

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE BIOCÊNCIAS
CURSO DE GRADUAÇÃO EM BIOTECNOLOGIA

Fernanda Back de Souza

Análise genômica comparativa e funcional de *Humulus lupulus* e *Cannabis sativa*
para a prospecção de genes associados com as rotas de biossíntese de terpenóides
de interesse industrial.

Porto Alegre
2024

Fernanda Back de Souza

Análise genômica comparativa e funcional de *Humulus lupulus* e *Cannabis sativa* para a prospecção de genes associados com as rotas de biossíntese de terpenóides de interesses industriais.

Trabalho de Conclusão de Curso apresentado como requisito parcial para obtenção do título de bacharela em Biotecnologia da Universidade Federal do Rio Grande do Sul.

Área de habilitação: Biotecnologia Molecular

Orientação: Prof. Dr. Diego Bonatto

Porto Alegre

2024

CIP - Catalogação na Publicação

de Souza, Fernanda Back
Análise genômica comparativa e funcional de *Humulus lupulus* e *Cannabis sativa* para a prospecção de genes associados com as rotas de biossíntese de terpenóides de interesse industrial. / Fernanda Back de Souza. -- 2024.
32 f.
Orientador: Diego Bonatto.

Trabalho de conclusão de curso (Graduação) -- Universidade Federal do Rio Grande do Sul, Instituto de Biociências, Curso de Biotecnologia: Biotecnologia Molecular, Porto Alegre, BR-RS, 2024.

1. Terpenos. 2. Genômica Comparativa. 3. Genômica Funcional. 4. Lúpulo. 5. Cannabis. I. Bonatto, Diego, orient. II. Título.

Fernanda Back de Souza

Análise genômica comparativa e funcional de *Humulus lupulus* e *Cannabis sativa* para a prospecção de genes associados com as rotas de biossíntese de terpenóides de interesses industriais.

Trabalho de conclusão de Curso apresentado como requisito parcial para obtenção do título de bacharela em Biotecnologia da Universidade Federal do Rio Grande do Sul.

Aprovada em: 14/08/2024

BANCA EXAMINADORA

Prof. Dr. Diego Bonatto

Profa. Dra. Fabiana Quoos Mayer

Prof. Dr. Felipe Klein Ricachenevsky

AGRADECIMENTOS

A realização deste estudo deu-se por meio de diversos encontros que me estimularam a fazer ciência. Assim, agradeço ao meu orientador Prof. Dr. Diego Bonatto, que aceitou me guiar durante essa trajetória.

Agradeço aos meus amigos e colegas que me acompanharam de forma presencial ou remota, apoiando minha pesquisa e motivação.

Por fim, agradecer aos meus pais que sempre me motivaram durante minha graduação e se mostraram presentes e cuidadosos em toda minha jornada.

RESUMO

O *Humulus lupulus* é uma planta fundamental na produção de cerveja, compondo um dos seus quatro ingredientes principais. Ele desempenha várias funções essenciais, incluindo a adição de amargor, sabor e aroma. As moléculas responsáveis por essas características são os terpenóides, derivados do composto orgânico isopreno. Esses compostos são produzidos através de duas vias distintas: a via do ácido mevalônico (MVA) e a via do metileritritol fosfato (MEP). A partir dessas rotas de biossíntese podem ser obtidos hemiterpenos, monoterpenos e diterpenos, que são moléculas de grande interesse industrial presentes no óleo essencial extraído do lúpulo. A utilização de ferramentas de bioinformática é essencial para a identificação de genes responsáveis pela produção de terpenóides. Além dos programas, é necessário utilizar organismos filogeneticamente relacionados ao lúpulo, como a *Cannabis sativa*, para a caracterização e descrição dos genes, aplicando os princípios de ortologia. Ao aplicar programas que rastreiam genes que possuem ancestralidade e resgatam domínios proteicos comuns entre *Humulus lupulus* e *Cannabis sativa*, foi possível descrever sequências que apresentam relevância na síntese de terpenóides. Estes genes identificados estão presentes na via do metileritritol fosfato (MEP), responsável pela produção de limoneno, geraniol, mirceno, humuleno, linalol, cariofileno e entre outros. Estas moléculas são os elementos principais que compõem os óleos essenciais obtidos do lúpulo, garantindo diferentes aromas e sabores na cerveja. O conhecimento obtido dos genes descritos neste estudo pode ser utilizado como alvo para melhorar o perfil sensorial de diferentes cervejas, além de aumentar a eficiência na produção de lúpulo e seus óleos essenciais.

Palavras-chaves: Terpenóides, lúpulo, óleos essenciais, *Cannabis sativa*, *Humulus lupulus*.

ABSTRACT

Humulus lupulus (hop) is a fundamental plant in beer production and comprises one of its four main ingredients. It has several essential functions, including addition of bitterness, flavor, and aroma. Terpenoids derived from the organic compound isoprene are responsible for these characteristics of beer. These compounds are produced through two distinct pathways, mevalonate pathway (MVA) and methylerythritol phosphate (MEP) pathway. From these biosynthetic routes, it is possible to obtain hemiterpenes, monoterpenes, and diterpenes, which are molecules of great industrial interest present in essential oils extracted from hops. Bioinformatic tools are essential for the identification of genes responsible for the production of terpenoids. In addition to these programs, it is necessary to utilize phylogenetically similar organisms, such as *Cannabis sativa*, for the characterization and description of these genes by applying the principles of orthology. Applying programs that track genes that share ancestry and retrieve similar protein domains between hop and cannabis made it possible to describe sequences that are relevant to terpenoid synthesis. These identified genes are present in the MEP pathway and are responsible for the production of limonene, geraniol, myrcene, humulene, linalol, and caryophyllene. These molecules are the main elements that compose the essential oils obtained from hops, guaranteeing different scents and flavors in beer. The knowledge obtained from the genes described in this study can be utilized as a target for improving the sensorial profile of different beers, in addition to increasing the efficiency of lupulus and its essential oil production.

Keywords: Terpenoids, hop, essential oils, *Cannabis sativa*, *Humulus lupulus*.

LISTA DE FIGURAS

Figura 1: Etapas enzimáticas das vias MVA e MEP e sua regulação na biossíntese de precursores de isoprenóides. (THOLL, 2015).

Figura 2: Termos de ontologia gênica recuperados através do enriquecimento funcional utilizando o método Benjamini-Hochberg (BH) para ajustar os valores de p , considerando-se enriquecidos os termos de ontologia com p ajustado $< 0,01$ nos conjuntos de dados analisados entre as ontologias descritas.

Figura 3: Gráfico apresentando a completude, N50 e contigs para o genoma de *Humulus lupulus*, *Cannabis sativa* e *Lotus japonicus*.

Figura 4: Gráficos gerados a partir de métricas do QUASt. (A) Densidade versus tamanho do genoma. (B) Densidade versus conteúdo GC.

Figura 5: Rede interatômica entre as ontologias gênicas que possuem de p ajustado < 0.01 enriquecidos dentro das ontologias descritas.

Figura 6: Rede de interação desenvolvida através do programa Cytoscape representando a interação entres os genes da biossíntese de terpenóides em *Cannabis sativa* (A) e *Humulus lupulus* (B).

Figura 7: Modelos proteicos de *Arabidopsis thaliana* preditos através do programa AlphaFold. (A) Terpenoid cyclases/protein prenyltransferase. (B) Isoprenoid synthase.

LISTA DE TABELAS

Tabela 1: Tabela dos genes ortólogos presente na biossíntese de terpenóides em *Cannabis sativa* e *Humulus lupulus*.

LISTA DE ABREVIATURAS

MEP : Via do Metileritritol Fosfato

MVA : Via do Ácido Mevalônico

NCBI : National Center for Biotechnology Information

FASTA : Formato de texto para representação de sequências biológicas

BH : Benjamini-Hochberg Procedure

GO : *Gene Ontology*

DXR : 1-desoxi-D-xilulose 5-fosfato reductoisomerase

DXS/CLA1 : provável 1-desoxi-D-xilulose-5-fosfato sintase 2 cloroplástica

CMS : 2-C-metil-D-eritritol 4-fosfato citidiltransferase cloroplástica

CMK : 4-difosfocitidil-2-C-metil-D-eritritol quinase cloroplástica

MCS : 2-C-metil-D-eritritol 2,4-ciclodifosfato sintase cloroplástica

HDS : 4-hidroxi-3-metilbut-2-en-1-il difosfato sintase (ferredoxina) cloroplástica

HDR : 4-hidroxi-3-metilbut-2-enil difosfato redutase cloroplástica

AACT1 : acetil-CoA acetiltransferase 1

AACT2 : acetil-CoA acetiltransferase 2

KEGG Kyoto Encyclopedia of Genes and Genomes

IPP: difosfato de isopentenila

DMAPP: difosfato de dimetilalila

SUMÁRIO

1. INTRODUÇÃO.....	15
2.MATERIAIS E MÉTODOS.....	17
2.1 Seleção de Genomas e análises de qualidade e completude.....	17
2.2 Identificação e Seleção de Ortólogos.....	18
2.3 Enriquecimento Funcional dos Genes Seleccionados.....	19
2.4 Recuperação de Domínios Proteicos de <i>Humulus lupulus</i> e <i>Cannabis sativa</i>	21
3. RESULTADOS.....	21
3.1 Avaliação da Qualidade Genômica: Métricas N50 e Completude do BUSCO.....	21
3.2 Genes ortólogos de <i>Cannabis sativa</i> e <i>Humulus lupulus</i>	23
3.3 Domínios Proteicos Compartilhados entre <i>Cannabis sativa</i> e <i>Humulus lupulus</i>	27
3.4 Genes com Funções Biotecnológicas em <i>Humulus lupulus</i>	28
4. DISCUSSÃO.....	28
4.1 Genes Ortólogos e Redes de Interação.....	28
4.2 Domínios Proteicos Compartilhados.....	29
4.3 Genes com Potencial Biotecnológico.....	29
5. CONCLUSÃO.....	29
REFERÊNCIAS.....	30

1. INTRODUÇÃO

O gênero *Humulus* pertence à família *Cannabaceae*, assim como o gênero *Cannabis* e gêneros anteriormente classificados na antiga família *Celtidaceae* presentes também na família *Cannabaceae* (SIMPSON, 2010, ALMAGUER et al., 2014). Este gênero é composto de três espécies conhecidas como *H. lupulus*, *H. japonicus* e *H. yunnanensis*. No entanto, a espécie que possui valor industrial e econômico é a *H. lupulus*, conhecida como lúpulo comum. Esta trepadeira perene e dióica ocorre espontaneamente em sebes e bordas de regiões temperadas da Europa, Ásia e América do Norte (BOCQUET et al., 2018). As inflorescências femininas do lúpulo são organizadas em brácteas dentro dos estróbilos femininos, enquanto em plantas masculinas as flores se estruturam em cachos. Contudo apenas a parte feminina é visada para a indústria, visto que na base dos cones femininos são encontradas as glândulas de lupulinas, responsáveis pela biossíntese de diversos metabólitos secundários, como compostos fenólicos (ácidos α e β), flavonóides e principalmente terpenóides (óleos essenciais) (DE KEUKELEIRE et al., 2007).

Por sua vez, a *Cannabis* é um gênero pertencente à família das *Cannabaceae*, que inclui também plantas como o celtis e a trema. Este gênero é composto por apenas três espécies: *C. sativa*, *C. indica* e *C. ruderalis* (ANDERSON, 1980). Todas as variedades são originalmente nativas da Ásia Central e Meridional, sendo que estas espécies apresentam perenidade e plantas dioicas, semelhante ao lúpulo; porém, apresentam flores incompletas, caracterizadas pela falta de pétalas. Além disso, as flores femininas são as únicas que apresentam pistilo, enquanto aquelas que apenas possuem estames são as masculinas (SOMMANO et al., 2020). Diferentemente do lúpulo, os terpenóides da *Cannabis* são produzidos e armazenados nos tricomas das folhas e principalmente das flores. Além dos terpenóides, esta planta é capaz de produzir outros compostos secundários em seus tricomas, como o THC (tetrahydrocannabinol), CBD (canabidiol), compostos fenólicos e alcalóides (RADWAN et al., 2021).

Presente tanto em Lúpulo quanto em Cannabis, os terpenóides compõe os óleos essenciais extraídos de ambas as plantas. Essas moléculas são produzidas a partir de diferentes precursores, como acetil-CoA e gliceraldeído-3-fosfato, que alimentam a via biossintética de terpenóides e outras moléculas. As plantas utilizam duas diferentes vias para produzir diferentes variedades de terpenos, a via do ácido mevalônico (MVA), que ocorre principalmente no citosol e a via do metileritritol fosfato (MEP), que ocorre nos

plastídeos. A via MVA utiliza o acetil-CoA como precursor, fornecendo predominantemente precursores para a biossíntese citosólica de sesquiterpenos, poliprenóis, fitosteróis, brassinosteroides e triterpenoides, além de terpenóides nas mitocôndrias (como ubiquinonas e poliprenóis). Já as unidades de cinco carbonos derivadas da via MEP são preferencialmente utilizadas na biossíntese de hemiterpenoides, monoterpenoides, diterpenoides, carotenoides e seus produtos de decomposição, citocininas, giberelinas, clorofila, tocoferóis e plastoquinonas (THOLL, 2015) (**Figura 1**). No lúpulo, os componentes principais em seu óleo essencial são os monoterpenos (esqueleto C10) e sesquiterpenos (esqueleto C15), representando 80% dos voláteis presentes no lúpulo, principalmente pela produção de mirceno, α -humuleno e β -cariofileno (RETTBERG; BIENDL; GARBE, 2018).

A importância de estudar os terpenóides no lúpulo utilizando a Cannabis como modelo se torna evidente ao considerarmos a similaridade biossintética e estrutural entre os compostos dessas duas plantas. Ambas as espécies, pertencentes à família *Cannabaceae*, possuem vias metabólicas e precursoras comuns para a produção de terpenóides. Além da relação filogenética entre lúpulo e Cannabis, é importante destacar que o genoma de *Cannabis sativa* é consideravelmente mais descrito e estudado que o de *Humulus lupulus*, visto que essa planta apresenta outras moléculas de interesse além dos terpenóides.

Reconhecendo a importância do lúpulo na indústria cervejeira, considerando que é o ingrediente mais caro para as cervejarias e é valorizado por suas propriedades aromáticas, conservantes e amargor (RODOLFI et al., 2023). Esse ingrediente desempenha diversas funções, como a conservação, devido ao seu conteúdo natural de antioxidantes e compostos antibacterianos, além de suas propriedades organolépticas que realçam o aroma e o amargor do produto. (KORDIALIK-BOGACKA, 2022). Entender a biossíntese dos terpenóides é fundamental para otimização do aroma e sabor, selecionando aqueles que apresentam perfis sensoriais desejados, também a inovação e desenvolvimento de novos produtos, criando variedades de lúpulos únicas.

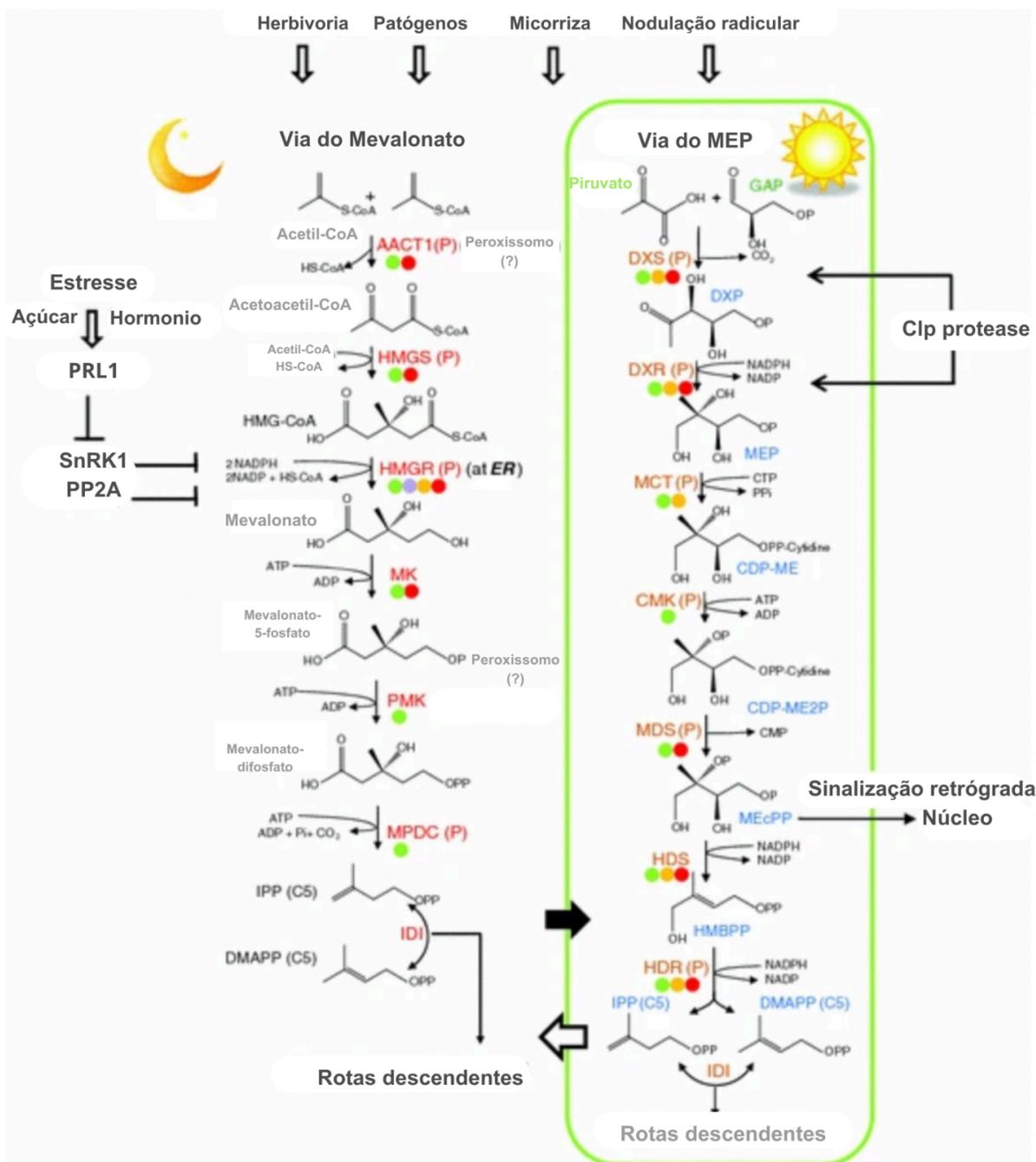


Figura 1: Etapas enzimáticas das vias MVA e MEP e sua regulação na biossíntese de precursores de isoprenóides. As setas indicam o tráfego preferencial de precursores isoprenoides entre o citosol e os plastídios em tecidos expostos claros (branco) e escuros (preto) Adaptado de (THOLL, 2015).

2. MATERIAIS E MÉTODOS

2.1 Seleção de Genomas e Análises de Qualidade e Completude

O genoma de *Humulus lupulus* (referência GenBank GCA_963169125.1; versão drHumLupu1.1) e *Cannabis sativa* (referência GenBank GCA_029168945.1; versão ASM2916894v1), os quais foram utilizados como objetivo de estudo, foram obtidos por meio do banco de dados *National Center for Biotechnology Information* (NCBI; <https://www.ncbi.nlm.nih.gov/genome>), utilizando como filtro a opção (“*Genome*”, [s.d.]). Foram selecionados mais dois genomas referências para uso como controle de qualidade e de anotação funcional, são eles os de *Lotus japonicus* e *Arabidopsis thaliana*. As estatísticas de qualidade genômica foram geradas para cada montagem de lúpulo, *Cannabis sativa* e lótus utilizando o programa QUAST (GUREVICH et al., 2013). A partir desse programa, foram obtidas métricas estatísticas como o valor de N50, o número de *contigs*, o tamanho do genoma, o conteúdo de GC, entre outras. O programa BUSCO (MANNI et al., 2021) foi empregado para estimar a completude do genoma, uma estatística que representa a porcentagem de ortólogos identificados entre a montagem de um determinado genoma em comparação a um genoma de referência, sendo útil para avaliar a qualidade do genoma.

2.2 Identificação e Seleção de Ortólogos

A identificação de genes ortólogos entre lúpulo e *Cannabis* foi realizada por meio do programa OrthoFinder (EMMS; KELLY, 2019), o qual utiliza algoritmos de similaridade entre os transcritos das espécies mapeando aqueles que possuem ancestralidade comum. As sequências proteicas foram obtidas pelo NCBI através da função (“*Genome*”, [s.d.]) em formato FASTA. Foram obtidos 42.852 ortólogos entre *Humulus lupulus* e *Cannabis sativa*, sendo que essas entradas forneceram informações como os identificadores dos genes além de seus respectivos ortogrupos através de um arquivo no formato tabular. (**Tabela S1**).

Para a seleção dos genes ortólogos de interesse primeiramente foi utilizado o genoma de *Cannabis sativa* (referência GenBank GCA_029168945.1; versão ASM2916894v1) como base para obter os identificadores de todas as enzimas descritas como ortólogas. Posteriormente, para selecionar e enriquecer as enzimas envolvidas na

biossíntese de terpenóides, foi realizada a convergência das enzimas da *Cannabis* com as enzimas responsáveis pela síntese de terpenóides em *Arabidopsis*, utilizando a plataforma RStudio para realizar essa união de nomes.

2.3 Enriquecimento Funcional dos Genes Selecionados

Por meio da plataforma Cytoscape versão 3.10.2 (SHANNON et al., 2003) foi possível visualizar redes de interação molecular e vias metabólicas, integrando redes com anotações e perfis de expressão gênica. A ferramenta BiNGO, disponível para o Cytoscape (MAERE; HEYMANS; KUIPER, 2005) foi utilizada para mapear termos funcionais predominantes no conjunto de genes através de ontologia gênica. Para o enriquecimento funcional foi usado o método *Benjamini-Hochberg* (BH) para o ajuste de valores p , sendo que os termos de ontologia com p ajustado < 0.01 foram considerados enriquecidos dentro dos conjuntos de dados analisados entre as ontologias descritas (**Tabela S2, Figura 2**).

Figura 2: Termos de ontologia gênica recuperados através do enriquecimento funcional utilizando o método Benjamini-Hochberg (BH) para ajustar os valores de p , considerando-se enriquecidos os termos de ontologia com p ajustado $< 0,01$ nos conjuntos de dados analisados entre as ontologias descritas.

2.4 Recuperação de Domínios Proteicos de *Humulus lupulus* e *Cannabis sativa*

Os dados proteômicos em formato FASTA foram obtidos através do NCBI pela função (“Genome”, [s.d.]) e suas análises funcionais foram realizadas pelo programa InterProScan, o qual integra informações preditivas sobre a função de proteínas a partir de diversos recursos, fornecendo uma visão geral das famílias às quais uma proteína pertence e dos domínios e sítios que ela possui (QUEVILLON et al., 2005). Os arquivos gerados no formato tabular foram posteriormente analisados através de scripts adaptados para linguagem R no RStudio a fim de selecionar aqueles domínios proteicos que estão presentes ou não em cânhamo e lúpulo.

3. RESULTADOS

3.1 Avaliação da Qualidade Genômica: Métricas N50 e Completude do BUSCO

A qualidade genômica foi avaliada utilizando métricas geradas pelo QUAST, como N50 e contigs, e pela completude calculada pelo BUSCO. Os genomas de *Humulus lupulus*, *Cannabis sativa* e *Lotus japonicus* demonstraram, em geral, alta qualidade. As espécies *Cannabis sativa* e *Humulus lupulus* apresentaram uma completude próxima a 80%, indicando uma boa conservação dos genes nos genomas analisados, enquanto o genoma de *Lotus japonicus* possui uma completude próxima aos 100%. Além disso, os valores de N50 foram excelentes, especialmente no lúpulo, que alcançou 251 Mb, refletindo uma representação precisa de regiões genômicas complexas. O número de contigs para as espécies foi distinto entre elas, onde *Lotus japonicus* apresentou o maior número de contigs e *Cannabis sativa* apresentou o menor. Apesar dos valores desiguais, eles ainda indicam valores aceitáveis, visto que para genomas maiores e mais complexos, valores entre 100-1000 contigs podem ser razoáveis (**Figura 3**).

A análise dos genomas de *Cannabis sativa* e *Humulus lupulus* revelou padrões distintos de densidade gênica em relação ao conteúdo de GC e ao tamanho do genoma. No gráfico de densidade gênica por tamanho do genoma os resultados indicaram que *H. lupulus* possui a maior densidade gênica em regiões de maiores *contigs*, seguido por *C. sativa*, que apresentou um pico em intervalos genômicos menores (**Figura 4A**). No gráfico de densidade por GC, observou-se que *H. lupulus* e *C. sativa* apresentaram uma distribuição uniforme ao longo das diferentes porcentagens de conteúdo GC (**Figura 4B**).

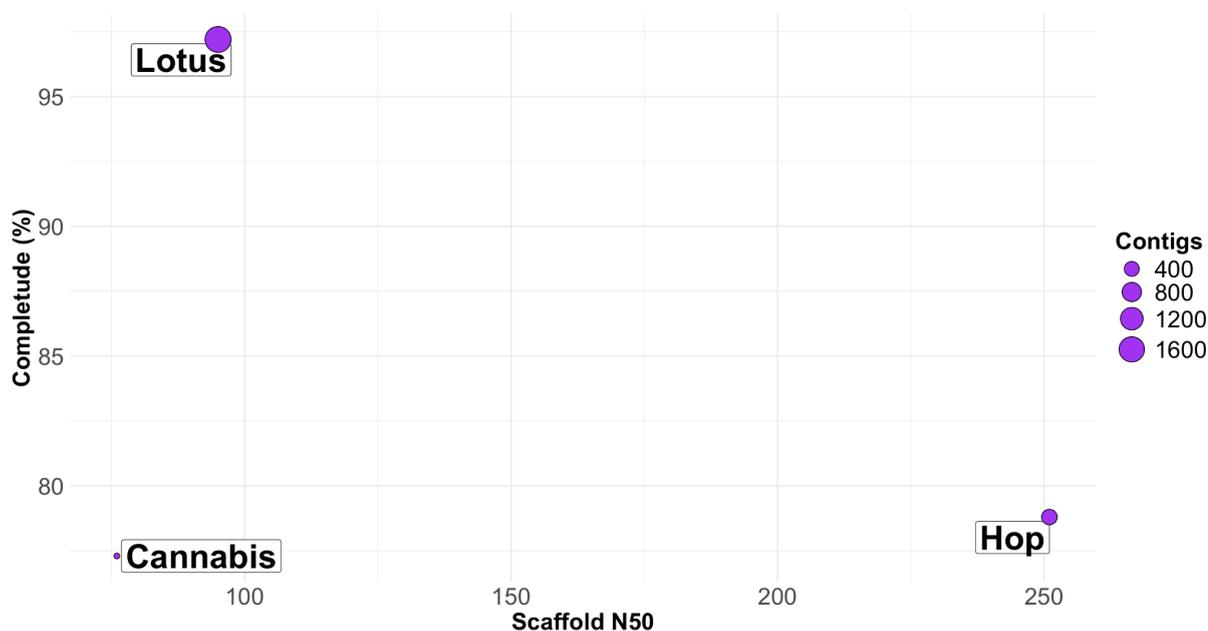


Figura 3: Gráfico apresentando a completude, N50 e contigs para o genoma de *Humulus lupulus*, *Cannabis sativa* e *Lotus japonicus*.

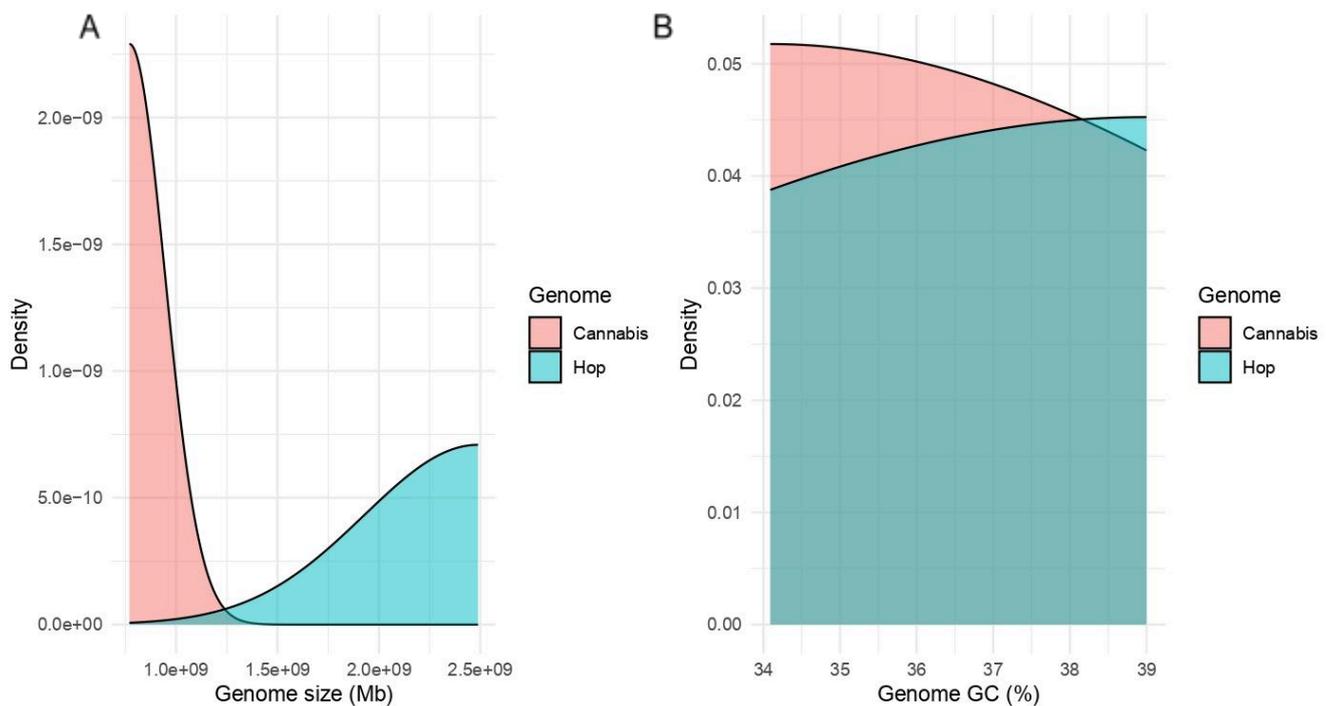


Figura 4: Gráficos gerados a partir de métricas do QUASt. (A) Densidade gênica versus tamanho do genoma. (B) Densidade versus conteúdo GC.

3.2 Genes ortólogos de *Cannabis sativa* e *Humulus lupulus*

A identificação de 42.852 ortólogos entre as espécies *Cannabis sativa* e *Humulus lupulus* por meio do programa Orthofinder revelou informações sobre suas ancestralidades comuns, corroborando a hipótese de que ambas compartilham rotas de síntese de terpenóides com alta semelhança genética. (BEDINI et al., 2016). Além disso, foi possível realizar a anotação funcional usando dados de ortologia com *Arabidopsis thaliana*, onde se verificou a presença de nove ortólogos entre *C. sativa* e *H. lupulus* (**Tabela 1**). Os genes previamente identificados pelo OrthoFinder e filtrados através dos genes ortólogos de *Arabidopsis* foram enriquecidos utilizando a ferramenta BiNGO presente na plataforma CytoScape e resultaram em uma tabela com 91 ontologias gênicas (**Tabela S2**) para essas entradas e uma rede interatômica (**Figura S1**). A análise dos resultados permitiu identificar a presença desses genes ortólogos em ontologias relacionadas à síntese de terpenóides (**Figura 5**).

Cannabis and Hop Orthologous Terpenoids

Name	Orthogroup	Hop Protein	Cannabis Protein	Chromosome
69 1-deoxy-D-xylulose 5-phosphate reductoisomerase, chloroplastic	OG0012904	XP_062116274.1	XP_030493319.2	3
137 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, chloroplastic	OG0015551	XP_062074424.1	XP_030501126.2	5
138 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase, chloroplastic	OG0014953	XP_062095393.1	XP_030499547.2	4
320 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase, chloroplastic	OG0016911	XP_062082389.1	XP_030506248.2	2
321 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (ferredoxin), chloroplastic	OG0001837	XP_062082072.1	XP_060964884.1	2
322 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (ferredoxin), chloroplastic	OG0001837	XP_062082070.1	XP_060964884.1	2
323 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (ferredoxin), chloroplastic	OG0001837	XP_062082071.1	XP_060964884.1	2
324 4-hydroxy-3-methylbut-2-enyl diphosphate reductase, chloroplastic	OG0010502	XP_062092843.1	XP_030486529.1	X
502 acetyl-CoA acetyltransferase 1	OG0012038	XP_062088076.1	XP_030491029.2	1
503 acetyl-CoA acetyltransferase 2	OG0002935	XP_062107124.1	XP_030482241.1	8
504 acetyl-CoA acetyltransferase 2	OG0002935	XP_062107125.1	XP_030482241.1	8
505 acetyl-CoA acetyltransferase 2	OG0002935	XP_062107126.1	XP_030482241.1	8
9504 probable 1-deoxy-D-xylulose-5-phosphate synthase 2, chloroplastic	OG0013892	XP_062100638.1	XP_030496468.1	4
9505 probable 1-deoxy-D-xylulose-5-phosphate synthase 2, chloroplastic	OG0018588	XP_062078592.1	XP_030511064.2	6

Tabela 1: Tabela dos genes ortólogos presente na biossíntese de terpenóides em *Cannabis sativa* e *Humulus lupulus*.

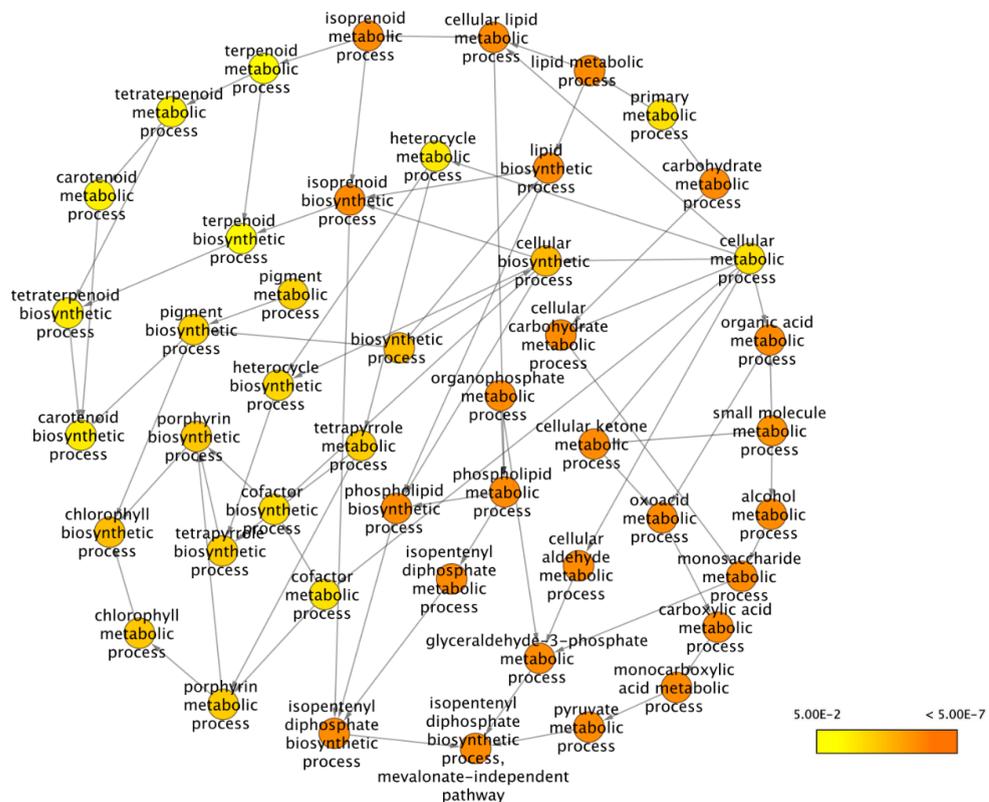


Figura 5: Rede interatômica entre as ontologias gênicas que possuem de p ajustado < 0.01 enriquecidos dentro das ontologias descritas.

Os genes identificados podem ser divididos em dois grupos principais. O primeiro grupo inclui genes que desempenham papéis na síntese de terpenoides, como *DXR*, *DXS*, *ISPD*, *ISPE*, *ISPF*, *ISPG* e *ISPH*. O segundo grupo consiste nos genes *AACT1* e *AACT2*, que são fundamentais para várias funções no metabolismo celular. Estes genes estão envolvidos na síntese de ácidos graxos, na produção de energia durante a respiração celular e na biossíntese de diversos compostos, como terpenóides, esteróides e outros lipídios essenciais para a função celular.

É importante destacar os genes que atuam na rota biossintética de terpenóides exclusivamente, o primeiro a ser identificado é o *DXR*, o qual codifica a proteína DXR (1-desoxi-D-xilulose 5-fosfato reductoisomerase) que desempenha o papel de catalisar uma reação crucial na via do Metileritritol Fosfato (MEP) da biossíntese de isoprenóides, convertendo 1-deoxi-D-xilulose-5-fosfato em 2-C-metil-D-eritritol-4-fosfato, além de desempenhar um papel importante no desenvolvimento dos cloroplastos (CARRETERO-PAULET et al., 2002). O gene *DXS* sintetiza a proteína DXS/CLA1 (1-desoxi-D-xilulose-5-fosfato sintase 2 cloroplástica) que atua na reação de condensação entre os átomos de carbono 2 e 3 do piruvato e do gliceroldeído-3-fosfato, resultando na formação de 1-deoxi-D-xilulose-5-fosfato (DXP). É uma enzima limitante para a biossíntese de isoprenóides plastídicos e essencial para o desenvolvimento dos cloroplastos (ESTÉVEZ et al., 2001). Outro gene importante é o *ISPD* que produz a enzima CMS (2-C-metil-D-eritritol 4-fosfato citidiltransferase cloroplástica), a qual desempenha o papel de catalisar a formação de 4-difosfocitidil-2-C-metil-D-eritritol, essa enzima também é essencial para o desenvolvimento dos cloroplastos e necessária para a biossíntese de diversos pigmentos e giberelinas (ROHDICH et al., 2000). O gene *ISPE* codifica a enzima CMK (4-difosfocitidil-2-C-metil-D-eritritol quinase cloroplástica), a qual catalisa a fosforilação do grupo hidroxila na posição 2 do 4-difosfocitidil-2-C-metil-D-eritritol, também essencial para o desenvolvimento dos cloroplastos (HSIEH et al., 2008). O gene consecutivo da rota é o *ISPF*, que codifica a enzimas MCS responsáveis pela conversão de 4-difosfocitidil-2-C-metil-D-eritritol 2-fosfato em 2-C-metil-D-eritritol 2,4-ciclodifosfato. Esse gene está envolvido na biossíntese de terpenoides a partir da via MEP, contribuindo para o desenvolvimento dos cloroplastos. (HSIEH; GOODMAN, 2006). Outro gene que também

está presente na MEP é o *ISPG* que codifica a enzima HDS (4-hidroxi-3-metilbut-2-en-1-il difosfato sintase (ferredoxina)) cloroplástica, responsável por converter 2-C-metil-D-eritritol 2,4-ciclodifosfato (ME-2,4cPP) em 1-hidroxí-2-metil-2-(E)-butenil 4-difosfato, essa enzima é essencial para o desenvolvimento dos cloroplastos e é necessária também para a resistência mediada por ácido salicílico contra patógenos biotróficos (SEEMANN et al., 2006). O último gene identificado, também descrito como *ISPH*, codifica a enzima HDR (4-hidroxi-3-metilbut-2-en-1-il difosfato sintase) presente na biossíntese de MEP para a biossíntese de isoprenóides que converte 1-hidroxí-2-metil-2-(E)-butenil 4-difosfato em difosfato de isopentenila (IPP) e difosfato de dimetilalila (DMAPP), moléculas importantes precursoras de terpenóides (HSIEH; GOODMAN, 2005).

Ao identificar os nove genes ortólogos foi possível descrever sequências de genes que ainda não possuíam anotação funcional de *Cannabis sativa* e *Humulus lupulus*. Para gerar as possíveis anotações funcionais foi utilizada a plataforma STRING (SZKLARCZYK et al., 2023), a qual fornece anotação funcional com base em interações conhecidas e preditas, além de visualização de redes, com exceção a proteína correspondente ao gene *AACT2*. As proteínas correspondentes aos genes ortólogos de *Arabidopsis thaliana* serviram como referência para gerar redes de interação proteína-proteína de *Cannabis* e lúpulo, que foram executadas e analisadas utilizando a plataforma Cytoscape (SHANNON et al., 2003) (**Figura 6**).

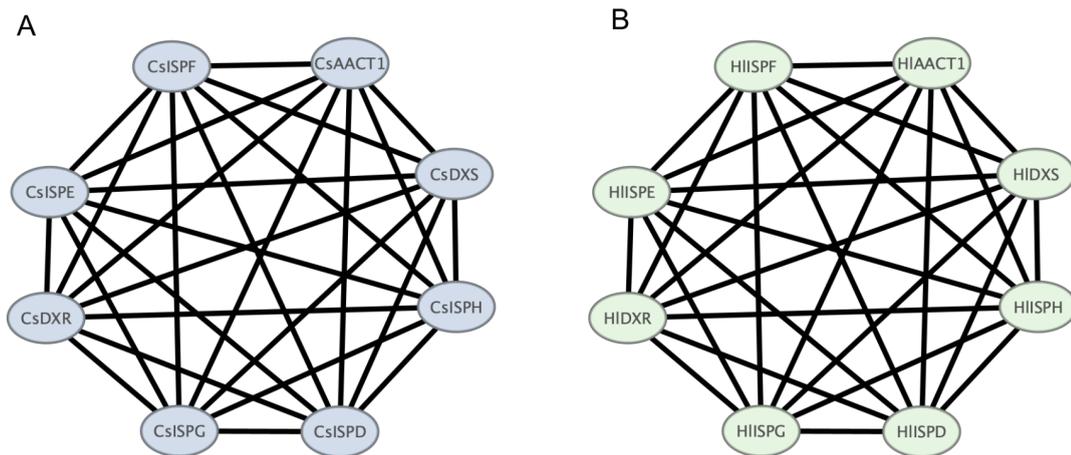


Figura 6: Rede de interação desenvolvida através do programa Cytoscape representando a interação entre os genes da biossíntese de terpenóides em *Cannabis sativa* (A) e *Humulus lupulus* (B).

3.3 Domínios Proteicos Compartilhados entre *Cannabis sativa* e *Humulus lupulus*

Os resultados obtidos pelas predições do InterProScan foram 952 domínios proteicos compartilhados entre *Cannabis sativa* e *Humulus lupulus* utilizando a base de dados Pfam (MISTRY et al., 2021), o qual utiliza informações sobre famílias de proteínas que incluem suas anotações e alinhamentos de sequências múltiplas derivados de modelos ocultos de Markov (LÖYTYNOJA; MILINKOVITCH, 2003). A fim de interesse ao nosso estudo, foi realizada uma filtragem manual desses 952 domínios. Como resultado da triagem manual, foi possível selecionar duas famílias proteicas, a “Terpenoid cyclases/protein prenyltransferase” (InterPro ID: IPR008930) e a família “Isoprenoid synthase” (InterPro ID: IPR008949) (**Figura 7**). Ambas as famílias possuem associação a síntese de terpenóides, além de apresentarem outros pontos em comum, como a estrutura “alfa-alfa toroide”, consistindo em várias alfa-hélices organizadas em um eixo, formando uma disposição semelhante a um anel. Além de sua organização espacial, ambas as famílias apresentam mecanismos enzimáticos próximos, consistindo na biossíntese de terpenóides e na prenilação de proteínas. Por fim, a organização do domínio funcional é similar, com regiões específicas da alfa-hélice desempenhando funções enzimáticas importantes.

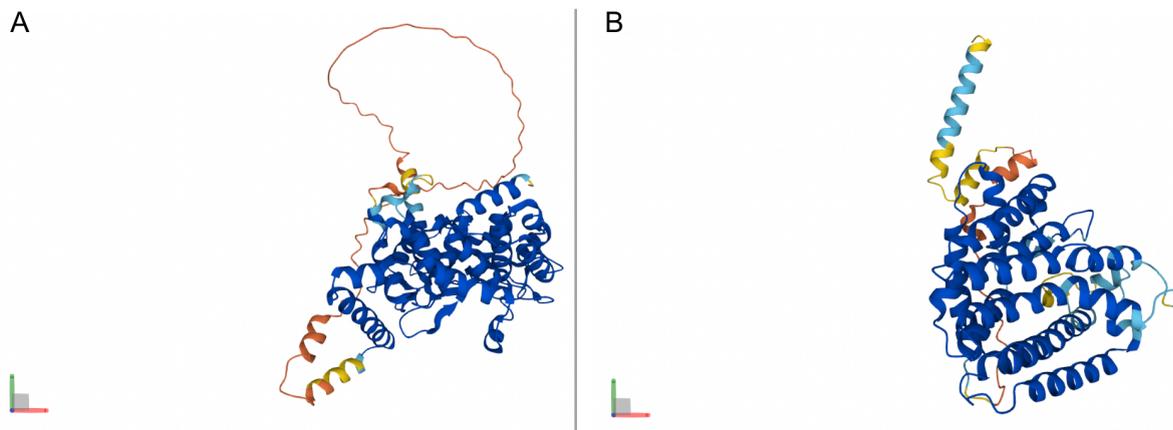


Figura 7: Modelos proteicos de *Arabidopsis thaliana* preditos através do programa AlphaFold. (A) Terpenoid cyclases/protein prenyltransferase. (B) Isoprenoid synthase.

3.4 Genes com Funções Biotecnológicas em *Humulus lupulus*

Na análise genômica de *Humulus lupulus*, foram identificados variados genes com funções biotecnológicas significativas para a síntese de terpenóides. As análises realizadas destacaram os genes *DXR*, *DXS*, *ISPD*, *ISPE*, *ISPF*, *ISPG* e *ISPH*, 4 que fazem parte da via biossintética MEP, responsável pela produção de sesquiterpenóides, hemi-terpenóides, monoterpenos e diterpenos, ou seja, as principais moléculas presentes nos óleos essenciais. Enquanto o gene *AACT1*, desempenha função crucial no metabolismo celular, atua como regulador da síntese de ácidos graxos, produção de energia na respiração celular, e biossíntese de diversos compostos essenciais, como terpenóides, esteróides e outros lipídios. O gene *AACT2* não possui interações claramente mapeadas, impossibilitando atribuir uma função específica a ele, visto que em análises em plataformas de anotações gênicas esse gene não foi encontrado em nenhum dos programas utilizados.

4. DISCUSSÃO

4.1 Genes Ortólogos e Redes de Interação

A identificação de 42.852 ortólogos entre as espécies *Cannabis sativa* e *Humulus lupulus* reforça a proximidade filogenética entre as espécies (MCPARTLAND, 2018). Tamanha similaridade foi capaz de identificar os genes *DXR*, *DXS*, *ISPD*, *ISPE*, *ISPF*, *ISPG*, *ISPH*, *AACT1* e *AACT2* conservados tanto em cânhamo quanto em lúpulo para biossíntese de terpenóides. A partir desses genes foi possível analisar redes de interação proteína-proteína através da plataforma Cytoscape para realizar sua anotação funcional e enriquecimento de ontologia gênica. O conhecimento das enzimas que constituem a via MEP é crucial para a produção de compostos voláteis de interesse industrial do lúpulo.

4.2 Domínios Proteicos Compartilhados

A identificação de 952 domínios proteicos compartilhados entre *Cannabis sativa* e *Humulus lupulus*, entre eles dois relacionados a síntese de terpenóides, como as famílias Terpenoid cyclases/protein prenyltransferase e Isoprenoid synthase. Estes domínios são fundamentais para a biossíntese de terpenóides e prenilação de proteínas (LIANG; KO; WANG, 2002). Esses domínios são fundamentais para a manutenção de funções essenciais, viabilidade e adaptação das plantas, reforçando a relevância da conservação evolutiva dessas vias metabólicas em espécies próximas.

4.3 Genes com Potencial Biotecnológico

Os genes identificados com funções biotecnológicas são apenas sete, sendo eles *DXR*, *DXS*, *ISPD*, *ISPE*, *ISPF*, *ISPG* e *ISPH*, todos estão presentes da biossíntese de terpenóides através da via MEP, responsável pela síntese de sesquiterpenóides, hemi-terpenóides, monoterpenos e diterpenos (BERGMAN; DAVIS; PHILLIPS, 2019). O gene identificado como *ISPH* na síntese de terpenóides, codifica a enzima HDR capaz de produzir os principais precursores da via biossintética MEP, IPP (difosfato de isopentenila) e DMAPP (difosfato de dimetilalila), ambos são essenciais para a síntese de todos terpenóides derivados dessa via metabólica (GUEVARA-GARCÍA et al., 2005). Assim, fica claro a importância do gene *ISPH*, pois a partir dos produtos que ele gera são obtidas todas as moléculas de relevância econômica e industrial produzida pelo lúpulo.

5. CONCLUSÃO

Esse estudo foi capaz de servir como início para o aprofundamento da pesquisa de genes que estão intrinsecamente ligados à rota de biossíntese de terpenóides em *Humulus lupulus*. O uso de uma planta filogeneticamente próxima do lúpulo, como a *Cannabis sativa*, foi essencial para a realização das análises de genômica comparativa e funcional, visto que esse organismo ainda possui um melhor nível de anotações funcionais em seu genoma.

As análises genômicas demonstraram que as rotas de biossíntese de terpenóides são muito similares entre ambas as espécies de plantas da família Cannabaceae, sendo que a presença de sequências não anotadas tanto em lúpulo quanto em Cannabis, associadas com a rota de formação de terpenóides, pode servir de fonte de novos genes para programas de melhoramento genético em lúpulo.

As implicações para a indústria cervejeira são significativas, uma vez que esses genes estão associados à formação de aromas e sabores para a cerveja. O conhecimento aprofundado dos genes que foram descritos nesse estudo podem ser utilizados de alvo para o melhoramento do perfil sensorial de variadas cervejas até o aumento da sustentabilidade e eficiência na produção de lúpulo.

REFERÊNCIAS

ALMAGUER, C. et al. *Humulus lupulus* – a story that begs to be told. A review. **Journal of the Institute of Brewing**, v. 120, n. 4, p. 289–314, 2014.

ANDERSON, L. C. Leaf Variation Among Cannabis Species from a Controlled Garden. **Botanical Museum Leaflets, Harvard University**, v. 28, n. 1, p. 61–69, 1980.

BEDINI, S. et al. *Cannabis sativa* and *Humulus lupulus* essential oils as novel control tools against the invasive mosquito *Aedes albopictus* and fresh water snail *Physella acuta*. **Industrial Crops and Products**, v. 85, p. 318–323, 1 jul. 2016.

BERGMAN, M. E.; DAVIS, B.; PHILLIPS, M. A. Medically Useful Plant Terpenoids: Biosynthesis, Occurrence, and Mechanism of Action. **Molecules**, v. 24, n. 21, p. 3961, jan. 2019.

BOCQUET, L. et al. *Humulus lupulus* L., a very popular beer ingredient and medicinal plant: overview of its phytochemistry, its bioactivity, and its biotechnology. **Phytochemistry Reviews**, v. 17, n. 5, p. 1047–1090, 1 out. 2018.

CARRETERO-PAULET, L. et al. Expression and molecular analysis of the Arabidopsis DXR gene encoding 1-deoxy-D-xylulose 5-phosphate reductoisomerase, the first committed enzyme of the 2-C-methyl-D-erythritol 4-phosphate pathway.

Plant Physiology, v. 129, n. 4, p. 1581–1591, ago. 2002.

DE KEUKELEIRE, J. et al. Relevance of Organic Farming and Effect of Climatological Conditions on the Formation of α -Acids, β -Acids, Desmethylxanthohumol, and Xanthohumol in Hop (*Humulus lupulus* L.). **Journal of Agricultural and Food Chemistry**, v. 55, n. 1, p. 61–66, 1 jan. 2007.

EMMS, D. M.; KELLY, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. **Genome Biology**, v. 20, n. 1, p. 238, 14 nov. 2019.

ESTÉVEZ, J. M. et al. 1-Deoxy-D-xylulose-5-phosphate synthase, a limiting enzyme for plastidic isoprenoid biosynthesis in plants. **The Journal of Biological Chemistry**, v. 276, n. 25, p. 22901–22909, 22 jun. 2001.

Genome. Disponível em: <<https://www.ncbi.nlm.nih.gov/datasets/genome/>>. Acesso em: 4 fev. 2024.

GUEVARA-GARCÍA, A. et al. Characterization of the Arabidopsis clb6 Mutant Illustrates the Importance of Posttranscriptional Regulation of the Methyl-d-Erythritol 4-Phosphate Pathway. **The Plant Cell**, v. 17, n. 2, p. 628–643, 1 fev. 2005.

HSIEH, M.-H. et al. Chloroplast localization of methylerythritol 4-phosphate pathway enzymes and regulation of mitochondrial genes in *ispD* and *ispE* albino mutants in Arabidopsis. **Plant Molecular Biology**, v. 66, n. 6, p. 663–673, abr. 2008.

HSIEH, M.-H.; GOODMAN, H. M. The Arabidopsis *IspH* homolog is involved in the plastid nonmevalonate pathway of isoprenoid biosynthesis. **Plant Physiology**, v. 138, n. 2, p. 641–653, jun. 2005.

HSIEH, M.-H.; GOODMAN, H. M. Functional evidence for the involvement of Arabidopsis *IspF* homolog in the nonmevalonate pathway of plastid isoprenoid biosynthesis. **Planta**, v. 223, n. 4, p. 779–784, mar. 2006.

KORDIALIK-BOGACKA, E. Biopreservation of beer: Potential and constraints. **Biotechnology Advances**, v. 58, p. 107910, 1 set. 2022.

LIANG, P.-H.; KO, T.-P.; WANG, A. H.-J. Structure, mechanism and function of prenyltransferases. **European Journal of Biochemistry**, v. 269, n. 14, p. 3339–3354, 2002.

LÖYTYNOJA, A.; MILINKOVITCH, M. C. A hidden Markov model for progressive multiple alignment. **Bioinformatics (Oxford, England)**, v. 19, n. 12, p. 1505–1513, 12 ago. 2003.

MAERE, S.; HEYMANS, K.; KUIPER, M. BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. **Bioinformatics**, v. 21, n. 16, p. 3448–3449, 15 ago. 2005.

MANNI, M. et al. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic,

and Viral Genomes. **Molecular Biology and Evolution**, v. 38, n. 10, p. 4647–4654, 1 out. 2021.

MCPARTLAND, J. M. Cannabis Systematics at the Levels of Family, Genus, and Species. **Cannabis and Cannabinoid Research**, v. 3, n. 1, p. 203–212, dez. 2018.

MISTRY, J. et al. Pfam: The protein families database in 2021. **Nucleic Acids Research**, v. 49, n. D1, p. D412–D419, 8 jan. 2021.

QUEVILLON, E. et al. InterProScan: protein domains identifier. **Nucleic Acids Research**, v. 33, n. Web Server issue, p. W116-120, 1 jul. 2005.

RADWAN, M. M. et al. Cannabinoids, Phenolics, Terpenes and Alkaloids of Cannabis. **Molecules**, v. 26, n. 9, p. 2774, jan. 2021.

RETTBERG, N.; BIENDL, M.; GARBE, L.-A. Hop Aroma and Hoppy Beer Flavor: Chemical Backgrounds and Analytical Tools—A Review. **Journal of the American Society of Brewing Chemists**, v. 76, n. 1, p. 1–20, 2 jan. 2018.

RODOLFI, M. et al. From Hop to Beer: Influence of Different Organic Foliar Fertilisation Treatments on Hop Oil Profile and Derived Beers' Flavour. **Plants**, v. 12, n. 9, p. 1861, jan. 2023.

ROHDICH, F. et al. Biosynthesis of terpenoids: 4-diphosphocytidyl-2C-methyl-D-erythritol synthase of *Arabidopsis thaliana*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 97, n. 12, p. 6451–6456, 6 jun. 2000.

SEEMANN, M. et al. Isoprenoid biosynthesis in plant chloroplasts via the MEP pathway: direct thylakoid/ferredoxin-dependent photoreduction of GcpE/IspG. **FEBS letters**, v. 580, n. 6, p. 1547–1552, 6 mar. 2006.

SHANNON, P. et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. **Genome Research**, v. 13, n. 11, p. 2498–2504, 11 jan. 2003.

SIMPSON, M. G. **Plant Systematics**. [s.l.] Academic Press, 2010.

SOMMANO, S. R. et al. The Cannabis Terpenes. **Molecules**, v. 25, n. 24, p. 5792, jan. 2020.

SZKLARCZYK, D. et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. **Nucleic Acids Research**, v. 51, n. D1, p. D638–D646, 6 jan. 2023.

THOLL, D. Biosynthesis and Biological Functions of Terpenoids in Plants. Em: SCHRADER, J.; BOHLMANN, J. (Eds.). **Biotechnology of Isoprenoids**. Advances in Biochemical Engineering/Biotechnology. Cham: Springer International Publishing, 2015. p. 63–106.

