).....	19
2.3 Superfamília Citocromo P450	23
2.4 Análise bioinformática e isolamento de genes.....	28
REFERÊNCIAS	34
3 IDENTIFICAÇÃO, ANÁLISE IN SÍLICO DE GENES E PROTEÍNAS E ELABORAÇÃO DE ESTRATÉGIAS DE ISOLAMENTO DE GENES POTENCIALMENTE ENVOLVIDOS NA BIOSÍNTESE E INATIVAÇÃO DE 20-HIDROXIECDISONA (20E) EM PLANTAS	42
RESUMO.....	42
ABSTRACT.....	43
3.1 Introdução	44
3.2 Material e Métodos	46
3.2.1 Ferramenta BLAST (Mineração de dados) e caracterização de proteínas	46
3.2.2 Alinhamento e filogenia de proteínas.....	47
3.2.3 Alinhamento de sequência de nucleotídeos dos genes.....	47
3.2.4 Desenho de oligonucleotídeos.....	48
3.3 Resultados	48
3.1 Identificação de proteínas similares por meio da ferramenta BLAST/NCBI (mineração de dados)	51
3.2 Caracterização físico-química de proteínas.....	63
4.3 Filogenia de proteínas	78
4.4 Desenho de oligonucleotídeos.....	82
3.4 Discussão	89

3.4.1 Proteína Neverland.....	92
3.4.2 Proteína Shade.....	93
3.4.3 Proteína Shadow.....	93
3.4.4 Proteína Disembodied	94
3.4.5 Proteína Phantom	95
3.4.6 Proteína Ecdysone Oxidase.....	95
3.4.7 Proteína 3-Dehydroecdysone 3beta-Reductase.....	96
3.4.8 Proteína 3-Dehydroecdysone 3alfa-Reductase.....	96
3.5 Conclusões	96
REFERÊNCIAS	97

1 CONTEXTUALIZAÇÃO

As plantas possuem uma rede ampla e diversificada de interações com outras espécies de seres vivos, que podem variar desde o antagonismo até o mutualismo (GIRON et al., 2018). A interação entre plantas e insetos coexiste há pelo menos 350 milhões de anos, sendo a principal interação entre os dois grupos, a herbivoria. Nem sempre as interações estão relacionadas a perdas para os vegetais, uma vez que muitas espécies de plantas dependem da polinização e dispersão de sementes mediadas por insetos (MITHÖFER; BOLAND; MAFFEI, 2018). Muitas plantas desenvolveram a habilidade de compensar a herbivoria, se recuperando e mantendo o crescimento e a boa forma após os danos causados pelos insetos (GARCIA; EUBANKS, 2018).

Sabe-se que desde o início das práticas na agricultura, diversas espécies de patógenos, animais vertebrados e invertebrados vêm causando perdas inestimáveis na área por meio da predação e competição com esses organismos considerados nocivos aos vegetais. Em especial, os insetos, que estão presentes em 3/4 de toda a fauna mundial, são os maiores causadores de perdas na agricultura atualmente, com mais de 16% de perda do total da produção global (OERKE, 2005; KUMAR; OMKAR, 2018). Em relação às culturas mais importantes, a perda varia, como por exemplo, de 50% para trigo, 80% para algodão, 31% para milho, 37% para arroz, 40% para batata e 26% para soja (NAZIR; KHAN, 2019).

Durante a herbivoria, o processo ocorre basicamente em três fases: o reconhecimento da praga, a transdução do sinal e a resposta de defesa da planta. Diversos genes são expressos como resposta ao ataque às plantas, ativando defesas pré-existente ou induzidas. Uma vasta quantidade de moléculas é produzida, como por exemplo, componentes tóxicos, como os alcaloides e glucosinolatos; bem como inibidores de vias celulares essenciais e inibidores de enzimas digestivas, como os inibidores de proteases (HEIDEL-FISCHER; MUSSER; VOGEL, 2018). Desta maneira, a planta produz esses tóxicos, conhecidos como metabólitos secundários, como uma vantagem adicional e como uma estratégia de sobrevivência (MAPLESTONE; STONE; WILLIAMS, 1992).

Dentre os compostos produzidos por plantas, a 20-hidroxicdisona (20E) vem sendo o alvo de diversas pesquisas ao redor do mundo devido ao seu grande potencial biotecnológico, medicinal e farmacêutico (FESTUCCI-BUSELLI et al., 2008). A molécula 20E é um tipo de ecdisteroide utilizado e produzido por diversas plantas como um metabólito secundário para proteção contra insetos não adaptados, que ao ingerir uma dieta rica em 20E, apresentam falhas em seu crescimento e desenvolvimento (BLACKFORD; DINAN, 1997; DINAN, 2001).

Curiosamente, o termo ecdisteroide é aplicado devido a sua função hormonal em artrópodes e nematoides, pois essa molécula é responsável pela regulação do desenvolvimento e fisiologia destes animais, cujos pulsos em determinadas fases de vida controlam a ecdise, metamorfose e diapausa (THUMMEL, 2001; URYU; AMEKU; NIWA, 2015).

Os ecdisteroides possuem uma grande variedade de formas devido a sua variação no número e na posição dos agrupamentos funcionais de hidroxila (-OH) em suas estruturas, que além dessas características, podem ainda ser conjugadas com outros compostos polares ou apolares (LAFONT et al., 2012). Atualmente estão catalogados mais de 520 tipos de ecdisteroides, segundo o último levantamento de mai/2021 na plataforma *Ecdybase.org* (LAFONT et al., 2002).

Ainda que os estudos avancem em torno da identificação, função e isolamento da molécula 20E, pouco ainda se sabe a biossíntese em plantas. Diversos estudos com a mosca das frutas (*Drosophila melanogaster*) identificaram pelo menos oito enzimas e suas moléculas intermediárias envolvidas na rota de biossíntese e inativação, entretanto ainda existem alguns passos que não estão caracterizados, conhecidos como “*Black-box*” (NIWA; NIWA, 2016). Em plantas a situação é ainda pior, somente um gene, o CYP71D443, foi avaliado como possível integrante da referida rota com função de 22-hidroxiase em raízes de plantas *Ajuga* (TSUKAGOSHI et al., 2016).

Para que se compreenda os processos biológicos e fisiológicos das plantas, como morfologia, resistência a pragas e doenças é necessário o isolamento e caracterização de genes e de rotas metabólicas de interesse (JANSSON; LOENNBORG, 1993). Por meio da bioinformática, é possível a realização de análises *in silico*, que possibilitam o estudo dos genes e proteínas possivelmente envolvidos na biossíntese e degradação de 20E em diversas espécies. Desta maneira, previamente o objeto de estudo pode ser definido e caracterizado, além de avaliar a viabilidade econômica para dar seguimentos aos estudos, economizando tempo e recursos laboratoriais.

As aplicabilidades biotecnológicas de genes e produtos gênicos dependem de um completo e amplo conhecimento dos aspectos moleculares, genéticos e fisiológicos. Deste modo, é essencial o conhecimento das rotas metabólicas e seus intermediários. Nesse contexto o objetivo deste trabalho consiste na identificação e análise *in silico* de genes e proteínas possivelmente envolvidos na rota de biossíntese e inativação de 20-hidroxiecdisona em plantas e na elaboração de estratégias de isolamentos dos genes envolvidos.

REFERÊNCIAS

- BLACKFORD, M. J. P.; DINAN, L. The effects of ingested 20-hydroxyecdysone on the larvae of *Aglais urticae*, *Inachis io*, *Cynthia cardui* (Lepidoptera: Nymphalidae) and *Tyria jacobaeae* (Lepidoptera: Arctiidae). **Journal of Insect Physiology**, v. 43, n. 4, p. 315-327, 1997.
- DINAN, L. Phytoecdysteroids: biological aspects. **Phytochemistry**, v. 57, n. 3, p. 325-339, 2001.
- FESTUCCI-BUSELLI, R. A. et al. Biosynthesis and potential functions of the ecdysteroid 20-hydroxyecdysone—a review. **Botany**, v. 86, n. 9, p. 978-987, 2008.
- GARCIA, L. C.; EUBANKS, M. D. Overcompensation for insect herbivory: a review and meta-analysis of the evidence. **Ecology**, v. 100, n. 3, p. e02585, 2019.
- GIRON, D. et al. Promises and challenges in insect–plant interactions. **Entomologia Experimentalis et Applicata**, v. 166, n. 5, p. 319-343, 2018.
- HEIDEL-FISCHER, H. M.; MUSSER, R. O.; VOGEL, H. Plant transcriptomic responses to herbivory. **Annual Plant Reviews online**, p. 155-196, 2018.
- KUMAR, B.; OMKAR, O. Insect Pest Management. In: **Pests and Their Management**. Springer, Singapore, 2018. p. 1015-1078.
- MAPLESTONE, R. A.; STONE, M. J.; WILLIAMS, D. H. The evolutionary role of secondary metabolites—a review. **Gene**, v. 115, n. 1-2, p. 151-157, 1992.
- MITHÖFER, A.; BOLAND, W.; MAFFEI, M. E. Chemical Ecology of Plant-Insect Interactions. **Annual Plant Reviews**, v. 34, p. 261–291, 2018.
- NAZIR, T.; KHAN, S.; QIU, D. Biological Control of Insect Pest. In: **Pests-Insects, Management, Control**. IntechOpen, 2019.
- NIWA, Y. S.; NIWA, R. Transcriptional regulation of insect steroid hormone biosynthesis and its role in controlling timing of molting and metamorphosis. **Development, growth & differentiation**, v. 58, n. 1, p. 94-105, 2016.
- OERKE, E. C. Crop losses to pests. **The Journal of Agricultural Science**, v. 144, n. 1, p. 31-43, 2006.
- THUMMEL, C. S. Molecular mechanisms of developmental timing in *C. elegans* and *Drosophila*. **Developmental cell**, v. 1, n. 4, p. 453-465, 2001.
- TSUKAGOSHI, Y. et al. Functional characterization of CYP71D443, a cytochrome P450 catalyzing C-22 hydroxylation in the 20-hydroxyecdysone biosynthesis of *Ajuga hairy* roots. **Phytochemistry**, v. 127, p. 23-28, 2016.

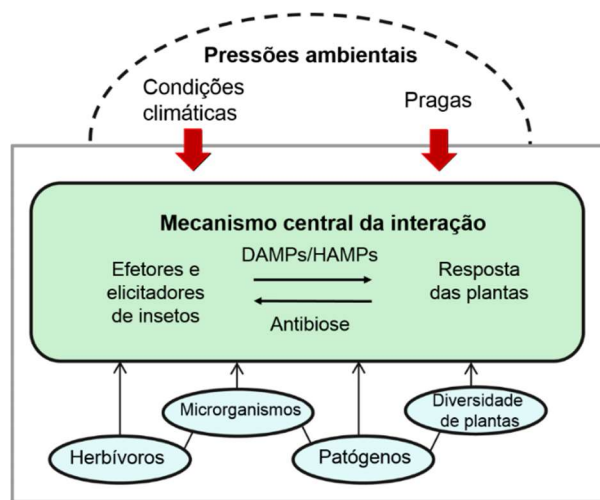
URYU, O.; AMEKU, T.; NIWA, R. Recent progress in understanding the role of ecdysteroids in adult insects: Germline development and circadian clock in the fruit fly *Drosophila melanogaster*. **Zoological letters**, v. 1, n. 1, p. 1-9, 2015.

2 REVISÃO DE LITERATURA

2.1 Interação Planta X Inseto

A interação planta-inseto (Figura 1) é antiga, esses dois grupos estão coexistindo e interagindo há mais de 350 milhões de anos. O desenvolvimento de várias relações entre esses grupos se originou desde as mais primordiais formas de plantas terrestres e insetos em todos os níveis, desde a bioquímica básica até a genética de populações (GATEHOUSE, 2002; SHARMA; MISHRA, 2021).

Figura 1 – Interação inseto planta. Mecanismos centrais e co-influência de outros organismos exibindo uma extensa rede multitrófica.



Fonte: Adaptado de Giron (2018).

Sabe-se que atualmente pelo menos um milhão de espécies de insetos são fitófagos e se alimentam de plantas. Estes animais desenvolveram diversas técnicas que permitem a localização dos vegetais e também desenvolveram artifícios para facilitar a alimentação e a oviposição usando sinais físicos ou químicos do hospedeiro (WU; BALDWIN, 2010; TOOKER; GIRON, 2020).

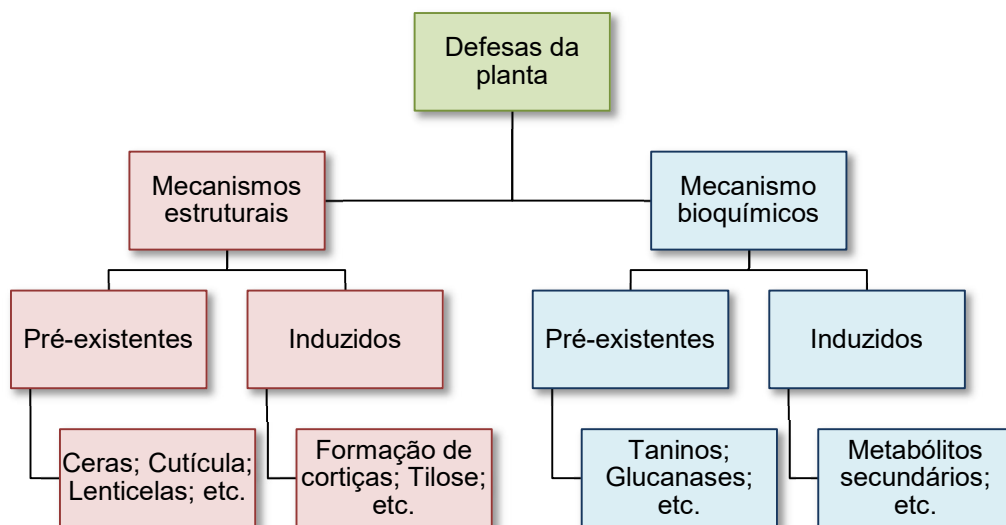
As relações entre os insetos e plantas podem ser benéficas a ambos os organismos, como a polinização e mutualismo (COHEN et al., 2020), entretanto a interação mais comum envolve a utilização das plantas como alimentação pelos insetos, bem como as defesas das plantas contra os insetos herbívoros (GATEHOUSE, 2002; SAM et al., 2020). A herbivoria atinge um total de 10 a 20% das principais culturas agrícolas em nível global, conseqüentemente, torna-se um fator limitante para a produção destinada ao mercado alimentício, o que mostra a importância do entendimento da relação entre plantas e insetos (FERRY et al., 2004).

Em relação ao nível de especiação de alimentação dos insetos, Schoonhoven et al. (2005) os classifica em dois níveis, os insetos especialistas e os generalistas. Os insetos

especialistas alimentam-se na natureza de uma ou poucas espécies de plantas da mesma família, sendo classificadas nesta categoria muitos representantes das ordens Lepidoptera, Hemiptera e Coleoptera. Os insetos generalistas se alimentam de plantas de variadas famílias. De maneira geral, aproximadamente 90% dos insetos herbívoros são restritos a plantas de uma única família taxonômica, enquanto que uma minoria é considerada generalista.

Devido os vegetais serem uma rica fonte de nutrientes para os herbívoros, foi necessário estruturar uma linha de defesa (Figura 2). As espécies vegetais desenvolveram, durante seu processo evolutivo, diversas estruturas responsáveis pelas defesas contra os ataques por insetos herbívoros e outros patógenos, sendo exemplo dessas barreiras físicas os espinhos, os tricomas e as cutículas (STIPANOVIC, 1983), que são classificadas como defesas pré-existentes. As plantas também apresentam defesa do tipo indutiva, que consiste na produção dos metabólitos secundários, os químicos tóxicos, as enzimas degradadoras de patógenos e o suicídio celular (FREEMAN; BEATTIE, 2008; MAZID; KHAN; MOHAMMAD, 2011; ZHOU; JANDER, 2021).

Figura 2 – Fluxograma da linha de defesa das plantas.



Fonte: Autor (2020).

Os metabólitos secundários são inúmeros compostos orgânicos que não possuem função direta no crescimento e desenvolvimento das plantas, mas sim na proteção contra os estresses bióticos e abióticos. Para Nahrstedt (1989), a acumulação desses compostos nas estruturas dos vegetais é resultante da necessidade de defesa química quando as plantas estão sobre ataque de insetos herbívoros.

Esses compostos orgânicos também podem ser classificados como constitutivos e indutivos, sendo que, os constitutivos são compostos pré-existentes nas plantas que funcionam como protetores aos ataques de insetos e patógenos, enquanto que os metabólitos indutivos são produzidos após o contato com os insetos com a planta hospedeira (DUFFEY; FELTON, 1989; GRAYER, 1994).

Neste contexto, é necessário que a planta saiba diferenciar um ataque por herbivoria de uma simples lesão mecânica, uma vez que há um grande gasto metabólico para ativar as defesas (WU; BALDWIN, 2010). Os trabalhos de Alborn et al. (2007), Paré e Tumlinson (1999) e Baldwin (1990) isolaram determinadas substâncias presentes nas secreções orais de insetos que eram efetivas para que a planta distinga a predação de outros tipos de lesões.

Muitos genes induzidos por lesões foram estudados em *Arabidopsis sp.* Para Reymond (2000), a expressão gênica nessas situações foram a partir de tratamentos induzidos com a presença do Ácido Jasmônico (JA), ou o precursor do JA, o Ácido Oxofitodienoico (OPDA); etileno e Ácido Salicílico (SA). Não se tem como quantificar o número exato de genes responsáveis pela defesa das plantas, sendo altamente incerto, uma vez que a maioria possui função dupla ou múltipla (REYMOND; FARMER, 1998).

O ácido jasmônico é ativado por uma via JA-dependente junto com o Etileno, que faz parte de outra cascata de reações relacionadas com a defesa de plantas. A via JA-dependente faz com que genes como PR3, PR4 e PR12, em *Arabidopsis thaliana* sejam expressos (PERVIEUX, et al. 2004). Os genes da família PR são responsáveis por desempenhar funções tanto no estresse biótico, quanto no abiótico, ou seja, são genes com grande potencial biotecnológico em função de desenvolver tolerância múltipla a estresses em espécies de interesse agrônomico e florestal (ALI et al., 2018).

Durante o percurso evolutivo diante de diversos estresses bióticos e abióticos, as plantas necessitaram desenvolver ferramentas para o rápido reconhecimento do distúrbio externo, decifrar os sinais intracelulares e também responder de forma eficiente tanto a nível celular, como a nível tecidual, uma vez que lesões mecânicas nos tecidos vegetais são consequências inevitáveis devido às atividades de herbivoria (MAFFEI; MITHÖFER; BOLAND, 2007). Nessa disputa, os insetos desenvolveram a capacidade de modificar e manipular a fisiologia do hospedeiro para seus próprios benefícios (PIETERSE; DICKE, 2007).

Os insetos que se alimentam de plantas possuem aparelhos bucais distintos para a forma de alimentação, em que usualmente classificam-se em dois grupos distintos, os mastigadores e

os sugadores (BONAVENTURE, 2012). Durante a alimentação, os insetos mastigadores tendem a liberar saliva ou secreções orais nos tecidos vegetais lesionados, mas também liberam componentes ativos, sendo eles os elicitores associados a insetos (HAEs), sendo o mais conhecido o ácido N-acilamino, ou padrões moleculares associados à herbivoria (HAMPs), que são capazes de sinalizar o ataque às plantas (MITHÖFER; BOLAND, 2008; BONAVENTURE; VANDOORN; BALDWIN, 2011; FRAGO; DICKE; GODFRAY, 2012).

Foram identificados em Lepidoptera dois tipos de elicitores presentes nas secreções orais e regurgitantes que são capazes de desencadear as defesas das plantas, sendo eles o β -glucosidase, um elicitor enzimático isolado na espécie *Pieris brassicae*, e o Volicitina N-(17-hidróxi linoleoíl)-L-Gln, um conjugado amino ácido graxo (FAC) identificado na espécie *Spodoptera exigua* (HALITSCHKE, 2001). Segundo Yoshinaga et al. (2008), nos insetos, os FACs possuem função direta na assimilação de nitrogênio no intestino a partir da regulação da glutamina.

À nível celular, a percepção das plantas durante a herbivoria desencadeia uma série de sinais, que geram a resposta celular apropriada. Essa rede de sinais é composta por diversos elementos como mensageiros não proteicos, enzimas e fatores de transcrição. O cálcio livre no citosol ($[Ca^{2+}]_c$) é um dos principais mensageiros, sendo envolvido em vários eventos de transdução para uma enorme variedade de sinais bióticos e abióticos (SANDERS et al., 2002; RANTY et al., 2016).

Além do cálcio livre, os hormônios são essenciais para a rede de sinalização das defesas contra herbivoria. A participação das vias hormonais na regulação do sinal primário em plantas era principalmente conhecida com o ácido salicílico, o ácido jasmônico e o etileno, entretanto os fitohormônios ABA, auxinas, giberelinas, citocininas e brassinosteroides também tem papéis importantes durante a ativação das defesas (PIETERSE et al., 2009; ERB; MELDAU; HOWE, 2012; CHECKER, 2018).

As plantas também possuem um grande número de moléculas que são análogas aos hormônios comumente encontrados em insetos, que são capazes de impedir a sincronização precisa do desenvolvimento destes animais, assim como suprimir a sua capacidade de sobrevivência e reprodução. Essas substâncias hormonais ativas em plantas se classificam de duas maneiras: 1) Substâncias com atividade hormonal juvenil em insetos (sesquiterpenos alifáticos ou monocíclicos); 2) Substâncias com atividade hormonal em ecdise (esteroides poli hidroxilados) (SLÁMA, 1969). Uma das principais moléculas polihidroxiladas encontradas em

plantas é a 20-hidroxiecdisona, que possui papel essencial para a ecdise em insetos (BECKSTEAD; LAM; THUMMEL, 2005).

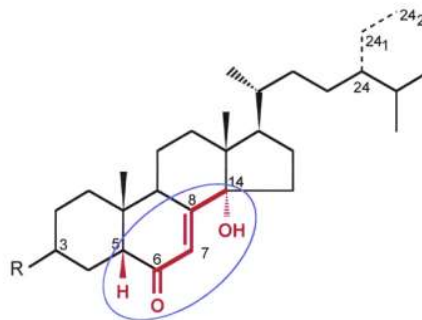
Em relação aos estudos da interação inseto-plantas, as interações com a *Arabidopsis thaliana* são as mais utilizadas como modelo para demonstrar a resistência dos vegetais aos insetos desfolhadores, bem como é também utilizada para avaliar as mudanças a nível celular e a níveis proteicos e gênicos (KROES et al., 2017). Há poucos estudos relacionados ao tema com espécies selvagens, que ainda se resumem a variedades não-domesticadas de culturas agrícolas.

2.2 Caracterização e biossíntese de 20-Hidroxiecdisona (20E)

Os esteroides poli hidroxilados, ou ecdisteroides, são uma família de elicitores essenciais para a classe Insecta, pois devido a presença de um exoesqueleto composto por quitina, para que os organismos possam crescer e se desenvolver, é necessário mudanças periódicas (ou ecdises) guiadas por esses hormônios (WARREN; GILBERT, 1998).

Os ecdisteroides fazem parte de uma família composta de mais de 300 membros de esteroides com uma particularidade estrutural em comum, uma junção cis (5β -H) dos anéis A e B, um cromóforo 7-eno-6-ona e uma junção trans (14α -OH ou H) dos anéis C e D (Figura 3), além desta semelhança estrutural, os ecdisteroides compartilham a mesma função biológica, pois atuam no processo de ecdise (LAFONT et al., 2012). Além dos insetos, estas estruturas químicas podem ser encontradas em fungos (VOKÁČ et al., 1998), plantas (ADLER; GREBENOK, 1999; CHADIN et al., 2003) e outros artrópodes (LACHAISE et al., 1989; WOLLENBERGER; DINAN; BREITHOLTZ, 2005; NIWA; NIWA, 2014).

Figura 3 – Estrutura em comum dos ecdisteroides.



Fonte: Lafont et al., 2012.

Os ecdisteroides, além de permitir a ecdise, possuem outras funções de grande importância em insetos, pois fazem parte dos processos biológicos de ovogênese (SWEVERS; IATROU, 2009), de controle de metamorfose (PAN; CONNACHER; O'CONNOR, 2020), além de outras funções hormonais atuando na resposta imune e diferenciação (SPINDLER et

al., 2009), ou seja, atuam em todas as fases de vida do organismo, da embriogênese a fase reprodutiva e fase adulta.

Existe uma grande diversidade química de ecdisteroides resultantes das variações que podem ocorrer com o número e a posição dos agrupamentos funcionais de hidroxila (-OH) nas estruturas, que podem ser isoladas ou conjugadas com outros vários compostos polares ou apolares (LAFONT et al., 2012). Segundo os últimos dados atualizados em maio de 2021, existem 527 ecdisteroides registrados que compartilham a mesma estrutura base (<http://ecdybase.org/>).

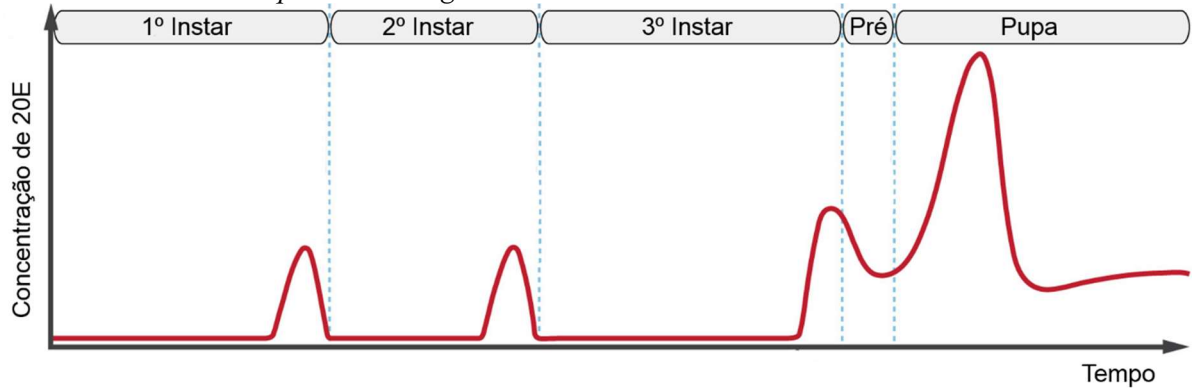
Dentre os hormônios responsáveis pela ecdise, o mais atuante em insetos é a 20-hidroxiectidisona (20E) (SVOBODA, 1997), este hormônio é um zooecdisteroide encontrado em insetos e em algumas plantas, cujas características permitem aplicabilidade em diversas áreas, como na agroquímica, biotecnologia, na farmacologia e para fins medicinais (FESTUCCI-BUSELLI et al., 2008).

O hormônio 20E difere dos outros esteroides devido a sua grande capacidade hidrofílica, em decorrência da sua maior quantidade de grupos hidroxilas em sua estrutura. Esse hormônio em insetos (Figura 4), é excretado pelas glândulas protorácicas com a função principal de regulação das mudanças no desenvolvimento e também da metamorfose (SLAMA et al., 1996), entretanto alterações desordenadas da concentração de 20E durante o desenvolvimento são letais para o inseto (FESTUCCI-BUSELLI et al., 2008). Para que ocorra a biossíntese deste hormônio na glândula protorácica, o colesterol e os fitoesteroides, provenientes da alimentação do inseto, atuam como precursores (IGARASHI et al., 2018). A biossíntese em insetos depende da regulação pela proteína *Neverland* do domínio Rieske e das enzimas da classe dos citocromos P450, ambas codificadas pelos genes “Halloween” (SCHUMANN et al., 2018).

Para a biossíntese do hormônio 20E em *Drosophila melanogaster* (Figura 5), o primeiro passo ocorre por meio da enzima *Neverland* (NVD), uma enzima do tipo Oxigenase Rieske, que catalisa a conversão do colesterol em 7-deidrocolesterol (7dC) (YOSHIYAMA-YANAGAWA et al., 2011). Em seguida, o 7dC é convertido em 5 β -ketodiol (25,22,2-trideoxiectidisona, KD), entretanto as reações contidas nesta conversão ainda não são caracterizadas, conhecidas pelo termo “*Black-Box*” (SAITO et al., 2016). A partir da conversão em 5 β -ketodiol, sucessivas hidroxilações em posições específicas em cada molécula intermediária são feitas por meio das enzimas codificadas pelos genes “*Halloween*”, sendo eles:

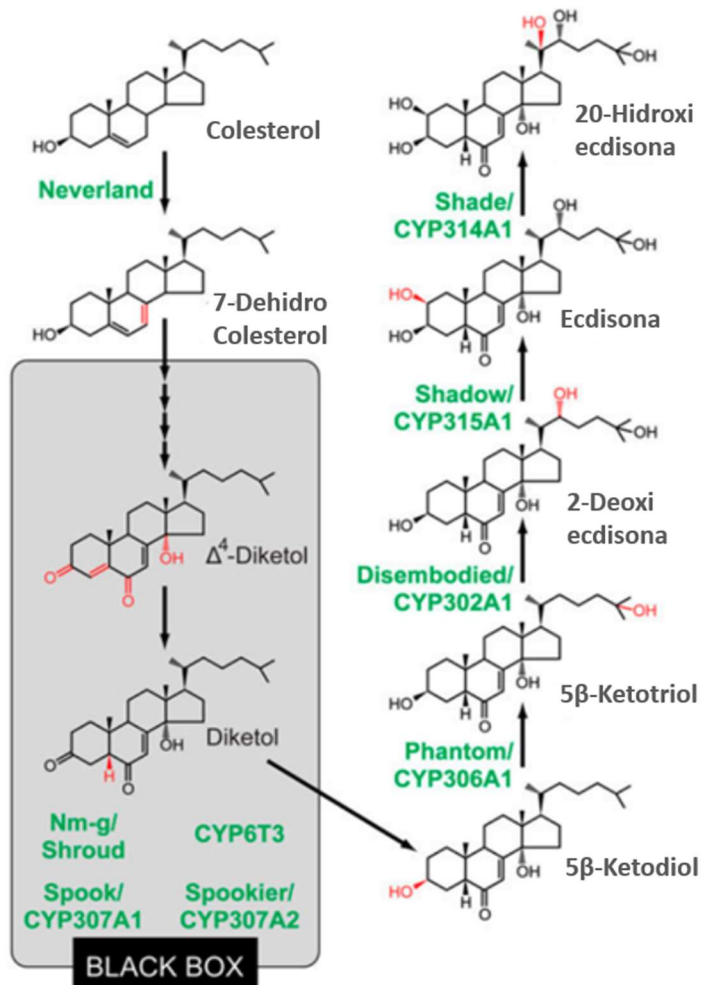
CYP306A2 (*Phantom*), CYP302A1 (*Disembodied*), CYP315A1 (*Shadow*) e CYP314A1 (*Shade*) (GILBERT; WARREN, 2005). Esses genes são essenciais para a rota de biossíntese do hormônio 20E.

Figura 4 – Alteração nas concentrações do hormônio 20-hidroxiecdisona durante o desenvolvimento da mosca dos frutos *Drosophilla melanogaster*.



Fonte: Adaptado de Schumann et al. (2018).

Figura 5 – A rota de biossíntese da molécula 20-Hidroxiecdisona e as enzimas ecdisteroidogênicas atuantes na rota em insetos. A caixa-preta (Black Box) representa os passos ainda desconhecidos na rota.



Fonte: Adaptado de Niwa e Niwa (2014).

Os genes *Phantom*, *Disembodied* e *Shadow* catalisam as reações que produzem a molécula Ecdisona (E) nas glândulas protorácicas em *D. melanogaster* e então liberado na hemolinfa. Nos tecidos periféricos às glândulas protorácicas, a ecdisona é convertida em 20-hidroiecdisona pelo gene *Shade*, também conhecido como ecdisona 20-monooxigenase (GILBERT; WARREN, 2005; LAFONT et al., 2005; FORTINI et al., 2020).

Dada a importância deste hormônio nos insetos, é necessário um controle correto durante os eventos biológicos em que ele atua, sendo assim a sua síntese é extremamente controlada para que não haja danos ao indivíduo durante o seu ciclo de vida (NIWA; NIWA, 2014). O mecanismo da biossíntese dos ecdisteroides a nível molecular é estudada tanto em insetos, como em nematoides (BARKER; REES, 1990; CHITWOOD, 2020; YAMANAKA, 2021), entretanto a rota ainda não é totalmente elucidada.

Diferente dos insetos, que necessitam dos precursores provenientes da alimentação para produzir 20E, as plantas já possuem a capacidade de sintetizar os esteróis (HEFTMANN, 1975; ADLER; GREBENOK, 1999; BATTH et al., 2020; GROENEVELD, 2020). Em plantas, o experimento com espinafre (*Spinacia oleracea* L.) mostrou que a biossíntese de 20E é altamente regulada por uma via com o precursor sendo o ácido mevalônico produzido em folhas mais velhas e então, após a síntese, distribuído pela planta (BAKRIM et al., 2008; BOKOV et al., 2020).

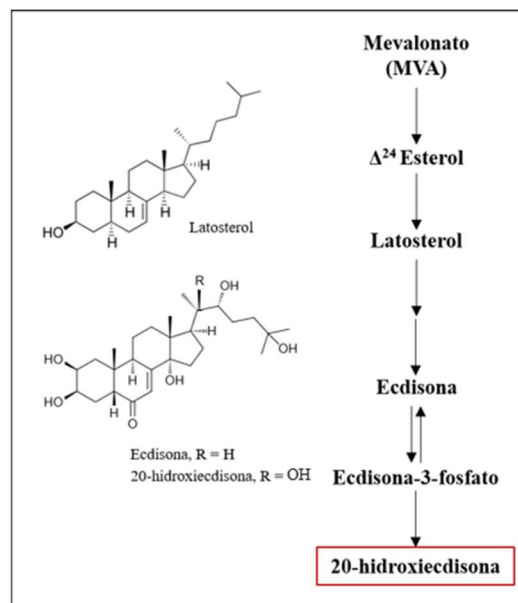
Os fitoecdisteroides (com ênfase no 20E), assim como os zooecdisteroides, são moléculas formadas a partir dos fitosteróis contendo de 27 a 29 carbonos em sua estrutura, que foram modificadas para gerar uma junção em anel A/B-cis, uma cetona a,b-insaturada em anel B e posterior inclusão de grupos funcionais de hidroxila à estrutura (SREEJIT et al., 2019). A função em plantas ainda não está totalmente elucidada. Para Dilan (2009), os fitoecdisteroides assumem função de defesa a predadores como compostos tóxicos quando ingeridos por insetos fitófagos. Essa colocação se sustenta devido a função destes hormônios durante a ecdise nos artrópodes, além disso, outros de estudos identificaram problemas no desenvolvimento quando alguns insetos consomem altas doses destes compostos.

Não somente a função em plantas ainda continua em suposições, a rota de biossíntese de 20E e os genes envolvidos em plantas ainda não é conhecida, mas assume-se que as enzimas da classe dos citocromos P450 estão envolvidas. Recentemente foi proposto a participação da enzima CYP71D443 de *Ajuga reptans*, uma pequena planta herbácea da família Lamiaceae. Esta enzima possui a capacidade de hidroxilação do C-22 com o substrato 5 β -cetona, gerando

o produto 22-hidroxi-5 β -ketona e posteriormente convertido em 20E em raízes fasciculadas de *A. reptans* (TSUKAGOSHI et al., 2016).

Nas plantas, os processos de biossíntese dos ecdisteroides ocorre no citosol a partir do precursor mevalonato (TARKOWSKA, 2016) (Figura 6), mas também é possível ocorrer nos plastídeos, uma vez que os compostos necessários para síntese de triterpenoide podem ser tanto produzidos via rota do mevalonato, quanto pela rota 2-C-metil-D-eritritol 4-fosfato nos plastídios (ROHMER, 1999).

Figura 6 – Rota de biossíntese de 20-hidroxiecdisona a partir do Mevalonato (MVA) em espinafre (*Spinacia oleracea* L.).



Fonte: Adaptado de Adler e Grebenok (1995).

2.3 Superfamília Citocromo P450

Os citocromos P450 (CYP) são uma superfamília de proteínas enzimáticas do tipo heme-tiolato monoxigenase extremamente diversificadas, cujos genes codificantes são amplamente encontrados em todos os organismos vivos, sejam plantas, animais, bactérias e fungos; apresentando algumas poucas exceções (OMURA, 1999; NELSON; GOLDSTONE; STEGEMAN, 2013). Nos eucariotos, o local de ação das enzimas CYP é no retículo endoplasmático, além disso necessitam de uma flavoproteína NADPH citocromo P450 redutase para completar seu ciclo catalítico (SCOTT; WEN, 2001).

Embora existam estudos desde a década de 50 (KLINGENBERG, 1958) referenciando a existência de uma molécula desconhecida que não apresentava quaisquer semelhanças com outras proteínas e era capaz de apresentar um pico de absorção óptica de 450 nm, somente quase

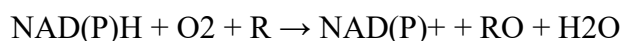
uma década depois, Omura e Saito (1962) publicaram um artigo confirmando a existência de um novo Citocromo P450 isolado em microsoma de fígado de ratos.

As limitações de técnicas de metodologia na bioquímica ainda eram determinantes, na década de 80, um quantitativo de 78 proteínas P450 divididas em 14 famílias era registrado (NEBERT; NELSON; FEYEREISEN, 1989). Nos dias atuais, estima-se a existência de pelo menos 21.000 genes codificantes destas proteínas, sendo somente em plantas, mais de 7.000 sequências conhecidas (NELSON, 2009, balanço atualizado em 2013).

Em relação a forma de identificação e classificação destas enzimas, utiliza-se da abreviação CYP seguida de um número que determina a família. Para a classificação a nível de família, aplica-se um mínimo de 40% de identidade nas sequências. As letras são utilizadas para representar uma subfamília específica, quando existem mais de 55% de identidade e, por último, a utilização de um número que discrimina o gene individual (HANNEMANN et al., 2007). Em relação às famílias, desde 1993 são classificadas com numerações de acordo com o tipo de organismo. Para as plantas, assume-se uma cobertura de CYP71 a CYP99 e CYP701 a CYP772 (NELSON, 2009), esta quebra de sequências ocorre devido o grande quantitativo de genes codificantes de proteínas que ainda estão sendo descobertos.

A função do citocromo P450 foi descoberta por meio de estudo comparativo, segundo Omura (1999) no ano de 1963 um grupo de cientistas estudavam a hidroxilação do C-21 do 17-hidroxiprogesterona catalisado pelo microsoma do córtex adrenal, conhecido por ser inibido pelo monóxido de carbono. A partir dessa pesquisa, alguns anos mais tarde, observou-se que o pigmento de ligação monóxido de carbono microsomal era uma hemoproteína, tendo características semelhantes de espectro ao citocromo P450, confirmando a sua atividade de hidroxilação.

Os citocromos P450 são enzimas heme-dependentes capazes de formar diversos produtos, com funções mistas de oxidorredutase que necessitam dos cofatores NADPH e/ou NADH para realizar a transferência de elétrons e assim clivar o O₂ para produzir um produto orgânico funcional e uma molécula de água (SCHULER, 1996). De maneira geral, a reação de monoxigenase por uma CYP em um substrato “R” ocorre da seguinte maneira (GUENGERICH, 2018):



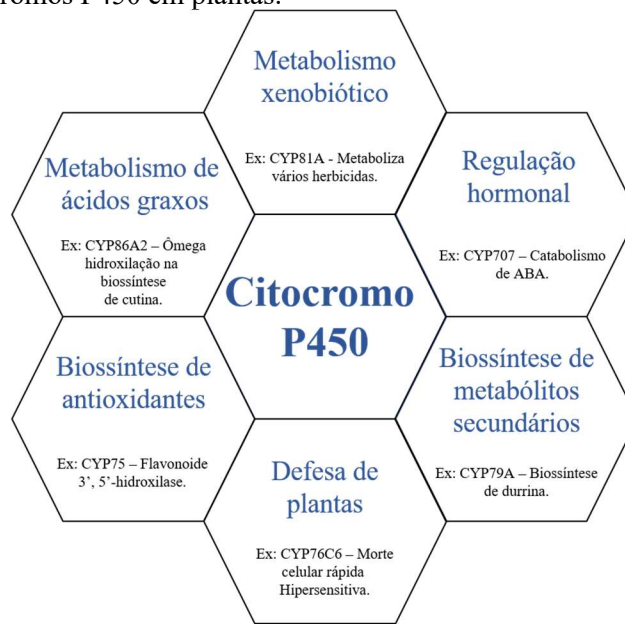
Sabe-se que atualmente as enzimas P450 estão envolvidas em uma série de reações bioquímicas essenciais para o organismo, sendo que pelo menos 95% de todas as reações de

oxidação/redução de químicos orgânicos estão relacionadas com as enzimas do citocromo P450, o que gera um grande interesse nas mais diversas áreas, como a biorremediação, agricultura, bacteriologia e desenvolvimento de medicamentos (GUENGERICH; WATERMAN; EGLI, 2016).

Especificamente em plantas (Figura 7), as enzimas pertencentes ao complexo do citocromo P450 estão envolvidas em reações na biossíntese de antioxidantes, fito-hormônios e principalmente, metabólitos secundários, como por exemplo, nas reações de metabolismo de lipídios, fenóis, terpenoides, diterpenos e alcaloides, ou seja, possui um papel essencial para a defesa das plantas durante o estresse (BOLWELL; BOZAK. ZIMMERLIN, 1994; BATHE; TISSIER, 2019; PANDIAN et al., 2020). As plantas utilizam os metabólitos secundários para agir diretamente no ato da herbivoria, como uma toxina direta ou funcionando também como inibidor digestivo, dessa maneira a planta possui mais chances de sobrevivência. Observa-se que as enzimas do citocromo P450 são extremamente necessárias para a produção destes compostos (MORANT et al., 2003; CALLA et al., 2017).

A grande variedade de genes codificadores de P450 é em decorrência da estratégia de sobrevivência das plantas, o que permite vantagens na competição evolucionária contínua e durante o mutualismo com outros organismos. Este autor classifica as famílias das proteínas do complexo P450 em quatro categorias de acordo com a função geral, sendo elas: Categoria I, reações essenciais conservadas em todo o reino vegetal; Categoria II, referente ao metabolismo central em todas as plantas terrestres; Categoria III, metabolismo essencial emergido durante a evolução de angiospermas; Categoria IV, produção de metabólitos secundários de acordo com cada espécie (MIZUTANI, 2012).

Figura 7 – Função dos citocromos P450 em plantas.



Fonte: Adaptados de Pandian et al. (2020).

A produção de metabólitos secundários, que utilizam as enzimas do citocromo P450, é de grande importância, uma vez que as plantas são organismos sésseis sem sistema imunológico adaptativo ou sistema de células circulantes, a adaptação com produção de químicos que funcionam como toxinas diretas ou inibidores digestivos é essencial para a defesa e proteção contra herbivoria ou ataque de patógenos (CALLA, 2021).

Ressalta-se que em relação à interação planta-inseto, os citocromos P450 são componentes essenciais entre esses dois grupos diversificados de espécies, uma vez que essas enzimas definem o sucesso dos compostos de defesa sintetizados pelas plantas, bem como as toxinas vegetais metabolizadas pelos insetos (SCHULER, 1996; CALLA et al., 2017; CALLA, 2021). Essa versatilidade agregada ao sucesso evolutivo das estratégias de defesa entre as espécies é decorrente da flexibilidade bioquímica de reconhecimento de múltiplos substratos (ZUNJARRAO, 2020).

Ainda que estes grupos de organismos tenham coevoluído, o número de enzimas P450 é muito maior em plantas devido constantes ciclos de duplicação e divergência de cópias reiteradas dos genes, resultando em inúmeras P450 em substratos específicos (SCHULER, 2011). Embora exista uma grande variedade de enzimas desta superfamília capaz de catalisar reações de diversos substratos, as estruturas são altamente conservadas (DARABI; SEDDIGH; ABASHAHR, 2017), como por exemplo, pode-se citar os motivos AG×DT (I-helix), KETRL (K-helix) e PERF/W (BAK et al., 2011; VASAV; BARVKAR, 2019).

O ciclo catalítico desta enzima é demonstrado na Figura 8. O ferro III é ligado ao citocromo P450 por meio da ligação tiolato derivada de um resíduo altamente conservado em enzimas dessa família de cisteína, além de outros resíduos que flanqueiam essa região (DARABI; SEDDIGH; ABASHAHR, 2017; GUENGERICH, 2018). O ciclo catalítico descrito a seguir é derivado dos estudos de Meunier, De Visser e Shaik (2004) e Shaik e De Visser (2005):

a) O ciclo inicia-se a partir do estado de descanso da enzima, composto pelo complexo férrico FeIII e uma molécula de água como ligante distal.

b) A entrada do substrato (R-H) no sítio ativo da enzima remove a molécula de água e gera uma espécie férrica cinco-coordenada, resultando na remoção do Fe do plano porfirina. Desta maneira, muda-se o Fe para um estado de spin alto e o complexo passa a ser um receptor de elétrons.

c) Ocorre a transferência de um elétron de origem do domínio redutase formando o complexo ferroso HS (FeII).

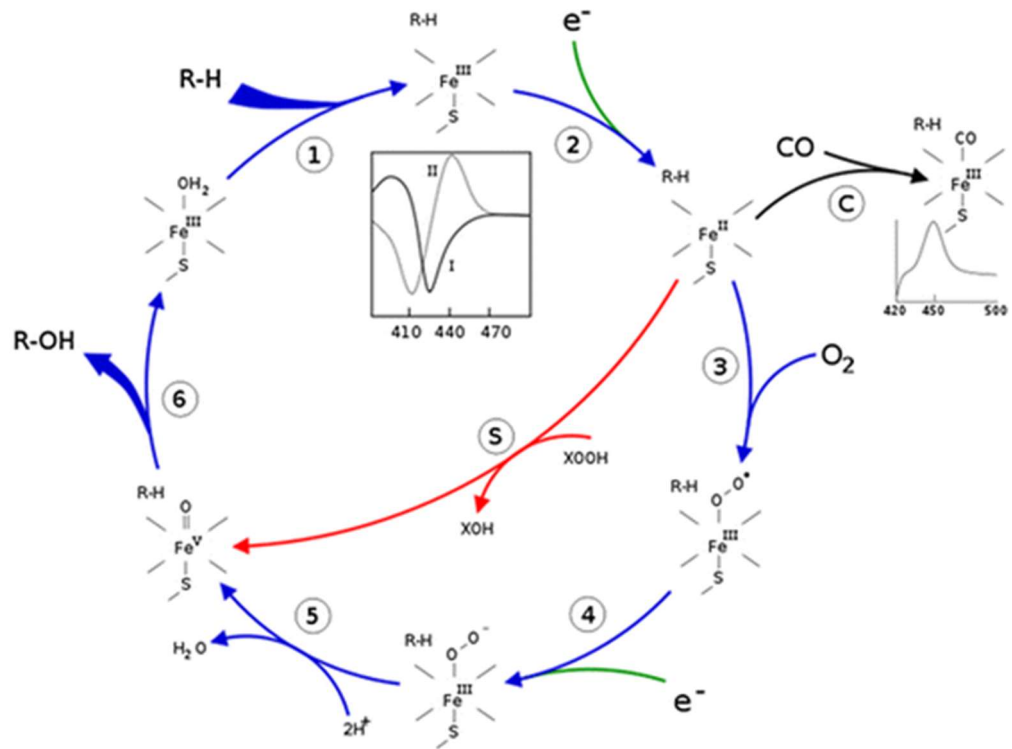
d) O resultado é um bom ligante de O₂ (porfirina ferrosa) e conseqüentemente gera a ligação do oxigênio molecular para a produção do complexo dioxigênio-ferroso LS.

e) A espécie formada recebe novamente um elétron da redutase e dá origem à espécies férricas-dioxo duas vezes reduzidas.

f) Esse complexo férrico-dioxo passa por protonação e transforma-se em complexo peróxido férrico (ou Cpd 0).

g) Ocorre uma segunda protonação e é liberado uma molécula de água, gerando uma espécie reativa conhecida como Cpd I, que por sua vez, transfere o oxigênio para o substrato, que é substituído por uma outra molécula de água, iniciando novamente o ciclo.

Figura 8 – Ciclo catalítico do citocromo P450.



Fonte: Darabi (2017).

2.4 Análise bioinformática e isolamento de genes

O campo das “ômicas” implica na análise das interações de objetos de informações biológicas e é aplicado como um sufixo a termos moleculares, como por exemplo, a genômica, proteômica ou transcriptômica (<http://omics.org>). O termo “ômica” tem sua origem a partir do sufixo latino “ome”, que significa “massa” ou “muito”, neste contexto, um estudo envolvendo uma ômica envolve uma grande quantidade de dados (LAY JR. et al., 2006).

A partir das ômicas, o acesso a conjunto de dados em grande escala de genomas, transcriptomas e proteomas completos transformou a biologia e trouxe novas abordagens para avançar os estudos nos processos biológicos, além de diminuir o tempo e o custo para gerar esses gigantescos bancos de dados (MISRA et al., 2019). A primeira ômica a surgir foi a voltada para o estudo completo do genoma, a genômica, que surgiu num contexto em que as pesquisas focavam apenas em variantes individuais ou poucos genes envolvidos (HASIN; SELDIN; LUSIS, 2017). As sequências biológicas como DNA, RNA e sequências de aminoácidos de proteínas são os objetos fundamentais a nível molecular, quando se trata de plantas, os genomas pioneiros sequenciados e disponibilizados para a comunidade científica em alta qualidade foram de *Arabidopsis thaliana* e arroz (RHEE; DICKERSON; XU, 2006).

Com o advento de novas tecnologias na biologia molecular e a alta taxa de produção de dados biológicos, a bioinformática se tornou indispensável para a pesquisa científica. Isso se dá pelo fato de os computadores lidarem com facilidade a uma enorme quantidade de dados e complexos dinâmicos observados na natureza (LUSCOMBE; GREENBAUM; GERSTEIN, 2001). Segundo Schneider e Orchard (2011) os bioinformatas usam a estatística computacional para estudar as extensivas pesquisas nos campos das ômicas, realizando a busca em bancos de dados biológicos e a comparação entre sequências de nucleotídeos de genes e de aminoácidos de proteínas obtidas em análises para diversas finalidades, como por exemplo homologia de proteínas (SAIGO et al, 2011), ontologia gênica (LAMBRIX; HABBOUCHE; PEREZ et al., 2003), entre outros. A utilização de computadores e softwares para realizar experimentações e simulações é chamada de análise *in silico*.

As análises *in silico* antecede as análises realizadas em bancadas (*in vivo*) e são importantes para definição e caracterização do objeto estudado, além de definir se será viável economicamente dar seguimentos aos estudos, economizando tempo e recursos laboratoriais. Segundo Dana et al. (2020), essas análises são utilizadas para, por exemplo, caracterizar as propriedades e estruturas de proteínas, podendo prever propriedades físico-químicas, estruturas, estabilidade entre outros, a partir de ferramentas de bioinformática e servidores computacionais.

De acordo com Zhang, Zhou e Wang (2010) novos termos científicos são sempre complexos de definir devido à grande variabilidade de aplicação. De uma maneira geral, a bioinformática é aplicada aos tópicos relacionados com armazenamento dos dados de DNA e modelagem matemática de sequências biológicas, ou seja, é definida como aplicação de técnicas computacionais aplicadas a organização e interpretação da vida a nível celular e molecular. As análises computacionais também envolvem a análise de dados por métodos matemáticos, estatísticos e algoritmos assistidos por computadores.

Diversos *softwares* e plataformas *on line* foram desenvolvidas para a integração e análise desses grandes bancos de dados. Para que os dados biológicos possam ser analisados e interpretados é necessário familiaridade com sistemas computacionais, além disso, tem-se como principal desafio a armazenagem, a pesquisa e recuperação desses dados, etapas importantes para a análise geral. Neste contexto, uma grande variedade de ideias algorítmicas é aplicada a uma ampla variedade de sistemas biológicos (BERGER; PENG; SINGH, 2013).

Além dos trabalhos envolvendo genes e genômica, a bioinformática é muito utilizada nos estudos de proteômica, que vem ganhando gigantescos bancos de dados nos últimos anos,

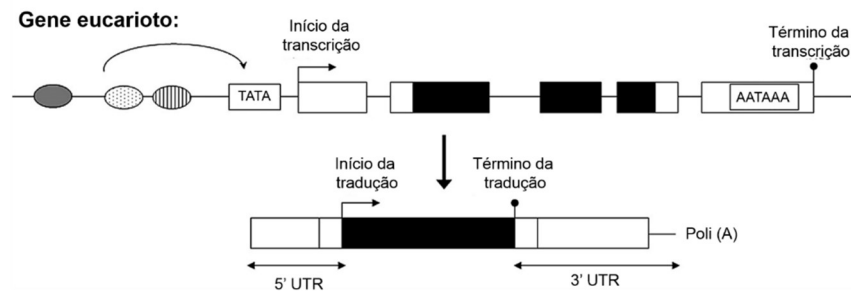
uma vez que as proteínas são estruturas complexas que possuem grande interesse em várias áreas em estudos biológicos, como sítio direto de mutagênese, homologia e estruturas bases de inibidores específicos (ARNOLD et al., 2006). Em um balanço geral dos principais bancos de dados, o *Protein Data Bank* tem registradas 188.929 proteínas cadastradas (atualização abr./2021) (<http://www wwpdb.org>), enquanto que o banco de dados *UniProtKB* possui 565.254 (atualização jun./2021) (<https://www.uniprot.org/>).

A busca por homologia entre as sequências de aminoácidos de proteínas com as diversas ferramentas de bioinformática existentes permite a inferências da função, estrutura e evolução de genes e genoma. Entretanto é preciso ter critérios e avaliações profundas com os resultados, pois nem sempre alta similaridade corresponde a homologia entre proteínas, principalmente quando se trata de resultados com valores não significativos. Além disso, a homologia entre proteínas não significa conservação de funções (RHEE; DICKERSON; XU, 2006).

As sequências de aminoácidos são importantes para diversas finalidades, principalmente para a genômica funcional, uma vez que a classificação dos genes codificadores de proteínas em famílias é possível a partir da estrutura, função e evolução das proteínas que codificam. As famílias são classificações de conjuntos de genes evolutivamente relacionados e presentes em várias espécies diferentes e com funções biológicas semelhantes. Os genes contidos dentro de uma determinada família também são chamados de genes homólogos (MARTINEZ, 2011).

Os genes são estruturas compostas de segmentos contínuos de ácidos nucleicos que são capazes de codificar diversos produtos em um organismo vivo (Figura 9). Embora exista uma rica diversidade de espécies, elas compartilham a mesma estrutura básica (LYNCH, 2005). Pearson (2006) destaca três conceitos na história da definição do gene: Na genética clássica, o gene era visto como um conceito abstrato, capaz de conduzir a hereditariedade de um progenitor a uma prole, entretanto a partir dos avanços bioquímicos, o gene assumiu características de codificar enzimas ou proteínas, uma por gene. Atualmente, com os estudos moleculares, os genes se tornaram estruturas físicas reais, com sequências de nucleotídeos convertidas em uma fita única, chamada de RNA mensageiro, usadas como base para construção de proteínas, ou assumir estruturas terciárias ativas e funcionais.

Figura 9 – Estrutura geral do gene de um organismo eucarioto. Os genes, após a transcrição, apresentam grandes regiões não traduzidas (sigla UTR, em inglês) e uma extensa cauda poli-A, proveniente de eventos pós-transcricionais.



Fonte: Adaptado de Lynch (2005).

O isolamento e a caracterização de genes são cruciais para compreender, em uma planta, por exemplo, o seu desenvolvimento, morfologia, resistência a patógenos ou herbívoros, resposta ao estresse, dentre outros (JANSSON; LOENNEBORG, 1993). Segundo Martinez (2011) algumas famílias gênicas são dinâmicas durante a evolução e possuem membros de genes espécie-específicos, entretanto outros são extremamente conservados e ortólogos, ou seja, compartilham um ancestral em comum que sofreu divergência durante especiação. No caso, esses genes ortólogos são úteis para a identificação de proteínas não-caracterizadas a partir da busca por identidades de sequências em ferramentas bioinformáticas do tipo BLAST ou FASTA.

As famílias de genes e proteínas podem ser representadas em árvores filogenéticas, permitindo inferir acerca das relações evolucionárias. As análises filogenéticas, usualmente iniciada a partir de alinhamento entre as sequências, também podem ser utilizadas em trabalhos de genômicas comparativas, predição de função gênica e inferir da transferência lateral entre os genes (DOOLITTLE, 1999). A árvore filogenética possui esse nome devido a semelhança com uma árvore, além disso os termos que se utilizam para este diagrama também remetem a partes de uma árvore (raiz, ramo, nó e folhas). Os nós terminais, ou folhas, representam os táxons existentes, também chamados de unidades taxonômicas operacionais (OTU, em inglês), enquanto os nós internos são chamados de unidades taxonômicas hipotéticas (HTU, em inglês) dos OTUs. Os chamados *clusters* ou grupos de taxa que possuem um ramo em comum, eles têm origem monofilética, se já não há ramos compartilhados, são de origem parafilética. O padrão dos ramos é chamado de topologia da árvore filogenética (LEMEY; SALEMI; VANDAMME, 2009).

Os estudos que envolvem caracterização e isolamento de genes e suas famílias é relativamente recente e ainda há muitas dificuldades em desenvolver técnicas. Nos anos 90,

para isolar utilizando a clonagem era necessário informações completas do produto gênico, ou seja, o transcrito ou o polipeptídeo, pois a partir dessas sequências é possível identificar o cDNA ou clones genômicos em uma biblioteca de DNA, mas para a maioria dos genes de plantas com interesse agrônomico não se tem informações de seus produtos polipeptídicos, dificultando o processo (JANSSON; LOENNEBORG, 1993). Devido ao processo evolutivo das angiospermas ter ocorrido recente em escala evolutiva, muito dos genes isolados de angiospermas podem ser usados para isolar genes correspondentes de outras plantas por hibridização heteróloga. O grupo das angiospermas inclui quase todas as plantas economicamente úteis (GIBSON; SOMERVILLE, 1993).

Em se tratando da interação planta-inseto, diversos genes de famílias não homólogas de inibidoras de proteases já foram isolados e caracterizados (LASKOWSKI JR. et al., 1987). De acordo com Ryan (1989) os inibidores de protease possuem importante função durante a herbivoria, pois esses inibidores são capazes de proteger as proteínas de uma célula, fluidos ou tecidos de enzimas proteolíticas externas.

A liberação de voláteis de planta induzidos por herbivoria (sigla HIPVs, em inglês) compõe o mecanismo de defesa dos vegetais durante o ataque de insetos, que podem atrair insetos predadores dos herbívoros, bem como repelir os insetos que a atacam. Os terpenoides fazem parte de uma grande e diversa família de HIPVs, sendo estes sintetizados pela Terpeno Sintase (sigla TPS, em inglês), sendo que já foram isolados 12 genes TPS (números de acesso *GenBank* GhTPS4-15) por meio de *primers* e amplificação por PCR, e oito proteínas expressas em *Escherichia coli* revelaram ação catalítica para esses GhTPS para *Gossypium hirsutum* e *G. raimondii* (HUANG et al., 2018).

Qi et al. (2016) isolaram, a partir da RT-PCR e clonagem total do cDNA, e caracterizaram o gene OsJMST1 em arroz (*Oryza sativa*) após infestação de larvas de *Spodoptera frugiperda*. Esse gene, expresso durante herbivoria, codifica a metiltransferase que é capaz de catalisar a metilação do fitohormônio Ácido Jasmônico. Segundo Howe (2001), o ácido jasmônico está relacionado na cascata de sinalização celular durante a herbivoria com outros compostos capazes de induzir a expressão de diversos genes.

Embora a maior parte dos trabalhos envolvam espécies vegetais consideradas modelos, Luck et al. (2016) isolaram quatro genes (CYP79) responsáveis pelos voláteis induzidos pelo Ácido Jasmônico em duas espécies de *Erythroxylum sp.*, sendo uma sul-americana e outra de origem africana. Esses quatro genes foram expressos (expressão heteróloga) em leveduras para a caracterização bioquímica. Existem várias metodologias para realizar o isolamento de um

gene, entretanto um grande problema para as técnicas de isolamento está relacionado com a pureza da sonda de ácido nucleico, uma vez que o gene a ser detectado é apenas uma pequena fração do DNA (BROWN, 1974).

Ainda que haja diversas técnicas para isolar os genes, todas elas exploram uma ou mais característica que definem os genes, como: os genes possuem uma estrutura primária definida, ou seja, sequência de nucleotídeos; os genes estão localizados em um lugar específico do genoma; os genes codificam o RNA com um padrão de expressão peculiar; os genes possuem funções no organismo. Além disso, existem técnicas generalistas, usadas para muitas espécies de plantas, enquanto outras, apenas para poucas espécies (GIBSON; SOMERVILLE, 1993).

Muitos genes foram isolados e caracterizados a partir de organismos que possuem seu genoma completo sequenciado, como o caso do arroz, do sorgo e da *Arabidopsis*, o que ainda é um desafio para as plantas selvagens, consideradas não-modelos, pois dificilmente possuem disponível o seu genoma em qualidade e integridade. Segundo Makkar et al. (2019), o conhecimento molecular da estrutura e função, além da clonagem dos genes, é base para o sucesso de programas de melhoramento para produção de vegetais superiores resistentes a pragas e doenças.

REFERÊNCIAS

- ADLER, J. H.; GREBENOK, R. J. Occurrence, biosynthesis, and putative role of ecdysteroids in plants. **Critical reviews in biochemistry and molecular biology**, v. 34, n. 4, p. 253-264, 1999.
- ALBORN, H. T. et al. An elicitor of plant volatiles from beet armyworm oral secretion. **Science**, v. 276, n. 5314, p. 945-949, 1997.
- ALI, S. et al. Pathogenesis-related proteins and peptides as promising tools for engineering plants with multiple stress tolerance. **Microbiological research**, v. 212, p. 29-37, 2018.
- APWEILER, R. et al. UniProt: the universal protein knowledgebase <http://www.uniprot.org>. **Nucleic acids research**, v. 32, n. suppl_1, p. D115-D119, 2004.
- ARNOLD, K. et al. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. **Bioinformatics**, v. 22, n. 2, p. 195-201, 2006.
- BAK, S. et al. Cytochromes P450. **The Arabidopsis Book/American Society of Plant Biologists**, v. 9, 2011.
- BALDWIN, I. T. Herbivory simulations in ecological research. **Trends in Ecology & Evolution**, v. 5, n. 3, p. 91-93, 1990.
- BARKER, G. C.; REES, H. H. Ecdysteroids in nematodes. **Parasitology Today**, v. 6, n. 12, p. 384-387, 1990.
- BATHE, U.; TISSIER, A. Cytochrome P450 enzymes: a driving force of plant diterpene diversity. **Phytochemistry**, v. 161, p. 149-162, 2019.
- BATTH, R. et al. Biosynthesis and Industrial Production of Androsteroids. **Plants**, v. 9, n. 9, p. 1144, 2020.
- BECKSTEAD, R. B.; LAM, G.; THUMMEL, Carl S. The genomic response to 20-hydroxyecdysone at the onset of *Drosophila* metamorphosis. **Genome biology**, v. 6, n. 12, p. 1-13, 2005.
- BERGER, B.; PENG, J.; SINGH, M. Computational solutions for omics data. **Nature reviews genetics**, v. 14, n. 5, p. 333-346, 2013.
- BERMAN, H.; NAKAMURA, H.; HENRICK, K. The Protein Data Bank (PDB) and the Worldwide Protein Data Bank (<http://www.wwpdb.org>). **Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics**, 2004.
- BOKOV, D. O. et al. Prospects for the use of spinach (*Spinacia oleracea* L.) containing phytoecdysteroids and polyphenols. **Pharmacognosy Journal**, v. 12, n. 2, 2020.
- BOLWELL, G. P.; BOZAK, K.; ZIMMERLIN, A. Plant cytochrome P450. **Phytochemistry**, v. 37, n. 6, p. 1491-1506, 1994.

BONAVENTURE, G.; VANDOORN, A.; BALDWIN, I. T. Herbivore-associated elicitors: FAC signaling and metabolism. **Trends in plant science**, v. 16, n. 6, p. 294-299, 2011.

BROWN, D. D.; STERN, R. Methods of gene isolation. **Annual review of biochemistry**, v. 43, n. 1, p. 667-693, 1974.

CALLA, B. et al. Cytochrome P450 diversification and hostplant utilization patterns in specialist and generalist moths: Birth, death and adaptation. **Molecular ecology**, v. 26, n. 21, p. 6021-6035, 2017.

CALLA, B. Signatures of selection and evolutionary relevance of cytochrome P450 s in plant-insect interactions. **Current Opinion in Insect Science**, v. 43, p. 92-96, 2021.

CHADIN, I. et al. Ecdysteroid content and distribution in plants of genus *Potamogeton*. **Biochemical systematics and ecology**, v. 31, n. 4, p. 407-415, 2003. VOKÁČ, Karel et al. Ecdysteroid constituents of the mushroom *Tapinella panuoides*. **Phytochemistry**, v. 49, n. 7, p. 2109-2114, 1998.

CHECKER, V. G. et al. Role of phytohormones in plant defense: signaling and cross talk. In: **Molecular aspects of plant-pathogen interaction**. Springer, Singapore, p. 159-184, 2018.

CHITWOOD, D. J. Biochemistry and Function of Nematode Steroids. In: *Biochemistry and Function of Sterols*. CRC Press, 2020. p. 169-180.

COHEN, H. et al. The relationship between pollinator community and pollination services is mediated by floral abundance in urban landscapes. *Urban Ecosystems*, v. 24, n. 2, p. 275-290, 2021.

DANA, H. et al. In silico analysis, molecular docking, molecular dynamic, cloning, expression and purification of chimeric protein in colorectal cancer treatment. **Drug design, development and therapy**, v. 14, p. 309, 2020.

DARABI, M.; SEDDIGH, S.; ABARSHAHR, M. Structural, functional, and phylogenetic studies of cytochrome P450 (CYP) enzyme in seed plants by bioinformatics tools. **Caryologia**, v. 70, n. 1, p. 62-76, 2017.

DOOLITTLE, W. F. Phylogenetic classification and the universal tree. **Science**, v. 284, n. 5423, p. 2124-2128, 1999.

DUFFEY, S. S.; FELTON, G. W. Plant Enzymes in Resistance to Insects. **Biocatalysis in Agricultural Biotechnology**, p. 289–313, 1989.

ERB, M.; MELDAU, S.; HOWE, G. A. Role of phytohormones in insect-specific plant reactions. **Trends in plant science**, v. 17, n. 5, p. 250-259, 2012.

FERRY, N. et al. Plant–insect interactions: molecular approaches to insect resistance. **Current opinion in Biotechnology**, v. 15, n. 2, p. 155-161, 2004.

FESTUCCI-BUSELLI, R. A. et al. Biosynthesis and potential functions of the ecdysteroid 20-hydroxyecdysone—a review. **Botany**, v. 86, n. 9, p. 978-987, 2008.

FORTINI, E. A. et al. Photoperiod modulates growth and pigments and 20-hydroxyecdysone accumulation in Brazilian ginseng [*Pfaffia glomerata* (Spreng.) Pedersen] grown in vitro. *Plant Cell, Tissue and Organ Culture (PCTOC)*, v. 142, n. 3, p. 595-611, 2020.

FRAGO, E.; DICKE, M.; GODFRAY, H. C. J. Insect symbionts as hidden players in insect–plant interactions. **Trends in Ecology & Evolution**, v. 27, n. 12, p. 705-711, 2012.

FREEMAN, B. C.; BEATTIE, G. A. An Overview of Plant Defenses against Pathogens and Herbivores. ***Plant Pathology and Microbiology Publications***, n. 94, 2008.

GATEHOUSE, J. A. Plant resistance towards insect herbivores: a dynamic interaction. **New phytologist**, v. 156, n. 2, p. 145-169, 2002.

GIBSON, S.; SOMERVILLE, C. Isolating plant genes. **Trends in biotechnology**, v. 11, n. 7, p. 306-313, 1993.

GILBERT, L. I.; WARREN, J. T. A molecular genetic approach to the biosynthesis of the insect steroid molting hormone. **Vitamins & Hormones**, v. 73, p. 31-57, 2005.

GRAYER, R. J.; HARBORNE, J. B. A survey of antifungal compounds from higher plants, 1982–1993. **Phytochemistry**, v. 37, n. 1, p. 19-42, 1994.

GROENEVELD, H. W. Tracing Steroid Synthesis in Plants. In: *Biochemistry and Function of Sterols*. **CRC Press**, p. 73-83, 2020.

GUENGERICH, F. P. Mechanisms of cytochrome P450-catalyzed oxidations. **ACS catalysis**, v. 8, n. 12, p. 10964-10976, 2018.

GUENGERICH, F. P.; WATERMAN, M. R.; EGLI, M. Recent structural insights into cytochrome P450 function. **Trends in pharmacological sciences**, v. 37, n. 8, p. 625-640, 2016.

HALITSCHKE, R. et al. Molecular interactions between the specialist herbivore *Manduca sexta* (Lepidoptera, *Sphingidae*) and its natural host *nicotiana attenuata*. III. Fatty acid-amino acid conjugates in herbivore oral secretions are necessary and sufficient for herbivore-specific plant responses. **Plant physiology**, v. 125, n. 2, p. 711-717, 2001.

HANNEMANN, F. et al. Cytochrome P450 systems—biological variations of electron transport chains. **Biochimica et Biophysica Acta (BBA)-General Subjects**, v. 1770, n. 3, p. 330-344, 2007.

HASIN, Y.; SELDIN, M.; LUSIS, A. Multi-omics approaches to disease. **Genome biology**, v. 18, n. 1, p. 1-15, 2017.

HOWE, G. A. Cyclopentenone signals for plant defense: remodeling the jasmonic acid response. *Proceedings of the National Academy of Sciences*, v. 98, n. 22, p. 12317-12319, 2001.

- HUANG, X-Z. et al. The terpene synthase gene family in *Gossypium hirsutum* harbors a linalool synthase GhTPS12 implicated in direct defence responses against herbivores. **Plant, cell & environment**, v. 41, n. 1, p. 261-274, 2018.
- JANSSON, C.; LÖNNEBORG, A. Isolation of plant genes by T-DNA and transposon mutagenesis - gene tagging. In: **Progress in Botany/Fortschritte der Botanik**. Springer, Berlin, Heidelberg, p. 295-305, 1993.
- KLINGENBERG, M. Pigments of rat liver microsomes. **Archives of biochemistry and biophysics**, v. 75, n. 2, p. 376-386, 1958.
- KROES, A. et al. *Brevicoryne brassicae* aphids interfere with transcriptome responses of *Arabidopsis thaliana* to feeding by *Plutella xylostella* caterpillars in a density-dependent manner. **Oecologia**, v. 183, n. 1, p. 107-120, 2017.
- LACHAISE, F. et al. Ecdysteroid synthesis by crab Y-organs. **Journal of Experimental Zoology**, v. 252, n. 3, p. 283-292, 1989.
- LAFONT, R. et al. Ecdysteroid chemistry and biochemistry. In: **Insect endocrinology**. Academic Press, 2012. p. 106-176.
- LAMBRIX, P.; HABBOUCHE, M.; PEREZ, M. Evaluation of ontology development tools for bioinformatics. **Bioinformatics**, v. 19, n. 12, p. 1564-1571, 2003.
- LASKOWSKI, M. et al. Positive Darwinian selection in evolution of protein inhibitors of serine proteinases. In: **Cold Spring Harbor symposia on quantitative biology**. Cold Spring Harbor Laboratory Press, 1987. p. 545-553.
- LAY JR, J. O. et al. Problems with the “omics”. **TrAC Trends in Analytical Chemistry**, v. 25, n. 11, p. 1046-1056, 2006.
- LEMEY, P.; SALEMI, M.; VANDAMME, A. **The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing**. Cambridge University Press, 2009.
- LUCK, K. et al. CYP79D enzymes contribute to jasmonic acid-induced formation of aldoximes and other nitrogenous volatiles in two *Erythroxylum* species. **BMC plant biology**, v. 16, n. 1, p. 215, 2016.
- LUSCOMBE, N. M.; GREENBAUM, D.; GERSTEIN, M. What is bioinformatics? A proposed definition and overview of the field. **Methods of information in medicine**, v. 40, n. 04, p. 346-358, 2001.
- LYNCH, M. The origins of eukaryotic gene structure. **Molecular biology and evolution**, v. 23, n. 2, p. 450-468, 2005.
- MAFFEI, M. E.; MITHÖFER, A.; BOLAND, W. Before gene expression: early events in plant–insect interaction. **Trends in plant science**, v. 12, n. 7, p. 310-316, 2007.

- MAKKAR, G. S. et al. Insect resistance in Rice (*Oryza sativa* L.): overview on current breeding interventions. **International Journal of Tropical Insect Science**, v. 39, n. 4, p. 259-272, 2019.
- MARTINEZ, M. Plant protein-coding gene families: emerging bioinformatics approaches. **Trends in plant science**, v. 16, n. 10, p. 558-567, 2011.
- MAZID, M. et al. Role of secondary metabolites in defense mechanisms of plants. **Biology and medicine**, v. 3, n. 2, p. 232-249, 2011.
- MISRA, B. B. et al. Integrated omics: tools, advances and future approaches. **Journal of molecular endocrinology**, v. 62, n. 1, p. 21-45, 2019.
- MITHÖFER, A.; BOLAND, W. Recognition of herbivory-associated molecular patterns. **Plant Physiology**, v. 146, n. 3, p. 825-831, 2008.
- MIZUTANI, M. Impacts of diversification of cytochrome P450 on plant metabolism. **Biological and Pharmaceutical Bulletin**, v. 35, n. 6, p. 824-832, 2012.
- MORANT, M. et al. Plant cytochromes P450: tools for pharmacology, plant protection and phytoremediation. **Current opinion in biotechnology**, v. 14, n. 2, p. 151-162, 2003.
- NAHRSTEDT, A. The significance of secondary metabolites for interactions between plants and insects. **Planta Medica**, v. 55, n. 04, p. 333-338, 1989.
- NELSON, D. R. The cytochrome p450 homepage. **Human genomics**, v. 4, n. 1, p. 1-7, 2009.
- NELSON, D. R.; GOLDSTONE, J. V.; STEGEMAN, J. J. The cytochrome P450 genesis locus: the origin and evolution of animal cytochrome P450s. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 368, n. 1612, p. 20120474, 2013.
- NIWA, R.; NIWA, Y. S. Enzymes for ecdysteroid biosynthesis: their biological functions in insects and beyond. **Bioscience, biotechnology, and biochemistry**, v. 78, n. 8, p. 1283-1292, 2014.
- Omics.org. Disponível em: <http://Omics.org/index.php?title=Main_Page&oldid=37397>. Acesso em: 20 jun. 2021.
- OMURA, T. A new cytochrome in liver microsomes. **J. biol. Chem.**, v. 237, p. 1375-1376, 1962.
- OMURA, T. Forty years of cytochrome P450. **Biochemical and biophysical research communications**, v. 266, n. 3, p. 690-698, 1999.
- PAN, X.; CONNACHER, R. P.; O'CONNOR, M. B. Control of the insect metamorphic transition by ecdysteroid production and secretion. **Current Opinion in Insect Science**, 2020.
- PANDIAN, B. A. et al. Role of cytochrome P450 enzymes in plant stress response. **Antioxidants**, v. 9, n. 5, p. 454, 2020.

PARÉ, P. W.; TUMLINSON, J. H. Plant volatiles as a defense against insect herbivores. **Plant physiology**, v. 121, n. 2, p. 325-332, 1999.

PEARSON, H. What is a gene? *Nature* **441**, 398–401, 2006.

PERVIEUX, I. et al. A spruce defensin showing strong antifungal activity and increased transcript accumulation after wounding and jasmonate treatments. **Physiological and Molecular Plant Pathology**, v. 64, n. 6, p. 331-341, 2004.

PIETERSE, C. M. J. et al. Networking by small-molecule hormones in plant immunity. **Nature chemical biology**, v. 5, n. 5, p. 308-316, 2009.

PIETERSE, C. M. J.; DICKE, M. Plant interactions with microbes and insects: from molecular mechanisms to ecology. **Trends in plant science**, v. 12, n. 12, p. 564-569, 2007.

RANTY, B. et al. Calcium sensors as key hubs in plant responses to biotic and abiotic stresses. **Frontiers in Plant Science**, v. 7, p. 327, 2016.

REYMOND, P. et al. Differential gene expression in response to mechanical wounding and insect feeding in *Arabidopsis*. **The Plant Cell**, v. 12, n. 5, p. 707-719, 2000.

REYMOND, P.; FARMER, E. E. Jasmonate and salicylate as global signals for defense gene expression. **Current opinion in plant biology**, v. 1, n. 5, p. 404-411, 1998.

RHEE, S. Y.; DICKERSON, J.; XU, D. Bioinformatics and its applications in plant biology. **Annu. Rev. Plant Biol.**, v. 57, p. 335-360, 2006.

ROHMER, M. The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Nat. Prod. Rep.* **1999**, 16, 565–574.

SAIGO, H. et al. Protein homology detection using string alignment kernels. **Bioinformatics**, v. 20, n. 11, p. 1682-1689, 2004.

SAITO, J. et al. Characterization of candidate intermediates in the Black Box of the ecdysone biosynthetic pathway in *Drosophila melanogaster*: Evaluation of molting activities on ecdysteroid-defective larvae. **Journal of insect physiology**, v. 93, p. 94-104, 2016.

SAM, K. et al. Insect herbivory and herbivores of *Ficus* species along a rain forest elevational gradient in Papua New Guinea. *Biotropica*, v. 52, n. 2, p. 263-276, 2020.

SANDERS, D. et al. Calcium at the crossroads of signaling. **The Plant Cell**, v. 14, n. suppl 1, p. S401-S417, 2002.

SCHNEIDER, M. V.; ORCHARD, S. Omics technologies, data and bioinformatics principles. **Bioinformatics for omics Data**, p. 3-30, 2011.

SCHOONHOVEN, L. M. et al. **Insect-plant biology**. Oxford University Press on Demand, 2005.

- SCHULER, M. A. The role of cytochrome P450 monooxygenases in plant-insect interactions. **Plant physiology**, v. 112, n. 4, p. 1411, 1996.
- SCOTT, J. G.; WEN, Z. Cytochromes P450 of insects: the tip of the iceberg. **Pest management science**, v. 57, n. 10, p. 958-967, 2001.
- SHARMA, A. V.; MISHRA, V. Simplified Perspective of Complex Insect–Plant Interactions. In: *Plant-Pest Interactions: From Molecular Mechanisms to Chemical Ecology*. Springer, Singapore, 2021. p. 399-415.
- SLÁMA, K. Plants as a source of materials with insect hormone activity. **Entomologia experimentalis et applicata**, v. 12, n. 5, p. 721-728, 1969.
- SPINDLER, K-D. et al. Ecdysteroid hormone action. **Cellular and molecular life sciences**, v. 66, n. 24, p. 3837-3850, 2009.
- SREEJIT, C. M. et al. Isolation, quantification and chemical characterisation of ecdysterone from medicinal plants of Kerala, Western Ghats. **Journal of Pharmacognosy and Phytochemistry**, v. 8, n. 2, p. 254-257, 2019.
- STIPANOVIC, R. D. Function and chemistry of plant trichomes and glands in insect resistance: protective chemicals in plant epidermal glands and appendages. In: HEDIN, P. **Plant Resistance to Insects**. ACS Symposium Series. American Chemical Society: Washington, DC, 1983.
- SVOBODA, J. A. Variability of metabolism and function of sterols in insects. **Biochemistry and function of sterols**, v. 55, 1997.
- SWEVERS, L.; IATROU, K. Ecdysteroids and ecdysteroid signaling pathways during insect oogenesis. In: **Ecdysone: Structures and Functions**. Springer, Dordrecht, 2009. p. 127-164.
- TARKOWSKÁ, D.; STRNAD, M. Plant ecdysteroids: Plant sterols with intriguing distributions, biological effects and relations to plant hormones. *Planta* **2016**, *244*, 545–555.
- TOOKER, J. F.; GIRON, David. The evolution of endophagy in herbivorous insects. *Frontiers in Plant Science*, v. 11, 2020.
- TSUKAGOSHI, Y. et al. Functional characterization of CYP71D443, a cytochrome P450 catalyzing C-22 hydroxylation in the 20-hydroxyecdysone biosynthesis of *Ajuga hairy roots*. **Phytochemistry**, v. 127, p. 23-28, 2016.
- VASAV, A. P.; BARVKAR, V. T. Phylogenomic analysis of cytochrome P450 multigene family and their differential expression analysis in *Solanum lycopersicum* L. suggested tissue specific promoters. **BMC genomics**, v. 20, n. 1, p. 1-13, 2019.
- WARREN, J. T.; GILBERT, L. I. Radioimmunoassay: ecdysteroids. In: **Immunological techniques in insect biology**. Springer, New York, NY, 1998. p. 181-214.
- WOLLENBERGER, L.; DINAN, L.; BREITHOLTZ, M. Brominated flame retardants: activities in a crustacean development test and in an ecdysteroid screening

assay. **Environmental Toxicology and Chemistry: An International Journal**, v. 24, n. 2, p. 400-407, 2005.

WU, J.; BALDWIN, I. T. New insights into plant responses to the attack from insect herbivores. **Annual review of genetics**, v. 44, p. 1-24, 2010.

YAMANAKA, N. Ecdysteroid signalling in insects—From biosynthesis to gene expression regulation. *Genes and Endocrine Signalling in Development and Homeostasis*, p. 1, 2021.

YOSHINAGA, N. et al. Active role of fatty acid amino acid conjugates in nitrogen metabolism in *Spodoptera litura* larvae. **Proceedings of the National Academy of Sciences**, v. 105, n. 46, p. 18058-18063, 2008.

YOSHIYAMA-YANAGAWA, Takuji et al. The conserved Rieske oxygenase DAF-36/Neverland is a novel cholesterol-metabolizing enzyme. **Journal of Biological Chemistry**, v. 286, n. 29, p. 25756-25762, 2011.

ZHANG, X.; ZHOU, X.; WANG, X. Basics for bioinformatics. In: **Basics of Bioinformatics**. Springer, Berlin, Heidelberg, p. 1-25, 2013.

ZHOU, S.; JANDER, G. Engineering insect resistance using plant specialized metabolites. *Current Opinion in Biotechnology*, v. 70, p. 115-121, 2021.

ZUNJARRAO, S. S. et al. Plant-insect interaction: the saga of molecular coevolution. **Co-evolution of secondary metabolites**, p. 19-45, 2020.

3 IDENTIFICAÇÃO, ANÁLISE IN SÍLICO DE GENES E PROTEÍNAS E ELABORAÇÃO DE ESTRATÉGIAS DE ISOLAMENTO DE GENES POTENCIALMENTE ENVOLVIDOS NA BIOSÍNTESE E INATIVAÇÃO DE 20-HIDROXIECDISONA (20E) EM PLANTAS

RESUMO

A caracterização e o isolamento de genes e proteínas, envolvidas em mecanismos de defesas de plantas é essencial para se elaborar estratégias que visam o controle de insetos que atacam espécies de interesse econômico. A molécula 20-hidroxiecdisona (20E) é sintetizada em insetos e algumas plantas. Em insetos, 20E está relacionada a processos de desenvolvimento, sendo parcialmente conhecida sua rota biossintética. Em plantas, 20E pode atuar em mecanismos de defesa, sendo desconhecida sua rota de biossintese. A molécula 20E pode apresentar aplicabilidade farmacológica, medicinal, agroquímica e biotecnológica. Deste modo, o objetivo deste trabalho foi realizar a identificação e análise *in silico* de genes e proteínas e elaboração de estratégias de isolamento de genes potencialmente envolvidos na biossintese e inativação de 20E em plantas. As sequências de nucleotídeos dos genes e as sequências de aminoácidos das proteínas (enzimas) que atuam na rota parcialmente conhecida de biossintese (Neverland, Shade, Shadow, Disembodied e Phantom) e inativação (Ecdysone oxidase, 3-Dehydroecdysone 3beta-reductase e 3-Dehydroecdysone 3alfa-reductase) de 20E em insetos foram identificadas e utilizadas para se realizar uma busca em três bancos de dados (Landmark; UniProtKB/Swiss-Prot; e, *Protein Data Bank*). Foram identificadas e analisadas *in silico* 239 potenciais proteínas que apresentaram similaridade entre 25 a 45% e apresentaram um valor E de pelo menos E^{-5} . Em seguida, foram realizados alinhamentos, predição de parâmetros físico-químicos (peso molecular, ponto isoelétrico teórico, número de resíduos negativos e positivos, coeficiente de extinção, índice de instabilidade, índice alifático e índice de hidrofobicidade), a localização subcelular e filogenia. Foi observado que 40 espécies de plantas produzem proteínas similares as que atuam na biossintese e inativação de 20E. As proteínas similares a Shade, Shadow e 3DE 3beta-reductase foram detectadas em maior quantidade e em várias espécies de plantas, enquanto as proteínas similares a Disembodied e Neverland foram detectadas, respectivamente, em uma e duas espécies. Estratégias que visam o isolamento de genes potencialmente envolvidos com a biossintese de 20E em plantas foram propostas por meio do uso de oligonucleotídeos, construídos a partir de regiões conservadas, obtidas por meio de alinhamentos múltiplos. Várias proteínas encontradas em vegetais apresentam função ainda desconhecidas, podendo estar envolvidas na biossintese de 20E em plantas; entretanto, seu envolvimento e real função ainda precisam ser analisadas.

Palavras-chave: Bioinformática. Defesa de Plantas. Genes. Proteínas. Metabólitos Secundários.

3 IDENTIFICATION, IN SILICO ANALYSIS OF GENES AND PROTEINS AND DEVELOPMENT OF GENES ISOLATION STRATEGIES POTENTIALLY INVOLVED IN THE BIOSYNTHESIS AND INACTIVATION OF 20-HYDROXYECDISONE (20E) IN PLANTS

ABSTRACT

The characterization and isolation of genes and proteins involved in plant defense mechanisms is essential to develop strategies aimed at controlling insects that attack species of economic interest. The 20-hydroxyecdysone (20E) molecule is synthesized in insects and some plants. In insects, 20E is related to developmental processes, its biosynthetic route being partially known. In plants, 20E can act in defense mechanisms, its biosynthetic route being unknown. The 20E molecule may have pharmacological, medicinal, agrochemical and biotechnological applicability. Thus, the objective of this work was to carry out the identification and in silico analysis of genes and proteins and the elaboration of strategies for the isolation of genes potentially involved in the biosynthesis and inactivation of 20E in plants. The nucleotide sequences of genes and amino acid sequences of proteins (enzymes) that act in the partially known pathway of biosynthesis (Neverland, Shade, Shadow, Disembodied and Phantom) and inactivation (Ecdysone oxidase, 3-Dehydroecdysone 3beta-reductase and 3- 20E dehydroecdysone 3alpha-reductase) in insects were identified and used to perform a search in three databases (Landmark; UniProtKB/Swiss-Prot; and Protein Data Bank). 239 potential proteins that showed similarity between 25 and 45% and presented an E value of at least e^{-5} were identified and analyzed in silico. Then, alignments, prediction of physicochemical parameters (molecular weight, theoretical isoelectric point, number of negative and positive residues, extinction coefficient, instability index, aliphatic index and hydrophobicity index), subcellular location and phylogeny were performed. It was observed that 40 species of plants produce proteins similar to those that act in the biosynthesis and inactivation of 20E. Shade, Shadow and 3DE 3beta-reductase-like proteins were detected in greater amounts and in several plant species, while Disembodied and Neverland-like proteins were detected in one and two species, respectively. Strategies aimed at isolating genes potentially involved in 20E biosynthesis in plants have been proposed through the use of oligonucleotides, constructed from conserved regions, obtained through multiple alignments. Several proteins found in vegetables have a function that is still unknown and may be involved in the biosynthesis of 20E in plants; however, its involvement and actual role still need to be analyzed.

Key-words: Bioinformatics. Plant Defense. Genes. Proteins. Secondary metabolites.

3.1 Introdução

A molécula 20-hidroxiecdisona (20E) é um hormônio presente nos insetos e em alguns nematoides que possui papel essencial na fisiologia e desenvolvimento desses organismos, participando de processos biológicos como metamorfose, ecdise e resposta imune (URYU; AMEKU; NIWA, 2015; SWEVERS, 2019; HAN et al., 2020). Nos insetos, além deste hormônio ser capaz de afetar a expressão gênica se ligando a dois receptores hormonais nucleares (receptor de ecdisona e ultraespiráculo), ele inclui diversos outros receptores hormonais e fatores de transcrição (LI et al., 2015; ABLES et al., 2016).

Esta molécula é um tipo de hormônio esteroide sinalizador que se difere dos outros devido a sua característica de possuir diversos agrupamentos hidroxila em sua estrutura, atuando em funções essenciais durante o desenvolvimento dos artrópodes, como na proliferação celular, diferenciação e apoptose (SCHWEDES; CARNEY, 2012). A rota de biossíntese e inativação desse hormônio é parcialmente conhecida em insetos, por meio de estudos moleculares e bioquímicos com a *Drosophila melanogaster* e a lepidóptera *Bombyx mori*, com pelo menos oito enzimas e seus intermediários conhecidos, entretanto uma parte da rota com o intermediário 7-Deidro colesterol até a conversão em 5beta-Ketodiol continua uma incógnita, chamada de “*Black Box*” (NIWA; NIWA, 2015).

Esse hormônio é principalmente ativado durante a ecdise e a metamorfose de insetos, a partir do Hormônio Protoracicotrópico Neuropeptídeo (PTTH, em inglês) liberado pelas células cerebrais neurosecretoras na hemolinfa, ativando a produção e secreção de 20E (SCHWEDES; CARNEY, 2012). Durante a biossíntese, os genes *Halloween* (*Neverland*, *Shade*, *Shadow*, *phantom* e *Disembodied*), codificadores de enzimas do complexo citocromo P450, são ativados para catalisar a rota da produção do hormônio 20E (WARREN et al., 2002; GILBERT, 2005; REWITZ et al., 2006).

Embora este hormônio seja conhecido por suas importantes funções em artrópodes, desde a década de 60 há identificação de plantas capazes de produzir a 20E (KAPLANIS et al., 1967). Atualmente o fitoecdisteroide mais comum presente nos vegetais é a 20E, que em plantas foi observado sua produção como metabólito secundário e participando da rede de defesas a insetos fitófagos não adaptados a dieta rica em 20E (CAO; RIU; BOO, 2018; PARDEDE et al., 2018; YACTAYO-CHANG et al., 2020), além de propiciar resistência a estresses abióticos, como por exemplo a salinidade (LI et al., 2016).

Logo, a molécula 20E adquiriu nos últimos anos uma importância para usos na meio agrícola e biotecnológico devido estar envolvido em importantes processos biológicos (FESTUCCI-BUSELLI et al., 2008). Além disso, os fitoecdisteroides estão sendo utilizados por suas propriedades farmacológica no combate de diversas doenças (BOUYAHYA et al., 2020; DINAN et al., 2021).

A produção de 20E em plantas advém da necessidade desses organismos em produzir fitotoxinas e se adaptar às pragas presentes no ambiente, uma vez que são organismos sesséis e precisam sobreviver e se reproduzir (LARIGUET; DUNAND, 2005). Para isto, durante o curso evolutivo as plantas desenvolveram um vasto número de metabolitos secundários com complexas rotas de biossíntese, que utilizam as enzimas do complexo citocromo P450 para catalisar diversas reações (WERCK-REICHHART; BAK; PAQUETTE, 2002). Essas enzimas são extremamente diversas e adaptadas a diferentes substratos, com atuação em importantes processos de biossíntese de várias outras moléculas como os metabolitos secundários, antioxidantes, fito hormônios, polímeros estruturais e moléculas sinalizadoras (RENAULT et al., 2013; DISTÉFANO et al., 2020; HANSEN et al., 2021).

A descoberta da produção e função de fitoecdisteroides em plantas possibilitou o desenvolvimento de diversos inseticidas comerciais com base em ecdisona DBH-46. Os inseticidas biorracionais e ecológicos atuam nos insetos focando os receptores de ecdisona (EcR, em inglês) testados em alguns lepidópteros e dípteros (NAKAGAWA, 2005; ZOTTI et al., 2013). Além da aplicação externa como inseticida, é possível desenvolver cultivares resistentes a uma variedade de insetos por meio da biossíntese natural de 20E pela planta, entretanto é necessário um conhecimento a nível molecular do ponto de vista das próprias defesas das plantas, além da necessidade de isolar e caracterizar os genes envolvidos nas rotas metabólicas (MAKKAR et al., 2019).

Entretanto um desafio para essa aplicabilidade biotecnológica do hormônio 20E ainda é o desconhecimento acerca da sua rota de biossíntese em plantas (FESTUCCI-BUSELLI et al., 2008). Nesse contexto, objetivo deste trabalho é realizar a identificação e análise *in silico* de genes e proteínas e elaborar estratégias de isolamento de genes potencialmente envolvidos na biossíntese e inativação de 20E em plantas.

3.2 Material e Métodos

3.2.1 Ferramenta BLAST (Mineração de dados) e caracterização de proteínas

Foram obtidas sequências de aminoácidos completas das proteínas envolvidas na rota parcialmente conhecida da biossíntese da biomolécula 20-hidroxicdisona (20E) e sua inativação (NIWA; NIWA, 2014; PAN; CONNACHER; O’CONNOR, 2020), em *Drosophila melanogaster* Meigen (1830) e *Spodoptera littoralis* Boisduval (1833), a partir da molécula de colesterol (C₂₇H₄₆O), por meio da plataforma NCBI (<https://www.ncbi.nlm.nih.gov/protein/>). As proteínas utilizadas como referência e os dados referentes as mesmas estão dispostas na Tabela 1.

Tabela 1 – Proteínas isoladas e caracterizadas da rota parcial da biossíntese de 20-hidroxicdisona em *D. melanogaster* M. (1830) e *S. littoralis* Boisduval (1833).

N.	Proteína	Organismo	Banco de dados	Tam.	Acesso	Referência
1	Neverland	<i>D. melanogaster</i> Meigen, 1830	GenBank	429	ABW08586	Adams et al. (2000)
2	Shade	<i>D. melanogaster</i> Meigen, 1831	NCBI	540	NP_001261843	Matthews et al. (2015)
3	Shadow	<i>D. melanogaster</i> Meigen, 1832	NCBI	520	NP_650123	Matthews et al. (2015)
4	Disembodied	<i>D. melanogaster</i> Meigen, 1833	NCBI	489	NP_524810	Matthews et al. (2015)
5	Phantom	<i>D. melanogaster</i> Meigen, 1834	NCBI	574	NP_573319	Matthews et al. (2015)
6	Ecdysone oxidase	<i>D. melanogaster</i> Meigen, 1835	NCBI	657	NP_572986	Matthews et al. (2015)
7	3-Dehydroecdysone 3beta-reductase	<i>S. littoralis</i> Boisduval, 1833	GenBank	345	CAB41997	Chen; Turner; Rees, 1999
8	3-Dehydroecdysone 3alfa-reductase	<i>S. littoralis</i> Boisduval, 1834	GenBank	249	AAF70499	Takeuchi et al. (2000)

Fonte: Autor, 2021.

Utilizou-se a ferramenta *Filter protein* presente no *Sequence Manipulation Suite* – SMS (<https://www.bioinformatics.org/sms2/>) para remoção de caracteres não-proteicos do texto copiado da plataforma de dados brutos no NCBI. Cada sequência de aminoácidos das proteínas atuante na biossíntese de 20E foram comparadas com proteínas potencialmente homólogas de espécies vegetais em três bancos de dados distintos, sendo eles: 1) Landmark – organismos modelos; 2) UniProtKB/Swiss-Prot e 3) *Protein Data Bank* – PDB). O algoritmo empregado nesta busca é o *Protein BLAST/NCBI* (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

A partir do resultado da busca por proteínas em espécies vegetais similares às que participam da biossíntese de 20E em insetos, foram selecionadas as sequências de aminoácidos de proteínas candidatas, que possuíam um *valor-E* apresentando similaridade confiável de pelo menos e^{-5} e construiu-se uma tabela geral com as informações geradas para cada proteína da

rota. Os textos brutos de proteínas foram copiados e filtrados com a ferramenta *Filter protein* conforme realizado anteriormente.

A caracterização das proteínas foi realizada por meio da ferramenta *ProtParam* disponível no servidor *Expasy* (<https://web.expasy.org/protparam/>) (GASTEIGER et al., 2005). Os parâmetros físico-químicos calculados das proteínas avaliadas foram: 1) Peso Molecular (em Daltons – Da); 2) Ponto Isoelétrico teórico – PI; 3) Número de Resíduos negativos (-) e positivos (+); 4) Coeficiente de Extinção; 5) Índice de Instabilidade (II); 6) Índice Alifático (IA); e 7) Índice de hidrofobicidade (mensurado pelo GRAVY - *Grand Average of Hydropathy*). A localização subcelular foi simulada utilizando a ferramenta *Plant-mPLOC* (<http://www.csbio.sjtu.edu.cn/cgi-bin/PlantmPLOC.cgi>).

3.2.2 Alinhamento e filogenia de proteínas

Foi realizado o alinhamento em pares por meio da ferramenta *Pairwise Sequence Alignment* disponível no *EMBOSS Needle* (https://www.ebi.ac.uk/Tools/psa/emboss_needle/), enquanto que um alinhamento múltiplo também foi construído por meio da ferramenta *Multiple Sequence Alignment* (<https://www.genome.jp/tools-bin/clustalw>) por CLUSTALW.

Uma árvore filogenética foi montada por meio do *software Molecular Evolutionary Genetics Analysis* – MEGA versão 10.8.1 (KUMAR et al., 2018) a partir da exportação, em arquivos *FASTA* para a plataforma MEGA, dos alinhamentos múltiplos formados pelo CLUSTALW, utilizando o método estatístico de Máxima Verossimilhança. A árvore foi construída utilizando o método heurístico NNI (*Nearest-Neighbor-Interchange*) para tornar a probabilidade máxima mais exaustiva. Aplicou-se o teste de filogenia de *Bootstrap* com 1000 replicações para validação das ramificações (FELSENSTEIN, 1985).

3.2.3 Alinhamento de sequência de nucleotídeos dos genes

As sequências de nucleotídeos dos genes que codificam as proteínas selecionadas a partir da busca *Protein BLAST* foram resgatadas por meio da página principal de cada proteína no NCBI e posterior acesso ao NCBI/*Nucleotide* visando a correspondência em cada sequência de nucleotídeos com a proteína selecionada. O acesso de nucleotídeos se deu por cópia das regiões codificadoras (CDS) de cada sequência de mRNA disponível.

Assim como realizado com as proteínas, os textos de nucleotídeos copiados foram filtrados com a ferramenta *Filter DNA* disponível no *Sequence Manipulation Suite* – SMS (<https://www.bioinformatics.org/sms2/>). O alinhamento em pares das sequências de

nucleotídeos dos genes homólogos foi realizada também pela ferramenta *Pairwise Sequence Alignment* (https://www.ebi.ac.uk/Tools/psa/emboss_needle/), enquanto que o alinhamento múltiplo foi a partir da ferramenta *Multiple Sequence Alignment* (<https://www.genome.jp/tools-bin/clustalw>) por CLUSTALW.

3.2.4 Desenho de oligonucleotídeos

A partir dos alinhamentos em pares e múltiplos das sequências de nucleotídeos de genes homólogos aos genes que codificam proteínas atuantes na biossíntese de 20E, foram propostos quatro pares de oligonucleotídeos para cada gene da rota metabólica. Dois oligonucleotídeos *forward* nas regiões 5'P, mais dois localizados na região 3'OH de forma reversa. Para cada par de oligonucleotídeo, variações entre bases nitrogenadas foram propostas para maximizar a área de cobertura com base nos genes homólogos alinhados.

As informações referentes às condições de cada oligonucleotídeo como conteúdo guanina-citosina (GC%), temperatura fusão (*Temperature of melting* – Tm) e testes de adequação para Reação em Cadeia da Polimerase (PCR) foram delimitados por meio da ferramenta *PCR Primer Status* disponível no *Sequence Manipulation Suite* – SMS (<https://www.bioinformatics.org/sms2/>).

3.3 Resultados

A partir da busca por meio da ferramenta BLAST/NCBI, com base nas sequências de aminoácidos de proteínas que atuam na rota de síntese e neutralização de 20-hidroxiecdisona em insetos, segundo modelo parcial de rota propostos para a mosca-das-frutas (*Drosophila melanogaster*) e para a lagarta desfolhadora (*Spodoptera littoralis*). Foram encontradas diversas proteínas candidatas na atuação da biossíntese e neutralização da molécula 20-hidroxiecdisona (20E) em diferentes espécies de plantas. A Tabela 2 apresenta o número de espécies de plantas encontradas, além do número de proteínas por espécie.

Tabela 2 – Diversidade de espécies de plantas e número de proteínas por espécie provenientes da busca em banco de dados pela ferramenta Blast/NCBI a partir de proteínas atuantes na rota de biossíntese e neutralização de 20-hidroxiecdisona em *Drosophila melanogaster* e *Spodoptera littoralis*. Total de proteínas encontradas = 239.

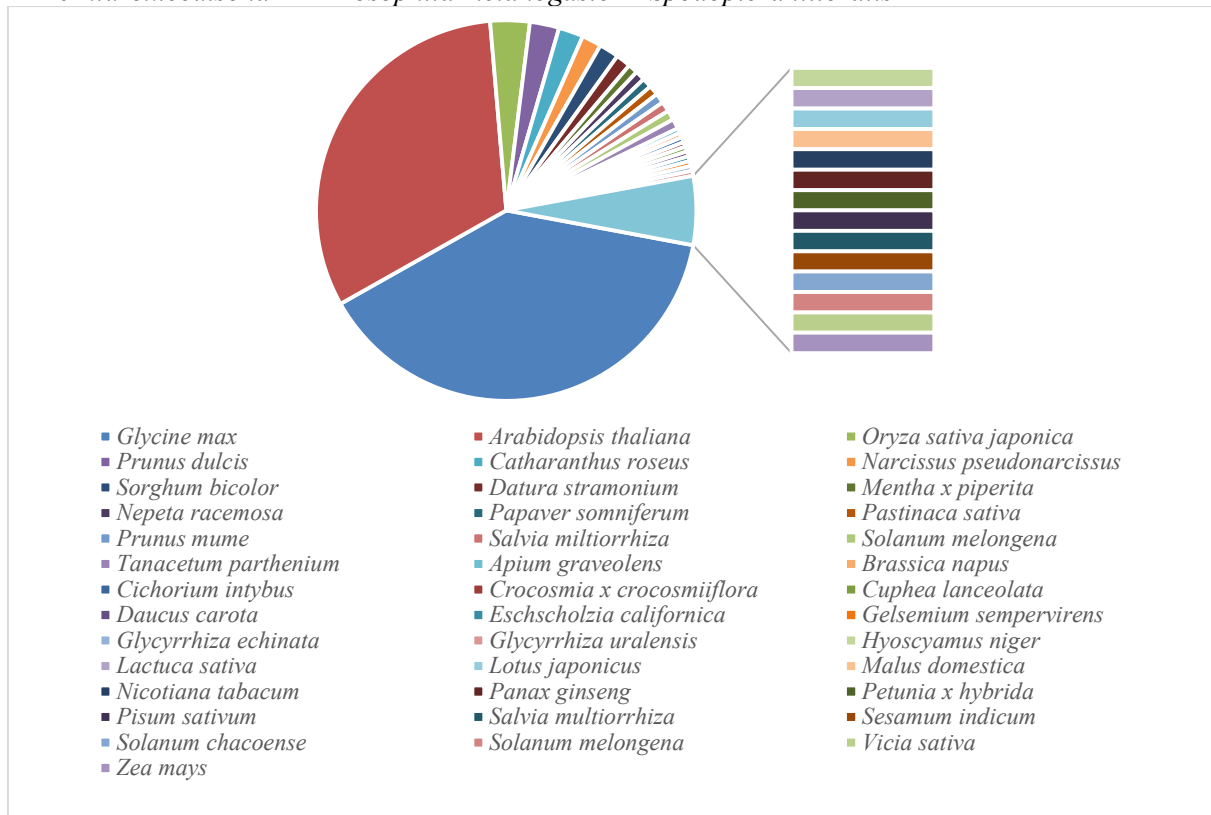
N.	Organismo	Número de proteínas	N.	Organismo	Número de proteínas
1	<i>Glycine max</i>	93	21	<i>Cuphea lanceolata</i>	1
2	<i>Arabidopsis thaliana</i>	76	22	<i>Daucus carota</i>	1
3	<i>Oryza sativa japonica</i>	8	23	<i>Eschscholzia californica</i>	1
4	<i>Prunus dulcis</i>	6	24	<i>Gelsemium sempervirens</i>	1

5	<i>Catharanthus roseus</i>	5	25	<i>Glycyrrhiza echinata</i>	1
6	<i>Narcissus pseudonarcissus</i>	4	26	<i>Glycyrrhiza uralensis</i>	1
7	<i>Sorghum bicolor</i>	4	27	<i>Hyoscyamus niger</i>	1
8	<i>Datura stramonium</i>	3	28	<i>Lactuca sativa</i>	1
9	<i>Mentha x piperita</i>	2	29	<i>Lotus japonicus</i>	1
10	<i>Nepeta racemosa</i>	2	30	<i>Malus domestica</i>	1
11	<i>Papaver somniferum</i>	2	31	<i>Nicotiana tabacum</i>	1
12	<i>Pastinaca sativa</i>	2	32	<i>Panax ginseng</i>	1
13	<i>Prunus mume</i>	2	33	<i>Petunia x hybrida</i>	1
14	<i>Salvia multiorrhiza</i>	2	34	<i>Pisum sativum</i>	1
15	<i>Solanum melongena</i>	2	35	<i>Salvia multiorrhiza</i>	1
16	<i>Tanacetum parthenium</i>	2	36	<i>Sesamum indicum</i>	1
17	<i>Apium graveolens</i>	1	37	<i>Solanum chacoense</i>	1
18	<i>Brassica napus</i>	1	38	<i>Solanum melongena</i>	1
19	<i>Cichorium intybus</i>	1	39	<i>Vicia sativa</i>	1
20	<i>Crocsmia x crocosmiflora</i>	1	40	<i>Zea mays</i>	1

Fonte: Autor (2021).

Observa-se um quantitativo de 40 espécies de plantas que produzem proteínas similares as que atuam na biossíntese de 20E. Em relação às proteínas, um total de 239, sendo a maior quantidade de proteínas identificadas nas espécies soja (*Glycine max*) e arábido (*Arabidopsis thaliana*), com 93 e 76 proteínas, respectivamente. Dentre as espécies encontradas, além da existência de plantas consideradas modelos para estudos genéticos e fisiológicos, observa-se a presença de culturas agrícolas, como o arroz (*Oryza sativa* var. japonica) e milho (*Zea mays*), mas a maior parte pertence a pequenas plantas herbáceas distribuídas por diferentes biomas no mundo. A Figura 10 mostra uma análise visual do quantitativo por espécie e da quantidade de proteínas encontradas pela ferramenta BLAST/NCBI.

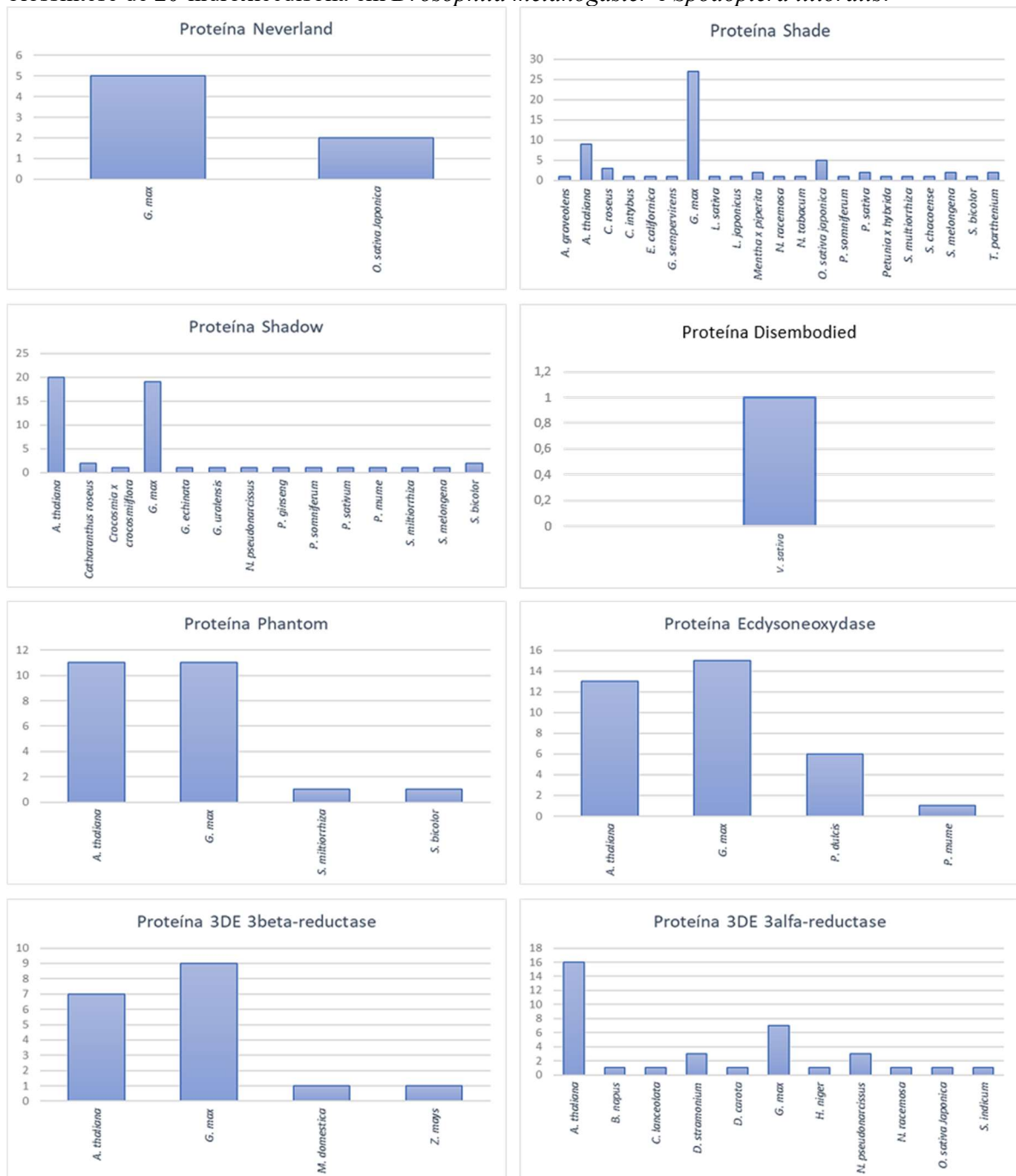
Figura 10 – Quantidade de proteínas similares e diversidade de espécies de plantas provenientes da busca em banco de dados pela ferramenta Blast/NCBI a partir de proteínas atuantes na rota de biossíntese de 20-hidroxiecdisona em *Drosophila melanogaster* e *Spodoptera littoralis*.



Fonte: Autor (2021).

A Figura 11 mostra os gráficos referentes as quantidades de proteínas similares e espécies de plantas, por proteína atuante na rota de biossíntese e neutralização de 20E em insetos. As proteínas *Shade* (SHD), *Shadow* (SAD) e *3DE 3beta-reductase* (3DE3BR) apresentam uma maior quantidade de espécies de plantas e, também, de proteínas encontradas. Dentre os outros grupos de proteínas da rota de biossíntese e neutralização, as proteínas *Disembodied* (DIB) e *Neverland* (NVD) apresenta, respectivamente, uma e duas espécies, com o quantitativo de uma proteína similar para DIB, enquanto que somente sete proteínas similares para NVD na busca em diferentes bancos de dados pela ferramenta BLAST/NCBI.

Figura 11 – Quantidade de proteínas similares e diversidade de espécies de plantas provenientes da busca em banco de dados pela ferramenta BLAST/NCBI a partir de proteínas atuantes na rota de biossíntese de 20-hidroxicdisona em *Drosophila melanogaster* e *Spodoptera littoralis*.



Fonte: Autor (2021).

3.1 Identificação de proteínas similares por meio da ferramenta BLAST/NCBI (mineração de dados)

A partir da busca na plataforma BLAST/NCBI, utilizando a sequência de aminoácido das proteínas que atuam na rota parcialmente descoberta de biossíntese e neutralização da molécula 20-hidroxicdisona (20E) nos insetos *Drosophila melanogaster* e *Spodoptera*

littoralis, foi possível identificar proteínas similares em diferentes bancos de dados, sendo que o acesso cadastrado na plataforma NCBI está disponível em cada tabela de resultado para total acesso às informações gerais e sequências de aminoácidos de proteínas utilizadas. Os resultados demonstrados apresentam a Porcentagem de Cobertura, a Similaridade, o Valor-E, o acesso cadastrado, a presença ou ausência de 20E (+ ou -, respectivamente, de acordo com Lafont, 2019) e os bancos de dados de origem de cada proteína. A análise do valor-E é apresentada individualmente para cada banco de dados encontrado.

Para a proteína Neverland (NVD) (Tabela 3), identificou-se sete proteínas distribuídas em duas espécies de plantas. Para a soja (*Glycine max*), a proteína Chlorophyllide a oxygenase e suas isoformas foram similares à proteína NVD, com uma taxa de cobertura variando de 25% a 49%, enquanto que a porcentagem de similaridade variou de 26,01% a 32,14%. Para o arroz (*Oriza sativa* var. japônica), apenas duas proteínas similares, a Pheophorbide a oxygenase e também uma Chlorophyllide a oxygenase, a cobertura foi de 24% e 50% e a similaridade foi de 29,91 e 25,78%, respectivamente. Para o valor-E, no banco de dados Modelos (LANDMARK) o máximo foi de $1,00e^{-06}$, referente a proteína Chlorophyllide a oxygenase (em suas isoformas), enquanto que para o banco de dados UNIPROT foi de $2,00e^{-07}$ para a proteína Chlorophyllide a oxygenase. Somente o arroz possui a produção da molécula de 20E, enquanto que a soja não apresentou a produção de 20E.

Tabela 3 – Proteínas vegetais que apresentam similaridade à proteína Neverland de *Drosophila melanogaster* a partir da busca em banco de dados pela ferramenta BLAST/NCBI, apresentando a porcentagem de Cobertura (Cob.), de Similaridade (Sim.), o Valor-E, o acesso cadastrado, a presença ou ausência de 20E (+ ou -, respectivamente) e os bancos de dados de origem.

N.	Proteína	Organismo	Cob.	Sim.	Valor-E	Acesso	20E	Banco de dados
1	Neverland	<i>Drosophila melanogaster</i>	100%	100%	0,0	NP_0010976 70.1	+	Modelos (LANDMARK)
2	Chlorophyllide a oxygenase X2	<i>Glycine max</i>	26%	32,14%	$1,00e^{-06}$	XP_0146226 98.1	-	Modelos (LANDMARK)
3	Chlorophyllide a oxygenase X1	<i>Glycine max</i>	26%	32,14%	$1,00e^{-06}$	XP_0035446 91.1	-	Modelos (LANDMARK)
4	Chlorophyllide a oxygenase	<i>Glycine max</i>	26%	32,14%	$2,00e^{-06}$	XP_0035430 47.1	-	Modelos (LANDMARK)
5	Chlorophyllide a oxygenase X2	<i>Glycine max</i>	25%	31,48%	$2,00e^{-06}$	XP_0259846 64.1	-	Modelos (LANDMARK)
6	Chlorophyllide a oxygenase X1	<i>Glycine max</i>	49%	26,01%	$3,00e^{-06}$	XP_0035266 33.1	-	Modelos (LANDMARK)
7	Chlorophyllide a oxygenase	<i>Oryza sativa</i> Japonica	50%	25,78%	$2,00e^{-07}$	Q8S7E1.1	+	UNIPROT
8	Pheophorbide a oxygenase	<i>Oryza sativa</i> Japonica	24%	29,91%	$2,00e^{-05}$	Q0DV66.1	+	UNIPROT

Fonte: Autor (2021).

Em relação à proteína Shade (SHD) (Tabela 4), foram encontradas um total de 65 proteínas similares nos bancos de dados Modelos (LANDMARK), UniProt e Protein Data

Base (PDB) distribuídas em 21 espécies de plantas. A porcentagem de cobertura variou entre 26% a 87%, enquanto que a similaridade variou de 21,21% a 31,18%. Para o valor-E, o máximo observado para o banco de dados Modelos (LANDMARK) foi de $9,00e^{-24}$, referente à proteína *Cytochrome P450 71A26* (variação 1), para o banco de dados UNIPROT, foi de $5,00e^{-21}$ para a proteína *Cytochrome P450 71AJ4*, e por último, de $8,00e^{-11}$ para a proteína *Cytochrome P450 76AH1* no banco de dados PDB. Dentre as espécies encontradas, apenas *Arabidopsis thaliana* e *Lactuca sativa* apresentaram produção de 20E, enquanto que *Glicine max* e *Sorghum bicolor* não foi detectado. As demais espécies encontradas não possuem registro de presença ou ausência de 20E, classificadas como Desconhecido (DS).

Tabela 4 – Proteínas vegetais que apresentam similaridade à proteína Shade de *Drosophila melanogaster* a partir da busca em banco de dados pela ferramenta BLAST/NCBI, apresentando a porcentagem de Cobertura (Cob.), de Similaridade (Sim.), o Valor-E, o acesso cadastrado, a presença ou ausência de 20E (+ ou -, respectivamente) (DS = Desconhecido) e o banco de dados de origem.

N.	Proteína	Organismo	Cob.	Sim.	Valor-E	Acesso	20E	Banco de dados
1	Shade	<i>Drosophila melanogaster</i>	100%	1	0	NP_001261843	+	Modelos (LANDMARK)
2	Cytochrome P450 71A19	<i>Arabidopsis thaliana</i>	82%	23,55%	$3e^{-19}$	NP_193065.1	+	Modelos (LANDMARK)
3	Cytochrome P450 71A20	<i>Arabidopsis thaliana</i>	82%	23,23%	$7e^{-19}$	NP_193067.3	+	Modelos (LANDMARK)
4	Cytochrome P450 76C6	<i>Arabidopsis thaliana</i>	82%	21,72%	$1e^{-18}$	NP_174633.1	+	Modelos (LANDMARK)
5	Cytochrome P450 71	<i>Arabidopsis thaliana</i>	82%	22,68%	$9e^{-18}$	NP_180635.3	+	Modelos (LANDMARK)
6	Cytochrome P450 82C2	<i>Arabidopsis thaliana</i>	69%	25,31%	$1e^{-17}$	NP_194925.1	+	Modelos (LANDMARK)
7	Cytochrome P450 79A2	<i>Arabidopsis thaliana</i>	82%	22,01%	$4e^{-17}$	NP_568153.2	+	Modelos (LANDMARK)
8	Cytochrome P450 71A25	<i>Arabidopsis thaliana</i>	82%	22,25%	$6e^{-17}$	NP_680107.1	+	Modelos (LANDMARK)
9	Cytochrome P450 87A6	<i>Arabidopsis thaliana</i>	87%	24,95%	$6e^{-17}$	NP_176674.1	+	Modelos (LANDMARK)
10	Cytochrome P450 706A5	<i>Arabidopsis thaliana</i>	43%	25,69%	$3e^{-16}$	NP_192968.3	+	Modelos (LANDMARK)
11	Cytochrome P450 71A26.1	<i>Glycine max</i>	87%	24,54%	$9e^{-24}$	XP_003533612.2	-	Modelos (LANDMARK)
12	Cytochrome P450 71A26.2	<i>Glycine max</i>	82%	24,25%	$1e^{-21}$	XP_003548205.1	-	Modelos (LANDMARK)
13	Cytochrome P450 82A2.1	<i>Glycine max</i>	71%	25,62%	$2e^{-19}$	NP_001240077.1	-	Modelos (LANDMARK)
14	Cytochrome P450 71A26.3	<i>Glycine max</i>	82%	24,62%	$5e^{-19}$	XP_014617658.1	-	Modelos (LANDMARK)
15	Cytochrome P450 93A3-like	<i>Glycine max</i>	82%	23,77%	$5e^{-19}$	XP_003537188.1	-	Modelos (LANDMARK)
16	Cytochrome P450 84A1	<i>Glycine max</i>	87%	22,67%	$7e^{-19}$	XP_003547953.1	-	Modelos (LANDMARK)
17	Flavone synthase II	<i>Glycine max</i>	63%	23,96%	$8e^{-19}$	NP_001241129.1	-	Modelos (LANDMARK)
18	Cytochrome P450 71A9 precursor	<i>Glycine max</i>	86%	24,06%	$1e^{-18}$	NP_001239692.1	-	Modelos (LANDMARK)
19	Cytochrome P450 82A4-like.1	<i>Glycine max</i>	69%	22,41%	$4e^{-18}$	XP_003554859.1	-	Modelos (LANDMARK)
20	Cytochrome P450 71A1-like.1	<i>Glycine max</i>	87%	23,78%	$7e^{-18}$	NP_001304406.1	-	Modelos (LANDMARK)

21	Cytochrome P450 71D8-like.1	<i>Glycine max</i>	82%	24,16%	1e-17	XP_003520239.1	-	Modelos (LANDMARK)
22	Cytochrome P450 71A24	<i>Glycine max</i>	78%	22,97%	2e-17	XP_003529310.2	-	Modelos (LANDMARK)
23	Cytochrome P450 93B16	<i>Glycine max</i>	63%	23,96%	2e-17	NP_001241007.1	-	Modelos (LANDMARK)
24	Cytochrome P450 71D8-like.2	<i>Glycine max</i>	82%	23,16%	3e-17	XP_003516584.1	-	Modelos (LANDMARK)
25	Cytochrome P450 71A26.4	<i>Glycine max</i>	78%	23,97%	6e-17	XP_014617657.1	-	Modelos (LANDMARK)
26	Cytochrome P450 82A2.2	<i>Glycine max</i>	66%	24,06%	7e-17	XP_003546624.1	-	Modelos (LANDMARK)
27	Cytochrome P450 71D8-like.3	<i>Glycine max</i>	84%	23,66%	9e-17	XP_014618600.1	-	Modelos (LANDMARK)
28	Cytochrome P450 71D8-like.4	<i>Glycine max</i>	84%	22,63%	9e-17	XP_003535171.1	-	Modelos (LANDMARK)
29	Cytochrome P450 71D8-like.5	<i>Glycine max</i>	84%	23,66%	1e-16	XP_003535172.1	-	Modelos (LANDMARK)
30	Cytochrome P450 71A1.2	<i>Glycine max</i>	55%	24,36%	1e-16	XP_003522006.3	-	Modelos (LANDMARK)
31	Cytochrome P450 71A26	<i>Glycine max</i>	67%	25,00%	1e-16	XP_003549108.1	-	Modelos (LANDMARK)
32	Cytochrome P450 71D8-like.6	<i>Glycine max</i>	84%	23,67%	2e-16	XP_003537208.2	-	Modelos (LANDMARK)
33	Cytochrome P450 71D8-like.7	<i>Glycine max</i>	84%	23,47%	2e-16	XP_014618299.1	-	Modelos (LANDMARK)
34	Cytochrome P450 82A4-like.2	<i>Glycine max</i>	65%	25,13%	2e-16	XP_003554118.2	-	Modelos (LANDMARK)
35	Cytochrome P450 71D8-like.8	<i>Glycine max</i>	82%	23,53%	3e-16	XP_003520242.1	-	Modelos (LANDMARK)
36	Cytochrome P450 71A1.3	<i>Glycine max</i>	77%	24,03%	3e-16	XP_003549825.1	-	Modelos (LANDMARK)
37	Abietadienol/Abieta dienal oxidase	<i>Glycine max</i>	33%	31,18%	3e-16	XP_006587411.1	-	Modelos (LANDMARK)
38	Cytochrome P450 71A6	<i>Nepeta racemosa</i>	64%	24,06%	2e-16	O04164	DS	UNIPROT
39	Cytochrome P450 71D20	<i>Nicotiana tabacum</i>	82%	24,24%	9e-17	Q94FM7	DS	UNIPROT
40	Cytochrome P450 99A2	<i>Oryza sativa japonica</i>	86%	23,68%	2e-17	Q7X7X4	DS	UNIPROT
41	Cytochrome P450 703A2	<i>Oryza sativa japonica</i>	64%	21,89%	3e-17	Q7EZR4	DS	UNIPROT
42	Cytochrome P450 76M8	<i>Oryza sativa japonica</i>	67%	23,97%	3e-17	Q6YTF1	DS	UNIPROT
43	Cytochrome P450 78A11	<i>Oryza sativa japonica</i>	85%	23,76%	6e-17	Q7Y1V5	DS	UNIPROT
44	Cytochrome P450 76M7	<i>Oryza sativa japonica</i>	78%	22,65%	3e-16	Q69X58	DS	UNIPROT
45	Cytochrome P450 71AJ4	<i>Pastinaca sativa</i>	86%	23,60%	5e-21	C0SJS3	DS	UNIPROT
46	Cytochrome P450 71AJ3	<i>Pastinaca sativa</i>	82%	23,23%	3e-19	C0SJS2	DS	UNIPROT
47	Cytochrome P450 75B2	<i>Petunia x hybrida</i>	65%	24,87%	2e-16	Q9SBQ9	DS	UNIPROT
48	Cytochrome P450 71D6	<i>Solanum chacoense</i>	82%	23,72%	3e-16	P93530	DS	UNIPROT
49	Cytochrome P450 71A4	<i>Solanum melongena</i>	84%	23,16%	3e-17	P37117	DS	UNIPROT
50	Cytochrome P450 71A2	<i>Solanum melongena</i>	86%	22,87%	1e-16	P37118	DS	UNIPROT
51	Cytochrome P450 Tp4149	<i>Tanacetum parthenium</i>	84%	22,93%	4e-19	X2ESR4	DS	UNIPROT
52	Cytochrome P450 71BL2	<i>Tanacetum parthenium</i>	87%	24,03%	6e-18	X2JE85	DS	UNIPROT
53	Cytochrome P450 71AJ2	<i>Apium graveolens</i>	82%	23,59%	1e-18	C0SJS4.1	DS	UNIPROT

54	Cytochrome P450 71D1V2	<i>Catharanthus roseus</i>	65%	26,68%	4e ⁻¹⁸	IITEM1.1	DS	UNIPROT
55	Cytochrome P450 V19H	<i>Catharanthus roseus</i>	82%	23,13%	4e ⁻¹⁷	P0DO22.1	DS	UNIPROT
56	Cytochrome P450 71D521	<i>Catharanthus roseus</i>	85%	21,21%	3e ⁻¹⁶	A0A343URW6.1	DS	UNIPROT
57	Cytochrome P450 71BL3	<i>Cichorium intybus</i>	87%	23,51%	8e ⁻¹⁷	G3GBK0.1	DS	UNIPROT
58	Protopine 6-monooxygenase	<i>Eschscholzia californica</i>	40%	26,99%	5e ⁻¹⁷	F2Z9C1.1	DS	UNIPROT
59	Cytochrome P450 71AY5	<i>Gelsemium sempervirens</i>	82%	24,04%	1e ⁻¹⁹	P0DO14.1	DS	UNIPROT
60	Cytochrome P450 71BL2	<i>Lactuca sativa</i>	87%	22,89%	4e ⁻¹⁶	F8S1I0.1	+	UNIPROT
61	Cytochrome P450 71D11	<i>Lotus japonicus</i>	34%	30,89%	3e ⁻¹⁶	O22307.1	DS	UNIPROT
62	Cytochrome P450 71A8	<i>Mentha x piperita</i>	67%	23,59%	1e ⁻¹⁸	Q42716.1	DS	UNIPROT
63	(+)-pulegone 9-hydroxylase	<i>Mentha x piperita</i>	77%	24,02%	3e ⁻¹⁸	Q947B7.1	DS	UNIPROT
64	Protopine 6-hydroxylase	<i>Papaver somniferum</i>	45%	27,13%	3e ⁻¹⁶	L7X0L7.1	DS	UNIPROT
65	Cytochrome P450 76AH1	<i>Salvia miltiorrhiza</i>	31%	26,74%	8e ⁻¹¹	5YLW_A	DS	PDB
66	Trans-cinnamate 4-monooxygenase	<i>Sorghum bicolor</i>	26%	27,21%	2e ⁻⁰⁶	6VBY_A	-	PDB

Fonte: Autor (2021).

Para a proteína Shadow (SAD) (Tabela 5), um quantitativo de 53 proteínas foi encontrado nos três bancos de dados (Modelos/LANDMARK, UNIPROT e PDB). A porcentagem de cobertura desta proteína variou de 28% a 89%, já a similaridade variou de 20,04% a 32,69%. O valor-E para o banco de dados Modelos (LANDMARK) foi de 1,00e⁻²⁵, referente a proteína *Geraniol 8-hydroxylase*, para o banco de dados UNIPROT foi de 1,00e⁻²³ para a proteína Cytochrome P450 76A2 e, por último, valor-E igual a 2,00e⁻¹⁷ para a proteína *Cytochrome P450 76AH1* no banco de dados PDB. Em relação a produção de 20E, os organismos que produzem foram *Arabidopsis thaliana*, *Prunus mume* e *Salvia miltiorrhiza*, enquanto que para *Glycine max* e o *Sorghum bicolor* a presença de 20E não foi detectada. As demais espécies encontradas para esta proteína são classificadas como DS para a produção desse fitoecdisteroide.

Tabela 5 – Proteínas vegetais que apresentam similaridade à proteína Shadow de *Drosophila melanogaster* a partir da busca em banco de dados pela ferramenta BLAST/NCBI, apresentando a porcentagem de Cobertura (Cob.), de Similaridade (Sim.), o Valor-E, o acesso cadastrado, a presença ou ausência de 20E (+ ou -, respectivamente) (DS = desconhecido) e o banco de dados de origem.

N.	Proteína	Organismo	Cob.	Sim.	Valor-E	Acesso	20E	Banco de dados
1	Shadow	<i>Drosophila melanogaster</i>	100%	100%	0.0	NP_650123	+	Modelos (LANDMARK)
2	Cytochrome P450 706A6	<i>Arabidopsis thaliana</i>	86%	23,92%	4e ⁻²²	NP_192969.4	+	Modelos (LANDMARK)
3	Cytochrome P450 76C3	<i>Arabidopsis thaliana</i>	78%	22,05%	1e ⁻²¹	NP_182082.2	+	Modelos (LANDMARK)
4	Cytochrome P450 91A2	<i>Arabidopsis thaliana</i>	76%	23,13%	4e ⁻²¹	NP_195459.1	+	Modelos (LANDMARK)

5	Cytochrome P450 81F4	<i>Arabidopsis thaliana</i>	77%	24,15%	3e-20	NP_195457.1	+	Modelos (LANDMARK)
6	Cytochrome P450 709B3	<i>Arabidopsis thaliana</i>	78%	23,29%	2e-19	NP_194501.1	+	Modelos (LANDMARK)
7	Cytochrome P450 76C3	<i>Arabidopsis thaliana</i>	32%	27,87%	4e-19	NP_001323696.1	+	Modelos (LANDMARK)
8	Cytochrome P450 705A13	<i>Arabidopsis thaliana</i>	41%	27,64%	4e-19	NP_179026.1	+	Modelos (LANDMARK)
9	Cytochrome P450 703A2	<i>Arabidopsis thaliana</i>	83%	23,54%	5e-19	NP_171635.1	+	Modelos (LANDMARK)
10	Cytochrome P450 81H1	<i>Arabidopsis thaliana</i>	55%	26,27%	9e-19	NP_568024.2	+	Modelos (LANDMARK)
11	Cytochrome P450 82C4	<i>Arabidopsis thaliana</i>	36%	28,23%	1e-18	NP_194922.1	+	Modelos (LANDMARK)
12	Cytochrome P450 705A22	<i>Arabidopsis thaliana</i>	31%	30,94%	2e-18	NP_188648.1	+	Modelos (LANDMARK)
13	Cytochrome P450 705A22	<i>Arabidopsis thaliana</i>	31%	30,94%	2e-18	NP_001078193.1	+	Modelos (LANDMARK)
14	Cytochrome P450 82C2	<i>Arabidopsis thaliana</i>	87%	22,91%	3e-18	NP_194925.1	+	Modelos (LANDMARK)
15	Cytochrome P450 705A30	<i>Arabidopsis thaliana</i>	55%	23,65%	3e-17	NP_188730.2	+	Modelos (LANDMARK)
16	Cytochrome P450 705A23	<i>Arabidopsis thaliana</i>	39%	26,20%	3e-17	NP_188649.1	+	Modelos (LANDMARK)
17	Cytochrome P450 711A1	<i>Arabidopsis thaliana</i>	28%	31,10%	4e-17	NP_850074.1	+	Modelos (LANDMARK)
18	Cytochrome P450 705A25	<i>Arabidopsis thaliana</i>	30%	28,09%	4e-17	NP_175471.1	+	Modelos (LANDMARK)
19	Geraniol 8-hydroxylase	<i>Glycine max</i>	82%	24,39%	1e-25	XP_003557080.2	-	Modelos (LANDMARK)
20	Geraniol 8-hydroxylase	<i>Glycine max</i>	81%	24,23%	9e-23	XP_003556053.1	-	Modelos (LANDMARK)
21	Geraniol 8-hydroxylase	<i>Glycine max</i>	81%	23,99%	6e-22	XP_006606065.1	-	Modelos (LANDMARK)
22	Geraniol 8-hydroxylase	<i>Glycine max</i>	77%	23,13%	9e-21	XP_003543189.1	-	Modelos (LANDMARK)
23	Flavone synthase II	<i>Glycine max</i>	30%	31,61%	1e-19	NP_001241129.1	-	Modelos (LANDMARK)
24	Cytochrome P450 93B16	<i>Glycine max</i>	31%	31,46%	4e-19	NP_001241007.1	-	Modelos (LANDMARK)
25	Geraniol 8-hydroxylase	<i>Glycine max</i>	77%	23,70%	4e-18	XP_003556289.1	-	Modelos (LANDMARK)
26	Cytochrome P450 93A3	<i>Glycine max</i>	36%	27,70%	4e-18	XP_014628608.2	-	Modelos (LANDMARK)
27	Cytochrome P450 82D47	<i>Glycine max</i>	36%	31,40%	5e-18	XP_003548649.3	-	Modelos (LANDMARK)
28	Cytochrome P450 82D47	<i>Glycine max</i>	36%	31,40%	5e-18	XP_003547786.2	-	Modelos (LANDMARK)
29	Isoflavone synthase 1 precursor	<i>Glycine max</i>	77%	21,65%	6e-18	NP_001236022.1	-	Modelos (LANDMARK)
30	Cytochrome P450 93A3-precursor	<i>Glycine max</i>	37%	27,16%	7e-18	NP_001304452.2	-	Modelos (LANDMARK)
31	Cytochrome P450 93C1	<i>Glycine max</i>	77%	20,04%	8e-18	NP_001238515.2	-	Modelos (LANDMARK)
32	Cytochrome P450 82C4	<i>Glycine max</i>	36%	31,71%	1e-17	XP_003523888.1	-	Modelos (LANDMARK)
33	Cytochrome P450 71A1	<i>Glycine max</i>	76%	23,46%	1e-17	XP_003522006.3	-	Modelos (LANDMARK)
34	Flavonoide 3'-monooxygenase	<i>Glycine max</i>	89%	23,50%	3e-17	XP_003547756.2	-	Modelos (LANDMARK)
35	Cytochrome P450 82D47	<i>Glycine max</i>	45%	28,31%	3e-17	XP_003523886.1	-	Modelos (LANDMARK)
36	Cytochrome P450 71A1	<i>Glycine max</i>	36%	26,57%	3e-17	XP_003528553.1	-	Modelos (LANDMARK)
37	Cytochrome P450 73A11	<i>Glycine max</i>	85%	24,40%	4e-17	NP_001237317.1	-	Modelos (LANDMARK)

38	Cytochrome P450 81F1	<i>Arabidopsis thaliana</i>	76%	23,13%	3e ⁻²¹	O65790.2	+	UNIPROT
39	Cytochrome P450 81F3	<i>Arabidopsis thaliana</i>	82%	21,66%	3e ⁻¹⁷	Q0WTF4.1	+	UNIPROT
40	Cytochrome P450 81D11	<i>Arabidopsis thaliana</i>	48%	26,22%	5e ⁻¹⁷	Q9LHA1.1	+	UNIPROT
41	Cytochrome P450 76B6	<i>Catharanthus roseus</i>	82%	21,76%	2e ⁻²⁰	Q8VWZ7.1	DS	UNIPROT
42	Cytochrome P450C4H	<i>Catharanthus roseus</i>	85%	24,50%	3e ⁻¹⁷	P48522.1	DS	UNIPROT
43	Cytochrome P450 75B137	<i>Crocasmia x crocosmiflora</i>	82%	24,30%	3e ⁻¹⁷	A0A4D6Q415.1	DS	UNIPROT
44	Cytochrome P450 93C2	<i>Glycyrrhiza echinata</i>	78%	22,05%	1e ⁻¹⁹	Q9SXS3.1	DS	UNIPROT
45	Cytochrome P450 93C2	<i>Glycyrrhiza uralensis</i>	78%	22,49%	9e ⁻²¹	G4XV71.2	DS	UNIPROT
46	Cytochrome P450 C3H	<i>Narcissus pseudonarcissus</i>	37%	28,91%	9e ⁻¹⁸	A0A2H5AIX6.1	DS	UNIPROT
47	Cytochrome P450 736A12	<i>Panax ginseng</i>	84%	23,93%	9e ⁻²⁰	H2DH18.1	DS	UNIPROT
48	N-methylstylopine hydroxylase	<i>Papaver somniferum</i>	88%	21,80%	2e ⁻¹⁹	L7X3S1.1	DS	UNIPROT
49	Cytochrome P450 82A1	<i>Pisum sativum</i>	36%	32,69%	9e ⁻²¹	Q43068.2	DS	UNIPROT
50	Cytochrome P450 71AU50	<i>Prunus mume</i>	31%	30,00%	7e ⁻¹⁷	A0A068Q5V6.1	+	UNIPROT
51	Cytochrome P450 76A2	<i>Solanum melongena</i>	83%	24,63%	1e ⁻²³	P37122.1	DS	UNIPROT
52	Cytochrome P450 98A1	<i>Sorghum bicolor</i>	37%	28,91%	7e ⁻¹⁹	O48956.1	-	UNIPROT
53	Cytochrome P450 76AH1	<i>Sorghum bicolor</i>	82%	23,12%	2e ⁻¹⁷	6VBY_A	-	PDB
54	Cytochrome P450 76AH1	<i>Salvia miltiorrhiza</i>	81%	22,32%	8e ⁻¹⁷	5YLW_A	+	PDB

Fonte: Autor (2021).

Dentre as proteínas utilizadas para a busca em bancos de dados, a proteína Disembodied (DIB) foi a única a apresentar apenas um resultado (Tabela 6), proveniente do banco de dados UNIPROT. A proteína Cytochrome P450 94A1 de *Vicia sativa* apresentou uma cobertura de 70% e similaridade de 25%, aproximadamente. O valor-E referente ao banco de dados encontrado (UNIPROT) foi de 5,00e⁻²⁴. Nessa espécie não foi detectada a presença de 20E.

Tabela 6 – Proteína vegetal que apresenta similaridade à proteína Disembodied de *Drosophila melanogaster* a partir da busca em banco de dados pela ferramenta BLAST/NCBI, apresentando a porcentagem de Cobertura (Cob.), de Similaridade (Sim.), o Valor-E, o acesso cadastrado, a presença ou ausência de 20E (+ ou -, respectivamente) (DS = desconhecido) e o banco de dados de origem.

N.	Proteína	Organismo	Cob.	Sim.	Valor-E	Acesso	20E	Banco de dados
1	Disembodied	<i>Drosophila melanogaster</i>	100%	100%	0.0	NP_524810.2	+	LANDMARK
2	Cytochrome P450 94A1	<i>Vicia sativa</i>	69%	25,21%	5e ⁻²⁴	O81117.2	-	UNIPROT

Fonte: Autor (2021).

Para a proteína Phantom (PHT) foram encontradas 24 proteínas distribuídas em quatro organismos distintos (Tabela 7). A porcentagem de cobertura para esta proteína variou de 30% a 95%, enquanto que a similaridade variou de 23,32% a 28,44%. O valor-E no banco de dados

Modelos (LANDMARK) atingiu um valor máximo de $6,00e^{-41}$ para a proteína Cytochrome P450 736A12 e sua isoforma, em relação ao banco de dados UNIPROT, o valor-E foi de $8,00e^{-39}$ para a proteína Cytochrome P450 71B11 e, por último, valor-E de $1,00e^{-29}$ para a proteína Cinnamate 4-hydroxylase (C4H1) no banco de dados PDB. Para a produção de 20E, apenas *Arabidopsis thaliana* e *Salvia miltiorrhiza* são confirmadas como produtoras, enquanto que em *Glycine max* e *Sorghum bicolor* não foi detectado a presença de 20E.

Tabela 7 – Proteínas vegetais que apresentam similaridade à proteína Phantom de *Drosophila melanogaster* a partir da busca em banco de dados pela ferramenta BLAST/NCBI, apresentando a porcentagem de Cobertura (Cob.), de Similaridade (Sim.), o Valor-E, o acesso cadastrado, a presença ou ausência de 20E (+ ou -, respectivamente) (DS = desconhecido) e o banco de dados de origem.

N.	Proteína	Organismo	Cob.	Sim.	Valor-E	Acesso	20E	Banco de dados
1	Phantom	<i>Drosophila melanogaster</i>	100%	100%	0.0	NP_001285414.1	+	Modelos (LANDMARK)
2	Cytochrome P450 706A4	<i>Arabidopsis thaliana</i>	91%	27,21%	$6e^{-40}$	NP_192967.1	+	Modelos (LANDMARK)
3	Cytochrome P450 71B11	<i>Arabidopsis thaliana</i>	90%	26,36%	$1e^{-38}$	NP_197894.1	+	Modelos (LANDMARK)
4	Cytochrome P450 76C1	<i>Arabidopsis thaliana</i>	87%	25,10%	$3e^{-38}$	NP_850439.1	+	Modelos (LANDMARK)
5	Cytochrome P450 706A6	<i>Arabidopsis thaliana</i>	91%	26,87%	$9e^{-38}$	NP_192969.4	+	Modelos (LANDMARK)
6	Cytochrome P450 736A12 -X1	<i>Glycine max</i>	94%	26,85%	$6e^{-41}$	XP_003531375.1	-	Modelos (LANDMARK)
7	Cytochrome P450 736A12	<i>Glycine max</i>	93%	25,99%	$7e^{-41}$	XP_003530005.1	-	Modelos (LANDMARK)
8	Uncharacterized protein	<i>Glycine max</i>	95%	27,12%	$1e^{-40}$	XP_003534174.2	-	Modelos (LANDMARK)
9	Cytochrome P450 736A12	<i>Glycine max</i>	92%	25,50%	$1e^{-40}$	XP_003528941.2	-	Modelos (LANDMARK)
10	Cytochrome P450 71D11-X2	<i>Glycine max</i>	92%	24,64%	$7e^{-38}$	XP_014633547.1	-	Modelos (LANDMARK)
11	Cytochrome P450 71D11-X1	<i>Glycine max</i>	92%	24,64%	$7e^{-38}$	XP_003529194.2	-	Modelos (LANDMARK)
12	Cytochrome P450 736A12-like	<i>Glycine max</i>	87%	27,75%	$1e^{-37}$	XP_003524281.3	-	Modelos (LANDMARK)
13	Cytochrome P450 71D8-like	<i>Glycine max</i>	86%	25,88%	$8e^{-37}$	XP_003516584.1	-	Modelos (LANDMARK)
14	Cytochrome P450 736A12	<i>Glycine max</i>	91%	24,45%	$3e^{-36}$	XP_003531374.1	-	Modelos (LANDMARK)
15	Cytochrome P450 736A12	<i>Glycine max</i>	86%	25,98%	$2e^{-35}$	XP_014617743.1	-	Modelos (LANDMARK)
16	Cytochrome P450 736A12	<i>Glycine max</i>	92%	23,78%	$3e^{-35}$	XP_003532846.1	-	Modelos (LANDMARK)
17	Cytochrome P450 71B11	<i>Arabidopsis thaliana</i>	90%	26,36%	$8e^{-39}$	P58049.1	+	UNIPROT
18	Cytochrome P450 76C1	<i>Arabidopsis thaliana</i>	87%	25,10%	$2e^{-38}$	O64636.1	+	UNIPROT
19	Cytochrome P450 97A3	<i>Arabidopsis thaliana</i>	35%	28,44%	$2e^{-14}$	6J94_A	+	PDB
20	Cytochrome P450 97A3	<i>Arabidopsis thaliana</i>	35%	28,44%	$2e^{-14}$	6L8I_A	+	PDB
21	Cytochrome P450 90B1.1	<i>Arabidopsis thaliana</i>	30%	24,46%	$3e^{-07}$	6A17_A	+	PDB
22	Cytochrome P450 90B1.2	<i>Arabidopsis thaliana</i>	30%	24,46%	$4e^{-07}$	6A15_A	+	PDB
23	Cytochrome P450 90B1.3	<i>Arabidopsis thaliana</i>	30%	23,32%	$1e^{-06}$	6A16_A	+	PDB

24	Cytochrome P450 76AH1	<i>Salvia miltiorrhiza</i>	89%	24,10%	8e ⁻²⁹	5YLW_A	+	PDB
25	Cinnamate 4-hydroxylase (C4H1)	<i>Sorghum bicolor</i>	86%	26,65%	1e ⁻²⁹	6VBY_A	-	PDB

Fonte: Autor (2021).

A busca utilizando a sequência da proteína Ecdysone Oxidase (EO) obteve um resultado de 35 proteínas similares em dois bancos de dados (Modelos – LANDMARK e PDB) distribuídas em quatro espécies (Tabela 8). A porcentagem de cobertura variou de 43% a 87%, enquanto que a similaridade foi de 23,99% a 32,79%. O valor-E máximo foi 1,00e⁻³¹, correspondente a Hydroxynitrile lyase 1 de *Prunus dulcis*, para o banco de dados PDB, e de 1,00e⁻²⁹, correspondente a proteína (R)-mandelonitrile lyase-like, de *Glycine max*, para o banco de dados Modelos (LANDMARK). Apenas em *Prunus mume* e *Arabidopsis thaliana* foi detectada a presença da molécula 20E.

Tabela 8 – Proteínas vegetais que apresentam similaridade à proteína Ecdysone Oxidase de *Drosophila melanogaster* a partir da busca em banco de dados pela ferramenta BLAST/NCBI, apresentando a porcentagem de Cobertura (Cob.), de Similaridade (Sim.), o Valor-E, o acesso cadastrado, a presença ou ausência de 20E (+ ou -, respectivamente) (DS = desconhecido) e o banco de dados de origem.

N.	Proteína	Organismo	Cob.	Sim.	Valor-E	Acesso	20E	Banco de dados
1	Ecdysone Oxidase	<i>Drosophila melanogaster</i>	100%	100%	0.0	NP_572986.1	+	Modelos (LANDMARK)
2	Glucose-methanolcholine 1	<i>Arabidopsis thaliana</i>	85%	26,30%	3e ⁻²⁷	NP_565050.1	+	Modelos (LANDMARK)
3	Glucose-methanolcholine 2	<i>Arabidopsis thaliana</i>	84%	25,21%	6e ⁻²⁷	NP_172718.2	+	Modelos (LANDMARK)
4	Glucose-methanolcholine 3	<i>Arabidopsis thaliana</i>	85%	26,59%	3e ⁻²⁶	NP_001154469.1	+	Modelos (LANDMARK)
5	Glucose-methanolcholine 4	<i>Arabidopsis thaliana</i>	86%	24,28%	9e ⁻²⁵	NP_172871.1	+	Modelos (LANDMARK)
6	Glucose-methanolcholine 5	<i>Arabidopsis thaliana</i>	85%	26,71%	1e ⁻²⁴	NP_567032.1	+	Modelos (LANDMARK)
7	Glucose-methanolcholine 6	<i>Arabidopsis thaliana</i>	86%	24,11%	1e ⁻²³	NP_001321001.1	+	Modelos (LANDMARK)
8	Glucose-methanolcholine 7	<i>Arabidopsis thaliana</i>	43%	32,79%	7e ⁻²¹	NP_177448.1	+	Modelos (LANDMARK)
9	Glucose-methanolcholine 8	<i>Arabidopsis thaliana</i>	86%	25,21%	1e ⁻²⁰	NP_001332516.1	+	Modelos (LANDMARK)
10	Glucose-methanolcholine 9	<i>Arabidopsis thaliana</i>	75%	27,79%	5e ⁻¹⁷	NP_563939.1	+	Modelos (LANDMARK)
11	Glucose-methanolcholine 10	<i>Arabidopsis thaliana</i>	85%	24,22%	2e ⁻¹⁵	NP_001330130.1	+	Modelos (LANDMARK)
12	Glucose-methanolcholine 11	<i>Arabidopsis thaliana</i>	44%	28,53%	4e ⁻¹⁴	NP_001318783.1	+	Modelos (LANDMARK)
13	Glucose-methanolcholine 12	<i>Arabidopsis thaliana</i>	44%	28,53%	4e ⁻¹⁴	NP_200008.1	+	Modelos (LANDMARK)
14	Glucose-methanolcholine 13	<i>Arabidopsis thaliana</i>	83%	24,54%	2e ⁻¹³	NP_200006.1	+	Modelos (LANDMARK)
15	(R)-mandelonitrile lyase-like	<i>Glycine max</i>	87%	28,57%	1e ⁻²⁹	XP_014633027.2	-	Modelos (LANDMARK)
16	Protein HOTHEAD 1	<i>Glycine max</i>	84%	26,50%	3e ⁻²⁷	XP_003543593.1	-	Modelos (LANDMARK)
17	Protein HOTHEAD 2	<i>Glycine max</i>	85%	25,42%	7e ⁻²⁷	XP_003531615.1	-	Modelos (LANDMARK)
18	Protein HOTHEAD 3	<i>Glycine max</i>	84%	26,92%	1e ⁻²⁶	XP_003546687.1	-	Modelos (LANDMARK)

19	Protein HOTHEAD-iX2	<i>Glycine max</i>	85%	25,70%	2e ⁻²⁶	XP_006590535.1	-	Modelos (LANDMARK)
20	Protein HOTHEAD-iX1	<i>Glycine max</i>	85%	25,70%	2e ⁻²⁶	XP_006590534.1	-	Modelos (LANDMARK)
21	Protein HOTHEAD 4	<i>Glycine max</i>	85%	25,04%	2e ⁻²⁵	XP_003549865.1	-	Modelos (LANDMARK)
22	Protein HOTHEAD 5	<i>Glycine max</i>	85%	23,99%	1e ⁻²³	XP_003525734.1	-	Modelos (LANDMARK)
23	(R)-mandelonitrile lyase-like 1	<i>Glycine max</i>	72%	29,55%	3e ⁻²⁰	XP_003517995.2	-	Modelos (LANDMARK)
24	Protein HOTHEAD 6	<i>Glycine max</i>	85%	25,25%	4e ⁻²⁰	XP_006606089.1	-	Modelos (LANDMARK)
25	(R)-mandelonitrile lyase-like 2	<i>Glycine max</i>	72%	29,55%	7e ⁻²⁰	XP_003520079.1	-	Modelos (LANDMARK)
26	Protein HOTHEAD 7	<i>Glycine max</i>	69%	24,59%	5e ⁻¹⁸	XP_014633109.1	-	Modelos (LANDMARK)
27	Protein HOTHEAD isoform X1	<i>Glycine max</i>	44%	28,48%	3e ⁻¹³	XP_003557066.1	-	Modelos (LANDMARK)
28	Protein HOTHEAD isoform X2	<i>Glycine max</i>	44%	28,48%	3e ⁻¹³	XP_006607079.1	-	Modelos (LANDMARK)
29	Protein HOTHEAD 8	<i>Glycine max</i>	44%	28,29%	2e ⁻¹¹	XP_014628783.1	-	Modelos (LANDMARK)
30	Hydroxynitrile lyase 1	<i>Prunus dulcis</i>	0,86	26,06%	1e ⁻³¹	1JU2_A	-	PDB
31	Hydroxynitrile lyase 2	<i>Prunus dulcis</i>	0,86	26,06%	1e ⁻³¹	3GDN_A	-	PDB
32	Endo-deglycosylated hydroxynitrile lyase 1	<i>Prunus dulcis</i>	0,86	26,01%	2e ⁻²⁹	6LQY_A	-	PDB
33	Deglycosylated hydroxynitrile lyase	<i>Prunus dulcis</i>	0,86	26,01%	3e ⁻²⁹	6LR8_A	-	PDB
34	Endo-deglycosylated hydroxynitrile lyase 2	<i>Prunus dulcis</i>	0,86	26,09%	3e ⁻²⁹	6JBY_A	-	PDB
35	Hydroxynitrile lyase 3	<i>Prunus dulcis</i>	0,86	26,01%	4e ⁻²⁹	5EB4_A	-	PDB
36	Hydroxynitrile lyase 4	<i>Prunus mume</i>	0,85	25,60%	2e ⁻²⁹	3RED_A	+	PDB

Fonte: Autor (2021).

A proteína 3-dehydroecdysone 3beta-reductase (3DE3BR) foi isolada e caracterizada em *Spodoptera littoralis* (Tabela 9). Para a busca em banco de dados a partir de sua sequência de aminoácidos, foram encontradas 18 proteínas similares, em dois bancos de dados (Modelo – LANDMARK e UNIPROT), distribuídas em quatro espécies. A cobertura variou de 79% a 87% e a similaridade de 32,78% a 38,36%. Em relação ao valor-E, para o banco de dados Modelos (LANDMARK), o valor máximo foi de 2,00e⁻⁶⁴, enquanto que para o banco de dados UNIPROT foi de 4,00e⁻⁶¹. Apenas *Arabidopsis thaliana* apresentou a produção de 20E, enquanto que *Malus domestica* foi classificada como DS. Nas espécies *Glycine max* e *Zea mays* não foi detectado a presença de 20E.

Tabela 9 – Proteínas vegetais que apresentam similaridade à proteína 3-dehydroecdysone 3beta-reductase de *Spodoptera littoralis* a partir da busca em banco de dados pela ferramenta BLAST/NCBI, apresentando a porcentagem de Cobertura (Cob.), de Similaridade (Sim.), o Valor-E, o acesso cadastrado, a presença ou ausência de 20E (+ ou -, respectivamente) (DS = desconhecido) e o banco de dados de origem.

N.	Proteína	Organismo	Cob.	Sim.	Valor-E	Acesso	20E	Banco de dados
1	3-dehydroecdysone 3beta-reductase	<i>Spodoptera littoralis</i>	100%	100%	0	CAB41997	+	GenBank
2	NAD(P)-linked oxidoreductase	<i>Arabidopsis thaliana</i>	83%	38,36%	2e ⁻⁶⁴	NP_179722.1	+	LANDMARK (MODELO)
3	NAD(P)-linked oxidoreductase	<i>Arabidopsis thaliana</i>	84%	37,58%	3e ⁻⁶⁴	NP_179721.1	+	LANDMARK (MODELO)
4	NAD(P)-linked oxidoreductase	<i>Arabidopsis thaliana</i>	83%	33,67%	3e ⁻⁵⁵	NP_001118465.1	+	LANDMARK (MODELO)
5	NAD(P)-linked oxidoreductase	<i>Arabidopsis thaliana</i>	81%	33,92%	7e ⁻⁵⁵	NP_181315.2	+	LANDMARK (MODELO)
6	NAD(P)-linked oxidoreductase	<i>Arabidopsis thaliana</i>	83%	33,67%	4e ⁻⁵⁴	NP_001078019.1	+	LANDMARK (MODELO)
7	NAD(P)-linked oxidoreductase	<i>Arabidopsis thaliana</i>	85%	32,78%	2e ⁻⁵³	NP_565871.1	+	LANDMARK (MODELO)
8	Uncharacterized protein	<i>Glycine max</i>	84%	36,36%	1e ⁻⁶²	NP_001240166.1	-	LANDMARK (MODELO)
9	Uncharacterized protein	<i>Glycine max</i>	79%	35,59%	1e ⁻⁵⁶	XP_025982382.1	-	LANDMARK (MODELO)
10	Methylecgonone reductase	<i>Glycine max</i>	80%	36,93%	4e ⁻⁵⁶	XP_003534120.1	-	LANDMARK (MODELO)
11	Chalcone reductase CHR1	<i>Glycine max</i>	87%	33,76%	3e ⁻⁵⁵	NP_001235973.1	-	LANDMARK (MODELO)
12	NAD(P)H-dependent 6'-deoxychalcone synthase	<i>Glycine max</i>	80%	37,50%	1e ⁻⁵⁴	XP_003547552.1	-	LANDMARK (MODELO)
13	NAD(P)H-dependent 6'-deoxychalcone synthase isoform X1	<i>Glycine max</i>	87%	33,97%	3e ⁻⁵⁴	XP_014622008.1	-	LANDMARK (MODELO)
14	NADPH-dependent aldo-keto reductase, chloroplastic	<i>Glycine max</i>	83%	32,88%	5e ⁻⁵⁴	XP_003516897.1	-	LANDMARK (MODELO)
15	Uncharacterized protein	<i>Glycine max</i>	79%	34,33%	2e ⁻⁵³	XP_025982381.1	-	LANDMARK (MODELO)
16	Aldo-keto reductase family 4 member C10	<i>Arabidopsis thaliana</i>	81%	33,92%	5e ⁻⁵⁵	Q84TF0.1	+	UNIPROT
17	NAD(P)H-dependent 6'-deoxychalcone synthase	<i>Glycine max</i>	87%	33,76%	2e ⁻⁵⁵	P26690.1	-	UNIPROT
18	NADP-dependent D-sorbitol-6-phosphate dehydrogenase	<i>Malus domestica</i>	83%	36,75%	4e ⁻⁶¹	P28475.1	DS	UNIPROT
19	Deoxymugineic acid synthase 1	<i>Zea mays</i>	86%	35,56%	3e ⁻⁵⁵	B4F9A4.1	-	UNIPROT

Fonte: Autor (2021).

Assim como a proteína 3DE3BR, a proteína 3-dehydroecdysone 3alfa-reductase (3DE3AR) (Tabela 10) foi isolada e caracterizada em *Sposoptera littoralis*. Foram encontradas 36 proteínas nos três bancos de dados disponíveis (Modelos – LANDMARK, UNIPROT e PDB), distribuídas em 11 espécies de plantas. A porcentagem de cobertura variou de 75% a

97% e a similaridade de 30,50% a 37,38%. O valor-E para o banco de dados Modelos (LANDMARK) teve máximo de $6,00e^{-33}$, para UNIPROT foi de $5,00e^{-33}$ e do PDB foi de $5,00e^{-32}$. As espécies *Arabidopsis thaliana*, *Daucus carota*, *Oryza sativa Japonica* e *Datura stramonium* são confirmadas como produtoras de 20E, entretanto nas espécies *Glycine max*, *Hyoscyamus niger* *Brassica napus* não foi detectado a presença de 20E. As demais espécies são consideradas como DS a presença da molécula.

Tabela 10 – Proteínas vegetais que apresentam similaridade à proteína 3-dehydroecdysone 3alfa-reductase de *Spodoptera littoralis* a partir da busca em banco de dados pela ferramenta BLAST/NCBI, apresentando a porcentagem de Cobertura (Cob.), de Similaridade (Sim.), o Valor-E, o acesso cadastrado, a presença ou ausência de 20E (+ ou -, respectivamente) (DS = desconhecido) e o banco de dados de origem.

N.	Proteína	Organismo	Cob.	Sim.	Valor-E	Acesso	20E	Banco de dados
1	3-dehydroecdysone 3alfa-reductase	<i>Spodoptera littoralis</i>	100%	100%	0	AAF70499.1	+	GenBank
2	NAD(P)-binding Rossmann-fold superfamily protein	<i>Arabidopsis thaliana</i>	96%	34,26%	$6e^{-33}$	NP_187177.1	+	Modelos (LANDMARK)
3	Hydroxysteroid dehydrogenase 3	<i>Arabidopsis thaliana</i>	82%	37,38%	$2e^{-31}$	NP_190320.1	+	Modelos (LANDMARK)
4	NAD(P)-binding Rossmann-fold superfamily protein	<i>Arabidopsis thaliana</i>	97%	33,74%	$2e^{-31}$	NP_564216.1	+	Modelos (LANDMARK)
5	NAD(P)-binding Rossmann-fold superfamily protein	<i>Arabidopsis thaliana</i>	97%	32,68%	$3e^{-31}$	NP_564670.2	+	Modelos (LANDMARK)
6	NAD(P)-binding Rossmann-fold superfamily protein	<i>Arabidopsis thaliana</i>	97%	32,68%	$3e^{-31}$	NP_001322466.1	+	Modelos (LANDMARK)
7	Hydroxysteroid dehydrogenase 4	<i>Arabidopsis thaliana</i>	80%	36,82%	$9e^{-29}$	NP_001331862.1	+	Modelos (LANDMARK)
8	Hydroxysteroid dehydrogenase 5	<i>Arabidopsis thaliana</i>	88%	34,68%	$2e^{-28}$	NP_199871.1	+	Modelos (LANDMARK)
9	Hydroxysteroid dehydrogenase 6	<i>Arabidopsis thaliana</i>	81%	35,78%	$2e^{-28}$	NP_190319.1	+	Modelos (LANDMARK)
10	Hydroxysteroid dehydrogenase 7	<i>Arabidopsis thaliana</i>	75%	37,23%	$3e^{-28}$	NP_001154667.1	+	Modelos (LANDMARK)
11	Uncharacterized protein LOC100809384	<i>Glycine max</i>	96%	33,47%	$2e^{-32}$	NP_001241158.1	-	Modelos (LANDMARK)
12	Tropinone reductase	<i>Glycine max</i>	97%	33,20%	$3e^{-31}$	XP_003539545.1	-	Modelos (LANDMARK)
13	3-oxoacyl-[acyl-carrier-protein] reductase 4	<i>Glycine max</i>	97%	32,66%	$5e^{-30}$	XP_003538564.1	-	Modelos (LANDMARK)
14	Uncharacterized protein LOC100785449	<i>Glycine max</i>	97%	31,97%	$3e^{-29}$	NP_001239850.1	-	Modelos (LANDMARK)
15	Tropinone reductase 1	<i>Glycine max</i>	97%	33,61%	$5e^{-29}$	XP_003539556.1	-	Modelos (LANDMARK)
16	Tropinone reductase 2	<i>Glycine max</i>	97%	32,79%	$7e^{-29}$	XP_003538424.1	-	Modelos (LANDMARK)
17	3-oxoacyl-[acyl-carrier-protein] reductase 4	<i>Glycine max</i>	97%	32,52%	$2e^{-28}$	XP_003531181.1	-	Modelos (LANDMARK)
18	Glucose and ribitol dehydrogenase homolog 2	<i>Arabidopsis thaliana</i>	96%	34,26%	$5e^{-33}$	Q9MA93.1	+	UNIPROT

19	NADPH-dependent aldehyde reductase 1	<i>Arabidopsis thaliana</i>	97%	32,68%	9e ⁻³²	Q9FZ42.1	+	UNIPROT
20	11-beta-hydroxysteroid dehydrogenase-like 3	<i>Arabidopsis thaliana</i>	82%	37,38%	1e ⁻³¹	Q9STY7.1	+	UNIPROT
21	3-oxoacyl-[acyl-carrier-protein] reductase	<i>Arabidopsis thaliana</i>	97%	33,74%	1e ⁻³¹	P33207.2	+	UNIPROT
22	11-beta-hydroxysteroid dehydrogenase-like 4A	<i>Arabidopsis thaliana</i>	88%	34,68%	1e ⁻²⁸	P0DKC7.1	+	UNIPROT
23	11-beta-hydroxysteroid dehydrogenase-like 2	<i>Arabidopsis thaliana</i>	81%	35,78%	2e ⁻²⁸	Q9STY8.1	+	UNIPROT
24	Tropinone reductase 3	<i>Arabidopsis thaliana</i>	96%	31,28%	1e ⁻²⁷	O49332.1	+	UNIPROT
25	3-oxoacyl-[acyl-carrier-protein] reductase	<i>Cuphea lanceolata</i>	97%	34,27%	4e ⁻³⁰	P28643.1	DS	UNIPROT
26	Glucose and ribitol dehydrogenase	<i>Daucus carota</i>	97%	33,73%	2e ⁻²⁹	Q5KTS5.1	+	UNIPROT
27	Tropinone reductase 4	<i>Hyoscyamus niger</i>	95%	32,79%	3e ⁻³⁰	P50164.1	-	UNIPROT
28	Noroxomaritidine/norcoarugsodine reductase	<i>Narcissus pseudonarcissus</i>	95%	32,51%	4e ⁻²⁸	A0A1A9TAK5.1	DS	UNIPROT
29	(+)-cis,trans-nepetalactol synthase NEPS1	<i>Nepeta racemosa</i>	97%	32,27%	5e ⁻²⁸	A0A3Q8GL18.1	DS	UNIPROT
30	Glucose and ribitol dehydrogenase	<i>Oryza sativa Japonica</i>	97%	30,50%	2e ⁻²⁸	Q75KH3.2	+	UNIPROT
31	11-beta-hydroxysteroid dehydrogenase A	<i>Sesamum indicum</i>	78%	35,53%	8e ⁻³³	Q93W57.1	DS	UNIPROT
32	Beta-Keto Acyl Carrier Protein Reductase	<i>Brassica napus</i>	97%	33,33%	5e ⁻³²	1EDO_A	-	PDB
33	Tropinone Reductase-II	<i>Datura stramonium</i>	95%	32,38%	3e ⁻³¹	2AE1_A	+	PDB
34	Tropinone Reductase-II	<i>Datura stramonium</i>	95%	32,38%	3e ⁻³¹	1IPE_A	+	PDB
35	Tropinone Reductase-I	<i>Datura stramonium</i>	95%	31,84%	4e ⁻²⁹	1AE1_A	+	PDB
36	Noroxomaritidine/Norcoarugsodine Reductase	<i>Narcissus pseudonarcissus</i>	95%	32,51%	8e ⁻²⁹	5FF9_A	DS	PDB
37	Noroxomaritidine/Norcoarugsodine Reductase	<i>Narcissus pseudonarcissus</i>	95%	32,51%	2e ⁻²⁸	5FEU_A	DS	PDB

Fonte: Autor (2021).

3.2 Caracterização físico-química de proteínas

A partir das sequências de aminoácidos das proteínas similares às proteínas que atuam na rota de biossíntese e inativação de 20E, foi realizada a estimativa dos parâmetros físico-químicos por meio da ferramenta ExPASy ProtParam. As tabelas exibem os parâmetros de Peso Molecular (em Da), ponto isoelétrico teórico (valor de pH em que a proteína apresenta carga

elétrica igual a zero), o número de resíduos negativos e positivos, o coeficiente de extinção (em $M^{-1} cm^{-1}$ a 280 nm medidos em água) (referente ao quanto de luz a proteína absorve a 280 nm), o índice de instabilidade (estimativa da estabilidade da proteína em um tubo de teste, abaixo de 40 é considerada estável), o índice alifático (volume relativo ocupado por cadeias alifáticas laterais), a grande média de hidropatia – GRAVY (soma dos valores de hidropatia de todos os aminoácidos) e a localização subcelular prevista.

A proteína NVD (Tabela 11) possui 429 aminoácidos, enquanto que as proteínas similares variaram em tamanho com valores de 434 a 541, com uma média de 505,71 aminoácidos. Em relação ao Peso Molecular, NVD apresenta 20106,29 Da, já as proteínas similares variam entre 49280,84 Da e 60855,04 Da, com uma média de 57268,95 Da. O Ponto Isoelétrico teórico para NVD é de 9,13, enquanto que as proteínas similares apresentaram valores mais baixos, variando de 7,06 a 8,47 e média de 7,91. Para o número de resíduos negativos e positivos, a proteína NVD apresentou um total de 48 e 53 resíduos, respectivamente. As proteínas similares apresentaram uma média de 55,28 e 56,14 resíduos, respectivamente.

Para o Coeficiente de Extinção (CE), a proteína NVD apresentou um elevado valor de $124955 M^{-1} cm^{-1}$, enquanto que as proteínas similares variaram de $85995 M^{-1} cm^{-1}$ a $110975 M^{-1} cm^{-1}$ (média de $88666,42 M^{-1} cm^{-1}$). O Índice de Instabilidade (II) de NVD ficou em 36,45, para as proteínas similares atingiu um mínimo de 41,73 e máximo de 48,57 (média de 45,31). O índice alifático de NVD foi de 87,02, com as proteínas similares variando de 74,74 a 85,87 (média de 82,99). A Grande Média de Hidropatia (GRAVY) para NVD foi de -155, para as proteínas similares, o GRAVY variou de -0,382 a -0,261 (média de -0,336). A localização prevista (LocSC) para todas as proteínas similares foi o Cloroplasto.

Tabela 11 – Parâmetros físico-químicos computados das proteínas vegetais que apresentam similaridade à proteína Neverland de *Drosophila melanogaster*, por meio da ferramenta ExPASy ProtParam, apresentando o tamanho das sequências de aminoácidos, o peso molecular, o Ponto Isoelétrico teórico (PI), número de resíduos negativos (-) e positivos (+), Coeficiente de Extinção (CE), Índice de Instabilidade (II), Índice Alifático (IA), Grande média de hidropatia (*Grand Average of Hydropathy* – GRAVY). A localização subcelular (LocSC) foi prevista pela ferramenta Plant-mPloc.

N.	Proteína	Tam.	PM	PI	(-) R	(+) R	CE	II	IA	GRAVY	LocSC
1	Neverland	429	20106,29	9,13	39	53	124955	36,45	87,02	-155,0	Transmembrana
2	Chlorophyllide a oxygenase X2	440	50056,47	7,06	48	47	73505	48,57	82,86	-0,382	Cloroplasto
3	Chlorophyllide a oxygenase X1	534	60845,92	7,99	61	63	87485	45,9	84,18	-0,370	Cloroplasto
4	Chlorophyllide a oxygenase	532	60443,41	7,65	62	63	85995	45,86	74,74	-0,380	Cloroplasto
5	Chlorophyllide a oxygenase X2	434	49280,84	7,63	47	48	73505	45,39	82,77	-0,352	Cloroplasto

6	Chlorophyllide a oxygenase X1	530	60326,56	8,47	59	64	91955	41,73	84,87	-0,344	Cloroplasto
7	Chlorophyllide a oxygenase	541	60855,04	8,43	56	51	97245	43,64	85,82	-0,268	Cloroplasto
8	Pheophorbide a oxygenase	529	59074,39	8,20	54	57	110975	46,13	85,75	-0,261	Cloroplasto

Fonte: Autor (2021).

A proteína SHD (Tabela 12) apresenta um total de 540 aminoácidos, enquanto que as proteínas similares variaram em tamanho com valores de 473 a 555 (média de 509 aminoácidos). Para o Peso Molecular, SHD apresenta 61689,83 Da, já as proteínas similares variam entre 53246,35 Da e 62154,18 Da (média de 57529,22 Da). O Ponto Isoelétrico teórico para SHD é de 7,17, enquanto que as proteínas similares variam de 5,93 a 9,34 e média de 7,73. Para o número de resíduos negativos e positivos, a proteína SHD apresentou um total de 67 e 68 resíduos, respectivamente. As proteínas similares apresentaram uma média de 59,68 e 61,20 resíduos, respectivamente.

O Coeficiente de Extinção (CE) apresentou um valor de $57925 \text{ M}^{-1} \text{ cm}^{-1}$, enquanto que as proteínas similares variaram de $35660 \text{ M}^{-1} \text{ cm}^{-1}$ a $84825 \text{ M}^{-1} \text{ cm}^{-1}$ (média de $59434,38 \text{ M}^{-1} \text{ cm}^{-1}$). O Índice de Instabilidade (II) de SHD ficou em 54,51, para as proteínas similares atingiu um mínimo de 26,3 e máximo de 60,13 (média de 39,56). O índice alifático de SHD foi de 103,31, com as proteínas similares variando de 83,86 a 103,68 (média de 95,86). A Grande Média de Hidropatia (GRAVY) para SHD foi de -0,014, para as proteínas similares o GRAVY variou de -0,269 a 0,096 (média de -0,120). A localização prevista (LocSC) para todas as proteínas similares foi o Retículo endoplasmático.

Tabela 12 – Parâmetros físico-químicos computados das proteínas vegetais que apresentam similaridade à proteína Shade de *Drosophila melanogaster*, por meio da ferramenta ExPasy ProtParam, apresentando o tamanho das sequências de aminoácidos, o peso molecular, o Ponto Isoelétrico teórico (PI), número de resíduos negativos (-) e positivos (+), Coeficiente de Extinção (CE), Índice de Instabilidade (II), Grande média de hidropatia (*Grand Average of Hydropathy* – GRAVY). A localização subcelular (LocSC) foi prevista pela ferramenta Plant-mPloc.

N.	Proteína	Tam.	PM	PI	(-) R	(+) R	CE	II	AI	GRAVY	LocSC
1	Shade	540	61689,83	6,69	67	66	57925	54,51	103,3	-0,014	Membrana mitocondrial
2	Cytochrome P450 71A19	490	55614,82	8,92	58	65	57785	43,58	102,3	-0,106	Retículo endoplasmático
3	Cytochrome P450 71A20	497	56497,95	9,07	60	69	56295	40,25	100,2	-0,155	Retículo endoplasmático
4	Cytochrome P450 76C6	511	57481,13	6,3	60	55	39795	44,35	94,01	-0,133	Retículo endoplasmático
5	Cytochrome P450 71	497	56086,12	8,8	61	66	52160	38,97	101	-0,17	Retículo endoplasmático
6	Cytochrome P450 82C2	523	59002,45	8,68	61	66	60765	45,38	88,72	-0,153	Retículo endoplasmático
7	Cytochrome P450 79A2	529	59811,53	9,16	59	69	62130	41,84	89,17	-0,238	Retículo endoplasmático

8	Cytochrome P450 71A25	489	55682,56	8,37	64	67	68785	42,86	91,68	-0,229	Retículo endoplasmático
9	Cytochrome P450 87A6	511	59277,88	9	65	72	69120	60,13	96,3	-0,269	Retículo endoplasmático
10	Cytochrome P450 706A5	520	59045,61	6,25	61	57	64080	45,71	98,4	-0,069	Retículo endoplasmático
11	Cytochrome P450 71A26.1	540	60846,58	6,68	58	55	63745	36,25	102,3	-0,037	Retículo endoplasmático
12	Cytochrome P450 71A26.2	517	58026,94	6,42	58	53	64650	42,15	96,15	-0,109	Retículo endoplasmático
13	Cytochrome P450 82A2.1	522	58715,97	6,54	60	58	76360	36,18	100,1	-0,059	Retículo endoplasmático
14	Cytochrome P450 71A26.3	523	58777,86	5,95	59	50	61920	40,3	92,6	-0,056	Retículo endoplasmático
15	Cytochrome P450 93A3-like	513	58265,74	5,93	67	59	58495	44,68	91,64	-0,143	Retículo endoplasmático
16	Cytochrome P450 84A1	518	58306,25	6,05	66	59	53775	41,72	96,02	-0,102	Retículo endoplasmático
17	Flavone synthase II	527	59686,65	8,86	60	68	41410	36,04	97,51	-0,123	Retículo endoplasmático
18	Cytochrome P450 71A9 precursor	499	56395,81	7,18	59	59	43805	40,73	101,4	-0,026	Retículo endoplasmático
19	Cytochrome P450 82A4-like.1	528	59851,39	7,17	63	63	70860	31,13	96,74	-0,064	Retículo endoplasmático
20	Cytochrome P450 71A1-like.1	517	59352,07	8,48	60	64	68215	39,18	92,71	-0,156	Retículo endoplasmático
21	Cytochrome P450 71D8-like.1	503	56970,32	7	62	61	47245	38	95,05	-0,138	Retículo endoplasmático
22	Cytochrome P450 71A24	517	58003,72	6,42	61	56	59275	38,21	92,73	-0,133	Retículo endoplasmático
23	Cytochrome P450 93B16	527	59580,4	8,29	61	64	39920	29,61	98,99	-0,122	Retículo endoplasmático
24	Cytochrome P450 71D8-like.2	505	57208,83	8,61	61	65	55140	34,17	103,7	-0,046	Retículo endoplasmático
25	Cytochrome P450 71A26.4	523	58899,13	6,22	58	52	63285	38,94	94,63	-0,057	Retículo endoplasmático
26	Cytochrome P450 82A2.2	521	58599,95	6,81	59	58	70985	34,97	98,56	-0,056	Retículo endoplasmático
27	Cytochrome P450 71D8-like.3	501	56714,9	6,96	60	59	47245	38,93	97,19	-0,136	Retículo endoplasmático
28	Cytochrome P450 71D8-like.4	501	56713,96	7,67	59	60	47245	40,33	97,19	-0,137	Retículo endoplasmático
29	Cytochrome P450 71D8-like.5	501	56714,95	7,25	60	60	47245	39,38	97,19	-0,137	Retículo endoplasmático
30	Cytochrome P450 71A1.2	509	58435,1	8,6	57	62	58370	43,91	99,55	-0,123	Retículo endoplasmático
31	Cytochrome P450 71A26	517	58085,25	6,42	55	50	50795	35,4	99,88	0,011	Retículo endoplasmático
32	Cytochrome P450 71D8-like.6	501	56743,99	7,67	59	60	47245	39,38	97,19	-0,141	Retículo endoplasmático
33	Cytochrome P450 71D8-like.7	503	56913,22	8,52	57	61	47245	40,36	97	-0,143	Retículo endoplasmático

34	Cytochrome P450 82A4-like.2	548	62154,18	8,22	59	62	77850	34,65	97,83	-0,01	Retículo endoplasmático
35	Cytochrome P450 71D8-like.8	502	56903,1	6,42	62	57	47245	40,09	96,79	-0,101	Retículo endoplasmático
36	Cytochrome P450 71A1.3	514	58888,77	8,99	58	67	72225	38,59	91,4	-0,126	Retículo endoplasmático
37	Abietadienol/Abietadienal oxidase	487	55829,2	8,3	56	56	73380	37,99	83,86	-0,21	Retículo endoplasmático
38	Cytochrome P450 71A6	511	57954,9	8,36	63	66	62255	54,72	93,82	-0,194	Retículo endoplasmático
39	Cytochrome P450 71D20	504	57083,32	8,68	58	63	60765	31,92	93,41	-0,094	Retículo endoplasmático
40	Cytochrome P450 99A2	532	59533,09	8,43	57	61	57005	49,31	92,73	-0,053	Retículo endoplasmático
41	Cytochrome P450 703A2	525	59631,46	8,42	65	69	77515	39,95	90,1	-0,2	Retículo endoplasmático
42	Cytochrome P450 76M8	500	55933,24	7,84	59	60	59025	38,74	98,9	0,068	Retículo endoplasmático
43	Cytochrome P450 78A11	555	59083,65	6,58	57	54	84825	40,95	89,68	0,096	Retículo endoplasmático
44	Cytochrome P450 76M7	500	55521,72	8,69	58	61	56045	39,98	99,5	0,05	Retículo endoplasmático
45	Cytochrome P450 71AJ4	478	54134,54	6,81	60	59	71070	36,5	97,47	-0,152	Retículo endoplasmático
46	Cytochrome P450 71AJ3	473	53246,35	7,17	57	57	68090	38,22	91,08	-0,158	Retículo endoplasmático
47	Cytochrome P450 75B2	512	56936,27	7,71	54	55	55140	26,3	99,84	0,026	Retículo endoplasmático
48	Cytochrome P450 71D6	501	57274,81	8,84	61	70	63995	33,54	92,24	-0,156	Retículo endoplasmático
49	Cytochrome P450 71A4	507	57637,62	6,72	64	62	69245	42,43	100	-0,186	Retículo endoplasmático
50	Cytochrome P450 71A2	505	57092,01	6,72	63	61	56755	38,52	101	-0,139	Retículo endoplasmático
51	Cytochrome P450 Tp4149	495	55480,4	8,5	49	52	71515	38,42	99,21	-0,01	Retículo endoplasmático
52	Cytochrome P450 71BL2	496	56266,5	8,4	57	61	65360	35,32	92,38	-0,174	Retículo endoplasmático
53	Cytochrome P450 71AJ2	476	53542,86	6,01	59	55	62590	31,35	96,64	-0,078	Retículo endoplasmático
54	Cytochrome P450 71D1V2	506	57443,44	8,34	59	62	61225	38,68	97,55	-0,167	Retículo endoplasmático
55	Cytochrome P450 V19H	506	57502,63	8,14	62	64	63620	38,6	98,7	-0,183	Retículo endoplasmático
56	Cytochrome P450 71D521	506	57512,73	8,89	58	65	54110	32,64	97,08	-0,21	Retículo endoplasmático
57	Cytochrome P450 71BL3	494	55681,73	8,26	56	59	62380	38,22	94,51	-0,16	Retículo endoplasmático
58	Protopine 6-monooxygenase	523	59350,67	7,27	64	64	58245	32,26	92,64	-0,158	Retículo endoplasmático
59	Cytochrome P450 71AY5	501	56593,17	8,59	57	62	52870	37,38	98,88	-0,093	Retículo endoplasmático
60	Cytochrome P450 71BL2	491	55380,46	8,58	56	61	62380	41,09	95,5	-0,153	Retículo endoplasmático
61	Cytochrome P450 71D11	490	56023,79	6,51	65	62	61350	46,18	94,31	-0,212	Retículo endoplasmático
62	Cytochrome P450 71A8	502	57213,42	9,34	61	71	54680	52,51	95,14	-0,252	Retículo endoplasmático
63	(+)-pulegone 9-hydroxylase	493	55360,07	9,24	59	67	46535	40,82	92,94	-0,159	Retículo endoplasmático

64	Protopine 6-hydroxylase	541	61437,98	6,93	66	65	70735	31,76	92,24	-0,202	Retículo endoplasmático
65	Cytochrome P450 76AH1	495	55520,29	7,73	58	59	35660	44,78	90,28	-0,146	Retículo endoplasmático
66	Trans-cinnamate 4-monooxygenase	507	57942,22	8,85	63	68	52160	46	95,92	-0,209	Retículo endoplasmático

Fonte: Autor (2021).

Em relação a proteína SAD (Tabela 13), esta apresenta um total de 520 aminoácidos, já as proteínas similares variaram em tamanho com valores de 361 a 544 (média de 508 aminoácidos). O Peso Molecular de SAD foi de 59631,08 Da, já as proteínas similares variam entre 40867,06 Da e 62055,15 Da (média de 57633,16 Da). O Ponto Isoelétrico teórico para SAD foi de 9,08, enquanto que as proteínas similares variam de 5,8 a 9,36 e média de 8,31. O número de resíduos negativos e positivos para a proteína SHD foi de 56 e 66 resíduos, respectivamente. As proteínas similares apresentaram uma média de 59,2 e 63,5 resíduos, respectivamente.

O Coeficiente de Extinção (CE) apresentou um valor de 68005 M⁻¹ cm⁻¹, enquanto que as proteínas similares variaram de 32555 M⁻¹ cm⁻¹ a 89060 M⁻¹ cm⁻¹ (média de 55298,01 M⁻¹ cm⁻¹). O Índice de Instabilidade (II) de SAD ficou em 58,7, para as proteínas similares atingiu um mínimo de 29,6 e máximo de 48,6 (média de 40,39). O índice alifático de SAD foi de 94,73, com as proteínas similares variando de 81,11 a 103,73 (média de 95,64). A Grande Média de Hidropatia (GRAVY) para SAD foi de -0,245, para as proteínas similares o GRAVY variou de -0,308 a -0,032 (média de -0,155). A localização prevista (LocSC) para todas as proteínas similares foi o Retículo endoplasmático.

Tabela 13 – Parâmetros físico-químicos computados das proteínas vegetais que apresentam similaridade à proteína Shadow de *Drosophila melanogaster*, por meio da ferramenta ExPasy ProtParam, apresentando o tamanho das sequências de aminoácidos, o peso molecular, o Ponto Isoelétrico teórico (PI), número de resíduos negativos (-) e positivos (+), Coeficiente de Extinção (CE), Índice de Instabilidade (II), Grande média de hidropatia (*Grand Average of Hydropathy* – GRAVY). A localização subcelular (LocSC) foi prevista pela ferramenta Plant-mPloc.

N.	Proteína	Tam.	PM	PI	(-)R	(+)R	CE	II	AI	GRAVY	LocSC
1	Shadow	520	59631,08	9,08	56	66	68005	58,68	94,73	-0,245	Retículo Endoplasmático
2	Cytochrome P450 706A6	518	59028,72	6,85	62	61	62590	43,17	102,9	-0,094	Retículo Endoplasmático
3	Cytochrome P450 76C3	515	58143,35	8,99	53	59	48025	45,41	92,78	-0,061	Retículo Endoplasmático
4	Cytochrome P450 91A2	500	57555,48	8,07	59	61	60655	48,56	90,06	-0,308	Retículo Endoplasmático
5	Cytochrome P450 81F4	501	57168,51	8,77	59	65	56185	42,8	97,68	-0,14	Retículo Endoplasmático
6	Cytochrome P450 709B3	518	59769,16	9,26	53	69	80830	38,98	95,06	-0,168	Retículo Endoplasmático

7	Cytochrome P450 76C3	361	40867,06	5,8	46	41	32555	38,84	91,25	-0,127	Retículo Endoplasmático
8	Cytochrome P450 705A13	518	58814,69	8,59	63	69	40505	37,18	93,69	-0,122	Retículo Endoplasmático
9	Cytochrome P450 703A2	510	57854,32	8,63	62	67	71890	34,79	93,69	-0,185	Retículo Endoplasmático
10	Cytochrome P450 81H1	518	58227,62	6,96	59	58	46785	43,97	99,34	-0,085	Retículo Endoplasmático
11	Cytochrome P450 82C4	524	58919,39	7,64	61	62	58245	40,19	88,02	-0,102	Retículo Endoplasmático
12	Cytochrome P450 705A22	515	58708,69	8,92	59	67	44265	44,45	95,18	-0,091	Retículo Endoplasmático
13	Cytochrome P450 705A22	487	55511,86	8,86	58	65	39795	44,94	92,05	-0,156	Retículo Endoplasmático
14	Cytochrome P450 82C2	523	59002,45	8,68	61	66	60765	45,38	88,72	-0,153	Retículo Endoplasmático
15	Cytochrome P450 705A30	523	59570,01	7,64	63	64	41745	45,7	91,38	-0,152	Retículo Endoplasmático
16	Cytochrome P450 705A23	510	57525,45	8,88	56	65	43485	36	94,98	-0,049	Retículo Endoplasmático
17	Cytochrome P450 711A1	439	49951,1	9,36	48	59	40005	37,14	98,41	-0,17	Retículo Endoplasmático
18	Cytochrome P450 705A25	519	58472,13	7,1	59	59	39015	30,17	92,24	-0,066	Retículo Endoplasmático
19	Geraniol 8-hydroxylase	483	54265,01	8,96	54	60	36565	37,43	98,7	-0,218	Retículo Endoplasmático
20	Geraniol 8-hydroxylase	501	55855,08	8,13	56	58	33710	36,62	103,5	-0,097	Retículo Endoplasmático
21	Geraniol 8-hydroxylase	498	55601,81	8,74	53	58	35200	37,25	100,8	-0,117	Retículo Endoplasmático
22	Geraniol 8-hydroxylase	485	54595,29	9,07	51	57	35075	38,99	103,7	-0,153	Retículo Endoplasmático
23	Flavone synthase II	527	59686,65	8,86	60	68	41410	36,04	97,51	-0,123	Retículo Endoplasmático
24	Cytochrome P450 93B16	527	59580,4	8,29	61	64	39920	29,61	98,99	-0,122	Retículo Endoplasmático
25	Geraniol 8-hydroxylase	490	55660,61	9,08	52	60	41620	46,25	94,51	-0,206	Retículo Endoplasmático
26	Cytochrome P450 93A3	522	59362,18	7,21	65	65	50600	38,74	90,38	-0,164	Retículo Endoplasmático
27	Cytochrome P450 82D47	530	60027,76	8,91	57	66	69830	39,47	96,96	-0,151	Retículo Endoplasmático
28	Cytochrome P450 82D47	541	61242,29	8,74	58	66	70080	41,25	97,34	-0,114	Retículo Endoplasmático
29	Isoflavone synthase 1 precursor	521	58929,81	9,11	62	71	48150	35,03	99,9	-0,113	Retículo Endoplasmático
30	Cytochrome P450 93A3-precursor	510	57986,24	8,26	62	65	43930	46,59	94,59	-0,165	Retículo Endoplasmático
31	Cytochrome P450 93C1	521	58868,86	9,1	62	71	48150	34,99	101,4	-0,098	Retículo Endoplasmático
32	Cytochrome P450 82C4	526	59340,37	8,08	59	61	84255	44,24	93,29	-0,17	Retículo Endoplasmático
33	Cytochrome P450 71A1	509	58435,1	8,6	57	62	58370	43,91	99,55	-0,123	Retículo Endoplasmático
34	Flavonoide 3'-mono oxygenase	536	61101,94	6,98	66	65	87570	40,06	97,87	-0,184	Retículo Endoplasmático
35	Cytochrome P450 82D47	526	59102,33	7,23	59	59	74620	34,86	96,56	-0,032	Retículo Endoplasmático
36	Cytochrome P450 71A1	516	59023,6	8,78	62	68	89060	42,31	98,43	-0,214	Retículo Endoplasmático
37	Cytochrome P450 73A11	506	58011,2	9,05	61	68	62005	44,44	97,65	-0,245	Retículo Endoplasmático

38	Cytochrome P450 81F1	500	57555,48	8,07	59	51	60655	48,56	90,06	-0,308	Retículo Endoplasmático
39	Cytochrome P450 81F3	501	56636,55	8,6	56	61	51840	47,45	89,34	-0,268	Retículo Endoplasmático
40	Cytochrome P450 81D11	509	57692,37	9,05	57	67	47370	45,15	102,2	-0,07	Retículo Endoplasmático
41	Cytochrome P450 76B6	493	55657,54	9,08	57	65	49180	45,49	95,7	-0,226	Retículo Endoplasmático
42	Cytochrome P450C4H	505	58279,73	9,13	63	72	60975	43,59	96,28	-0,272	Retículo Endoplasmático
43	Cytochrome P450 75B137	508	56799,76	8,72	53	58	62255	38	101,5	-0,055	Retículo Endoplasmático
44	Cytochrome P450 93C2	523	59429,11	9,03	64	72	49640	40,76	98,93	-0,155	Retículo Endoplasmático
45	Cytochrome P450 93C2	523	59241,9	8,95	64	71	48150	40,59	99,29	-0,149	Retículo Endoplasmático
46	Cytochrome P450 C3H	509	57710,98	7,71	63	64	75650	35,39	92,69	-0,187	Retículo Endoplasmático
47	Cytochrome P450 736A12	500	56042,12	7,24	57	57	55015	35,21	100,6	-0,083	Retículo Endoplasmático
48	N-methyl stylophine hydroxylase	526	59346,4	6,26	67	62	70860	32,51	91,35	-0,203	Retículo Endoplasmático
49	Cytochrome P450 82A1	544	62055,15	7,02	67	66	80370	32,1	96,34	-0,116	Retículo Endoplasmático
50	Cytochrome P450 71AU50	497	56803,02	8,1	64	66	64775	42,68	92,58	-0,207	Retículo Endoplasmático
51	Cytochrome P450 76A2	505	57807,08	7,65	67	68	71640	45,38	81,11	-0,278	Retículo Endoplasmático
52	Cytochrome P450 98A1	512	58271,58	8,91	61	67	77140	31,76	95,55	-0,249	Retículo Endoplasmático
53	Cytochrome P450 76AH1	507	55520,29	7,73	58	59	35660	44,78	90,28	-0,146	Retículo Endoplasmático
54	Cytochrome P450 76AH1	495	57942,22	8,85	63	68	52160	46	95,92	-0,209	Retículo Endoplasmático

Fonte: Autor (2021).

A proteína DIB (Tabela 14) apresenta um total de 489 aminoácidos, a única proteína similar encontrada possui 514 aminoácidos. O Peso Molecular de DIB foi de 55573,97 Da, já da proteína similar foi de 59110,86 Da. O Ponto Isoelétrico teórico para DIB foi de 9,24, enquanto que a proteína similar foi de 8,5. O número de resíduos negativos e positivos para a proteína DIB foi de 55 e 63 resíduos, respectivamente. A proteína similar apresentou 59 e 62 resíduos, respectivamente.

O Coeficiente de Extinção (CE) apresentou um valor de $64205 \text{ M}^{-1} \text{ cm}^{-1}$, enquanto que a proteína similar foi de $82975 \text{ M}^{-1} \text{ cm}^{-1}$. O Índice de Instabilidade (II) de DIB foi igual a 48,68, para a proteína encontrada foi de 39,44. O Índice Alifático de DIB foi de 94,74, a proteína similar foi de 86,44. A Grande Média de Hidropatia (GRAVY) para DIB foi de -0,306, para a proteína similar o GRAVY foi igual a -0,270. A localização prevista (LocSC) para todas as proteínas foi o Retículo endoplasmático.

Tabela 14 – Parâmetros físico-químicos computados das proteínas vegetais que apresentam similaridade à proteína Disembodied de *Drosophila melanogaster*, por meio da ferramenta ExPasy ProtParam,

apresentando o tamanho das sequências de aminoácidos, o peso molecular, o Ponto Isoelétrico teórico (PI), número de resíduos negativos (-) e positivos (+), Coeficiente de Extinção (CE), Índice de Instabilidade (II), Grande média de hidropatia (*Grand Average of Hydropathy* – GRAVY). A localização subcelular (LocSC) foi prevista pela ferramenta Plant-mPloc.

N.	Proteína	Tam.	PM	PI	(-)R	(+)R	CE	II	AI	GRAVY	LocSC
1	Disembodied	489	55573,97	9,24	52	63	64205	48,68	94,74	-0,306	Retículo Endoplasmático
2	Cytochrome P450 94A1	514	59110,86	8,5	59	62	82975	39,44	86,44	-0,27	Retículo Endoplasmático

Fonte: Autor (2021).

Em relação a proteína PHT (Tabela 15), esta apresenta um total de 574 aminoácidos, já as proteínas similares variaram em tamanho com valores de 456 a 981 (média de 562 aminoácidos). O Peso Molecular de PHT foi de 65177,22 Da, já as proteínas similares variam entre 52692,87 Da e 111016,55 Da (média de 63835,10 Da). O Ponto Isoelétrico teórico para PHT foi de 5,81, enquanto que as proteínas similares variam de 5,29 a 9,19 (média de 7,88). O número de resíduos negativos e positivos para a proteína SHD foi de 81 e 69 resíduos, respectivamente. As proteínas similares apresentaram uma média de 68,04 e 70,41 resíduos, respectivamente.

O Coeficiente de Extinção (CE) apresentou um valor de $65610 \text{ M}^{-1} \text{ cm}^{-1}$, enquanto que as proteínas similares variaram de $35660 \text{ M}^{-1} \text{ cm}^{-1}$ a $107760 \text{ M}^{-1} \text{ cm}^{-1}$ (média de $66702,29 \text{ M}^{-1} \text{ cm}^{-1}$). O Índice de Instabilidade (II) de PHT ficou em 56,47, para as proteínas similares atingiu um mínimo de 30,47 e máximo de 56,47 (média de 41,45). O índice alifático de PHT foi de 92,94, com as proteínas similares variando de 84,22 a 104,63 (média de 96,37). A Grande Média de Hidropatia (GRAVY) para PHT foi de -0,229, para as proteínas similares o GRAVY variou de -0,416 a -0,046 (média de -0,158). A localização prevista (LocSC) para todas as proteínas similares foi o Retículo endoplasmático.

Tabela 15 – Parâmetros físico-químicos computados das proteínas vegetais que apresentam similaridade à proteína Phantom de *Drosophila melanogaster*, por meio da ferramenta ExPASy ProtParam, apresentando o tamanho das sequências de aminoácidos, o peso molecular, o Ponto Isoelétrico teórico (PI), número de resíduos negativos (-) e positivos (+), Coeficiente de Extinção (CE), Índice de Instabilidade (II), Grande média de hidropatia (*Grand Average of Hydropathy* – GRAVY). A localização subcelular (LocSC) foi prevista pela ferramenta Plant-mPloc.

N.	Proteína	Tam.	PM	PI	(-)R	(+)R	CE	II	IA	GRAVY	LocSC
1	Phantom	574	65177,22	5,81	81	69	65610	56,47	92,94	-0,229	Retículo Endoplasmático
2	Cytochrome P450 706A4	516	58658,94	6,31	63	60	66600	39,53	98,76	-0,144	Retículo Endoplasmático
3	Cytochrome P450 71B11	496	56898,69	8,59	60	65	68215	30,47	98,79	-0,079	Retículo Endoplasmático
4	Cytochrome P450 76C1	512	56800,55	7,12	58	58	43680	42,99	99,59	-0,048	Retículo Endoplasmático
5	Cytochrome P450 706A6	518	59028,72	6,85	62	61	62590	43,17	102,92	-0,094	Retículo Endoplasmático

6	Cytochrome P450 736A12 - X1	498	56942,55	7,71	62	63	71640	36,36	97,11	-0,116	Retículo Endoplasmático
7	Cytochrome P450 736A12	503	56871,24	8,17	55	57	54110	43,12	104,63	-0,059	Retículo Endoplasmático
8	Uncharacterized protein	977	111016,5	9,19	105	124	107760	39,6	99,05	-0,164	Retículo Endoplasmático
9	Cytochrome P450 736A12	981	110970,2	7,87	119	121	102510	43,97	98,62	-0,131	Retículo Endoplasmático
10	Cytochrome P450 71D11-X2	517	58919,46	8,22	65	68	58370	44,44	96,91	-0,175	Retículo Endoplasmático
11	Cytochrome P450 71D11-X1	517	58629	8,39	64	68	58370	44,25	95,78	-0,194	Retículo Endoplasmático
12	Cytochrome P450 736A12-like	536	61210,45	8,03	70	72	74745	37,01	93,32	-0,192	Retículo Endoplasmático
13	Cytochrome P450 71D8-like	505	57208,83	8,61	61	65	55140	34,17	103,68	-0,046	Retículo Endoplasmático
14	Cytochrome P450 736A12	493	56328,87	8,68	62	67	66265	36,96	94,73	-0,141	Retículo Endoplasmático
15	Cytochrome P450 736A12	978	109527,2	8,26	112	116	99405	43,22	99,65	-0,097	Retículo Endoplasmático
16	Cytochrome P450 736A12	500	57199,82	8,69	63	69	70860	42,93	95,92	-0,204	Retículo Endoplasmático
17	Cytochrome P450 71B11	496	56898,69	8,59	60	65	68215	30,47	98,79	-0,079	Retículo Endoplasmático
18	Cytochrome P450 76C1	512	56800,55	7,12	58	58	43680	42,99	99,59	-0,048	Retículo Endoplasmático
19	Cytochrome P450 97A3	519	58397,06	5,35	71	59	63495	44,5	92,81	-0,175	Retículo Endoplasmático
20	Cytochrome P450 97A3	519	58364,08	5,29	72	59	57995	42,93	94,12	-0,162	Retículo Endoplasmático
21	Cytochrome P450 90B1.1	457	52806,97	9,15	53	62	73130	45,87	88,77	-0,342	Retículo Endoplasmático
22	Cytochrome P450 90B1.2	456	52692,87	9,15	53	62	73130	46,14	88,97	-0,335	Retículo Endoplasmático
23	Cytochrome P450 90B1.3	491	56408,57	7,32	64	64	73130	48,93	84,22	-0,416	Retículo Endoplasmático
24	Cytochrome P450 76AH1	495	55520,29	7,73	58	59	35660	44,78	90,28	-0,146	Retículo Endoplasmático
25	Cinnamate 4-hydroxylase (C4H1)	507	57942,22	8,85	63	68	52160	46	95,92	-0,209	Retículo Endoplasmático

Fonte: Autor (2021).

Em relação a proteína EO (Tabela 16), esta apresenta um total de 657 aminoácidos, já as proteínas similares variaram em tamanho com valores de 392 a 594 (média de 553 aminoácidos). O Peso Molecular de EO foi de 72559,99 Da, já as proteínas similares variam entre 42559,68 Da e 65415,87 Da (média de 60596,32 Da). O Ponto Isoelétrico teórico para EO foi de 8,93, enquanto que as proteínas similares variam de 4,71 a 9,74 (média de 7,28). O número de resíduos negativos e positivos para a proteína EO foi de 58 e 67 resíduos, respectivamente. As proteínas similares apresentaram uma média de 51,51 e 51,22 resíduos, respectivamente.

O Coeficiente de Extinção (CE) apresentou um valor de 86915 M⁻¹ cm⁻¹, enquanto que as proteínas similares variaram de 45755 M⁻¹ cm⁻¹ a 76000 M⁻¹ cm⁻¹ (média de 59766,57 M⁻¹ cm⁻¹). O Índice de Instabilidade (II) de EO ficou em 43,63, para as proteínas similares atingiu um mínimo de 29,34 e máximo de 47,42 (média de 37,50). O índice alifático de EO foi de 90,09, com as proteínas similares variando de 79,95 a 86,85 (média de 88,38). A Grande Média de Hidropatia (GRAVY) para EO foi de -0,256, para as proteínas similares o GRAVY variou de -0,277 a 0,024 (média de -0,127). A localização prevista (LocSC) para as proteínas similares em sua totalidade foi para o vacúolo, com presença também em algumas proteínas no cloroplasto, conforme visualizado na Tabela 16.

Tabela 16 – Parâmetros físico-químicos computados das proteínas vegetais que apresentam similaridade à proteína Ecdysone Oxidase de *Drosophila melanogaster*, por meio da ferramenta ExPASy ProtParam, apresentando o tamanho das sequências de aminoácidos, o peso molecular, o Ponto Isoelétrico teórico (PI), número de resíduos negativos (-) e positivos (+), Coeficiente de Extinção (CE), Índice de Instabilidade (II), Grande média de hidropatia (*Grand Average of Hydropathy* – GRAVY). A localização subcelular (LocSC) foi prevista pela ferramenta Plant-mPloc.

N.	Proteína	Tam.	PM	PI	(-)R	(+)R	EC	II	AI	GRAVY	LocSC
1	Ecdysone Oxidase	657	72559,99	8,93	58	67	86915	43,63	90,09	-0,256	
2	Glucose-methanolcholine 1	594	65344,07	9,74	44	72	50685	39,2	85,32	-0,19	Cloroplasto; Vacúolo
3	Glucose-methanolcholine 2	572	62000,38	8,82	47	53	69135	37,72	79,95	-0,176	Vacúolo
4	Glucose-methanolcholine 3	567	62209,42	9,73	42	68	49195	40,65	83,9	-0,201	Vacúolo
5	Glucose-methanolcholine 4	501	55138,46	6,12	59	54	48165	38,9	82,24	-0,277	Vacúolo
6	Glucose-methanolcholine 5	577	63357,42	6,25	57	52	56520	31,37	90,75	-0,032	Vacúolo
7	Glucose-methanolcholine 6	522	57506,29	6,25	59	55	48165	39,39	84,16	-0,222	Cloroplasto; Vacúolo
8	Glucose-methanolcholine 7	552	60782,62	9,55	47	57	56965	43,47	94,76	0,009	Vacúolo
9	Glucose-methanolcholine 8	584	64544,72	6,74	54	52	76000	32,15	87,52	-0,104	Vacúolo
10	Glucose-methanolcholine 9	503	55420,24	8,31	50	53	46675	34,18	86,16	-0,163	Vacúolo
11	Glucose-methanolcholine 10	551	60459,04	6,29	55	50	56060	29,34	90,42	-0,118	Vacúolo
12	Glucose-methanolcholine 11	551	60632,05	6,41	53	49	66030	30,87	86,73	-0,147	Cloroplasto; Vacúolo
13	Glucose-methanolcholine 12	586	64804,05	6,74	54	52	76000	31,95	87,22	-0,106	Vacúolo
14	Glucose-methanolcholine 13	582	64164,68	6,41	54	50	62145	30,94	95,64	0,024	Vacúolo
15	(R)-mandelonitrile lyase-like	542	58841,81	6,37	48	44	63175	40,04	96,42	0,020	Cloroplasto; Vacúolo
16	Protein HOTHEAD 1	585	64466,04	9,59	47	70	60195	34,15	88,44	-0,198	Cloroplasto; Vacúolo
17	Protein HOTHEAD 2	591	65415,87	9,28	46	61	65820	36,24	84,72	-0,167	Cloroplasto; Vacúolo
18	Protein HOTHEAD 3	585	64351,76	9,54	47	68	61685	38,47	88,94	-0,199	Vacúolo
19	Protein HOTHEAD-iX2	579	63375,75	9,37	48	62	68675	45,87	85,18	-0,135	Vacúolo

20	Protein HOTHEAD-iX1	581	63575,99	9,37	48	62	68675	45,6	95,39	-0,128	Vacúolo
21	Protein HOTHEAD 4	581	62965,68	6,37	54	50	64665	41,67	86,28	-0,109	Cloroplasto; Vacúolo
22	Protein HOTHEAD 5	581	63251,06	6,37	55	51	70165	47,42	85,59	-0,129	Vacúolo
23	(R)-mandelonitrile lyase-like 1	571	62995,31	8,7	57	64	52425	42,05	88,62	-0,153	Cloroplasto; Vacúolo
24	Protein HOTHEAD 6	570	62585,94	8,91	46	54	64665	42,69	91,82	-0,054	Cloroplasto; Vacúolo
25	(R)-mandelonitrile lyase-like 2	559	61658,51	8,67	56	62	52300	39,07	86,85	-0,213	Cloroplasto; Vacúolo
26	Protein HOTHEAD 7	483	53801,86	9,22	42	54	45755	41,66	86,71	-0,244	Vacúolo
27	Protein HOTHEAD isoform X1	594	64702,76	6,13	57	50	60655	37,65	90,56	-0,056	Vacúolo
28	Protein HOTHEAD isoform X2	593	64631,68	6,13	57	50	60665	37,7	90,54	-0,059	Vacúolo
29	Protein HOTHEAD 8	392	42559,68	8,46	33	36	50100	44,5	94,03	0,015	Vacúolo
30	Hydroxynitrile lyase 1	536	58039,14	5,31	54	38	59500	30,56	87,07	-0,107	Vacúolo
31	Hydroxynitrile lyase 2	521	56486,22	5,23	53	36	59500	30,9	84,15	-0,144	Vacúolo
32	Endo-deglycosylated hydroxynitrile lyase 1	538	58737,65	5,05	55	33	60865	36,05	90,39	-0,153	Vacúolo
33	Deglycosylated hydroxynitrile lyase	538	58695,57	5,05	55	33	60865	36,05	89,85	-0,156	Vacúolo
34	Endo-deglycosylated hydroxynitrile lyase 2	542	59138,04	4,97	57	33	60865	35,7	90,09	-0,158	Vacúolo
35	Hydroxynitrile lyase 3	532	57860,67	4,78	55	33	60865	37,23	90,32	-0,131	Vacúolo
36	Hydroxynitrile lyase 4	521	56371,01	4,71	58	32	58010	31,41	85,3	-0,117	Vacúolo

Fonte: Autor (2021).

A proteína 3DE3BR (Tabela 17), esta apresenta um total de 249 aminoácidos, já as proteínas similares variaram em tamanho com valores de 250 a 337 (média de 294 aminoácidos). O Peso Molecular de 3DE3BR foi de 26079,09 Da, já as proteínas similares variam entre 28019,56 Da e 36967,11 Da (média de 31927,50 Da). O Ponto Isoelétrico teórico para 3DE3BR foi de 5,82, enquanto que as proteínas similares variam de 5,18 a 9,48 (média de 7,55). O número de resíduos negativos e positivos para a proteína 3DE3BR foi de 44 e 42 resíduos, respectivamente. As proteínas similares apresentaram uma média de 38,47 e 35,79 resíduos, respectivamente.

O Coeficiente de Extinção (CE) apresentou um valor de 45505 M⁻¹ cm⁻¹, enquanto que as proteínas similares variaram de 45170 M⁻¹ cm⁻¹ a 67170 M⁻¹ cm⁻¹ (média de 53245,26 M⁻¹ cm⁻¹). O Índice de Instabilidade (II) de 3DE3BR ficou em 49,05, para as proteínas similares atingiu um mínimo de 24,9 e máximo de 44,87 (média de 38,99). O índice alifático de 3DE3BR

foi de 90,70, com as proteínas similares variando de 81,08 a 96,23 (média de 91,63). A Grande Média de Hidropatia (GRAVY) para 3DE3BR foi de 0,282, para as proteínas similares o GRAVY variou de -0,298 a 0,282 (média de 0,0305). A localização prevista (LocSC) para todas as proteínas similares foi o cloroplasto.

Tabela 17 – Parâmetros físico-químicos computados das proteínas vegetais que apresentam similaridade à proteína 3-dehydroecdysone 3beta-reductase de *Spodoptera littoralis*, por meio da ferramenta ExPasy ProtParam, apresentando o tamanho das sequências de aminoácidos, o peso molecular, o Ponto Isoelétrico teórico (PI), número de resíduos negativos (-) e positivos (+), Coeficiente de Extinção (CE), Índice de Instabilidade (II), Grande média de hidropatia (Grand Average of Hydropathy – GRAVY). A localização subcelular (LocSC) foi prevista pela ferramenta Plant-mPloc.

N.	Proteína	Tan.	PM	PI	(-)R	(+)R	CE	II	IA	GRAVY	LocSC
1	3-dehydroecdysone 3beta-reductase	249	26079.09	5.82	44	42	45505	49.05	90.70	0.282	citoplasma
2	NAD(P)-linked oxidoreductase	289	31451.84	6.09	44	35	46785	28.53	94.01	-0.137	cloroplasto
3	NAD(P)-linked oxidoreductase	309	34947.04	8.66	43	39	46785	26.29	91.46	0.267	cloroplasto
4	NAD(P)-linked oxidoreductase	319	33548.08	9.46	32	34	45170	40.24	95.15	0.210	cloroplasto
5	NAD(P)-linked oxidoreductase	335	36764.84	8.78	37	34	60180	42.64	89.36	-0.276	cloroplasto
6	NAD(P)-linked oxidoreductase	337	36967.11	8.78	31	34	45170	40.05	95.48	-0.264	cloroplasto
7	NAD(P)-linked oxidoreductase	250	28019.56	8.90	38	36	50670	42.34	90.93	0.179	cloroplasto
8	Uncharacterized protein LOC100806500	299	33247.78	9.07	40	35	48275	24.90	94.98	0.168	cloroplasto
9	Uncharacterized protein LOC100788951 isoform X4	308	34683.56	7.67	42	35	67170	42.85	89.69	0.269	cloroplasto
10	methylecgonone reductase	321	35715.64	7.69	36	37	54555	39.66	90.47	0.282	cloroplasto
11	chalcone reductase CHR1	294	31661.03	6.39	38	35	47565	43.70	94.41	-0.232	cloroplasto
12	NAD(P)H-dependent 6'-deoxychalcone synthase	267	28431.27	5.77	34	31	64775	42.44	91.62	0.007	cloroplasto
13	NAD(P)H-dependent 6'-deoxychalcone synthase isoform X1	320	33712.73	9.17	39	33	47565	44.87	94.11	0.058	cloroplasto
14	NADPH-dependent aldo-keto reductase, chloroplastic	269	28966.90	5.99	36	33	63035	34.14	87.30	-0.032	cloroplasto
15	uncharacterized protein LOC100788951 isoform X3	265	28477.44	5.18	43	39	67170	41.02	90.21	0.098	cloroplasto
16	Aldo-keto reductase family 4 member C10	272	29362.70	8.33	37	34	60180	42.64	89.36	0.092	cloroplasto
17	NAD(P)H-dependent 6'-deoxychalcone synthase	299	31746.63	9.48	38	35	47565	43.70	94.41	0.044	cloroplasto

18	NADP-dependent D-sorbitol-6-phosphate dehydrogenase	289	31451.84	6.09	39	38	47120	30.75	96.23	-0.137	cloroplasto
19	Deoxymugineic acid synthase 1	288	31387.46	6.11	40	41	56420	41.08	81.08	-0.298	cloroplasto

Fonte: Autor (2021).

A proteína 3DE3AR (Tabela 18), esta apresenta um total de 249 aminoácidos, já as proteínas similares variaram em tamanho com valores de 244 a 348 (média de 291 aminoácidos). O Peso Molecular de 3DE3AR foi de 26079,09 Da, já as proteínas similares variam entre 25332,45 Da e 39567,57 Da (média de 31586,19 Da). O Ponto Isoelétrico teórico para 3DE3AR foi de 5,82, enquanto que as proteínas similares variam de 5,18 a 9,48 (média de 7,28). O número de resíduos negativos e positivos para a proteína 3DE3AR foi de 24 e 21 resíduos, respectivamente. As proteínas similares apresentaram uma média de 28,94 e 29,58 resíduos, respectivamente.

O Coeficiente de Extinção (CE) apresentou um valor de 4595 M⁻¹ cm⁻¹, enquanto que as proteínas similares variaram de 13325 M⁻¹ cm⁻¹ a 52495 M⁻¹ cm⁻¹ (média de 26505,41 M⁻¹ cm⁻¹). O Índice de Instabilidade (II) de 3DE3AR ficou em 24,18, para as proteínas similares atingiu um mínimo de 22,68 e máximo de 43,32 (média de 33,41). O índice alifático de 3DE3BR foi de 107,71, com as proteínas similares variando de 77,57 a 112,33 (média de 93,85). A Grande Média de Hidropatia (GRAVY) para 3DE3AR foi de 0,282, para as proteínas similares o GRAVY variou de -0,298 a 0,282 (média de 0,0517). A localização prevista (LocSC) para todas as proteínas similares foi o cloroplasto.

Tabela 18 – Parâmetros físico-químicos computados das proteínas vegetais que apresentam similaridade à proteína 3-dehydroecdysone 3alpha-reductase de *Spodoptera littoralis*, por meio da ferramenta Expasy ProtParam, apresentando o tamanho das sequências de aminoácidos, o peso molecular, o Ponto Isoelétrico teórico (PI), número de resíduos negativos (-) e positivos (+), Coeficiente de Extinção (CE), Índice de Instabilidade (II), Grande média de hidropatia (Grand Average of Hydropathy – GRAVY). A localização subcelular (LocSC) foi prevista pela ferramenta Plant-mPloc.

N.	Proteína	Tam.	PM	PI	(-)R	(+)R	CE	II	IA	GRAVY	LocSC
1	3-dehydroecdysone 3alpha-reductase	249	26079,09	5,82	24	21	4595	24,18	107,71	0,282	Citoplasma
2	NAD(P)-binding Rossmann-fold superfamily protein	289	31451,84	6,09	32	28	19285	38,98	84,01	-0,137	Cloroplasto
3	Hydroxysteroid dehydrogenase 3	309	34947,04	8,66	30	34	37610	32,7	112,33	0,267	Cloroplasto
4	NAD(P)-binding Rossmann-fold superfamily protein	319	33548,08	9,46	27	37	18700	38,6	101,6	0,21	Cloroplasto
5	NAD(P)-binding Rossmann-fold superfamily protein	335	36764,84	8,78	34	39	19285	39,9	79,46	-0,276	Cloroplasto
6	NAD(P)-binding Rossmann-fold superfamily protein	337	36967,11	8,78	34	39	19285	40,09	79,29	-0,264	Cloroplasto

7	Hydroxysteroid dehydrogenase 4	250	28019,56	8,9	24	28	24535	35,04	102,12	0,179	Cloroplasto
8	Hydroxysteroid dehydrogenase 5	299	33247,78	9,07	29	36	38305	37,45	103,65	0,168	Cloroplasto
9	Hydroxysteroid dehydrogenase 6	308	34683,56	7,67	31	32	36120	26,72	107,31	0,269	Cloroplasto
10	Hydroxysteroid dehydrogenase 7	321	35715,64	7,69	32	33	24660	33,58	107,2	0,282	Cloroplasto
11	Uncharacterized protein	294	31661,03	6,39	29	27	14940	26,72	80,65	-0,232	Cloroplasto
12	Tropinone reductase	267	28431,27	5,77	25	21	25815	24,21	97,23	0,007	Cloroplasto
13	3-oxoacyl-[acyl-carrier-protein] reductase 4	320	33712,73	9,17	29	35	24075	36,62	95,41	0,058	Cloroplasto
14	Uncharacterized protein	269	28966,9	5,99	27	24	24325	29,39	94,31	-0,032	Cloroplasto
15	Tropinone reductase 1	265	28477,44	5,18	26	20	34295	27,08	96,83	0,098	Cloroplasto
16	Tropinone reductase 2	272	29362,7	8,33	21	24	34880	30	93,2	0,092	Cloroplasto
17	3-oxoacyl-[acyl-carrier-protein] reductase 4	299	31746,63	9,48	25	32	17085	32,74	94,65	0,044	Cloroplasto
18	Glucose and ribitol dehydrogenase homolog 2	289	31451,84	6,09	32	28	19285	38,98	84,01	-0,137	Cloroplasto
19	NADPH-dependent aldehyde reductase 1	288	31387,46	6,11	33	30	19160	36,91	77,57	-0,298	Cloroplasto
20	11-beta-hydroxysteroid dehydrogenase-like 3	309	34947,04	8,66	30	34	37610	32,7	112,33	0,267	Cloroplasto
21	3-oxoacyl-[acyl-carrier-protein] reductase	319	33548,08	9,46	27	37	18700	38,6	101,6	0,21	Cloroplasto
22	11-beta-hydroxysteroid dehydrogenase-like 4A	299	33247,78	9,07	29	36	38305	37,45	103,65	0,168	Cloroplasto
23	11-beta-hydroxysteroid dehydrogenase-like 2	308	34683,56	7,67	31	32	36120	26,72	107,31	0,269	Cloroplasto
24	Tropinone reductase 3	262	28025,95	5,41	24	19	28795	29,5	91,95	0,127	Cloroplasto
25	3-oxoacyl-[acyl-carrier-protein] reductase	320	33103,24	9,28	26	34	18700	34,51	95,78	0,184	Cloroplasto
26	Glucose and ribitol dehydrogenase	291	31500,57	6,25	33	30	20525	36,44	78,14	-0,282	Cloroplasto
27	Tropinone reductase 4	260	28437,41	5,46	30	24	31775	32,93	84,81	-0,038	Cloroplasto
28	Noroxomaritidine/norcraugsodine reductase	257	27605,33	5,61	29	24	24450	36,23	89,57	-0,027	Cloroplasto
29	(+)-cis,trans-nepetalactol synthase NEPS1	271	28683,21	6,37	25	23	13325	22,68	88,52	0,211	Cloroplasto
30	Glucose and ribitol dehydrogenase	300	32267,54	5,76	36	33	19160	36,03	78,13	-0,253	Cloroplasto
31	11-beta-hydroxysteroid dehydrogenase A	348	39567,57	5,7	42	39	52495	43,32	87,9	0,002	Cloroplasto

32	Beta-Keto Acyl Carrier Protein Reductase	244	25332,45	9,22	21	27	17085	29,35	102,42	0,254	Cloroplasto
33	Tropinone Reductase-II	260	28311,47	5,86	27	24	31900	29,77	91,92	0,053	Cloroplasto
34	Tropinone Reductase-II	259	28180,27	5,88	27	24	31900	29,37	92,28	0,046	Cloroplasto
35	Tropinone Reductase-I	273	29616,95	6,13	29	27	32805	28,56	91,14	-0,046	Cloroplasto
36	Noroxomaritidine/Norcraugsodine Reductase	257	27605,33	5,61	29	24	24450	36,23	89,57	-0,027	Cloroplasto
37	Noroxomaritidine/Norcraugsodine Reductase	291	31149,23	6,45	29	26	24450	35,43	81,79	-0,133	Cloroplasto

Fonte: Autor (2021).

4.3 Filogenia de proteínas

A partir do alinhamento múltiplo das sequências de aminoácidos das proteínas potencialmente homólogas de plantas às proteínas atuantes na rota de biossíntese e inativação do hormônio 20-hidroxicdisona (20E) foi possível construir árvores filogenéticas com validação por *Bootstrap*. Cada sequência de proteína proveniente de inseto foi utilizada como grupo externo para a formação das árvores filogenéticas.

Para a proteína Neverland (Figura 12 – A) observou-se a formação de três grupos distintos. As proteínas homólogas (Chlorophyllide A Oxygenase e suas isoformas) provenientes de *Glycine max* e *Oryza sativa* var. japonica estão enquadradas no Grupo I, enquanto que a proteína Pheophorbide A oxygenase ficou isolada no Grupo II.

Para a proteína Shade (Figura 12 – B) também se verificou três grupos formados. O grande grupo formado incluiu todas as espécies encontradas de vegetais para essa proteína (Grupo I). Ainda na Figura 12 observa-se, com marcações destacadas, os nós que formam o subgrupo de algumas superfamílias de proteínas. Estes subgrupos são bem definidos e estatisticamente satisfatórios pelo método *bootstrap* (90%>), distinguindo as superfamílias de citocromo P450 71, 93, 82 e 76. A proteína Abietadienol/Abietadienal Oxidase formou um grupo isolado às demais (II).

Apenas quatro grupos foram formados na árvore filogenética da proteína Shadow (Figura 12 – C). O maior grupo formado (Grupo I) englobou todas as espécies encontradas, sendo possível verificar a existência de nós estatisticamente satisfatórios para a formação do subgrupo contendo as superfamílias do citocromo P450 93, 70, 81 e 82. Dois grupos foram formados com apenas uma proteína, sendo ambos da espécie *Arabidopsis thaliana*, com as proteínas e CYP 711A1 (Grupo II) e CYP709B3 (Grupo III).

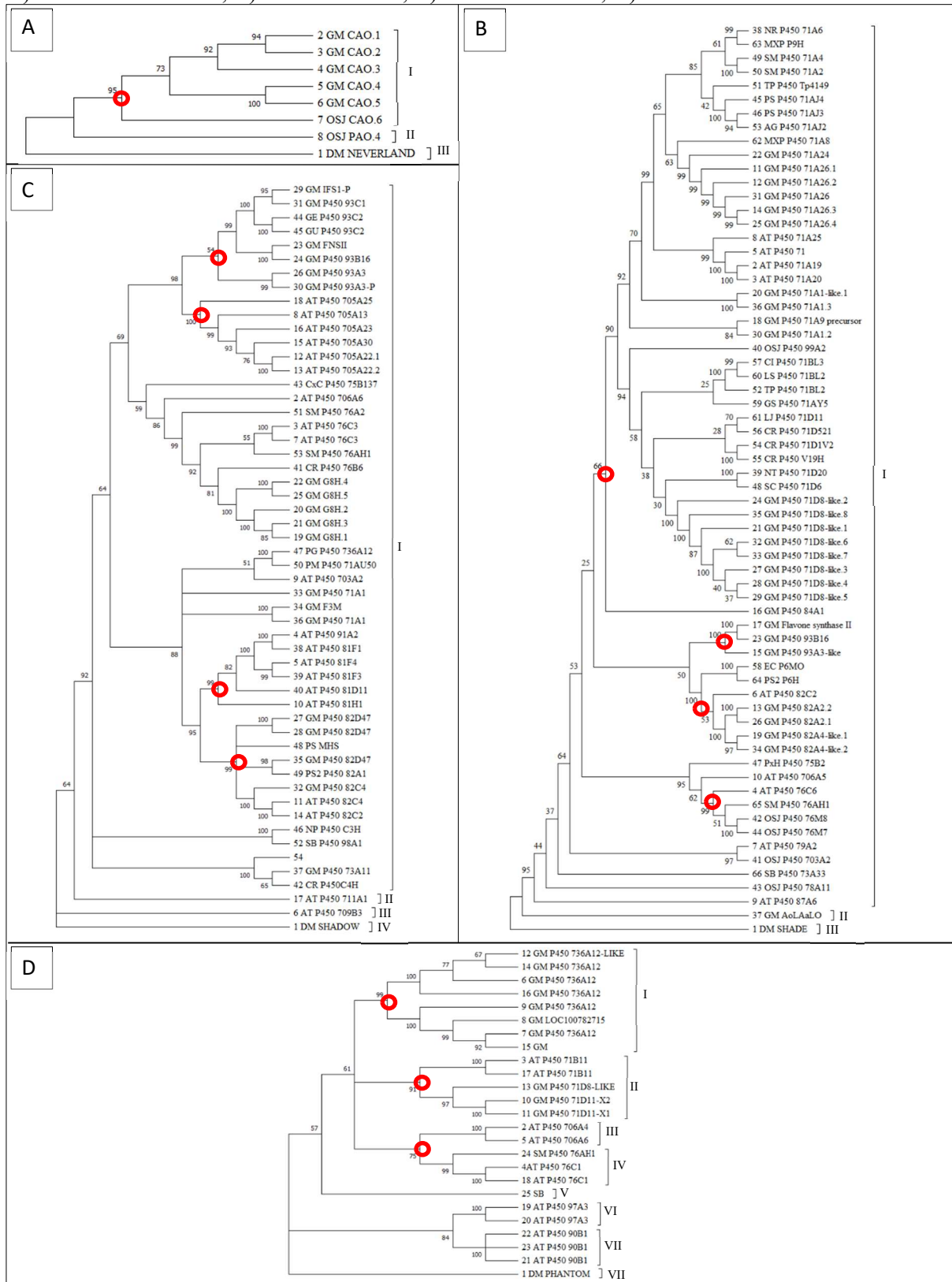
A proteína *Phantom* (Figura 12 – D) apresentou uma árvore filogenética diversificada, com formação de sete grupos. O Grupo I apresentou somente proteínas da espécie *G. max* da superfamília citocromo P450 73 (Nó destacado). O Grupo II já inclui, além da *G. max*, a espécie *A. thaliana*, e distingue a superfamília do citocromo P450 71 (Nó destacado). O Grupo III possui duas proteínas integrantes da superfamília 706 dos citocromos P450 de *A. thaliana*. O Grupo IV distingue três proteínas classificadas na superfamília 76 dos citocromos P450 de *A. thaliana* e *Salvia miltiorrhiza*. O Grupo V apresentou apenas uma proteína isolada, a Cinnamate 4-hydroxylase (C4H1), de *Sorghum bicolor*. Os Grupos VI e VII apresentaram somente proteínas provenientes de *A. thaliana*, especificamente das superfamílias 97 e 90 do citocromo P450, respectivamente.

A filogenia para a proteína Ecdysone Oxidase apresentou quatro grupos (Figura 13 – A). O Grupo I é formado pelas proteínas Glucose-methanol-choline de *A. thaliana* e Protein HOTHEAD de *G. max* (e suas isoformas). O Grupo II inclui três isoformas de *Glucose-methanol-choline* de *A. thaliana* e *(R)-mandelonitrile lyase-like* de *G. max*. O Grupo III para EO possui duas outras isoformas destas duas proteínas. O Grupo IV englobou as proteínas das espécies do gênero *Prunus* (nó destacado), sendo elas Hydroxynitrile lyase, Endoglycosylated hydroxynitrile lyase e Deglycosylated hydroxynitrile lyase.

Apenas três grupos foram formados para a filogenia de 3-dehydroecdysone 3beta-reductase (Figura 13 – B). A maior quantidade de proteínas de diferentes espécies se concentrou no Grupo I, sendo a maioria das proteínas a NAD(P)-linked oxidoreductase e suas isoformas. O Grupo II é composto de quatro proteínas das espécies *A. thaliana*, *Malus domestica* e *G. max*. Para a proteína 3DE3BR e suas proteínas similares, não há sub agrupamento específico de nenhuma família ou superfamília de proteínas.

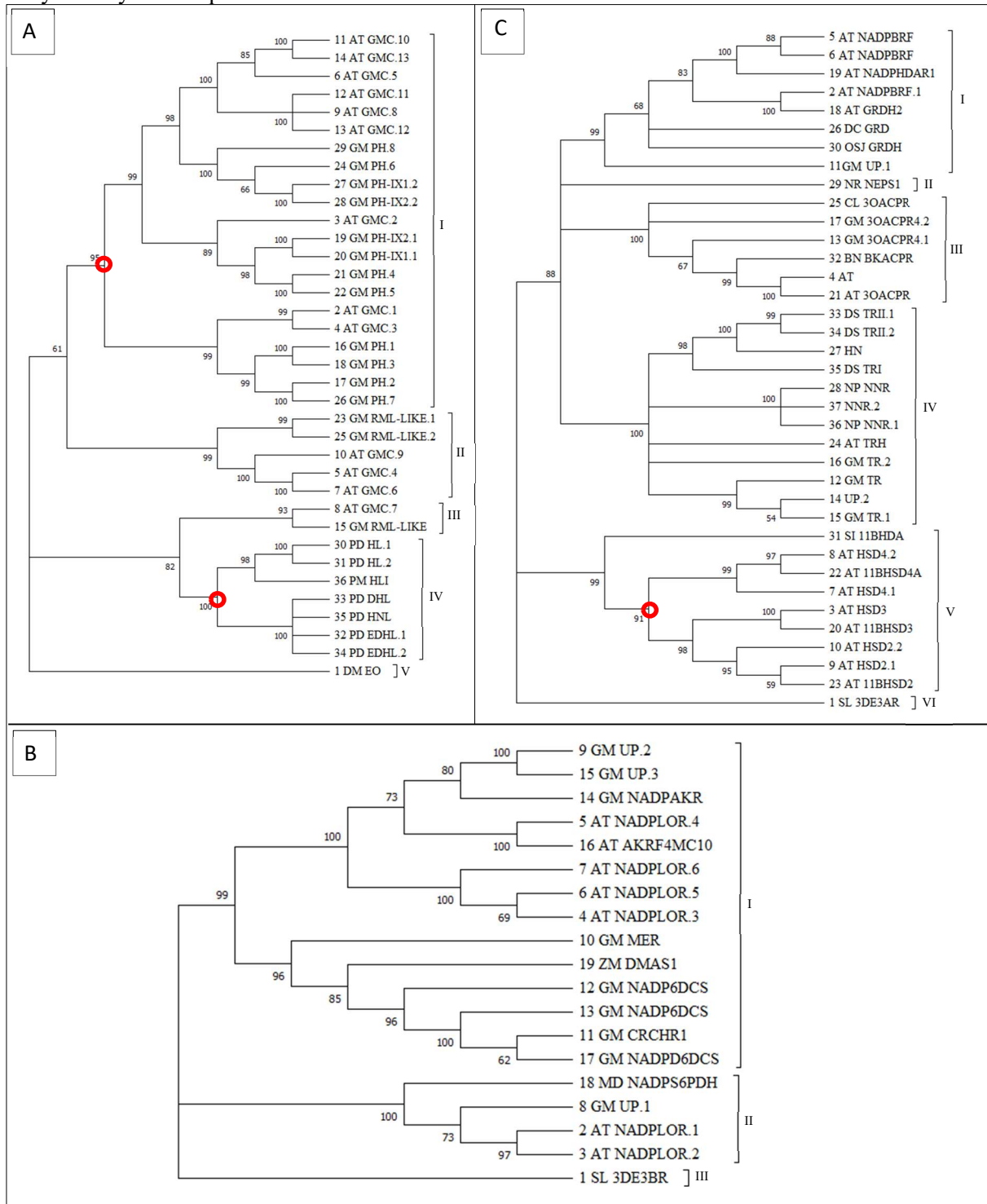
Por último, a filogenia da proteína 3-dehydroecdysone 3alpha-reductase (Figura 13 – C) apresentou cinco grupos diferentes. O Grupo I consiste em três espécies distintas de plantas: *A. thaliana*, *G. max*, *O. sativa* var. *japonica* e *Daucus carota*. Esse grupo é composto por proteínas NADP e NADPH dependentes, além de Glucose/Ribitol desidrogenases. O Grupo II é formado por uma proteína isolada da espécie *Nepeta racemosa*. Para o grupo III, inclui-se as espécies *Cuphea lanceolata* e *Brassica napus*, além das espécies *A. thaliana* e *G. max*. O grupo IV é o maior, além de englobar proteínas de diferentes espécies vegetais. O grupo V foi formado com maioria de proteínas provenientes da *A. thaliana*, exceto a espécie *Sesamum indicum*. Este grupo reuniu diversas proteínas do tipo *Hydroxysteroid dehydrogenase* (nó de origem destacado).

Figura 12 – Árvore filogenéticas construídas com proteínas vegetais similares às proteínas de *Drosophila melanogaster*, a partir do alinhamento múltiplo ClustalW utilizando o *software* MEGA ver. 10.8.1. A árvore foi construída pelo método de Máxima Verossimilhança e validada por análise Bootstrap (1000 repetições). As diferentes cores aplicadas representam o grupo formado por proteínas. A) Proteína Neverland; B) Proteína Shade; C) Proteína Shadow; D) Proteína Phantom.



Fonte: Autor (2021).

Figura 13 – Árvores filogenéticas construídas com proteínas vegetais similares às proteínas de *Drosophila melanogaster*, a partir do alinhamento múltiplo ClustalW utilizando o *software* MEGA ver. 10.8.1. A árvore foi construída pelo método de Máxima Verossimilhança e validada por análise Bootstrap (1000 repetições). As diferentes cores aplicadas representam o grupo formado por proteínas. A) Proteína Ecdysone oxidase; B) Proteína 3-dehydroecdysone 3alpha-reductase; C) Proteína 3-dehydroecdysone 3alpha-reductase.



Fonte: Autor (2021).

4.4 Desenho de oligonucleotídeos

A partir do alinhamento múltiplo e dos alinhamentos em pares das sequências de nucleotídeos dos genes de plantas potencialmente homólogos aos genes atuantes na rota de biossíntese e inativação da molécula 20-hidroxicidisona em insetos foi realizado o desenho de oligonucleotídeos em regiões que apresentam alta similaridade entre as sequências nucleotídicas analisadas para posteriores análises *in vitro* para localização e isolamento destes genes em plantas.

As Tabelas de número 19 a 26 apresentam os oligonucleotídeos desenhados com base em cada sequência de nucleotídeos dos genes atuantes na biossíntese em insetos, alinhadas com as sequências nucleotídicas de genes de plantas possivelmente homólogos. Cada oligonucleotídeo é formado em pares (*forward* e *reverse*), as tabelas apresentam as informações dos estados básicos como tamanho, temperatura de fusão e conteúdo GC%.

Os oligonucleotídeos (ou primers) foram confeccionados com tamanho variando de 18 a 25 pb. Com o objetivo de identificar os potenciais genes atuantes na rota de 20E em plantas. Os pares de *primers* foram alocados de maneira que seja possível obter um produto amplificado com tamanho o suficiente para analisar e identificar as regiões conservadas e compará-las com os genes que sintetizam 20E em insetos.

A temperatura de fusão (T_m) geral variou de 47 °C a 66 °C, enquanto que a concentração de guanina e citosina (GC%) geral variou de 31,58% a 85%. Durante a construção destes oligonucleotídeos, não houve formação de estruturas do tipo grampo (*hairpin*) e autoanelamento, nem formação de braçadeira GC (GC clamp) nas regiões terminais de cada *primer*.

Tabela 19 – Oligonucleotídeos *Forwards* e *Reverse*s propostos com base na identidade de sequências com o gene *Neverland* de *Drosophila melanogaster*. Estatística dos oligonucleotídeos realizada por meio da ferramenta PCR Primer Stats (Disponível em: https://www.bioinformatics.org/sms2/pcr_primer_stats.html).

Abreviações: pb = Par de bases; T_m = Temperatura de fusão (*Temperature of melting*) (em ° C); GC% = Conteúdo em porcentagem de Guanina e Citosina.

N.	Sentido		Oligonucleotídeos		Tam.	T_m	GC%
1	Forward	5'P	GCTTGGGCGAAGATCTAGTCAT	3'OH	22	55,00	50,00
	Reverse	5'P	ACCACAGCCTAAACTTCTTTATCG	3'OH	24	54,00	41,67
2	Forward	5'P	GCTTGGGCGAGGATCTAGTCAT	3'OH	22	57,00	54,55
	Reverse	5'P	ACCACAGCCTAAACTTCTCTATCG	3'OH	24	56,00	45,83
3	Forward	5'P	GCTTGGGCGAGCATCTAGTCAT	3'OH	22	57,00	54,55
	Reverse	5'P	ACCACAGCCTAAGCTTCTCTATCG	3'OH	24	57,00	50,00
4	Forward	5'P	GCTTGGGCGAGGATCGAGTCAT	3'OH	22	59,00	59,09
	Reverse	5'P	ACCACACCCTAAGCTTCTCTATCG	3'OH	24	57,00	50,00

5	Forward	5'P	GGTATCCTCAAATCGTCTC	3'OH	19	49,00	47,37
	Reverse	5'P	TCCAATGATGAGATCCAAG	3'OH	19	47,00	42,11
6	Forward	5'P	GGTATCCCCAAATCGTCTC	3'OH	19	52,63	51,00
	Reverse	5'P	TCCAATGATGAGATCCAGG	3'OH	19	49,00	47,37
7	Forward	5'P	GGTATCCCCAAGTCGTCTC	3'OH	19	53,00	57,89
	Reverse	5'P	TCCAATGACGAGATCCAGG	3'OH	19	51,00	52,63
8	Forward	5'P	GGTAGCCCCAAGTCGTCTC	3'OH	19	55,00	63,16
	Reverse	5'P	TCCAATGGCGAGATCCAGG	3'OH	19	53,00	57,89

Fonte: Autor (2021).

Tabela 20 – Oligonucleotídeos *Forwards* e *Reverse*s propostos com base na identidade de sequências com o gene *Shade* de *Drosophila melanogaster*. Estatística dos oligonucleotídeos realizada por meio da ferramenta PCR Primer Stats (Disponível em: https://www.bioinformatics.org/sms2/pcr_primer_stats.html).

Abreviações: pb = Par de bases; Tm = Temperatura de fusão (*Temperature of melting*) (em ° C); GC% = Conteúdo em porcentagem de Guanina e Citosina.

N.	Sentido		Oligonucleotídeos		Tam.	Tm	GC%
1	Forward	5'P	AGGACCAATGTGGCAGCGCCTACGA	3'OH	25	63	60
	Reverse	5'P	CGCTCTGGGGTAAACTGCTTGGC	3'OH	23	61	60,87
2	Forward	5'P	AAGACCAATGTGGCAGCGCCTACGA	3'OH	25	61	56
	Reverse	5'P	CGCTCTGGGGTAAACTGCTTGGG	3'OH	23	61	60,87
3	Forward	5'P	AAGTCCAATGTGGCAGCGCCTACGA	3'OH	25	61	56
	Reverse	5'P	CGCTCTGGGGTGAAGTCTTGGG	3'OH	23	62	65,22
4	Forward	5'P	AAGTCCGGTGTGGCAGCGCCTACGA	3'OH	25	64	64
	Reverse	5'P	CGCTCTGGGATGAAGTCTTGGG	3'OH	23	61	60,87
5	Forward	5'P	AAGTCCGGTGTGGCAGCGCCTACAA	3'OH	25	63	60
	Reverse	5'P	CGCTCTGGTTTGAAGTCTTGGG	3'OH	23	59	56,52
6	Forward	5'P	AAGTCCGGTGTGGCAGCGCTTACAA	3'OH	25	61	56
	Reverse	5'P	CGCTCTGGAATGAAGTCTTGGG	3'OH	23	59	56,52
7	Forward	5'P	AAGTCCGGTGTGGCTGCGCTTACAA	3'OH	25	61	56
	Reverse	5'P	CGCTCTGGTTTAAACTGCTTGGC	3'OH	23	57	52,17
8	Forward	5'P	AAGTCCGGTGTGGCTGCGGTTACAA	3'OH	25	61	56
	Reverse	5'P	CGCTCTGGAATAAACTGCTTGGC	3'OH	23	57	52,17
9	Forward	5'P	AAGAGGATTCCTTCCTGGG	3'OH	20	54	55
	Reverse	5'P	CACGATCTTCGACCCACGC	3'OH	19	55	63,16
10	Forward	5'P	GTGAGGATTCCTTCCTGGG	3'OH	20	56	60
	Reverse	5'P	CACGATCTTCGACCCAAGC	3'OH	19	53	57,89
11	Forward	5'P	GGGAGGATTCACTTCCTGGG	3'OH	20	56	60
	Reverse	5'P	CACGATCTCCGACCCAAGC	3'OH	19	55	63,16
12	Forward	5'P	GGGAGGCTTCACTTCCTGGG	3'OH	20	58	65
	Reverse	5'P	CACGATCTCCGACCAAAGC	3'OH	19	53	57,89
13	Forward	5'P	GGGAGCCTTCACTTCCTGGG	3'OH	20	58	65
	Reverse	5'P	CACGATCTCCGACCTAAGC	3'OH	19	53	57,89
14	Forward	5'P	GGGAGCCTTCACTTACTGGG	3'OH	20	56	60
	Reverse	5'P	CACGACCTCCGACCTAAGC	3'OH	19	55	63,16

15	Forward	5'P	GGGAGCCTTCACTTCCAGGG	3'OH	20	58	65
	Reverse	5'P	CACGACCTCCGTCCTAAGC	3'OH	19	55	63,16
16	Forward	5'P	GGGAGTCTTCACTTACAGGG	3'OH	20	54	55
	Reverse	5'P	CACGACCTCCGTCCTAATC	3'OH	19	53	57,89

Fonte: Autor (2021).

Tabela 21 – Oligonucleotídeos *Forwards* e *Reverse*s propostos com base na identidade de sequências com o gene *Shadow* de *Drosophila melanogaster*. Estatística dos oligonucleotídeos realizada por meio da ferramenta PCR Primer Stats (Disponível em: https://www.bioinformatics.org/sms2/pcr_primer_stats.html).

Abreviações: pb = Par de bases; Tm = Temperatura de fusão (*Temperature of melting*) (em ° C); GC% = Conteúdo em porcentagem de Guanina e Citosina.

N.	Sentido	Oligonucleotídeos	Tam.	Tm	GC%	
1	Forward	5'P GATTGGGCGGTACCCAGGATGC	3'OH	22	60	63,64
	Reverse	5'P ACTGAGCAGCACATCGGCCCA	3'OH	21	58	61,9
2	Forward	5'P GACTGGGCGGTACCCAGGATGC	3'OH	21	62	68,18
	Reverse	5'P ACTGAGCAGCACATCGGTCCA	3'OH	21	56	57,14
3	Forward	5'P GAGTGGGCGGTACCCAGGATGC	3'OH	21	62	68,18
	Reverse	5'P ACTGAGCCGCACATCGGTCCA	3'OH	21	58	61,9
4	Forward	5'P GAGTGGGCGGTACCCAGGAGGC	3'OH	21	62	68,18
	Reverse	5'P ACTGAGCCGCACGTCGGTCCA	3'OH	21	60	66,67
5	Forward	5'P GAGTGGGCGGTATCCAGGAGGC	3'OH	21	60	63,64
	Reverse	5'P ACTGTGCCGCACGTCGGTCCA	3'OH	21	60	66,67
6	Forward	5'P GAGTCGGCGGTATCCAGGAGGC	3'OH	21	62	68,18
	Reverse	5'P ACTGTGCGGCACGTCGGTCCA	3'OH	21	60	66,67
7	Forward	5'P GAGTCGACGGTATCCAGGAGGC	3'OH	21	60	63,64
	Reverse	5'P ACTGTGCGGCAAGTCGGTCCA	3'OH	21	58	61,9
8	Forward	5'P GAGTCGACGGTATGCAGGAGGC	3'OH	21	60	63,64
	Reverse	5'P ACTGTGCCGCAAGTCGGTCCA	3'OH	21	58	61,9
9	Forward	5'P GAGCCACGCACCTTCAT	3'OH	17	49	58,82
	Reverse	5'P GGTACAGACGCAGGGACTC	3'OH	19	55	63,16
10	Forward	5'P GGGCCACGCACCTTCAT	3'OH	17	52	64,71
	Reverse	5'P GGTACAGACGCAGGGATTC	3'OH	19	53	57,89
11	Forward	5'P GGGCCACGCACCTTCAT	3'OH	17	54	70,59
	Reverse	5'P GGTACAGACGCAGGGCCTC	3'OH	19	58	68,42
12	Forward	5'P GGGCCACACGCCTTCAT	3'OH	17	52	64,71
	Reverse	5'P GGTACAGACGCAGGGCTTC	3'OH	19	55	63,16
13	Forward	5'P GGGCCACACGCCGTCAT	3'OH	17	54	70,59
	Reverse	5'P GGTACAGACGCAAGGCTTC	3'OH	19	53	57,89
14	Forward	5'P GGGCCACACACCGTCAT	3'OH	17	52	64,71
	Reverse	5'P GGTACATACGCAAGGCTTC	3'OH	19	51	52,63
15	Forward	5'P GGGTCACACGCCGTCAT	3'OH	17	52	64,71
	Reverse	5'P GTTACATACGCAAGGCTTC	3'OH	19	49	47,37
16	Forward	5'P GGGTCACACGCCCCAT	3'OH	17	54	70,59
	Reverse	5'P GGTGCATACGCAAGGCTTC	3'OH	19	53	57,89

Fonte: Autor (2021).

Tabela 22 – Oligonucleotídeos *Forwards* e *Reverse*s propostos com base na identidade de sequências com o gene *Disembodied* de *Drosophila melanogaster*. Estatística dos oligonucleotídeos realizada por meio da ferramenta PCR Primer Stats (Disponível em: https://www.bioinformatics.org/sms2/pcr_primer_stats.html).

Abreviações: pb = Par de bases; Tm = Temperatura de fusão (*Temperature of melting*) (em ° C); GC% = Conteúdo em porcentagem de Guanina e Citosina.

N.	Sentido	Oligonucleotídeos	Tam.	Tm	GC%	
1	Forward	5'P GAGTGCACCAAAGAGTGTGCGGAAC	3'OH	25	61	56
	Reverse	5'P GACCGAAGGGAAGGACCAGATAGGGAT	3'OH	27	63	55,56
2	Forward	5'P GAGTGCACCAAAGACCGTGCGAAC	3'OH	25	60	63
	Reverse	5'P GACCGAAGGGAAGGACCGGATAGGAAT	3'OH	27	63	55,56
3	Forward	5'P CACACGCTTGATGGATGCGGCTGA	3'OH	24	61	58,33
	Reverse	5'P CACAAAGTAGCCACTGCAGATCGC	3'OH	19	59	54,17
4	Forward	5'P CACACGCTTGATGGATGCAACTGA	3'OH	24	57	50
	Reverse	5'P CACAAAGTAGCCACTGAAGATCGC	3'OH	24	57	50

Fonte: Autor (2021).

Tabela 23 – Oligonucleotídeos *Forwards* e *Reverse*s propostos com base na identidade de sequências com o gene *Phantom* de *Drosophila melanogaster*. Estatística dos oligonucleotídeos realizada por meio da ferramenta PCR Primer Stats (Disponível em: https://www.bioinformatics.org/sms2/pcr_primer_stats.html).

Abreviações: pb = Par de bases; Tm = Temperatura de fusão (*Temperature of melting*) (em ° C); GC% = Conteúdo em porcentagem de Guanina e Citosina.

N.	Sentido	Oligonucleotídeos	Tam.	Tm	GC%	
1	Forward	5'P CCGGGTCCCTGGGGCCTGCC	3'OH	20	66	85
	Reverse	5'P GGGTCCATGTGGATAGCCCA	3'OH	20	56	60
2	Forward	5'P CCGGGACCCTGGGGCCTGCC	3'OH	20	66	85
	Reverse	5'P GGGTCCATGTGGATTGCCCA	3'OH	20	56	60
3	Forward	5'P CCGGGACCCTGAGGCCTGCC	3'OH	20	64	80
	Reverse	5'P GGGTCCATGTGAATAGCCCA	3'OH	20	54	55
4	Forward	5'P CCGGGTCCCTGAGGCCTGCC	3'OH	20	64	80
	Reverse	5'P GGGTCCATGTGTATAGCCCA	3'OH	20	54	55
5	Forward	5'P CCGGGACCCTGCGGCCTGCC	3'OH	20	64	80
	Reverse	5'P GGGTCCATGTTCGATAGCCCA	3'OH	20	56	60
6	Forward	5'P CCGGGACCCTGCCGCCTGCC	3'OH	20	66	85
	Reverse	5'P GGGTCCATGTTCGATTGCCCA	3'OH	20	56	60
7	Forward	5'P CCGGGTCCCTGCCGCCTGCC	3'OH	20	66	85
	Reverse	5'P GGGTCTATGTTCGATTGCCCA	3'OH	20	54	55
8	Forward	5'P CCGGGACCCTGCCGCCTACC	3'OH	20	64	80
	Reverse	5'P GGGTCTATGTTGATTGCCCA	3'OH	20	52	50
9	Forward	5'P CCCTTTCTGGGTTACCTGC	3'OH	19	53	57,89
	Reverse	5'P CGATAGCCGGACGAGAATGG	3'OH	20	56	60
10	Forward	5'P CCCTTTCTGGGTTGCCTGC	3'OH	19	55	63,16
	Reverse	5'P CGATAGCCGGACGAGAACGG	3'OH	20	58	65
11	Forward	5'P CCCTTTCTGGGCTGCCTGC	3'OH	19	58	68,42

	Reverse	5'P	CGATAGCCGGACGAAAACGG	3'OH	20	56	60
12	Forward	5'P	CCCTTCCTGGGCTGCCTGC	3'OH	19	60	73,68
	Reverse	5'P	CGATAGCCGGACCAAAAACGG	3'OH	20	56	60
13	Forward	5'P	CCCTTCCTGGGGTGCCTGC	3'OH	19	60	73,68
	Reverse	5'P	CGATAGCCGGACCAAAAATGG	3'OH	20	54	55
14	Forward	5'P	CCCTTTCTGGGGTGCCTGC	3'OH	19	58	68,42
	Reverse	5'P	CGATAACCGGACCAAAAATGG	3'OH	20	54	55
15	Forward	5'P	CCCTTTCTGGGTTCCCTGC	3'OH	19	55	63,16
	Reverse	5'P	CGATAACCAGACCAAAAACGG	3'OH	20	52	50
16	Forward	5'P	CCCTTGCTGGGCTGCCTGC	3'OH	19	60	73,68
	Reverse	5'P	CGATAACCTGACCAAAAACGG	3'OH	20	52	50

Fonte: Autor (2021).

Tabela 24 – Oligonucleotídeos *Forwards* e *Reverse*s propostos com base na identidade de sequências com o gene *Ecdysone Oxidase* de *Drosophila melanogaster*. Estatística dos oligonucleotídeos realizada por meio da ferramenta PCR Primer Stats (Disponível em: https://www.bioinformatics.org/sms2/pcr_primer_stats.html).

Abreviações: pb = Par de bases; Tm = Temperatura de fusão (*Temperature of melting*) (em ° C); GC% = Conteúdo em porcentagem de Guanina e Citosina.

N.	Sentido		Oligonucleotídeos		Tam.	Tm	GC%
1	Forward	5'P	GGTGCAGGTAGTGCCGGAT	3'OH		55	63,16
	Reverse	5'P	GTCCTGCGCTTAGAATGACCTC	3'OH	23	57	52,17
2	Forward	5'P	GGGGCAGGTAGTGCCGGAT	3'OH		58	68,42
	Reverse	5'P	GTCCTGCGCTTAGGATGACCTC	3'OH	23	59	56,52
3	Forward	5'P	GGGGCTGGTAGTGCCGGAT	3'OH		58	68,42
	Reverse	5'P	GTCCTGCGCTTAGAATGACCTC	3'OH	23	55	47,83
4	Forward	5'P	GGGGCTGGGAGTGCCGGAT	3'OH		60	73,68
	Reverse	5'P	GTCCTGCGCTTAGTATGACCTC	3'OH	23	55	47,83
5	Forward	5'P	GGTGCTGGAAGTGCCGGAT	3'OH		55	63,16
	Reverse	5'P	GTCCTGCGCTTACAATGACTTC	3'OH	23	55	47,83
6	Forward	5'P	GGAGCTGGAAGTGCCGGAT	3'OH		58	68,42
	Reverse	5'P	GTCCTGCGCTTACTATGACTTC	3'OH	23	55	47,83
7	Forward	5'P	GGAGCTGGAAGTGCCGGAT	3'OH		58	68,42
	Reverse	5'P	GTCCTGCGCTTACTATGACCTC	3'OH	23	57	52,17
8	Forward	5'P	GGTGCTGGAAGTGCCGGAT	3'OH		55	63,16
	Reverse	5'P	GTCCTGCGCTTACTATGACCTC	3'OH	23	59	56,52
9	Forward	5'P	ACGGCTGCGTTTGGTGGCATGG	3'OH	22	60	63,64
	Reverse	5'P	CCCGAAAGCAGAAGCAACTGG	3'OH	22	57	54,55
10	Forward	5'P	ATGGCTGCGTTTGGTGGCATGG	3'OH	22	59	59,09
	Reverse	5'P	CCCGAAAGCAGAAGCAACTTG	3'OH	22	55	50
11	Forward	5'P	ATGGATGCGTTTGGTGGCATGG	3'OH	22	57	54,55
	Reverse	5'P	CCCGAAAGCAGAAGCAACTTAG	3'OH	22	55	50
12	Forward	5'P	ATGGTTGCGTTTGGTGGCATGG	3'OH	22	57	54,55
	Reverse	5'P	CCCGAAAGCAGAAGCAACTGAG	3'OH	22	57	54,55
13	Forward	5'P	ATGGTTGTGTTTGGTGGCATGG	3'OH	22	55	50

	Reverse	5'P	CCCGAAAGCAGAAGCAGCTGAG	3'OH	22	59	59,09
14	Forward	5'P	ACGGTTGTGTTTGGTGGCATGG	3'OH	22	57	54,55
	Reverse	5'P	CCCGAAAGCAGAAGAAGCTGAG	3'OH	22	57	54,55
15	Forward	5'P	ACGGCTGTGTTTGGTGGCATGG	3'OH	22	59	59,09
	Reverse	5'P	CCAGAAAGCAGAAGAAGCTGAG	3'OH	22	55	50
16	Forward	5'P	ACGGATGTGTTTGGTGGCATGG	3'OH	22	57	54,55
	Reverse	5'P	CCAGAAAGCAGAAGAAGCTGGG	3'OH	22	57	54,55

Fonte: Autor (2021).

Tabela 25 – Oligonucleotídeos *Forwards* e *Reverse*s propostos com base na identidade de sequências com o gene *3-dehydroecdysone 3beta-reductase* de *Spodoptera littoralis*. Estatística dos oligonucleotídeos realizada por meio da ferramenta PCR Primer Stats (Disponível em: https://www.bioinformatics.org/sms2/pcr_primer_stats.html).

Abreviações: pb = Par de bases; Tm = Temperatura de fusão (*Temperature of melting*) (em ° C); GC% = Conteúdo em porcentagem de Guanina e Citosina.

N.	Sentido		Oligonucleotídeos		Tam.	Tm	GC%
1	Forward	5'P	GACTTTCGACTGCTGCCATCT	3'OH	24	59	54,17
	Reverse	5'P	TTGATGTTTTCCAACAGTC	3'OH	19	45	36,84
2	Forward	5'P	GCCACTTCGACTGCTGCCATCT	3'OH	24	61	58,33
	Reverse	5'P	TTGATGTTCTCCAACAGTC	3'OH	19	47	42,11
3	Forward	5'P	GGCACTTCGACTGCTGCCATCT	3'OH	24	61	58,33
	Reverse	5'P	TTGAGGTTCTCCAACAGTC	3'OH	19	49	47,37
4	Forward	5'P	GGCACTTCGATACTGCTGCCATCT	3'OH	24	59	54,17
	Reverse	5'P	TTCAGGTTCTCCAACAGTC	3'OH	19	49	47,37
5	Forward	5'P	GCCACTTCGATACTGCTGCCATCT	3'OH	24	59	54,17
	Reverse	5'P	TTCAGGTTCTCCAACAGCC	3'OH	19	51	52,63
6	Forward	5'P	GCCACTTCGATACTGCTGCCGTCT	3'OH	24	61	58,33
	Reverse	5'P	TTCAGGTTCTCCACCAGCC	3'OH	19	53	57,89
7	Forward	5'P	GGCACTTCGATACTGCTGCCGTCT	3'OH	24	61	58,33
	Reverse	5'P	TTCAGGTTCTGCACCAGCC	3'OH	19	53	57,89
8	Forward	5'P	GACTTTCGATAGTGCTGCCGTCT	3'OH	24	59	54,17
	Reverse	5'P	TTCAGGTTCTGCACCAGTC	3'OH	19	51	52,63
9	Forward	5'P	TACAACACGGAGGCTGAAGTGGG	3'OH	23	59	56,52
	Reverse	5'P	CCAACGCAACACAATTTGGGC	3'OH	21	54	52,38
10	Forward	5'P	TACGACACGGAGGCTGAAGTGGG	3'OH	23	61	60,87
	Reverse	5'P	CCAACGCAACACAATCTGGGC	3'OH	21	56	57,14
11	Forward	5'P	TACGGCACGGAGGCTGAAGTGGG	3'OH	23	62	65,22
	Reverse	5'P	CCAACGCAACACAACCTGGGC	3'OH	21	58	61,9
12	Forward	5'P	TACGGCTCGGAGGCTGAAGTGGG	3'OH	23	62	65,22
	Reverse	5'P	CCAACGCAACACCACCTGGGC	3'OH	21	60	66,67
13	Forward	5'P	TACGGCTAGGAGGCTGAAGTGGG	3'OH	23	61	60,87
	Reverse	5'P	CCAACGCAACGCCACCTGGGC	3'OH	21	62	71,43
14	Forward	5'P	TACGGCTAGGAAGCTGAAGTGGG	3'OH	23	59	56,52
	Reverse	5'P	CCAACGCAAGGCCACCTGGGC	3'OH	21	62	71,43
15	Forward	5'P	TACGACTAGGAAGCTGAAGTGGG	3'OH	23	57	52,17

	Reverse	5'P	CCACCGCAAGGCCACCTGGGC	3'OH	21	64	76,19
16	Forward	5'P	TACAACTAGGAAGCTGAAGTGGG	3'OH	23	55	47,83
	Reverse	5'P	CCACCGCAAGGCCACTTGGGC	3'OH	21	62	71,43

Fonte: Autor (2021).

Tabela 26 – Oligonucleotídeos *Forwards* e *Reverse*s propostos com base na identidade de sequências com o gene *3-dehydroecdysone 3alpha-reductase* de *Spodoptera littoralis*. Estatística dos oligonucleotídeos realizada por meio da ferramenta PCR Primer Stats (Disponível em: https://www.bioinformatics.org/sms2/pcr_primer_stats.html).

Abreviações: pb = Par de bases; Tm = Temperatura de fusão (*Temperature of melting*) (em ° C); GC% = Conteúdo em porcentagem de Guanina e Citosina.

N.	Sentido		Oligonucleotídeos		Tam.	Tm	GC%
1	Forward	5'P	GCTGCCATCTACAACACGGAGGCTG	3'OH	25	63	60
	Reverse	5'P	GCAACACAATTTGGGCAGGAGT	3'OH	22	55	50
2	Forward	5'P	GTTGCCATCTACAACACGGAGGCTG	3'OH	25	61	56
	Reverse	5'P	GCAACACAATTTGGGCAGGAAT	3'OH	22	53	45,45
3	Forward	5'P	GTTGCCATCTACAACACGGAGGCAG	3'OH	25	61	56
	Reverse	5'P	GCAACACAATTTGGGCAGGTGT	3'OH	22	55	50
4	Forward	5'P	GTTGCCATCTACAACACGGAGTCAG	3'OH	25	59	52
	Reverse	5'P	GCAACACAATTTGGGCAGGGGT	3'OH	22	57	54,55
5	Forward	5'P	GTTGCCATCTACAACACGGCGTCAG	3'OH	25	61	56
	Reverse	5'P	GCAACACAATTTGGGCCGGGGT	3'OH	22	59	59,09
6	Forward	5'P	GTTGCCATCTACAACACTGCGTCAG	3'OH	25	59	52
	Reverse	5'P	GCAACACACTTTGGGCCGGGGT	3'OH	22	60	63,64
7	Forward	5'P	GTTGCCATCTACAAAACACTGCGTCAG	3'OH	25	58	48
	Reverse	5'P	GCACCACACTTTGGGCCGGGGT	3'OH	22	62	68,18
8	Forward	5'P	GCTGCCATCTACAAAACACTGCGTCAG	3'OH	25	59	52
	Reverse	5'P	GCACAACACTTTGGGCCGGGGT	3'OH	22	60	63,64
9	Forward	5'P	TCAGAATGAAGATTGATGAGGGAGTG	3'OH	26	56	42,31
	Reverse	5'P	ATAATTTGAGGATGAACCT	3'OH	19	42	31,58
10	Forward	5'P	TCGGAATGAAGATTGATGAGGGAGTG	3'OH	26	58	46,15
	Reverse	5'P	ATAATCTGAGGATGAACCT	3'OH	19	45	36,84
11	Forward	5'P	TCGGGATGAAGATTGATGAGGGAGTG	3'OH	26	60	50
	Reverse	5'P	AGAATCTGAGGATGAACCT	3'OH	19	47	42,11
12	Forward	5'P	TCGTGATGAAGATTGATGAGGGAGTG	3'OH	26	58	46,15
	Reverse	5'P	AGGATCTGAGGATGAACCT	3'OH	19	49	47,37
13	Forward	5'P	TCGTGATGGAGATTGATGAGGGAGTG	3'OH	26	60	50
	Reverse	5'P	AGGATCGGAGGATGAACCT	3'OH	19	51	52,63
14	Forward	5'P	TCGTGATGGAGTTTGTGAGGGAGTG	3'OH	26	60	50
	Reverse	5'P	AGGATCGGAGGATGAGCCT	3'OH	19	53	57,89
15	Forward	5'P	TCGTGATGGAGTTTGGTGAGGGAGTG	3'OH	26	61	53,85
	Reverse	5'P	AGGATCGGAGGTTGAGCCT	3'OH	19	53	57,89
16	Forward	5'P	TCGTGATGGAGTTTGGTGAGGAAGTG	3'OH	26	60	50
	Reverse	5'P	AGGATTGGAGGTTGAGCCT	3'OH	19	51	52,63

Fonte: Autor (2021).

3.4 Discussão

A comparação entre as sequências de nucleotídeos de genes e as sequências de aminoácidos de proteínas presentes entre organismos de diferentes espécies permite o estudo dos mecanismos evolutivos e identificação de sequências que se mantiveram conservadas ao longo da evolução das espécies. Além disso, entender esses mecanismos que ocorrem naturalmente possibilita a elaboração de estratégias visando a identificação, isolamento, caracterização e análise de genes e proteínas de interesse biotecnológico.

Desde a década de 90 a movimentação gênica entre espécies já era alvo de estudos para criação de vegetais superiores. A manipulação de genes presentes em alguns organismos viabiliza o aumento da resistência de cultivos agrícolas, como o trabalho de Perlak et al. (1991), que transferiu e aumentou a expressão do gene *cryIA(b)* da bactéria *Bacillus thuringiensis*, capaz de afetar o desenvolvimento de lepidópteros, nas espécies de planta tabaco e tomate.

A partir das sequências de aminoácidos das proteínas envolvidas na biossíntese e inativação da molécula 20-hidroxicidona (20E) em *Drosophila melanogaster* e *Spodoptera littoralis*, foi possível encontrar 239 proteínas similares distribuídas em diversas famílias de plantas, assim como o trabalho realizado por Egan et al. (1999), que avaliou a existência de diferentes proteínas do citocromo P450 com funções similares entre insetos, plantas e mamíferos, com até 50% de identidade (ou similaridade) entre as sequências.

Além da utilização da similaridade, outro importante parâmetro estatístico para a seleção de proteínas para este trabalho foi o *valor-E*, em que se selecionou as proteínas com o valor mais próximo de zero, com um corte de e^{-5} . De acordo com Kerfeld e Scott (2011), as sequências que compartilham de um ancestral em comum possuem altos valores de similaridade, baixo *valor-E* e, no alinhamento de suas sequências, terão maior quantidade de resíduos idênticos, baixas substituições e lacunas.

A similaridade das proteínas de plantas encontradas a partir das sequências das proteínas dos insetos foi variável, mas não superior a 40%. Domínios proteicos com pelo menos 30% de similaridade possuem a mesma estrutura no dobramento final, entretanto o mesmo dobramento de proteínas não significa estatisticamente que compartilham de sequências similares (ORENGO; JONES; THORNTON, 1994). A similaridade na estrutura traz fortes indícios de relações evolucionárias, uma vez que a estrutura de proteínas, durante o curso evolutivo, é bem

mais conservada que a sequência em si. A estrutura possui um papel essencial para a função de uma proteína (KOLODNY; PETREY; HONIG, 2006).

Muitas das espécies de plantas encontradas ainda não possuem registro de produção da molécula 20E. O banco de dados de plantas produtoras (disponível em *ecdybase.org*) (DINAN; LAFONT, 2019) ainda está em construção, entretanto ocorrem estudos no mundo inteiro com diferentes espécies de plantas, algas e fungos, capazes de gerar uma grande base de dados de referência para a área de concentração.

Em relação as proteínas encontradas, houve uma grande quantidade de isoformas nos resultados de todas as proteínas atuantes na rota metabólica de 20E, isso se deve ao fato que durante o curso evolutivo das famílias de gene se desenvolvem várias isoformas em diferentes organismos devido a pressão evolutiva seletiva, gerando especificidade funcional de acordo com a espécie e conservação de partes da proteína que atuam na interação proteína/proteína (ROSENQUIST et al., 2000; SEHNKE et al., 2002; CASTELLANO; SANTOS, 2011).

De maneira geral, a caracterização físico-química das proteínas mostrou diversificada composição entre os parâmetros estimados, isso se deve ao fato que pelo menos 70% das composições das sequências de aminoácidos dessas proteínas não possuem similaridade com as sequências de aminoácidos de proteínas dos insetos. As proteínas NVD, SAD, DIB e EO apresentaram pontos isoelétrico teóricos acima de 8, isto mostra a natureza básica destas proteínas, enquanto que para SAD, PHT, 3DE3BR e 3DE3AR apresentaram natureza mais ácida. Os parâmetros Peso Molecular, número de resíduos positivos e negativos e coeficiente de extinção foram variados para todas as proteínas atuantes na rota, isso ocorre devido estes parâmetros dependerem do tamanho das proteínas e, também, a composição das mesmas.

O Índice de instabilidade se refere a uma estimativa de estabilidade de proteína em ensaios de bancada, dentre as proteínas que atuam na rota, NVD e 3DE3AR foram as únicas que obtiveram valores abaixo de 40 e consideradas estáveis. Em relação às proteínas encontradas em plantas, verificou-se heterogeneidade de resultados, apresentando tantas proteínas estáveis, quanto instáveis, segundo as estimativas da ferramenta ProtParam. O índice alifático foi superior a 70 em todas as proteínas de insetos, e em grande maioria das proteínas similares, isto indica maior quantidade de resíduos alifáticos que aromáticos.

A variável GRAVY se refere a capacidade de solubilidade da proteína, valores positivos significam hidrofobicidade, enquanto que valores negativos significam que a proteína é mais

hidrofílica. Todas as proteínas apresentaram capacidades hidrofílicas elevadas, exceto para as proteínas de inseto 3DE3BR e 3DEAR e suas proteínas similares, que apresentaram natureza hidrofóbica.

Foram construídas árvores filogenéticas por meio das ferramentas de bioinformática para cada proteína atuante na rota de biossíntese de 20E. A partir da estatística de validação por *Bootstrap*, diversos grupos foram formados englobando as famílias do complexo citocromo P450 independentemente das espécies, demonstrando existência de ancestral em comum. Em alguns casos, proteínas distintas foram classificadas em clados com famílias definidas, isso se deve ao fato das diferentes isoformas apresentarem mudanças nas nomenclaturas no decorrer dos anos (DANIELSON; JOHANSON, 2010).

A filogenia de proteínas é uma análise robusta e possui alguns desafios principalmente para validação de famílias multigênicas, que exigem também de dados fora da análise filogenética, como por exemplo, a assinatura de *motifs*, ontologia gênica, parâmetros físico-químicos dos produtos entre outros (PAQUETTE; MØLLER; BAK, 2003). As distinções clássicas para separar as proteínas em grupos funcionais são o tamanho, número e organização dos domínios proteicos, a localização subcelular, os *motifs* e sítios ativos das atividades enzimáticas, desta forma, proteínas com atividades distintas (diferentes famílias) são agrupadas em clados diferentes, assim como foi observado nas análises realizadas com as diferentes proteínas de plantas neste trabalho.

A classificação das proteínas potencialmente homólogas em diferentes grupos, identificadas com diferentes funções enzimáticas, ocorre em decorrência de uma divergência funcional ocorrida após a duplicação dos genes, o que gera um papel crítico na diversificação primária das famílias gênicas (BUCK; ATCHLEY, 2005; HOUSTON et al., 2005). A ordenação em diferentes grupos na filogenia permite inferir que essas proteínas possuem um ancestral em comum, mas que adquiriram novas funções durante o curso evolutivo.

Foram desenhados oligonucleotídeos para cada enzima que atua na rota de biossíntese e inativação de 20E. Isso foi permitido devido a homologia entre as sequências de aminoácidos das proteínas e das sequências de nucleotídeos dos genes encontrados em insetos e plantas, em que estas apresentam regiões conservadas durante o curso evolutivo, identificadas e avaliadas durante os alinhamentos duplos (em pares) e múltiplos. Segundo Makkar et al. (2019), um passo primordial para o desenvolvimento de variedades superiores resistentes a herbivoria é o conhecimento dos genes a nível molecular, que necessitam serem clonados e caracterizados

para elucidar a sua estrutura e função nos organismos. O presente estudo apresentou vários pares de oligonucleotídeos para cada gene estudado que podem ser utilizados para se isolar os genes e se determinar a sua função e relação com a interação planta-inseto e com a rota de biossíntese e inativação de 20E em plantas.

Foram detectadas proteínas que apresentam funções conhecidas e também proteínas que apresentam funções desconhecidas. Também foram encontradas proteínas que aparentemente podem não estar relacionadas a defesa e/ou interação planta/inseto. O envolvimento destas proteínas na interação planta-inseto e na rota de biossíntese e inativação de 20E ainda precisam ser investigados. Também é possível a existência de outras funções ainda não caracterizadas para as proteínas encontradas ou a existência de uma rota alternativa desenvolvida pelos vegetais durante o curso evolutivo, com proteínas possuindo diferentes funções bioquímicas quando comparadas as proteínas atuantes na rota de insetos. O papel das proteínas analisadas e seu possível envolvimento com rota de biossíntese de 20E em plantas necessita de estudos adicionais e validação em trabalhos *in vivo*.

3.4.1 Proteína Neverland

Para a proteína Neverland, duas espécies de plantas foram encontradas, entretanto apenas *Oryza sativa* var. japonica apresenta a produção da molécula 20-hidroxiecdsona (BLACKFORD; CLARKE; DINAN, 1996). O arroz é uma das principais culturas agrícolas no mundo, principalmente no continente asiático. Por uma cultura susceptível a diversas pragas, chegando a ter perdas consideráveis de 15% ao ano (SHARMA et al., 2017), possui grandes programas de melhoramento genético a pragas e doenças.

A proteína Chlorophyllide a oxygenase (CAO) é uma proteína conhecida com uma grande importância no processo de fotossíntese em plantas, pois permite uma reação de oxigenação para transformar a Clorofila A em Clorofila B. Essa proteína é codificada por um único gene (TANAKA et al., 1998; NAGATA et al., 2004). Assim como a proteína CAO, a Pheophorbide a oxygenase (PAO) atua no processo da fotossíntese em plantas, especificamente na degradação da clorofila, catalizando a abertura do anel de porfirina da molécula *Pheophorbide a* (MA et al., 2012).

As duas proteínas encontradas possuem atuação nos cloroplastos e são amplamente estudadas, cujas sequências são extremamente conservadas com uma similaridade maior que 70%, com uma sequência N-terminal cloroplástica conservada em diversas espécies de plantas

como *Arabidopsis*, tomate e milho (LEI, 2008). Não há registros em literatura da atuação destas proteínas diretamente ligadas a defesa de plantas aos insetos, entretanto segundo Lorrain et al. (2003), plantas mutantes que possuem deficiência da proteína Red chlorophyll catabolite reductase (gene RCCR), que é essencial para a atuação enzimática da proteína PAO, geraram os fenótipos mimetizadores de lesão, com defeitos específicos na sinalização de morte celular e/ou rotas de defesa das plantas.

3.4.2 Proteína Shade

A proteína Shade foi a que mais obteve mais proteínas similares em plantas, bem como a maior variedade de espécies vegetais, como observado nos resultados gerais, entretanto de acordo com o banco de dados de presença/ausência de 20-hidroxiecdisona (20E), somente duas espécies foram confirmadas como produtoras de 20E. A maioria não possui dados ou estudos avaliando a existência do fitohormônio.

As espécies que apresentaram a produção de 20E foram *Arabidopsis thaliana* e *Lactuca sativa*, ambas são espécies com grande interesse econômico. A alface (*L. sativa*) já é alvo de estudos avaliando a sua resistência a herbivoria (HUANG; MCAUSLANE; NUSSLY, 2003; TIERRANEGRA-GARCÍA et al., 2011), além de programas de melhoramento com a inclusão de genes considerados anti-herbivoria e promover resistência a espécie (KAWAZU et al., 2012; SMITH; CHUANG, 2014).

Em relação as proteínas encontradas, todas fazem parte da família 71 do Citocromo P450. Esta família de proteínas possui um papel essencial na defesa das plantas, participando da biossíntese de compostos que estão relacionados com a proteção a herbivoria, como por exemplo genes CYP71 envolvidos na conversão de aldoxima em volátil nitrila, repelindo lagartas generalistas (IRMISCH et al., 2014) e produção de fitotóxicos do tipo aldeídos sesquiterpênicos (TIAN et al., 2018).

3.4.3 Proteína Shadow

Além da *A. thaliana*, foram encontradas duas espécies positivas com produção de 20E na busca por proteínas similares com a proteína *Shadow*. A espécie *Prunus mume* possui grande interesse econômico para as regiões da Ásia, a sua domesticação ocorreu a mais de 3.000 anos e foi um dos primeiros gêneros da subfamília *Prunus* a serem totalmente sequenciados e ter seu genoma disponível (ZHANG et al., 2012). Já a espécie *Salvia miltiorrhiza* é conhecida

principalmente por suas propriedades medicinais utilizada amplamente na medicina tradicional no continente Asiático (CHUN-YAN et al., 2015).

Diversas famílias do complexo citocromo P450 foram encontradas, além da CYP71 já abordada. A família CYP76 em *A. thaliana* está relacionada com a resistência a herbicidas, no metabolismo de monoterpênicos e herbicidas fenilureias (HÖFER et al., 2014). Em relação à família CYP81, segundo Wang et al. (2019), existem muitos alelos em decorrência da duplicação tandem de genes durante a evolução, adquirindo diversas funções em plantas, mas majoritariamente eles atuam na resposta a estresse abiótico, como por exemplo, tolerância a salinidade.

A família CYP82, por outro lado, atuam nos processos biológicos envolvendo a interação planta/inseto, conferindo defesa as plantas atuando na via de sinalização de ácido jasmônico e etileno em soja (YAN et al., 2016), biossíntese de voláteis em *A. thaliana* (LEE et al., 2010) e biossíntese de alcaloides (HORI et al., 2018). Foi encontrado a proteína CYP91A1 em *A. thaliana*, entretanto segundo Werck-Reichhart, Bak e Paquette (2002) essa classificação não existe mais e foi renomeada como integrante da família CYP82. As proteínas podem sofrer atualizações em suas funções e reclassificações de acordo com o avanço científico, entretanto a mudança nas configurações dos bancos de dados pode não acompanhar as atualizações, sendo necessário a pesquisa em literatura atualizada.

Ainda presente em *A. thaliana*, as proteínas das famílias CYP703, CYP705 e CYP711 ainda não possuem definidas as suas funções. Esses genes são específicos da família Brassicaceae e surgiram de repetitivas duplicações gênicas no genoma dessa espécie (WERCK-REICHHART; BAK; PAQUETTE, 2002). Os CYP705 talvez estejam envolvidos na biossíntese de derivados triterpenóides, que são compostos que tem função de defesa contra pragas e doenças nas plantas, bem como propriedades anticancerígenas em humanos (FIELD; OSBOURN, 2008).

3.4.4 Proteína Disembodied

Para única espécie encontrada durante a busca para a proteína Disembodied não detectada a presença de 20E. Desta forma, a utilização da sequência da proteína de inseto não foi significativa para identificar similaridade em proteínas vegetais. Além disso, o papel desta proteína, de hidroxilação do carbono 22 na rota de biossíntese em insetos, é, segundo

Tsukagoshi et al. (2016), possivelmente desempenhada por uma enzima CYP71D443, de acordo com estudos desenvolvidos com a espécie de planta do gênero *Ajuga*.

3.4.5 Proteína Phantom

O resultado para a proteína Phantom foi similar às proteínas anteriores, observou-se que somente as espécies *A. thaliana* e *Salvia miltiorrhiza* possuem produção de 20E. Além disso, as proteínas encontradas fazem parte do complexo citocromo P450, nas famílias CYP71, responsável pela defesa das plantas a agentes bióticos, bem como as famílias CYP706, que se supõe participar nos processos de defesa a agentes abióticos.

As famílias CYP90 e CYP97 possuem poucas citações em literatura, entretanto acredita-se que a CYP90 está envolvida na rota de biossíntese de brassinosteroide em tomates, enquanto que a família CYP97 possui integrantes envolvidos na hidroxilação de carotenoides (KIM; DELLAPENNA, 2006; OHNISHI et al., 2006; NELSON et al., 2008).

3.4.6 Proteína Ecdysone Oxidase

Esta proteína atua na inativação da rota da biossíntese de 20E, transformando a Ecdisona em 3-dehidroxiecdisona. Das espécies encontradas, apenas em *A. thaliana* e *Prunus mume* foi detectado a existência de 20E. Para a *A. thaliana*, foram encontradas 13 isoformas não identificadas da família de proteínas Glucose-methanol-choline, classificadas como proteínas flavino oxidases. Essas oxidorredutases são encontradas em diversos organismos, mas ainda não estão totalmente caracterizadas, possuindo um grande potencial biotecnológico. Em algas, essas enzimas são capazes de converter ácidos graxos em alcanos ou alcenos com a utilização da energia solar (ALEKSENKO et al., 2020). Embora os membros desta família de proteínas oxidorredutases compartilhem estruturas e possuam um sítio de ligação FAD idêntico, elas apresentam diferentes atividades catalíticas que ainda não estão elucidadas (MICHALSKI et al., 2008).

A proteína encontrada no damasco asiático (*P. mume*) é uma hydroxynitrile lyase, integrante do grupo de proteínas contendo Dinucleótido de Flavina e Adenina (FAD, em inglês). Embora essa proteína seja encontrada em outros organismos, é uma proteína sobretudo vegetal, com uma atividade enzimática capaz de catalisar o passo final da rota de biodegradação de glicosídeos cianogênicos (NANDA; KATO; ASANO, 2005; FUKUTA et al., 2011). Os glicosídeos cianogênicos fazem parte dos metabólitos secundários e são toxinas naturais produzidas por algumas plantas como forma de defesa (VETTER, 2000).

3.4.7 Proteína 3-Dehydroecdysone 3beta-Reductase

Para a proteína 3-Dehydroecdysone 3beta-Reductase somente a espécie *A. thaliana* foi confirmada como produtora da molécula 20E. As proteínas cadastradas como NAD(P)-linked oxidoreductase são uma superfamília englobando a família Aldo-keto reductase. Estas enzimas estão relacionadas com a redução de Nicotinamida Adenina Dinucleotídeo fosfato (NAD(P), em inglês) para a formação de álcoois primários e secundários a partir de uma grande quantidade de substratos como por exemplo, aldeídos e cetonas aromáticos, cetosteroides e xenobióticos (MINDNICH; PENNING, 2009). Assim, de acordo com Simpson et al. (2009), essas enzimas atuam durante os processos de estresse biótico e abiótico de plantas metabolizando os compostos tóxicos produzidos.

3.4.8 Proteína 3-Dehydroecdysone 3alfa-Reductase

A proteína 3-Dehydroecdysone 3alfa-Reductase apresentou quatro espécies com a produção de 20E, sendo elas *A. thaliana*, *Daucus carota* (cenoura selvagem), *Oryza sativa* var. japonica e *Datura stramonium* (erva trombeteira). Essas espécies possuem importância econômica, agrícola e medicinal. A arábida foi a espécie que mais apresentou proteínas similares dentre as citadas.

Em *A. thaliana*, a superfamília de proteína NAD(P)-binding Rossmann-fold, incluindo as proteínas Glucose e ribitol dehydrogenase, Tropinone reductase e 3-oxoacyl-[acyl-carrier-protein] reductase, possuem uma importante função durante a detoxificação de carbonilas reativas em plantas durante o processo de estresse biótico ou abiótico (YAMAUCHI et al., 2011). Já a proteína Hydroxysteroid Dehydrogenase possui função na regulação do crescimento e desenvolvimento de plantas (LI et al., 2007). As demais espécies também apresentaram proteínas integrantes da superfamília NAD(P) dependentes.

3.5 Conclusões

Este trabalho identificou e analisou *in silico* 239 proteínas em plantas possivelmente homólogas as proteínas que atuam na biossíntese de 20-hidroxiecdisona em insetos. Estas proteínas encontradas apresentaram similaridade entre 25% a 45%, altas taxas de cobertura e valor-E de pelo menos e^{-5} .

As proteínas similares em plantas apresentam funções conhecidas, mas possivelmente ainda existam funções não identificadas, desta maneira o envolvimento e real função destas

proteínas ainda precisam ser analisadas para identificar se há envolvimento na biossíntese e inativação de 20E.

A partir do alinhamento par a par e múltiplo, foi possível construir vários pares de oligonucleotídeos a partir das regiões conservadas entre as sequências de nucleotídeos dos genes, que poderão ser utilizados em estratégias de isolamento de genes com potencial envolvimento na interação planta-inseto e na biossíntese e inativação de 20E.

REFERÊNCIAS

ABLES, E. T. et al. A genetic mosaic screen reveals ecdysone-responsive genes regulating *Drosophila* oogenesis. **G3: Genes, Genomes, Genetics**, v. 6, n. 8, p. 2629-2642, 2016.

ALEKSENKO, V. A. et al. Phylogeny and structure of fatty acid photodecarboxylases and glucose-methanol-choline oxidoreductases. **Catalysts**, v. 10, n. 9, p. 1072, 2020.

BAILEY, T. L. et al. MEME SUITE: tools for motif discovery and searching. **Nucleic acids research**, v. 37, n. suppl_2, p. W202-W208, 2009.

BLACKFORD, M.; CLARKE, B.; DINAN, L. Tolerance of the Egyptian cotton leafworm *Spodoptera littoralis* (Lepidoptera: Noctuidae) to ingested phytoecdysteroids. **Journal of Insect Physiology**, v. 42, n. 10, p. 931-936, 1996.

BOUYAHYA, A. et al. Ethnomedicinal use, phytochemistry, pharmacology, and toxicology of *Ajuga iva* (L.) schreb. **Journal of ethnopharmacology**, v. 258, p. 112875, 2020.

BUCK, M. J.; ATCHLEY, W. R. Phylogenetic analysis of plant basic helix-loop-helix proteins. **Journal of molecular evolution**, v. 56, n. 6, p. 742-750, 2003.

CAO, V. D.; RIU, K.; BOO, K. Biosynthesis and accumulation of 20-hydroxyecdysone in individual male and female spinach plants during the reproductive stage. **Plant Physiology and Biochemistry**, v. 129, p. 394-399, 2018.

CASTELLANO, E.; SANTOS, E. Functional specificity of ras isoforms: so similar but so different. **Genes & cancer**, v. 2, n. 3, p. 216-231, 2011.

CHUN-YAN, S. U. et al. *Salvia miltiorrhiza*: Traditional medicinal uses, chemistry, and pharmacology. **Chinese journal of natural medicines**, v. 13, n. 3, p. 163-182, 2015.

DANIELSON, J. Å.; JOHANSON, U. Phylogeny of major intrinsic proteins. **MIPs and their role in the exchange of metalloids**, p. 19-31, 2010.

DINAN, L. et al. 20-Hydroxyecdysone, from Plant Extracts to Clinical Use: Therapeutic Potential for the Treatment of Neuromuscular, Cardio-Metabolic and Respiratory Diseases. **Biomedicines**, v. 9, n. 5, p. 492, 2021.

DISTÉFANO, A. M. et al. Roles of cytochromes P450 in plant reproductive development. **The International Journal of Developmental Biology**, 2020.

EGAN, E. S. et al. An extraretinally expressed insect cryptochrome with similarity to the blue light photoreceptors of mammals and plants. **Journal of Neuroscience**, v. 19, n. 10, p. 3665-3673, 1999.

FELSENSTEIN, J. Confidence limits on phylogenies: an approach using the bootstrap. **Evolution**, v. 39, n. 4, p. 783-791, 1985.

FUKUTA, Y. et al. Characterization of a new (R)-hydroxynitrile lyase from the Japanese apricot *Prunus mume* and cDNA cloning and secretory expression of one of the isozymes in *Pichia pastoris*. **Bioscience, biotechnology, and biochemistry**, p. 1012222310-1012222310, 2011.

GASTEIGER, E. et al. Protein identification and analysis tools on the ExPASy server. **The proteomics protocols handbook**, p. 571-607, 2005.

GILBERT, L. I. Halloween genes encode P450 enzymes that mediate steroid hormone biosynthesis in *Drosophila melanogaster*. **Molecular and cellular endocrinology**, v. 215, n. 1-2, p. 1-10, 2004.

HAN, P. et al. 20-Hydroxyecdysone enhances Immulectin-1 mediated immune response against entomogenous fungus in *Locusta migratoria*. **Pest management science**, v. 76, n. 1, p. 304-313, 2020.

HANSEN, C. C. et al. Plant cytochrome P450 plasticity and evolution. **Molecular Plant**, 2021.

HÖFER, R. et al. Geraniol hydroxylase and hydroxygeraniol oxidase activities of the CYP76 family of cytochrome P450 enzymes and potential for engineering the early steps of the (seco) iridoid pathway. **Metabolic engineering**, v. 20, p. 221-232, 2013.

HORI, K. et al. Mining of the uncharacterized cytochrome P450 genes involved in alkaloid biosynthesis in California poppy using a draft genome sequence. **Plant and Cell Physiology**, v. 59, n. 2, p. 222-233, 2018.

HOUSTON, N. L. et al. Phylogenetic analyses identify 10 classes of the protein disulfide isomerase family in plants, including single-domain protein disulfide isomerase-related proteins. **Plant Physiology**, v. 137, n. 2, p. 762-778, 2005.

HUANG, J.; MCAUSLANE, H. J.; NUSSLY, G. S. Resistance in lettuce to *Diabrotica balteata* (Coleoptera: Chrysomelidae): the roles of latex and inducible defense. **Environmental Entomology**, v. 32, n. 1, p. 9-16, 2003.

KAPLANIS, J. N. et al. Insect hormones: alpha ecdysone and 20-hydroxyecdysone in bracken fern. **Science**, v. 157, n. 3795, p. 1436-1438, 1967.

KAWAZU, K. et al. Evaluation of anti-herbivory genes using an Agrobacterium-mediated transient expression system. **Plant Biotechnology**, p. 12.0711 a, 2012.

- KERFELD, C. A.; SCOTT, K. M. Using BLAST to teach “E-value-tionary” concepts. **PLoS biology**, v. 9, n. 2, p. e1001014, 2011.
- KIM, J.; DELLAPENNA, D. Defining the primary route for lutein synthesis in plants: the role of *Arabidopsis* carotenoid β -ring hydroxylase CYP97A3. **Proceedings of the National Academy of Sciences**, v. 103, n. 9, p. 3474-3479, 2006.
- KOLODNY, R.; PETREY, D.; HONIG, B. Protein structure comparison: implications for the nature of ‘fold space’, and structure and function prediction. **Current opinion in structural biology**, v. 16, n. 3, p. 393-398, 2006.
- KUMAR, S. et al. MEGA X: molecular evolutionary genetics analysis across computing platforms. **Molecular biology and evolution**, v. 35, n. 6, p. 1547-1549, 2018.
- LARIGUET, P.; DUNAND, C. Plant photoreceptors: phylogenetic overview. **Journal of Molecular Evolution**, v. 61, n. 4, p. 559-569, 2005.
- LEE, S. et al. Herbivore-induced and floral homoterpene volatiles are biosynthesized by a single P450 enzyme (CYP82G1) in *Arabidopsis*. **Proceedings of the National Academy of Sciences**, v. 107, n. 49, p. 21205-21210, 2010.
- LEI, T. A. N. G. Pheophorbide a monoxygenase and chlorophyll degradation. **Chemistry of Life**, p. 05, 2008.
- LI, J. et al. 20-Hydroxyecdysone protects wheat seedlings from salt stress. **Archives of Biological Sciences**, v. 70, n. 2, p. 379-386, 2018.
- LI, K. et al. *Bombyx* E75 isoforms display stage-and tissue-specific responses to 20-hydroxyecdysone. **Scientific reports**, v. 5, n. 1, p. 1-13, 2015.
- LORRAIN, S. et al. Lesion mimic mutants: keys for deciphering cell death and defense pathways in plants? **Trends in plant science**, v. 8, n. 6, p. 263-271, 2003.
- MA, N. et al. Cloning and expression analysis of wheat pheophorbide a oxygenase gene TaPaO. **Plant Molecular Biology Reporter**, v. 30, n. 5, p. 1237-1245, 2012.
- MICHALSKI, C. et al. Salicyl alcohol oxidase of the chemical defense secretion of two chrysomelid leaf beetles: molecular and functional characterization of two new members of the glucose-methanol-choline oxidoreductase gene family. **Journal of Biological Chemistry**, v. 283, n. 28, p. 19219-19228, 2008.
- MINDNICH, R. D.; PENNING, T. M. Aldo-keto reductase (AKR) superfamily: genomics and annotation. **Human genomics**, v. 3, n. 4, p. 1-9, 2009.
- NAGATA, N. et al. Domain structures of chlorophyllide a oxygenase of green plants and *Prochlorothrix hollandica* in relation to catalytic functions. **Planta**, v. 218, n. 6, p. 1019-1025, 2004.
- NAKAGAWA, Y. Nonsteroidal ecdysone agonists. **Vitamins & Hormones**, v. 73, p. 131-173, 2005.

NANDA, S.; KATO, Y.; ASANO, Y. A new (R)-hydroxynitrile lyase from *Prunus mume*: asymmetric synthesis of cyanohydrins. **Tetrahedron**, v. 61, n. 46, p. 10908-10916, 2005.

NEEDLEMAN, S. B.; WUNSCH, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. **Journal of molecular biology**, v. 48, n. 3, p. 443-453, 1970.

NELSON, D. R. et al. Comparison of cytochrome P450 genes from six plant genomes. **Tropical Plant Biology**, v. 1, n. 3, p. 216-235, 2008.

NIWA, R.; NIWA, Y. S. Enzymes for ecdysteroid biosynthesis: their biological functions in insects and beyond. **Bioscience, biotechnology, and biochemistry**, v. 78, n. 8, p. 1283-1292, 2014.

NIWA, Y. S.; NIWA, R. Transcriptional regulation of insect steroid hormone biosynthesis and its role in controlling timing of molting and metamorphosis. **Development, growth & differentiation**, v. 58, n. 1, p. 94-105, 2016.

OHNISHI, T. et al. CYP724B2 and CYP90B3 function in the early C-22 hydroxylation steps of brassinosteroid biosynthetic pathway in tomato. **Bioscience, biotechnology, and biochemistry**, v. 70, n. 9, p. 2071-2080, 2006.

ORENGO, C. A.; JONES, D. T.; THORNTON, J. M. Protein superfamilies and domain superfolds. **Nature**, v. 372, n. 6507, p. 631-634, 1994.

PAN, X.; CONNACHER, R. P.; O'CONNOR, M. B. Control of the insect metamorphic transition by ecdysteroid production and secretion. **Current Opinion in Insect Science**, 2020.

PAQUETTE, S.; MØLLER, B. L.; BAK, S. On the origin of family 1 plant glycosyltransferases. **Phytochemistry**, v. 62, n. 3, p. 399-413, 2003.

PARDEDE, A. et al. Isolation of secondary metabolites from *Stenochlaena palustris* stems and structure-activity relationships of 20-hydroxyecdysone derivatives on antitermite activity. **Holzforschung**, v. 72, n. 10, p. 899-904, 2018.

PERLAK, F. J. et al. Modification of the coding sequence enhances plant expression of insect control protein genes. **Proceedings of the National Academy of Sciences**, v. 88, n. 8, p. 3324-3328, 1991.

RENAULT, H. et al. Cytochrome P450-mediated metabolic engineering: current progress and future challenges. **Current opinion in plant biology**, v. 19, p. 27-34, 2014.

REWITZ, K. F. et al. The Halloween genes code for cytochrome P450 enzymes mediating synthesis of the insect moulting hormone. **Biochemical Society Transactions**, v. 34, n. 6, p. 1256-1260, 2006.

ROSENQUIST, M. et al. Evolution of the 14-3-3 protein family: does the large number of isoforms in multicellular organisms reflect functional specificity? **Journal of molecular evolution**, v. 51, n. 5, p. 446-458, 2000.

SCHWEDES, C. C.; CARNEY, Ginger E. Ecdysone signaling in adult *Drosophila melanogaster*. **Journal of insect physiology**, v. 58, n. 3, p. 293-302, 2012.

SEHNKE, P. C. et al. Evolution and isoform specificity of plant 14-3-3 proteins. **Plant molecular biology**, v. 50, n. 6, p. 1011-1018, 2002.

SHARMA, H. C.; SHARMA, K. K.; CROUCH, J. H. Genetic transformation of crops for insect resistance: potential and limitations. **Critical Reviews in Plant Sciences**, v. 23, n. 1, p. 47-72, 2004.

SIMPSON, P. J. et al. Characterization of two novel aldo–keto reductases from Arabidopsis: expression patterns, broad substrate specificity, and an open active-site structure suggest a role in toxicant metabolism following stress. **Journal of molecular biology**, v. 392, n. 2, p. 465-480, 2009.

SMITH, C. M.; CHUANG, W-P. Plant resistance to aphid feeding: behavioral, physiological, genetic and molecular cues regulate aphid host selection and feeding. **Pest management science**, v. 70, n. 4, p. 528-540, 2014.

STOTHARD, P. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. **Biotechniques**, v. 28, n. 6, p. 1102-1104, 2000.

SWEVERS, L. An update on ecdysone signaling during insect oogenesis. **Current opinion in insect science**, v. 31, p. 8-13, 2019.

TANAKA, A. et al. Chlorophyll a oxygenase (CAO) is involved in chlorophyll b formation from chlorophyll a. **Proceedings of the National Academy of Sciences**, v. 95, n. 21, p. 12719-12723, 1998.

TIAN, X. et al. Characterization of gossypol biosynthetic pathway. **Proceedings of the National Academy of Sciences**, v. 115, n. 23, p. E5410-E5418, 2018.

TIERRANEGRA-GARCÍA, N. et al. Effect of foliar salicylic acid and methyl jasmonate applications on protection against pill-bugs in lettuce plants (*Lactuca sativa*). **Phytoparasitica**, v. 39, n. 2, p. 137-144, 2011.

URYU, O.; AMEKU, T.; NIWA, R. Recent progress in understanding the role of ecdysteroids in adult insects: Germline development and circadian clock in the fruit fly *Drosophila melanogaster*. **Zoological letters**, v. 1, n. 1, p. 1-9, 2015.

VETTER, János. Plant cyanogenic glycosides. **Toxicon**, v. 38, n. 1, p. 11-36, 2000.

WARREN, J. T. et al. Molecular and biochemical characterization of two P450 enzymes in the ecdysteroidogenic pathway of *Drosophila melanogaster*. **Proceedings of the National Academy of Sciences**, v. 99, n. 17, p. 11043-11048, 2002.

WERCK-REICHHART, D.; BAK, S.; PAQUETTE, S. Cytochromes P450. **The Arabidopsis book/American Society of Plant Biologists**, v. 1, 2002.

YACTAYO-CHANG, J. P. et al. Plant defense chemicals against insect pests. **Agronomy**, v. 10, n. 8, p. 1156, 2020.

YAMAUCHI, Y. et al. NADPH-dependent reductases involved in the detoxification of reactive carbonyls in plants. **Journal of biological chemistry**, v. 286, n. 9, p. 6999-7009, 2011.

YAN, Q. et al. GmCYP82A3, a soybean cytochrome P450 family gene involved in the jasmonic acid and ethylene signaling pathway, enhances plant resistance to biotic and abiotic stresses. **PloS one**, v. 11, n. 9, p. e0162253, 2016.

ZHANG, Q. et al. The genome of *Prunus mume*. **Nature communications**, v. 3, n. 1, p. 1-8, 2012.

ZOTTI, M. J. et al. A cell-based reporter assay for screening for EcR agonist/antagonist activity of natural ecdysteroids in Lepidoptera (Bm5) and Diptera (S2) cell cultures, followed by modeling of ecdysteroid-EcR interactions and normal mode analysis. **Pesticide biochemistry and physiology**, v. 107, n. 3, p. 309-320, 2013.