

UNIVERSIDADE FEDERAL DO OESTE DA BAHIA
CENTRO MULTIDISCIPLINAR DE LUÍS EDUARDO MAGALHÃES
ENGENHARIA DE BIOTECNOLOGIA

SARAH BARBOSA DE DEUS

**Prospecção de genes de suscetibilidade (S-genes) a doenças em Cafeeiro (*Coffea arabica*)
para edição via CRISPR/Cas9**

LUÍS EDUARDO MAGALHÃES

2025

SARAH BARBOSA DE DEUS

**Prospecção de genes de suscetibilidade (S-genes) a doenças em Cafeeiro (*Coffea arabica*)
para edição via CRISPR/Cas9**

Trabalho de Conclusão de Curso apresentado ao curso de Engenharia de Biotecnologia do Centro Multidisciplinar de Luís Eduardo Magalhães, da Universidade Federal do Oeste da Bahia, como requisito parcial para obtenção do título de bacharel em Engenharia de Biotecnologia.

Orientador: Prof. Dr. Samuel Mazzinghy Alvarenga.

LUÍS EDUARDO MAGALHÃES

2025

FICHA CATALOGRÁFICA

D486 Deus, Sarah Barbosa de.

Prospecção de genes de suscetibilidade (S-genes) a doenças em Cafeeiro (*Coffea arabica*) para edição via CRISPR/Cas9. / Sarah Barbosa de Deus. – 2025.

48 f.; il. color.

Orientador: Prof. Dr. Samuel Mazzinghy Alvarenga.

Trabalho de Conclusão de Curso: (graduação em Engenharia de Biotecnologia) – Universidade Federal do Oeste da Bahia. Centro Multidisciplinar de Luís Eduardo Magalhães, Luís Eduardo Magalhães, BA, 2025.

1. Agricultura – Aplicações biotecnológicas. 2. Café – Melhoramento genético. 3. Prevenção e Controle de doenças.

I. Alvarenga, Samuel Mazzinghy. II. Universidade Federal do Oeste da Bahia – Centro Multidisciplinar de Luís Eduardo Magalhães. III. Título.

CDD: 631.52

BIBLIOTECAS UFOB - Biblioteca Universitária de Luís Eduardo Magalhães

SARAH BARBOSA DE DEUS

Prospecção de genes de suscetibilidade (S-genes) a doenças em Cafeeiro (Coffea arabica) para edição via CRISPR/Cas9

Trabalho de Conclusão de Curso apresentado ao Centro Multidisciplinar de Luís Eduardo Magalhães da Universidade Federal do Oeste da Bahia – UFOB, como requisito parcial para obtenção do título de Bacharel em Engenharia de Biotecnologia.

Luís Eduardo Magalhães, BA, 01/08/2025

BANCA EXAMINADORA

Prof^{ta}. Dra. Aurizangela Oliveira de Sousa
Universidade Federal do Oeste da Bahia

Pesq. Dra. Eveline Teixeira Caixeta
Embrapa Café

Pesq. Dr. Diogo Paes da Costa
Universidade Federal do Agreste de Pernambuco

(A versão assinada deste documento encontra-se com a coordenação de curso)

AGRADECIMENTOS

À minha mãe, Samantha, razão por eu ter chegado onde estou e motivo pra eu continuar seguindo.

À minha avó, Gildete, base forte na minha criação e modelo de mulher que me ensinou a ser quem sou.

Aos meus tios Helton, Sabrina e Marisa, que sempre estiveram por mim quando eu precisei.

À Aurélio, meu companheiro e pilar principal da minha estrutura nos últimos 5 anos, que me acompanhou em toda a minha jornada.

Aos meus amigos João Paulo, Jardel, Guilherme e Antony que estão desde sempre e a todo momento disponíveis a me ouvir, sorrir e chorar comigo.

Às minhas companheiras de universidade Nubya e Giovana que formaram comigo uma rede de companheirismo para enfrentar os momentos de desespero e alegrias.

Aos meus mestres, em especial ao meu orientador, Professor Samuel, que me apoiou mesmo quando eu falhei; aos Professores Felipe e Pedro, que me ajudaram a crescer em aspectos pessoais e profissionais e enxergar as coisas de maneira mais ampla.

E especialmente, a mim mesma, por não ter desistido.

“Não prometo nada de completo; pois qualquer obra humana que se suponha completa, deve, por essa mesmíssima razão, infalivelmente revelar-se imperfeita.”

Herman Melville.

RESUMO

A ocorrência da ferrugem, doença causada pelo fungo *Hemileia vastatrix*, em cafeeiro (*Coffea arabica*) é uma das maiores preocupações dos produtores dessa cultura, pois é capaz de reduzir a produtividade em níveis que causam grandes prejuízos econômicos. Tradicionalmente, o controle da doença é feito com fungicidas e cultivares tolerantes, mas a adaptação do patógeno limita a eficácia dessas estratégias. Nesse contexto, a tecnologia CRISPR/Cas9 surge como uma alternativa para o desenvolvimento de variedades com resistência durável. Este identificou genes de suscetibilidade em *C. arabica* e desenvolveu gRNAs específicos para edição gênica via CRISPR. Foram selecionados três genes candidatos (MLO-like, DMR6-like e nsLTP-like) com base em sua descrição funcional em espécies como *Arabidopsis thaliana*, *Solanum lycopersicum*, *Nicotiana tabacum* e *Capsicum annuum*. A análise comparativa das sequências nucleotídicas e proteicas foi realizada utilizando a ferramenta EMBOSS Water e os resultados foram visualizados no *software* Jalview. Os genes MLO-like e DMR6-like apresentaram identidade nucleotídica moderada (39,8% a 44,4%) entre as espécies comparadas, enquanto o gene nsLTP-like mostrou maior conservação, especialmente com *A. thaliana*, com 66,7% de identidade proteica. Para o desenho dos gRNAs, utilizou-se a plataforma CRISPOR, focando no genoma de referência do *Coffea arabica* e parâmetros como conteúdo GC (40-60%) e especificidade (MIT Score ≥ 80). Foram gerados 14 gRNAs, sendo seis para MLO-like, cinco para DMR6-like e três para nsLTP-like. Os gRNAs para DMR6-like destacaram-se pelos altos *scores* de especificidade (95-99), indicando baixo risco de efeitos *off-target*. Em contraste, os gRNAs para o gene nsLTP-like apresentaram *scores* de especificidade mais baixos (82-89), sugerindo uma menor eficiência potencial de clivagem e um risco ligeiramente maior de edições indesejadas. O desenho de gRNAs para edição de genes de suscetibilidade em *C. arabica* por CRISPR representa uma abordagem ainda não explorada na literatura. A seleção de sequências-alvo que prioriza regiões com provável função relevante se mostra como uma estratégia potencialmente eficiente para a redução de vulnerabilidade de plantas a patógenos específicos. Os resultados demonstram a viabilidade da abordagem computacional para identificar genes de suscetibilidade e desenhar gRNAs específicos em *C. arabica*, apesar da complexidade do genoma tetraplóide da espécie. A validação experimental dos gRNAs propostos neste trabalho deve ser realizada em estudos futuros de edição dos genes-alvo, assim a funcionalidade das proteínas no contexto da suscetibilidade ao patógeno pode ser determinada, podendo contribuir para o desenvolvimento de variedades resistentes.

Palavras-chave: *bioinformática; cafeeiro; patógenos; melhoramento; gRNAs.*

ABSTRACT

The occurrence of coffee leaf rust, a disease caused by the fungus *Hemileia vastatrix*, in coffee plants (*Coffea arabica*) is one of the main concerns for producers of this crop, as it can reduce productivity to levels that cause major economic losses. Traditionally, the disease is controlled with fungicides and tolerant cultivars, but the pathogen's adaptability limits the effectiveness of these strategies. In this context, CRISPR/Cas9 technology emerges as an alternative for developing varieties with durable resistance. This study identified susceptibility genes in *C. arabica* and developed specific gRNAs for gene editing using CRISPR. Three candidate genes (*MLO-like*, *DMR6-like*, and *nsLTP-like*) were selected based on their functional descriptions in species such as *Arabidopsis thaliana*, *Solanum lycopersicum*, *Nicotiana tabacum*, and *Capsicum annuum*. Comparative analysis of nucleotide and protein sequences was performed using the EMBOSS Water tool, and results were visualized with Jalview software. The *MLO-like* and *DMR6-like* genes showed moderate nucleotide identity (39.8% to 44.4%) among the species compared, while the *nsLTP-like* gene showed greater conservation, especially with *A. thaliana*, with 66.7% protein identity. The CRISPOR platform was used to design gRNAs, focusing on the reference genome of *Coffea arabica* and parameters such as GC content (40–60%) and specificity (MIT Score \geq 80). A total of 14 gRNAs were generated—six for *MLO-like*, five for *DMR6-like*, and three for *nsLTP-like*. The gRNAs for *DMR6-like* stood out for their high specificity scores (95–99), while those for the *nsLTP-like* gene showed lower scores (82–89), indicating lower potential efficiency. The design of gRNAs for editing susceptibility genes in *C. arabica* via CRISPR represents an approach not yet explored in the literature. The selection of target sequences that prioritize regions with likely relevant function appears to be a potentially effective strategy for reducing plant vulnerability to specific pathogens. The results demonstrate the feasibility of a computational approach to identify susceptibility genes and design specific gRNAs in *C. arabica*, despite the complexity of the species' tetraploid genome. Experimental validation of the gRNAs proposed in this study should be carried out in future gene editing studies, so that protein functionality in the context of susceptibility to the pathogen can be determined, potentially contributing to the development of resistant varieties.

Keywords: *bioinformatics; coffee tree; pathogen; improvement; gRNAs.*

LISTA DE FIGURAS

Figura 2 – Árvore filogenética representando a relação evolutiva entre <i>Coffea canephora</i> , <i>Solanum lycopersicum</i> , <i>Vitis vinifera</i> e <i>Oryza sativa</i>	16
Figura 1 - Representação do funcionamento do sistema CRISPR.	18
Figura 3 - Fórmula recursiva do algoritmo de Smith-Waterman.	23
Figura 4 - Matriz de pontuação do algoritmo Smith-Waterman.	23
Figura 5 - Visualização do alinhamento local do gene MLO-like.	27
Figura 6 - Visualização do alinhamento local do gene DMR6-like.	28
Figura 7 - Visualização do alinhamento local do gene nsLTP-like.....	29
Figura 8 - Visualização do alinhamento local da proteína MLO-like.	32
Figura 9 - Visualização do alinhamento local da proteína DMR6-like.	33
Figura 10 - Visualização do alinhamento local da proteína nsLTP-like..	33

LISTA DE TABELAS

Tabela 1 - Identidade dos alinhamentos nucleotídicos entre genes de <i>Coffea arabica</i> e seus correspondentes em outras espécies.	30
Tabela 2 - Identidade e similaridade dos alinhamentos proteicos entre produtos gênicos de <i>Coffea arabica</i> e seus correspondentes em outras espécies.	31
Tabela 3 - RNAs guias propostos para a edição dos genes MLO-like, DMR6-like e nsLTP-like em <i>Coffea arabica</i>	35

LISTA DE SIGLAS

BLOSUM62	<i>Blocks Substitution Matrix 62</i>
Cas9	<i>CRISPR-associated protein 9</i>
CRISPR	<i>Clustered Regularly Interspaced Short Palindromic Repeats</i>
DCL	<i>Dicer-like</i>
EBE	<i>Effector Binding Elements</i>
EST	<i>Expressed Sequence Tags</i>
ETI	<i>Effector-Triggered Immunity</i>
GISH	<i>Genomic in situ hybridization</i>
HDR	<i>Homology-Directed Repair</i>
Hi-C	<i>High-throughput chromosome conformation capture</i>
KEGG	<i>Kyoto Encyclopedia of Genes and Genomes</i>
LRR	<i>Leucine-Rich Repeat</i>
NBS	<i>Nucleotide-Binding Site</i>
NCBI	<i>National Center for Biotechnology Information</i>
NHEJ	<i>Non-Homologous End Joining</i>
PAM	<i>Protospacer Adjacent Motif</i>
PAMPS	<i>Pathogen-Associated Molecular Patterns</i>
PDB	<i>Protein Data Bank</i>
QTL	<i>Quantitative trait loci</i>
RFLP	<i>Restriction Fragment Length Polymorphism</i>
RNAi	<i>Ácido Ribonucleico De Interferência</i>
SNP	<i>Single Nucleotide Polymorphisms</i>
TALEN	<i>Transcription Activator-Like Effector Nuclease</i>
TIR	<i>Toll/Interleukin-1 Receptor</i>
UniProt	<i>Universal Protein Resource</i>
ZFN	<i>Zinc Finger Nucleases</i>
GWAS	<i>Genome-Wide Association Study</i>

SUMÁRIO

1. INTRODUÇÃO.....	9
1.1. OBJETIVOS	10
1.1.1. OBJETIVO GERAL.....	10
1.1.2. OBJETIVOS ESPECÍFICOS.....	10
2. REFERENCIAL TEÓRICO	11
2.1. CAFEEIRO.....	11
2.2. SEQUENCIAMENTO DE GENES, ESTs e GENOMA EM <i>Coffea Arabica</i>	12
2.3. INTERAÇÃO PLANTA-PATÓGENO	13
2.3.1. S-GENES	14
2.3.1.1. PROSPECÇÃO DE S-GENES EM CAFEEIRO	15
2.4. SISTEMA CRISPR/Cas.....	17
2.5. BIOINFORMÁTICA.....	20
2.5.1. PLATAFORMAS PARA EDIÇÃO COM CRISPR	21
2.5.2. CRISPOR.....	22
2.5.3. EMBOSS Water.....	22
2.5.4. JALVIEW	24
3. METODOLOGIA.....	25
3.1. IDENTIFICAÇÃO DE GENES DE SUSCETIBILIDADE	25
3.2. ALINHAMENTO DE SEQUÊNCIAS	25
3.3. DESENHO E SELEÇÃO DOS gRNAs.....	26
4. RESULTADOS E DISCUSSÃO	27
5. CONSIDERAÇÕES FINAIS.....	36
6. REFERÊNCIAS	37
APÊNDICE A – CÓDIGO EM PYTHON PARA MINERAÇÃO DOS RNAs GUIA	48

1. INTRODUÇÃO

O cultivo de café é uma das atividades agrícolas mais importantes no contexto econômico mundial e nacional, com o café arábica sendo a espécie mais valorizada devido à qualidade da bebida (Souza et al., 2004). No entanto, a produção de café enfrenta desafios relacionados a doenças como a ferrugem-do-cafeeiro, causada pelo fungo *Hemileia vastatrix*. Essa doença é responsável por perdas entre 35% e 40% na produção da cultura (Vieira Júnior et al, 2008). O controle tradicional da ferrugem tem sido realizado por meio de fungicidas e do desenvolvimento de cultivares resistentes obtidos por meio do melhoramento genético convencional. Contudo, essas estratégias podem ser limitadas pela fácil adaptação do patógeno (Talhinhas et al., 2017). Dessa forma, a tecnologia CRISPR/Cas9 é uma alternativa com potencial para desenvolvimento de variedades de café com resistência durável a doenças.

Apesar dos avanços no sequenciamento do genoma do café arábica, o entendimento dos genes de suscetibilidade, conhecidos como S genes, ainda carece de aprofundamento. Segundo Cui et al. (2020), os S genes são genes do hospedeiro, que quando inativados, podem conferir resistência a doenças, pois nessa condição podem impedir a interação bem-sucedida entre o patógeno e a planta. No entanto, no café arábica, a identificação e caracterização desses genes ainda não são vastamente estudadas, o que representa uma oportunidade para pesquisas que possam explorar essa área e contribuir para o melhoramento genético da cultura.

Além disso, a aplicação da tecnologia CRISPR/Cas9 no café arábica enfrenta o desafio da ausência do seu genoma em ferramentas de *design* de RNAs guias como CRISPR-P e CHOPCHOP. Isso dificulta a criação de gRNAs eficientes e específicos, essenciais para fins de edição genética na espécie. Logo, o desenho de gRNAs específicos para o genoma do cafeeiro é um passo fundamental para o avanço das pesquisas na área.

A utilização de ferramentas de bioinformática no processo de aperfeiçoamento genético pode acelerar substancialmente o desenvolvimento de variedades de cafeeiro com características de resistência a patógenos. A implementação de uma *pipeline* que utilize essas ferramentas, com foco para o genoma do cafeeiro, permitirá a identificação de genes de suscetibilidade e o desenho de gRNAs com alta especificidade e eficiência. Além disso, a simulação computacional dos resultados da edição genética pode reduzir o tempo e os custos

associados aos experimentos *in vivo*, tornando o processo mais eficiente e acessível (Li et al., 2022).

Portanto, este trabalho visa preencher uma lacuna no campo do melhoramento genético do café arábica, focando na mineração de genes de suscetibilidade e no desenvolvimento de uma *pipeline* para o *design* de gRNAs específicos, com o objetivo de criar variedades menos suscetíveis a ferrugem-do-cafeeiro.

1.1. OBJETIVOS

1.1.1. OBJETIVO GERAL

Identificar genes de suscetibilidade em *Coffea arabica* por meio de análise comparativa com espécies-modelo e desenvolver RNAs guias (gRNAs) específicos para edição gênica utilizando a tecnologia CRISPR/Cas9.

1.1.2. OBJETIVOS ESPECÍFICOS

- Identificar e selecionar sequência gênicas do cafeeiro associadas à suscetibilidade a *Hemileia vastatrix*;
- Utilizar ferramentas computacionais para o desenho de gRNAs e analisar as sequências gênicas selecionadas;
- Selecionar gRNAs com alta especificidade e eficiência.

2. REFERENCIAL TEÓRICO

2.1. CAFEEIRO

O *Coffea arabica* é uma espécie aloploplóide resultante da hibridização natural entre *Coffea canephora* e *Coffea eugenioides*, com número cromossômico de $2n = 4x = 44$. Essa natureza tetraplóide torna o seu genoma particularmente complexo, tanto em termos de tamanho como de organização gênica (Denoeud et al., 2014).

Estima-se que o genoma do *C. arabica* possui aproximadamente 1,2 gigabases distribuídos em 4 pares de 11 cromossomos, com 80.883 genes anotados, dos quais 50.280 (62%) são classificados como codificadores de proteínas, segundo o RefSeq (2024). Dados do GenBank (NCBI, 2024) indicam que o genoma de referência mais completo atualmente disponível é o "Coffea Arabica ET-39 HiFi", depositado sob o acesso GCF_036785885.1. O valor de N50 para os *scaffolds* é de 53,7 Mb, com L50 igual a 10, indicando que metade do genoma está contida em 10 *scaffolds* de pelo menos 57,7 Mb cada.

A estrutura genética complexa da espécie, com alta redundância gênica e presença de múltiplas cópias alélicas, cria desafios à aplicação de técnicas de engenharia genética, uma vez que a edição eficaz requer a mutagênese simultânea de todos os alelos funcionais para que uma característica de interesse seja obtida (Prado et al., 2024; Abdelrahman et al., 2021).

Esse desafio pode ser observado em outras espécies aloploplóides, como algodão, onde múltiplos alelos requerem múltiplas guias ou edição multialélica para alcançar o fenótipo desejado. Portanto, para a configuração tetraplóide, a eficiência pode ser reduzida significativamente e exige rotas alternativas como o uso de múltiplos gRNA's (WANG et al., 2018).

A ferrugem-do-cafeeiro é causada pelo fungo *Hemileia vastatrix*, pertencente ao filo Basidiomycota e à ordem Puccinales, e é conhecida por acarretar perdas na produção cafeeira ao redor do mundo. A *H. vastatrix* foi registrada pela primeira vez no Sri Lanka em 1869, onde devastou as plantações de café obrigando os produtores do país a substituírem o cultivo por chá (Salazar-Navarro et al., 2024). A doença se manifesta inicialmente por manchas cloróticas nas folhas, que evoluem para lesões alaranjadas. Essas lesões, localizadas na face interna das folhas, causam sua queda precoce e comprometem o crescimento e produtividade das plantas (Talhinhas et al., 2017).

De acordo com Porto et al. (2019), o entendimento sobre o genoma desse patógeno ainda apresenta lacunas, especialmente quando se trata da bioquímica do processo de infecção baseado na interação com o hospedeiro. O genoma de referência presente no GenBank (2025),

identificado como Hv_R1_Cat_CENPAT (*Assembly* GCA_030280995.1), foi obtido na raça I do fungo, e possui 748 Mb. A montagem está distribuída em 794 *scaffolds*, com valor N50 de 20,4 Mb e L50 de 13, e a composição GC é de 34%.

O controle da ferrugem tem sido realizado por meio de fungicidas e do desenvolvimento de cultivares resistentes a partir de melhoramento genético tradicional. No entanto, a alta variabilidade do fungo pode superar a resistência genética ao longo do tempo, tornando necessário o desenvolvimento contínuo de novas cultivares. O controle químico é amplamente utilizado, mas envolve custos elevados e apresenta a limitação de exigir doses crescentes ao longo do tempo de uso, impactando a sustentabilidade do cultivo (Salazar-Navarro et al., 2024; Sera et al., 2007).

2.2. SEQUENCIAMENTO DE GENES, ESTs e GENOMA EM *Coffea Arabica*

A caracterização genômica e transcriptômica do cafeeiro tem sido tema de projetos e estudos ao longo dos últimos 20 anos. O *Brazilian Coffee Genome Project* (Vieira et al., 2006) gerou mais de 214 000 leituras EST de *C. arabica*, *C. canephora* e *C. racemosa*, identificando cerca de 33000 unigenes após a montagem. Mekbib et al. (2022) sequenciaram 90 amostras de *Coffea arabica* da Etiópia, identificando 11 milhões de variantes genéticas entre SNPs (Single Nucleotide Polymorphism) e inserções/deleções. Enquanto Mosharaf et al. (2019) identificaram cinco genes Dicer-like (DCL), componentes essenciais da via de RNAi, Campos et al. (2007) investigaram a presença de genes de resistência (Classe LRR/NBS/TIR) em *C. arabica* por meio de mineração de ESTs no banco CafEST, identificando 293 sequências de associadas às subclasses RPP4, RPP5 e RPS4, destacando a diversidade desses genes no cafeeiro.

Medrano et al. (2024) sequenciaram e montaram o genoma completo da variedade *Geisha* com tecnologias de longa leitura e Hi-C, com 97,7% de completude e mais de 47 mil *loci* gênicos. A comparação com outras variedades indicou alta identidade genômica, reforçando a hipótese de um único evento de aloploidização na origem da espécie. O trabalho de Fernandez et al. (2012) analisou o transcriptoma de *Hemileia vastatrix* durante infecção compatível com cafeeiro, gerando 22.774 *contigs* por sequenciamento 454, dos quais 30% foram atribuídos ao fungo. Já Andrade et al. (2007), no Projeto Brasileiro do Genoma Café (PBGC), usaram mineração de dados, BLAST e a ferramenta Electronic Northern para identificar 8.968 sequências relacionadas à resistência à ferrugem, com potencial uso em programas de melhoramento. Em Alvarenga et al (2010), foram identificadas 11.300 ESTs associadas a mecanismos de resistência do cafeeiro a doenças, que posteriormente foram

agrupados em 979 *contigs*. Desses *contigs*, 140 relacionados a ESTs correspondentes a proteínas do tipo NBS-LRR, foram funcionalmente classificados e apresentaram indicação de relação com processos metabólicos e biológicos, como apoptose, sinalização celular e mecanismos de defesa da planta.

Ferrão et al. (2015) desenvolveram 101 marcadores microssatélites derivados de ESTs (EST-SSRs) do PBGC, com alta taxa de transferibilidade (88%) e potencial para mapeamento de QTLs e seleção assistida por marcadores moleculares em diferentes espécies de *Coffea*, e Lashermes et al. (1999) confirmaram, por RFLP e GISH a origem híbrida recente de *C. arabica* a partir de *C. canephora* e *C. eugenioides*. Outros trabalhos com foco no genoma das 3 espécies são os de Scalabrin et al. (2020), que montou os sub genomas do *Coffea arabica* e seus progenitores, identificando baixa diversidade genética em 736 amostras, enquanto, mais recentemente, Salojärvi et al. (2024) montaram os mesmos genomas, revelando introgressões recentes.

2.3. INTERAÇÃO PLANTA-PATÓGENO

Para o desenvolvimento de tecnologias que auxiliem no combate aos danos causados pela ferrugem-do-cafeeiro à cultura do café, é importante compreender de que forma se dá a interação entre a planta hospedeira e o microrganismo causador da doença.

Como descrito por Florez Varon (2018), as plantas desenvolveram mecanismos de defesa contra patógenos, baseados em dois sistemas principais de detecção, sendo o primeiro o reconhecimento de padrões moleculares associados a patógenos (PAMPs), mediados por receptores de reconhecimento de padrões, que ativa a imunidade basal, e o segundo o reconhecimento específico de efeitos patogênicos por proteínas de resistência, desencadeando a imunidade efetora (ETI).

Além do reconhecimento molecular, as plantas também apresentam defesas estruturais e bioquímicas integradas contra patógenos. Entre as respostas estruturais estão o reforço da parede celular com calose e lignina. Já as respostas bioquímicas incluem a produção de espécies reativas de oxigênio, ativação de enzimas como quitinases e glucanases e compostos com ação microbiana (Florez Varon, 2018; Salazar-Navarro et al., 2024).

O processo de infecção do cafeeiro pela *H. vastatrix* inicia-se quando uredosporos aderem à superfície interior da folha, germinam e formam estruturas de penetração (apressórios), selecionando estômatos por meio de sinais na estrutura tridimensional da planta e enzimáticos, que facilitam a colonização do parênquima. Após a formação de haustórios nas

células hospedeiras, o fungo estabelece um relacionamento biotrófico estável, captando nutrientes diretamente da planta (Silva et al., 2022).

Algumas variedades de *Coffea spp.* expressam genes de resistência que detectam moléculas efetoras produzidas pelo patógeno para suprimir a resposta da imunidade basal, desencadeando a resposta hipersensível e impedindo a propagação da infecção através da morte celular, visto que o fungo é biotrófico obrigatório e necessita de células vivas do hospedeiro para completar seu ciclo de vida. No entanto, a elevada variabilidade genética da ferrugem e o surgimento frequente de novas raças impõem-se como desafios ao desenvolvimento de estratégias de controle duradouras, como a edição gênica (Silva et al., 2022).

2.3.1. S-GENES

Os genes de suscetibilidade, ou S-genes, são genes com potencial para ser alvos estratégicos no desenvolvimento de resistência genética em plantas frente a patógenos, como fungos, vírus e bactérias. Diferentemente dos genes de resistência, os S-genes são genes do hospedeiro em que a expressão é explorada pelo patógeno para facilitar o processo de infecção. Esses genes codificam proteínas essenciais para processos fisiológicos normais da planta, mas que ao serem ativados, contribuem para a suscetibilidade (Zaidi, Mukhtar e Mansoor, 2018; Garcia-Ruiz et al., 2021).

O gene MLO (*Mildew Locus O*) codifica uma proteína de membrana pertencente à família de proteínas de resposta ao estresse. Essa proteína atua como reguladora negativa da imunidade basal, suprimindo respostas associadas à produção de espécies reativas de oxigênio e à expressão de genes de defesa mediados por cálcio e ácido salicílico. Patógenos como fungos do oídio se aproveitam dessa supressão para penetrar na epiderme vegetal (Garcia-Ruiz et al., 2021; Jacott et al., 2021).

O gene DMR6 (*Downy Mildew Resistant 6*) codifica uma oxigenase 2-oxoglutarato/Fe(II)-dependente que hidroxila o ácido salicílico, convertendo-o em 2,5-dihidroxibenzóico (2,5-DHBA). Essa modificação reduz a concentração ativa de ácido salicílico, crucial para a ativação de genes de defesa, suprimindo a imunidade da planta. A mutação em DMR6 bloqueia essa rota de supressão, permitindo o acúmulo de ácido salicílico e o fortalecimento de defesas sistêmicas (Zhang et al., 2017; Thomazella et al., 2021).

Os genes nsLTPs (*non-specific lipid transfer proteins*) expressam proteínas que participam do transporte de cutina, suberina e outros lipídios estruturais importantes para a formação da cutícula e da parede celular. Embora tradicionalmente associadas à defesa,

algumas isoformas específicas, como a LTP3 de *Arabidopsis thaliana* foram identificadas como reguladoras negativas da resposta imune. A LTP3 modula o balanço entre ácido abscísico (ABA) e ácido salicílico (SA), promovendo a via de ABA em detrimento da via de AS, o que resulta em maior suscetibilidade à infecções (Gao et al., 2016).

A grande vantagem da abordagem baseada na edição de genes de suscetibilidade é a possibilidade de desenvolver resistência recessiva, que ocorre quando a planta perde ou modifica um gene essencial para a infecção do patógeno, dificultando a adaptação do microrganismo e tornando essa resistência geralmente mais ampla e menos propensa a ser superada por mutações no patógeno (Zaidi, Mukhtar e Mansoor, 2018).

No contexto do *Coffea arabica*, grande parte das estratégias de resistência à ferrugem tem se concentrado no uso de genes SH (SH1 a SH9). Esses genes estão associados ao reconhecimento de raças específicas de *Hemileia vastatrix* por meio da resposta de resistência específica, como apontado por Guerra-Guimarães et al. (2023). Entretanto, essa estratégia tem se mostrado vulnerável, devido à alta variabilidade genética do fungo, que já conta com dezenas de raças virulentas ao redor do mundo (Maia et al., 2017).

A proposta de utilizar edição gênica para desativar S-genes em café exige uma etapa inicial de caracterização funcional desses genes. Técnicas com RNA-seq e análise de expressão em folhas infectadas e não infectadas podem ajudar a identificar genes altamente expressos em resposta à infecção, como o trabalho realizado por Florez Varon (2018).

2.3.1.1. PROSPECÇÃO DE S-GENES EM CAFEEIRO

A identificação de genes de suscetibilidade em *Coffea arabica* é um desafio devido à lacuna na caracterização de genes com essa função na espécie. Para um estudo de genes candidatos a apresentar funções que facilitam a infecção por patógenos, como *H. vastatrix*, é possível se utilizar da documentação já existente dos S-genes em espécies de outras famílias, como Solanaceae e Brassicaceae, que apresentam validações experimentais da importância desses genes no processo de interação entre a planta e o agente infeccioso.

Além de as famílias Rubiaceae e Solanaceae pertencerem ao clado Asterid I (Figura 2), de Guyot et al. (2012) destaca que essas famílias são muitas vezes consideradas próximas no ponto de vista genético, o que é corroborado por trabalhos como o de Lin (2005), que demonstra grande sintonia entre *C. canephora* e *Solanum lycopersicum*, o último pertencente à família Solanaceae. Assim, pode-se criar uma perspectiva de compartilhamento de mecanismos

moleculares relacionados à suscetibilidade a patógenos possivelmente conservados em plantas de famílias próximas, como *Capsicum annuum*, *Nicotiana tabacum* e *Solanum lycopersicum*.

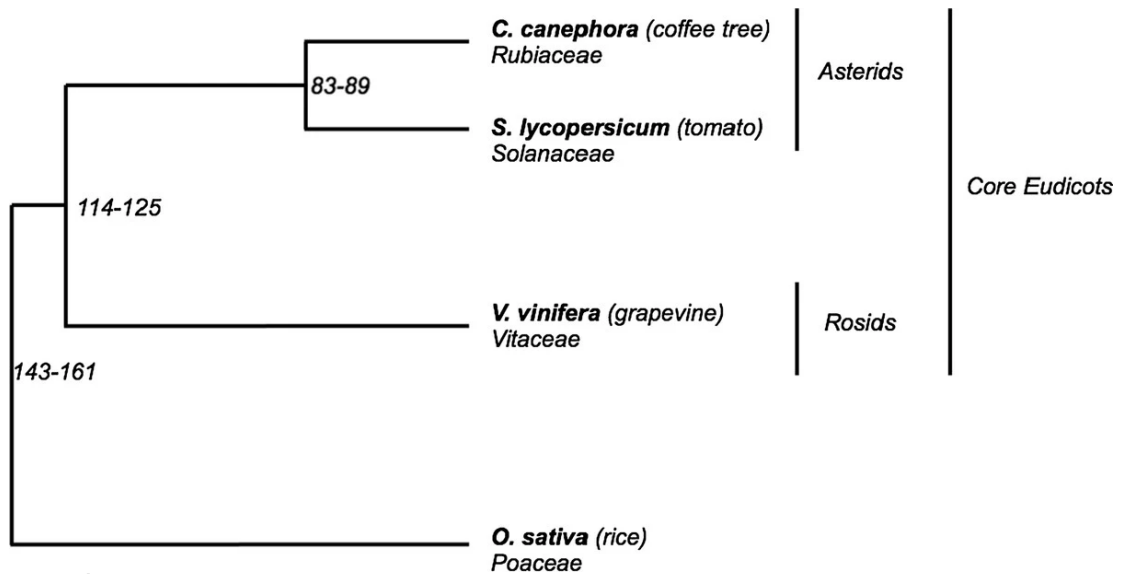


Figura 1 – Árvore filogenética representando a relação evolutiva entre *Coffea canephora*, *Solanum lycopersicum*, *Vitis vinifera* e *Oryza sativa*. Fonte: Guyot et al. (2012).

Embora *Arabidopsis thaliana* e o cafeeiro pertençam a famílias distantes filogeneticamente, a primeira é extensamente utilizada como espécie modelo para referência em estudos genômicos devido ao seu genoma completamente sequenciado, vasta disponibilidade em ferramentas moleculares e caracterização funcional de seus genes. Muitos mecanismos moleculares foram inicialmente reconhecidos em *A. thaliana*, o que permitiu identificações para outras espécies (Yaschenko, 2025).

Segundo os resultados obtidos por Van Damme et al. (2008), o gene DMR6 (Downy Mildew Resistant 6), em *A. thaliana*, atua como um supressor da resposta imune da planta, e sua perda de função confere resistência ao oídio (*Hyaloperonospora parasitica*). O gene ortólogo DMR6 em tomate, quando teve sua função perdida, conferiu resistência a fungos, bactérias e oomicetos na planta, como demonstrado por Thomazella et al. (2021).

De maneira semelhante, genes da família MLO (Mildew Locus O) foram reconhecidos como determinantes de suscetibilidade ao oídio em diversas espécies, incluindo NtMLO1 em tabaco (*Nicotiana tabacum*) e SIMLO1 em tomate, cuja inativação confere resistência durável (Appiano et al., 2015). Zheng et al. (2013) demonstraram que os genes dessa família também estão envolvidos na suscetibilidade de *Capsicum annuum* a *Leveillula taurica*, assim como Xuebo et al. (2023) demonstrou eficiência na aplicação de CRISPR/Cas9 na edição dos genes NtMLO1 e NtMLO2 em tabaco, resultando em resistência ao oídio.

Além disso, genes nsLTP-like (Non specific Lipid Transfer Protein) são candidatos promissores, visto que em *Arabidopsis thaliana*, o gene LTP3 foi caracterizado como um regulador negativo da imunidade, atuando na modulação da biossíntese de ácido abscísico e na supressão de respostas mediadas por ácido salicílico, aumentando a suscetibilidade a *Pseudomonas syringae* (Gao et al., 2016).

2.4. SISTEMA CRISPR/Cas

O sistema CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats*) veio à luz como um mecanismo de defesa em bactérias e arqueas durante a década de 80, quando Ishino et al. (1987) publicaram a sua descoberta, porém com pouco conhecimento a respeito do seu funcionamento em completo. A partir deste momento, muitas pesquisas foram realizadas até que, enfim, Jinek et al. (2012) publicaram os resultados de uma pesquisa que demonstraram a possibilidade de uso do mecanismo desses organismos para edição genômica. Desde então, a técnica CRISPR vem sendo utilizada como uma alternativa para pesquisas em engenharia genética, devido à sua eficiência e facilidade, além do baixo custo em comparação com outros métodos como ZFNs (Zinc Finger Nucleases) e TALENs (Transcription activator-like effector nucleases) (Ferreira e Choupina, 2022). O sistema CRISPR consiste em sequências repetitivas intercaladas por espaçadores derivados do DNA de vírus invasores, e tem como principal objetivo o reconhecimento para clivagem do material genético exógeno em infecções futuras. A quebra do DNA é feita pelo conjunto formado por uma proteína Cas (CRISPR associated protein) e um híbrido entre um fragmento de RNA transcrito pelo *locus* onde o material genético viral foi adicionado no genoma bacteriano, o pré-crRNA, e um tracrRNA - trans activating crRNA (Molla, Karmakar e Islam, 2020). A Figura 1 mostra um esquema de representação do funcionamento do sistema CRISPR.

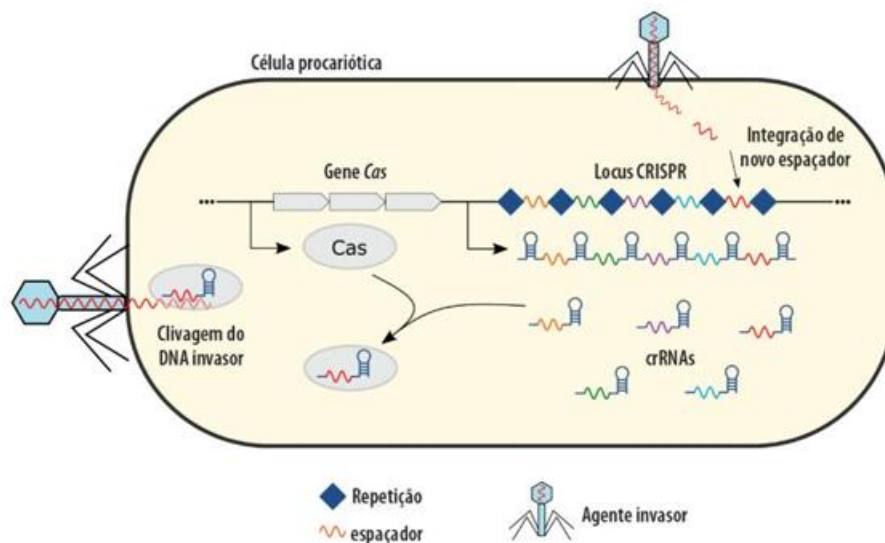


Figura 2 - Representação do funcionamento do sistema CRISPR. Fonte: Molinari et al. (2020).

Segundo Galucci (2023), o mecanismo é dividido em três etapas principais. Na fase de aquisição, uma célula bacteriana é infectada por um fago e fragmentos do genoma viral são incorporados ao *locus* CRISPR do genoma bacteriano como espaçadores. Esse processo permite que a bactéria desenvolva uma memória imunológica, conferindo proteção contra futuras infecções por um vírus já conhecido. Posteriormente, na fase de expressão e processamento, o *locus* CRISPR é transcrito em um RNA chamado pre-crRNA, que é processado em pequenos fragmentos de crRNA (CRISPR RNA). Esses crRNA se associam à enzima Cas para formar um complexo funcional que tem como objetivo a proteção da célula contra invasores virais. Na fase de interferência, o complexo crRNA-Cas reconhece no material genético viral as sequências complementares presentes no *locus* CRISPR, e age de forma com que a Cas clive o DNA invasor, impedindo a sua replicação. As endonucleases Cas comumente utilizadas, como a Cas9 (*CRISPR-associated protein 9*), possuem dois domínios, chamados RuvC e HNH, além de um domínio de interação conhecido com PAM (protospacer adjacente motif). Os domínios RuvC e HNH são responsáveis por clivar o DNA de fita dupla, gerando uma quebra da mesma. Para que esses domínios desempenhem sua função, a Cas precisa se ligar a um local específico do DNA, o que depende da presença da sequência PAM. Após o corte do DNA pela Cas, as quebras de fita dupla são reparadas pelos mecanismos de reparo da célula. Existem duas principais vias de reparo: NHEJ e HDR. O mecanismo NHEJ (Non-homologous end joining) pode introduzir inserções ou deleções de material genético no local da quebra, enquanto o HDR (Homology direct repair) requer um molde de DNA para a quebra de maneira precisa. Esse molde pode ser um DNA exógeno de fita simples ou dupla, o

cromossomo homólogo, a cromátide irmã ou até mesmo sequências altamente relacionadas ao DNA (Zhang et al., 2021; Ferreira e Choupina, 2022).

Embora seja uma opção eficaz para a edição genômica, a tecnologia CRISPR/Cas apresenta uma limitação pela possibilidade de efeitos *off-target* durante o processo (Cho et al., 2014). Os efeitos *off-target* referem-se a alterações não intencionais em regiões diferentes daquelas para onde foi direcionada a edição. Esse efeito ocorre porque a Cas pode interagir com sequências genômicas semelhantes ao alvo planejado. Esses eventos podem resultar em mutações indesejadas, que podem comprometer o funcionamento de genes essenciais (Manghwar et al., 2020).

A utilização da tecnologia CRISPR para o melhoramento de plantas vem sendo explorado na última década de forma extensiva, com estudos como Feng et al. (2013) relatando a aplicação bem sucedida do sistema para gerar mutações direcionadas em genes de *Arabidopsis thaliana* e arroz. Da mesma forma, a CRISPR/Cas9 foi utilizada por Schachtsiek e Stehle (2019) para desenvolver plantas de tabaco (*Nicotiana tabacum*) livres de nicotina, e por Rönspies et al. (2022), que descreveram protocolos para engenharia cromossômica em *Arabidopsis* com o método. A manipulação de características de interesse, como resistência a pragas, foi explorada em estudos como o realizado por Tripathi et al. (2019), que demonstrou a aplicação do sistema CRISPR/Cas9 para inativar vírus endógeno da banana integrado ao genoma da bananeira, impedindo a ativação do vírus sob condições de estresse. Na soja (*Glycine max*), Zhang et al. (2022) obtiveram resultados de aumento da resistência contra *H. armigera* e *S. litura* ao modificar a biossíntese de flavonoides por uma mutação direcionada do gene GmUGT. Pramanik et al. (2021) empregaram a tecnologia para gerar resistência ao vírus do mosaico amarelo do tomateiro (TYLCV) por meio da edição do gene SIPelo, enquanto em *Theobroma cacao*, a inativação do gene TcNPR3 que resultou em uma resposta imune aprimorada contra patógenos, como *Phytophthora tropicalis*, foi desenvolvida por Fister et al. (2018).

Outras características de interesse têm sido manipuladas a partir do sistema CRISPR/Cas9, como o aumento do rendimento foliar da alface demonstrada por Choi et al. (2022) ao editar o gene SOC1, aumentando o florescimento tardio e prolongando a fase vegetativa da planta. Além disso, Oz et al. (2021) utilizaram a tecnologia para induzir mutações nos alelos dos genes ALS em cana-de-açúcar, conferindo tolerância a herbicidas. Em trigo, Sánchez-Léon et al. (2018) obtiveram resultados de diminuição da quantidade de gliadinas, proteínas do glúten associadas à doença celíaca e gerando linhagens de trigo com menor imunogenicidade para pessoas sensíveis ao glúten.

Como descrito por Rojas-Vásquez e Gatica-Arias (2020), o uso do mecanismo tem sido utilizado em cultivares como cacau, arroz e tomates como ferramenta no combate aos danos causados por fatores e abióticos, como o melhoramento na tolerância ao estresse hídrico e a resistência a doenças fúngicas e bacterianas. Shi et al. (2017) desenvolveram variantes de milho com maior tolerância à seca e aumento da produtividade de grãos em condições de estresse hídrico a partir da modificação do gene ARGOS8 por meio de CRISPR/Cas.

No melhoramento genético do cafeeiro, o sistema se mostrou eficiente em estudos como o de Breitler et al. (2018), em que foi desenvolvida uma ferramenta genômica para identificar sequências-alvo de gRNA no genoma de *Coffea canephora*, permitindo a edição precisa do gene CcPDS (fitoeno desaturase), responsável pela biossíntese de carotenoides. A eficiência de mutação foi estimada em 30,4% com 22,8% das plantas regeneradas apresentando mutações heterozigotas e 7,6% homozigotas. Casarin et al. (2022), mostraram que a edição do gene CcPDS em *Coffea canephora* resultou em fenótipos albinos, variegados e verdes, com 77% das plantas apresentando mutações em pelo menos um alelo, indicando a CRISPR/Cas9 como uma técnica viável para o melhoramento do cafeeiro.

Para além dos benefícios obtidos pela alteração de atributos de interesse em plantas, a tecnologia permite a possibilidade de edição específica de genes sem a necessidade de inserção de DNA exógeno, o que pode facilitar o processo legal e regulatório de culturas modificadas para fins comerciais (Poletti, 2021).

2.5. BIOINFORMÁTICA

A bioinformática é uma área que combina biologia, computação e estatística para analisar e interpretar dados biológicos, principalmente DNA, RNA e proteínas, utilizando-se de ferramentas computacionais. Surgiu da necessidade de processar grandes volumes de dados gerados por técnicas de sequenciamento de genomas, como o Projeto Genoma Humano. Além da análise de sequências nucleotídicas, inclui a modelagem molecular, predição de estruturas proteicas e a simulação de interações moleculares (MOUNT, 2004; LESK, 2013; PEVSNER, 2015). Os bancos de dados são itens essenciais para a bioinformática, fornecendo conjuntos de dados organizados para armazenamento e acesso de informações biológicas. O GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>), mantido pelo National Center for Biotechnology Information (NCBI), é um dos maiores bancos de dados de sequências de DNA, contendo milhões de entradas de diversas espécies (Benson et al., 2013). Outro banco de dados importante é o Protein Data Bank (PDB) (<https://www.rcsb.org/>), que armazena estruturas

tridimensionais de proteínas determinadas por técnicas como cristalografia de raios-X e ressonância magnética (wwPDB consortium, 2019). Além desses, existem bancos de dados como o KEGG (Kyoto Encyclopedia of Genes and Genomes) (<https://www.genome.jp/kegg/>), que fornece informações sobre vias metabólicas e redes de interação gênica e o UniProt (<https://www.uniprot.org/>), que apresenta dados sobre proteínas, incluindo suas funções, localizações celulares e interações (UniProt Consortium, 2019; Kanehisa et al., 2017).

2.5.1. PLATAFORMAS PARA EDIÇÃO COM CRISPR

Uma das etapas mais importantes para a aplicação da tecnologia CRISPR na edição de genomas é a o design de RNAs guia. Ferramentas computacionais para esse fim permitem a identificação de sequências-alvo específicas no genoma, minimizando efeitos *off-target* e maximizando a eficiência da edição. Essas ferramentas utilizam algoritmos que levam em consideração fatores como a especificidade da sequência, a presença de motivos PAM e a acessibilidade da região genômica (Hsu et al., 2013). De acordo com Li et al. (2022), a ferramenta CRISPOR é abrangente e oferece múltiplos modelos preditivos para eficiência e especificidade de RNAs guias, além de suportar centenas de organismos e diferentes nucleases Cas. A ferramenta CHOPCHOP (<https://chopchop.cbu.uib.no/>), (Labun et al., 2019) é compatível com mais de 200 genomas e oferece opções personalizadas de PAM, suportando diferentes sistemas de edição como CRISPR/Cas e TALEN. CRISPR RGEN Tools (<https://www.rgenome.net/cas-designer/>) (Park; Bae; Kim, 2015) se destaca por oferecer ferramentas específicas para edição de base e *prime editing*, que, segundo Anzalone et al. (2019), é um método de edição genômica que permite modificar sequências de DNA sem necessidades de quebras de fita dupla, utilizando 11 uma Cas9 modificada acoplada a uma transcriptase reversa e guiada por um RNA especializado (*prime editing guide RNA*), tendo alcance de mais de 100 organismos, mas pode ser limitada ao SpCas9 em casos específicos de edição. Com foco em plantas, as ferramentas CRISPR-GE (<http://skl.scau.edu.cn/>), (Xie et al., 2017) e CRISPR-P (<http://crispr.hzau.edu.cn/CRISPR2/>), (Liu et al., 2017) atendem várias culturas agrícolas e diferentes nucleases Cas. Para a escolha da ferramenta, é importante a observação dos critérios a serem seguidos na pesquisa, buscando o software que atenda a todos os requisitos, como o suporte à nuclease e disponibilidade do genoma principal de interesse do estudo, assim como os critérios de seleção do gRNA, como a possibilidade de mudança ou não dos valores mínimos de eficiência e a avaliação de efeitos *off-target*.

2.5.2. CRISPOR

CRISPOR é uma ferramenta *web* para desenho de guias CRISPR/Cas, que automatiza a seleção de gRNAs otimizados para quase 150 genomas, incorporando modelos de predição de eficiência de edição e especificidade, além de sugerir primers para testes de *off-target* e opções de clonagem. A busca por *off-targets* é realizada com o alinhador Burrows-Wheeler, um algoritmo de alinhamento de sequências que mapeia sequências curtas de DNA contra um genoma de referência. Essa busca retorna regiões com distância especificada e filtragem conforme a PAM utilizada no desenho. As sequências selecionadas são classificadas de acordo com índices de especificidade, com o *MIT specificity score*, um modelo computacional que avalia os riscos de edições *off-target* proposto por Hsu et al. (2013). O cálculo desse *score* considera três fatores principais: (i) Posição e número de *mismatches*; (ii) Tipo de *mismatch*, com diferentes valores de pontuação a depender do nucleotídeo que difere e; (iii) Presença de PAMs alternativas (Hsu et al., 2013; Doench et al., 2016; Concordet e Haussler, 2018).

2.5.3. EMBOSS Water

O EMBOSS Water é uma ferramenta de alinhamento local de sequências baseada no algoritmo Smith-Waterman, utilizado para identificar regiões de similaridade entre sequências biológicas. Esse recurso faz parte do pacote EMBOSS (European Molecular Biology Open Software Suite), que oferece diversas ferramentas para análise de sequências gênicas e proteicas (Rice et al., 2000). No presente estudo, o EMBOSS Water foi crucial para realizar alinhamentos locais entre as sequências gênicas e proteicas de *Coffea arabica* e as espécies-modelo, permitindo a identificação de regiões de homologia e a quantificação das identidades e similaridades entre as sequências.

O algoritmo Smith-Waterman (SMITH e WATERMAN, 1981) é uma adaptação do método de programação dinâmica para alinhamento local. Diferentemente do alinhamento global, como o Needleman-Wunsch, que alinha sequências na totalidade, o primeiro busca subsequências com maiores similaridades, permitindo a identificação de domínios conservados e regiões homólogas mesmo em sequências com tamanhos muito divergentes (Madeira et al., 2024). A Figura 4 mostra a matriz de pontuação do algoritmo Smith-Waterman.

O funcionamento do algoritmo se baseia na criação de uma matriz bidimensional, na qual uma sequência é posicionada ao longo da primeira linha e a outra ao longo da primeira coluna. Inicialmente, todas as células da primeira linha e coluna são preenchidas com zeros, pois o alinhamento local pode começar em qualquer ponto. Para calcular os valores das demais

células, utiliza-se uma fórmula recursiva que compara quatro possibilidades para cada posição da matriz: (1) a pontuação vinda da diagonal superior esquerda somada ao *score* de similaridade entre os nucleotídeos ou aminoácidos comparados; (2) o maior valor obtido ao considerar um *gap* na sequência da linha; (3) o maior valor ao se inserir um *gap* na sequência da coluna; e (4) o valor zero, que representa o reinício do alinhamento (FARRAR, 2008; KHAJEH-SAEED; POOLE; PEROT, 2010). A Figura 3 demonstra a fórmula recursiva utilizada.

$$A_{i,j} = \max \begin{cases} A_{i-1,j-1} + sbt(S_0[i], S_1[j]) \\ A_{i-1,j} + G \\ A_{i,j-1} + G \\ 0 \end{cases}$$

Figura 3 - Fórmula recursiva do algoritmo de Smith-Waterman. Fonte: Figueirêdo Júnior (2021).

O valor zero evita pontuações negativas, fazendo com que o algoritmo considere apenas subsequências com similaridade relevante. Na etapa de *traceback*, o alinhamento ótimo é reconstruído a partir da célula com a maior pontuação na matriz, traçando o caminho inverso até que um valor zero seja encontrado.

	0	C	A	G	C	C	U	C	G	C	U	U	A	G
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A	0	0	5	0	0	0	0	0	0					
A	0	0	5	2	0	0	0	0	0					
U	0	0	0	2	0	0	5	0	0					
G	0	0	0	5	0	0	0	2	5					
C	0	5	0	0	10	5	0	5	0					
C	0	5	2	0	5	15	6	5	4					
A	0													
U	0													
U	0													
G	0													
C	0													
C	0													
G	0													
G	0													

Figura 4 - Matriz de pontuação do algoritmo Smith-Waterman. Fonte: Khajeh-Saeed; Poole; Perot (2010)

Os valores de similaridade utilizados na matriz são determinados por sistemas de pontuação que variam conforme o tipo de sequência. Para proteínas, a matriz BLOSUM62 é

a utilizada e foi construída com base em alinhamentos de sequências homólogas de diferentes espécies, atribuindo pontuações mais altas a substituições conservadas. No contexto do EMBOSS Water, enquanto o algoritmo calcula a melhor correspondência entre subsequências, preenchendo uma matriz com base em similaridades, penalidades por *gaps* e extensões de *gaps*, a BLOSUM62 fornece os valores específicos que indicam o quão provável é a substituição de um aminoácido por outro (HENIKOFF e HENIKOFF, 1992; ALTSCHUL, 2008).

2.5.4. JALVIEW

O Jalview é uma aplicação desenvolvida em Java voltada para a visualização, edição e análise de alinhamentos de sequências biológicas, tanto de DNA quanto de proteínas. Entre suas funcionalidades principais estão visualizar e editar alinhamentos múltiplos de sequências, avaliar a qualidade dos alinhamentos, identificar regiões conservadas com base em propriedades dos resíduos e representar essas informações de forma gráfica. Ele também oferece a construção de árvores filogenéticas com base nos alinhamentos, utilizando algoritmos de agrupamento hierárquico e de componentes principais, além de apresentar ferramentas para predição de estruturas secundárias de proteínas (Waterhouse et al., 2005; Procter et al., 2021).

3. METODOLOGIA

3.1. IDENTIFICAÇÃO DE GENES DE SUSCETIBILIDADE

A metodologia utilizada para a identificação dos genes candidatos seguiu a abordagem proposta por Cui et al. (2020), em que genes de suscetibilidade descritos em organismos modelo são utilizados como referência para a investigação em outras espécies de interesse agrônômico.

Considerando que os genes de suscetibilidade em *C. arabica* ainda não são anotados ou funcionalmente caracterizados de forma abrangente e que não se tratam necessariamente de ortólogos dos genes dos organismos utilizados neste trabalho, foi adotada uma estratégia de busca por genes com a mesma nomenclatura funcional em bancos de dados genômico públicos, como o NCBI e o Phytozome. Essa estratégia permitiu a identificação de possíveis genes com funções de suscetibilidade no cafeeiro, com base em similaridade de nome e função descrita, mesmo na ausência de uma relação ortóloga direta.

Para isso, foram selecionados 4 organismos com genes de suscetibilidade previamente estudados, sendo eles: *Solanum lycopersicum*, *Nicotiana tabacum*, *Capsicum annuum* e *Arabidopsis thaliana*.

3.2. ALINHAMENTO DE SEQUÊNCIAS

As sequências identificadas em *Coffea arabica* foram alinhadas com suas correspondentes nos organismos selecionados utilizando o software EMBOSS Water (Madeira et al., 2024) (https://www.ebi.ac.uk/jdispatcher/psa/emboss_water), que implementa o algoritmo de Smith-Waterman para alinhamento local. Cada gene do cafeeiro foi alinhado individualmente com o seu correspondente identificado nas outras espécies de comparação. No alinhamento das sequências, foi utilizada a matriz de substituição EDNAFULL, com penalidade de *gap* igual a 10 e penalidade de extensão de *gap* igual a 0,5.

Os resultados de identidade, similaridade, *gaps* e *scores* dos alinhamentos foram salvos localmente e utilizados para visualização e análise no *software* Jalview, em que foram coloridos com base na porcentagem de identidade nucleotídica, facilitando a identificação de regiões conservadas.

Para a análise das proteínas codificadas pelos genes candidatos, foi realizada uma etapa de alinhamento das sequências de aminoácidos, com o objetivo de avaliar a conservação estrutural e funcional entre as espécies estudadas. As sequências proteicas foram obtidas a partir das proteínas relacionadas aos respectivos genes no banco de dados NCBI. Essas sequências foram inicialmente comparadas por meio de alinhamentos par-a-par, também utilizando a

ferramenta EMBOSS Water. Esses alinhamentos utilizaram os seguintes parâmetros: penalidade de *gap* igual a 10 e penalidade de extensão de *gap* igual a 0,5 e matriz de substituição EBLOSUM62, padrão da aplicação.

Após a análise individual, as proteínas foram submetidas a um alinhamento múltiplo local no *software* Jalview, o que permitiu a visualização simultânea das regiões com maior conservação entre as sequências de aminoácidos.

3.3. DESENHO E SELEÇÃO DOS gRNAs

O desenho dos RNAs guias foi conduzido a partir das sequências previamente selecionadas como candidatas a genes de suscetibilidade. Utilizou-se o sistema CRISPOR (Concordet e Haussler, 2018) (<https://crispor.gi.ucsc.edu/crispor.py>) para a geração dos gRNAs, tendo como referência o genoma “*Coffea arabica* - Cara 1.0”. O sistema de edição adotado foi o baseado na endonuclease spCas9, com sequência PAM do tipo NGG e espaçador de 20 pares de base. Para cada sequência-alvo, a ferramenta retornou diversas opções de RNAs guias, que foram submetidas a uma etapa de filtragem baseada em critérios específicos de qualidade e eficiência.

Foi estabelecido um conteúdo GC entre 40% e 60%, como indicado por Konstantakos et al. (2022), para garantir estabilidade da molécula e eficiência de ligação à sequência-alvo. Além disso, foi adotado como valor mínimo de especificidade o *MIT Specificity Score* ≥ 80 , com intenção de manter a alta especificidade e minimizar a ocorrência de cortes em regiões não desejadas do genoma.

A fim de facilitar o processamento e análise de dados obtidos na plataforma CRISPOR, foi desenvolvido um código em linguagem de programação Python, responsável por realizar a mineração automática dos RNAs guias, filtrando as sequências pelos parâmetros de conteúdo guanina/citosina e valor de especificidade. Com isso, foi gerado um relatório contendo as sequências organizadas para posterior consulta e análise. O código se encontra no Apêndice A deste trabalho. Este *script* em Python, executado no ambiente Google Colab, processou um arquivo .tsv de saída da plataforma CRISPOR contendo informações sobre gRNAs candidatos. Utilizando a biblioteca pandas, o *script* extraiu as colunas *targetSeq* e *mitSpecScore*, e implementou uma função para calcular o percentual de conteúdo GC de cada sequência. Os dados foram então exportados para um novo arquivo .csv para posterior análise.

4. RESULTADOS E DISCUSSÃO

A análise comparativa entre genes de *Coffea arabica* e os genes com provável designação funcional entre as espécies vegetais *Arabidopsis thaliana*, *Capsicum annuum*, *Nicotiana tabacum* e *Solanum lycopersicum* foi realizada através de comparações pareadas por alinhamento de sequências.

A identidade nucleotídica entre os genes revelou variações significativas entre os três genes candidatos. No alinhamento com o gene MLO-like do cafeeiro, *Nicotiana tabacum* apresentou a maior identidade (41,1%) e o maior *score* (6256,5), indicando maior conservação da sequência e um alinhamento mais robusto, em comparação com os outros organismos. Em seguida, destacam-se os alinhamentos com *Capsicum annuum* (40,9% de identidade, *score* 8629,0), *Solanum lycopersicum* (40,0%, *score* 5850,5) e *Arabidopsis thaliana* (39,8%, *score* 4171,5). O maior valor de identidade e a menor proporção de *gaps* no alinhamento entre o gene MLO-like de *C. arabica* e *N. tabacum* sugerem que o gene da espécie-modelo pode compartilhar maior similaridade funcional e estrutural com o gene do cafeeiro. A Figura 5 mostra uma região do alinhamento, entre as posições 7390 e 7450 onde há similaridade maior em relação ao alinhamento total, em que a espécie que mais se difere é *Coffea arabica*, apesar de apresentar identidade com todas as outras espécies em diversas posições.

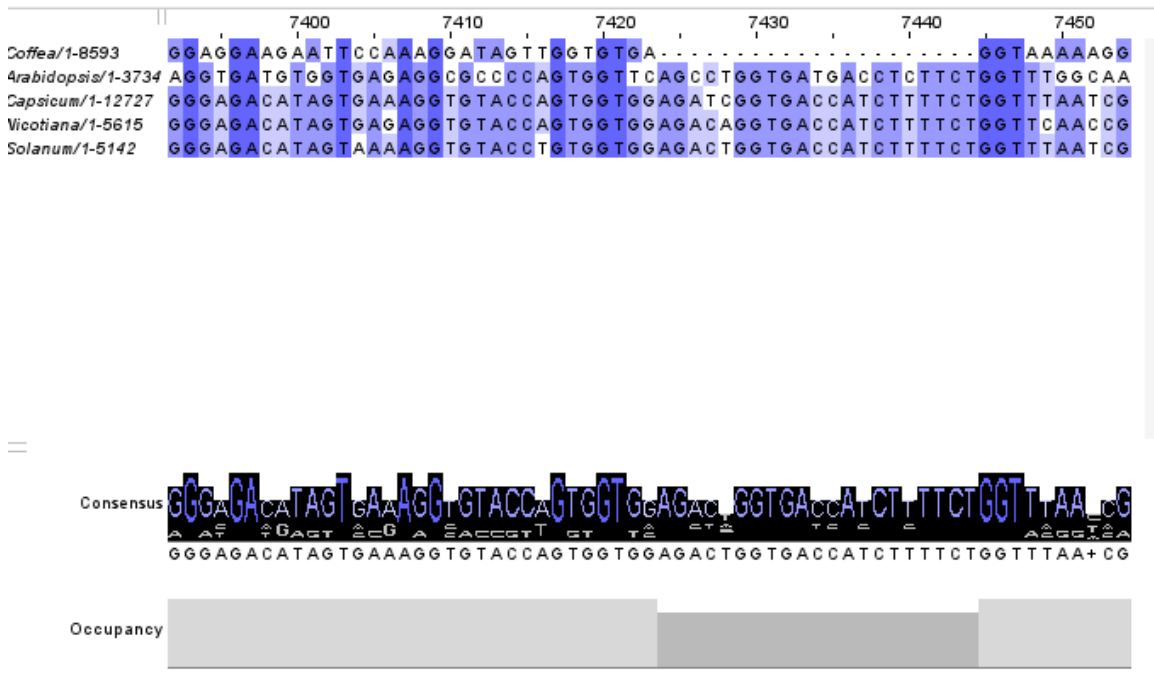


Figura 5 - Visualização do alinhamento local do gene MLO-like. Fonte: Autoria própria.

Para o gene DMR6-like, os resultados também apontam *Nicotiana tabacum* como o organismo com o gene correspondente mais similar entre as 4 espécies em estudo, com 44,4%

de identidade e *score* de 1581,0. *Capsicum annuum* apresentou 42,8% de identidade (*score* 1761,5), seguido de *Arabidopsis thaliana* com 39,2% (*score* 1777,5), e *Solanum lycopersicum*, com apenas 34,1% de identidade (*score* 1784,0), o menor valor entre os pares.

A figura 6 ilustra o alinhamento dos genes DMR6-like entre as espécies comparadas, em que se observa uma conservação razoável entre as espécies nas posições 630 a 650, com blocos alinhados e ricos em adeninas e timinas (A e T). A sequência consenso é bem definida nessas regiões, reforçando a conservação

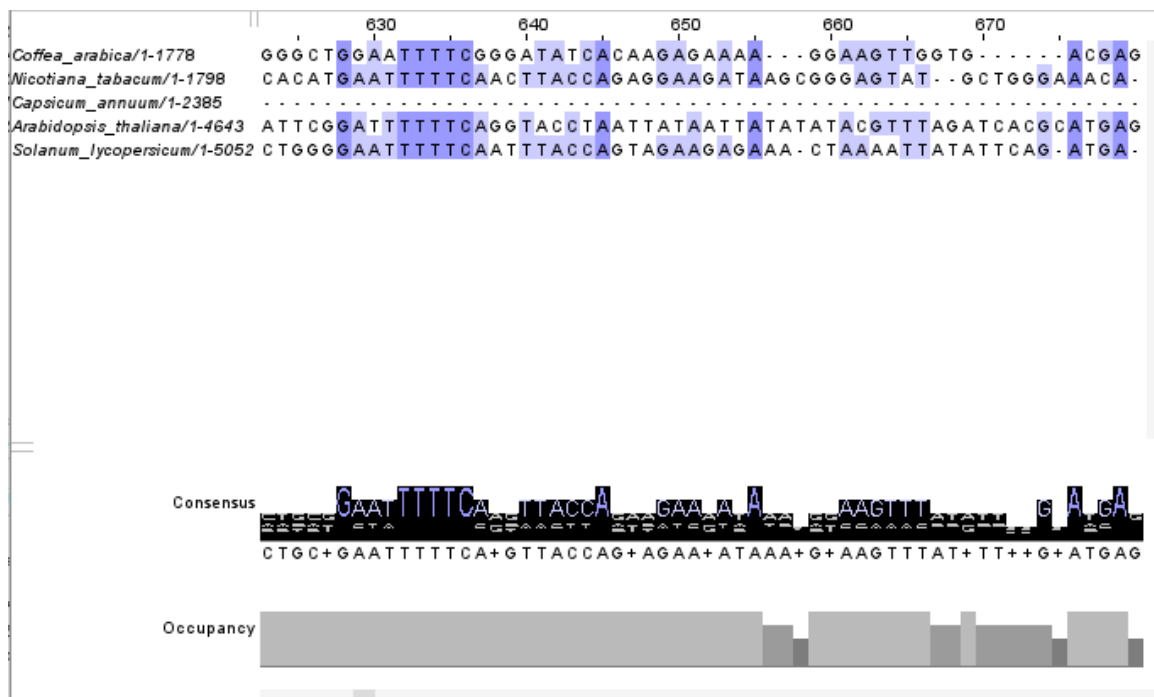


Figura 6 - Visualização do alinhamento local do gene DMR6-like. Fonte: Autoria própria.

No terceiro gene avaliado, da família das proteínas transportadoras de lipídios (non-specific lipid-transfer protein 1-like), novamente *Nicotiana tabacum* apresentou os melhores resultados, com 46,0% de identidade e *score* de 1347,0. *Arabidopsis thaliana* mostrou valor próximo (45,8%, *score* 980,5), mas o alinhamento com o tabaco apresentou menor proporção de *gaps* e maior extensão. *Capsicum annuum* (38,8%, *score* 1454,5) e *Solanum lycopersicum* (37,2%, *score* 723,0) ficaram notavelmente abaixo, o que indica maior divergência evolutiva para esse gene específico. Os valores de similaridade foram idênticos aos de identidade em todos os alinhamentos. Portanto, *N. tabacum* se destaca como a espécie mais próxima entre as analisadas para estudos a respeito da função de suscetibilidade do gene MLO-like, DMR6-like e nsLTP-like em cafeeiro. Os valores de identidade dos alinhamentos dos genes são apresentados na Tabela 1.

A Figura 7 mostra blocos conservados entre as posições 230 e 270, com destaque para sequências ricas em citosina e guanina, o que pode indicar uma região codificante estável. A sequência consenso é forte refletindo alta identidade entre as espécies analisadas nessa região do alinhamento.

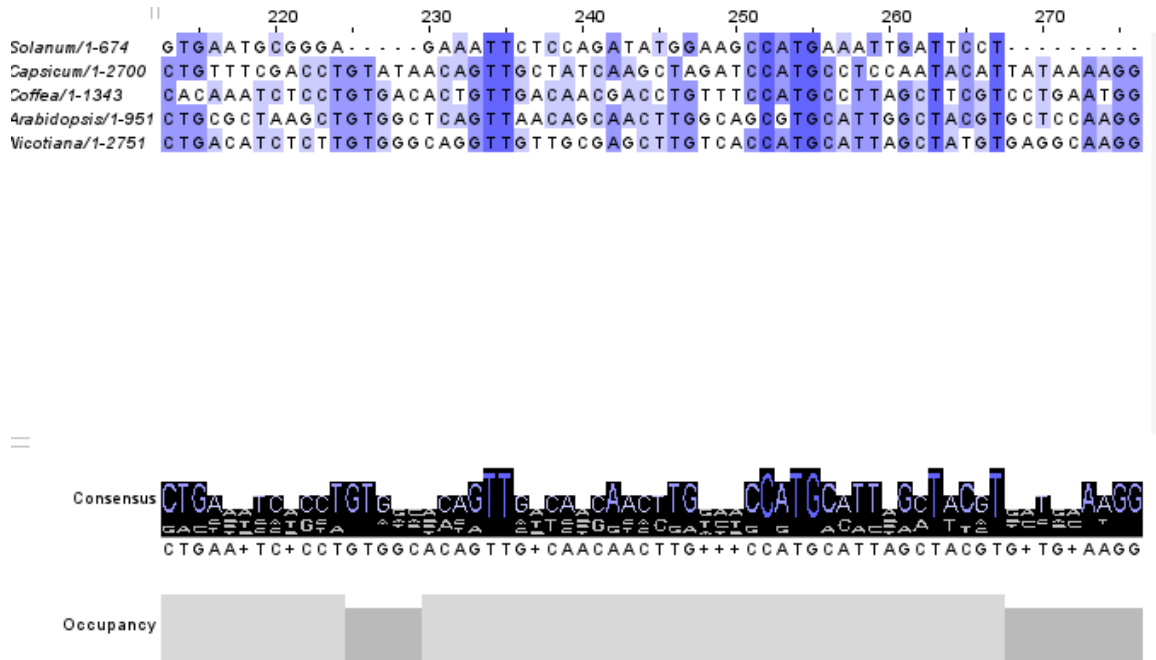


Figura 7 - Visualização do alinhamento local do gene nsLTP-like.

Tabela 1 - Identidade dos alinhamentos nucleotídicos entre genes de *Coffea arabica* e seus correspondentes em outras espécies.

Gene (<i>Coffea arabica</i>)	ID NCBI	Organismo comparado	ID NCBI	Identidade (%)
LOC113688250 MLO-like protein 11	113688250	<i>Solanum lycopersicum</i>	778228	40,0%
		<i>Capsicum annuum</i>	107871340	40,9%
		<i>Arabidopsis thaliana</i>	837673	39,8%
		<i>Nicotiana tabacum</i>	107791015	41,1%
LOC113741652 protein DMR6- LIKE OXYGENASE 2	113741652	<i>Solanum lycopersicum</i>	100736526	34,1%
		<i>Capsicum annuum</i>	107840347	42,8%
		<i>Arabidopsis thaliana</i>	832524	39,2%
		<i>Nicotiana tabacum</i>	107812220	44,4%
LOC113741701 non-specific lipid-transfer protein 1-like	113741701	<i>Solanum lycopersicum</i>	101258342	37,2%
		<i>Capsicum annuum</i>	107867449	38,8%
		<i>Arabidopsis thaliana</i>	836051	45,8%
		<i>Nicotiana tabacum</i>	107779117	46,0%

Fonte: Autoria própria.

A análise comparativa entre as sequências proteicas dos três genes candidatos de *Coffea arabica* e suas respectivas espécies de referência revelou diferentes padrões de conservação, tanto em termos de identidade quanto de similaridade.

Para a proteína MLO-like, o maior valor de identidade observado foi com *Capsicum annuum* (36,0%), seguido por *Solanum lycopersicum* (35,9%), *Nicotiana tabacum* (34,7%) e *Arabidopsis thaliana* (34,5%). A maior similaridade também foi com *Capsicum annuum* (54,4%), evidenciando que, além de maior conservação de resíduos idênticos, há uma manutenção considerável de resíduos com propriedades físico-químicas semelhantes.

Em contrapartida, a proteína DMR6-like apresentou baixos níveis de identidade e similaridade variável, sendo os melhores valores obtidos com *Solanum lycopersicum* (22,0% de identidade; 40,4% de similaridade). Os menores valores foram com *Capsicum annuum* (21,4%/34,0%) e *Nicotiana tabacum* (21,4%/37,7%). *Arabidopsis thaliana* apresentou o menor índice de identidade (20,8%).

A proteína associada o gene nsLTP-like mostrou padrões mais variados de conservação, com o alinhamento com *Arabidopsis thaliana* se destacando, já que mostrou identidade e similaridade elevadas (66,7%) em contraste com as outras espécies sob análise, o que indica conservação da proteína, apesar da distância entre *C. arabica* e *A. thaliana*. As outras espécies também mostraram valores relevantes: *Nicotiana tabacum* (39,8% de identidade; 57,6% de similaridade), *Capsicum annuum* (27,4%/52,1%) e *Solanum lycopersicum* (23,8%/43,6%).

Tabela 2 - Identidade e similaridade dos alinhamentos proteicos entre produtos gênicos de *Coffea arabica* e seus correspondentes em outras espécies.

Proteína (<i>Coffea arabica</i>)	ID NCBI	Organismo comparado	ID NCBI	Identidade (%)	Similaridade (%)
MLO-like protein 11 isoform X2	XP_071905 295	<i>Solanum lycopersicum</i>	AOL57956	35,9%	53,5%
		<i>Capsicum annuum</i>	NP_001311939	36,0%	54,4%
		<i>Arabidopsis thaliana</i>	NP_001320358	34,5%	50,4%
		<i>Nicotiana tabacum</i>	BAV69291	34,7%	53,1%
Protein DMR6- LIKE OXYGENASE 2	XP_027125 038	<i>Solanum lycopersicum</i>	NP_001233840	22,0%	40,4%
		<i>Capsicum annuum</i>	XP_016539669	21,4%	34,0%
		<i>Arabidopsis thaliana</i>	NP_197841	20,8%	38,0%
		<i>Nicotiana tabacum</i>	XP_016492755	21,4%	37,7%
Non-specific lipid-transfer protein 1-like	XP_027125 097	<i>Solanum lycopersicum</i>	NP_001306089	23,8%	43,6%
		<i>Capsicum annuum</i>	XP_016582189	27,4%	52,1%
		<i>Arabidopsis thaliana</i>	NP_001336504	66,7%	66,7%
		<i>Nicotiana tabacum</i>	XP_016454944	39,8%	57,6%

Fonte: Autoria própria.

Quando é feita a avaliação entre os níveis de identidade gênica e os níveis de identidade e similaridade proteica, o gene MLO-like apresentou similaridade proteica acima de 50% para

todas as espécies e identidades gênicas uniformes com os valores de identidade proteica, o que pode indicar uma conservação estrutural e funcional nas proteínas, baseada nas substituições conservativas de aminoácidos. A ordem das espécies em termos de identidade também se manteve de forma coerente entre as duas análises, indicando uma relação positiva entre os resultados gênicos e proteicos.

A Figura 8 ilustra uma região do alinhamento entre os aminoácidos da proteína MLO-like em que há uma alta conservação entre as 5 espécies vegetais analisadas, especialmente entre as posições 420 e 440, onde existe alta identidade. Fora dessa região central, há trechos com maior variabilidade, o que pode indicar regiões menos críticas ou sujeitas a diversificação funcional. O alinhamento sugere um possível domínio funcional na região.

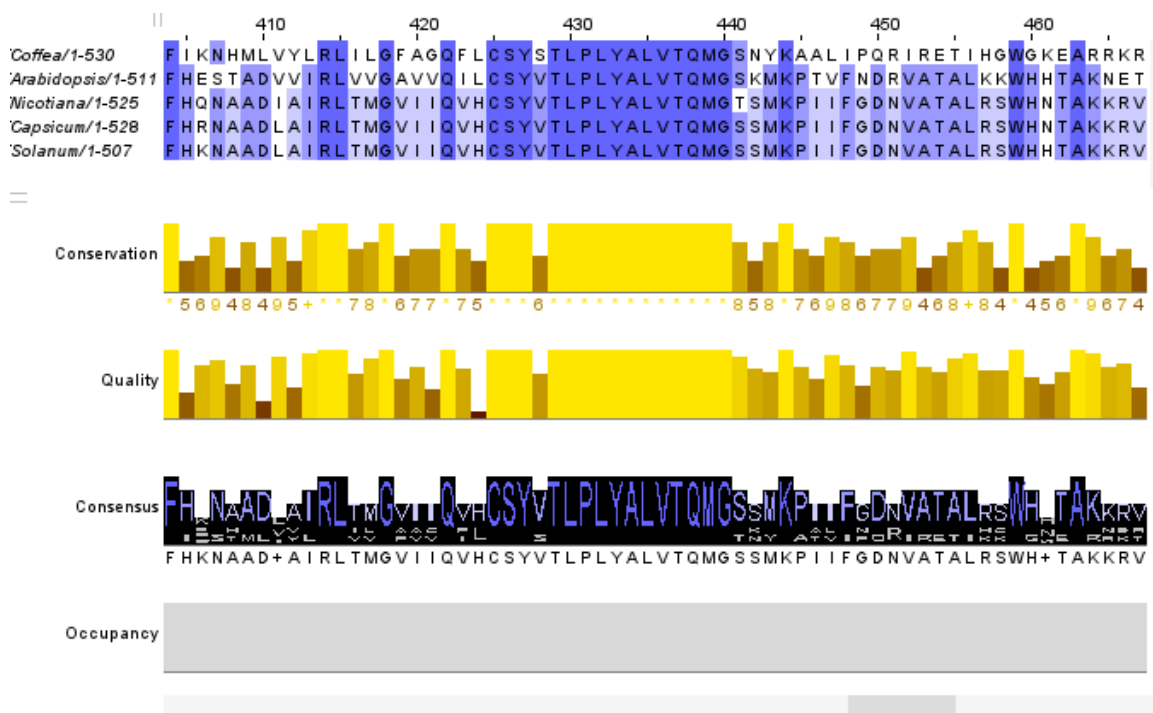


Figura 8 - Visualização do alinhamento local da proteína MLO-like. Fonte: Autoria própria.

Na Figura 9, é possível visualizar que a região entre as posições 340 e 365 mostram conservação razoável, e grande parte das sequências estão completas nesta região da proteína DMR6-like. Já no alinhamento da proteína LTP-like, a conservação é baixa, exceto por poucos resíduos específicos, como mostra a Figura 10. Isso indica baixa probabilidade de domínio funcional conservado na região.

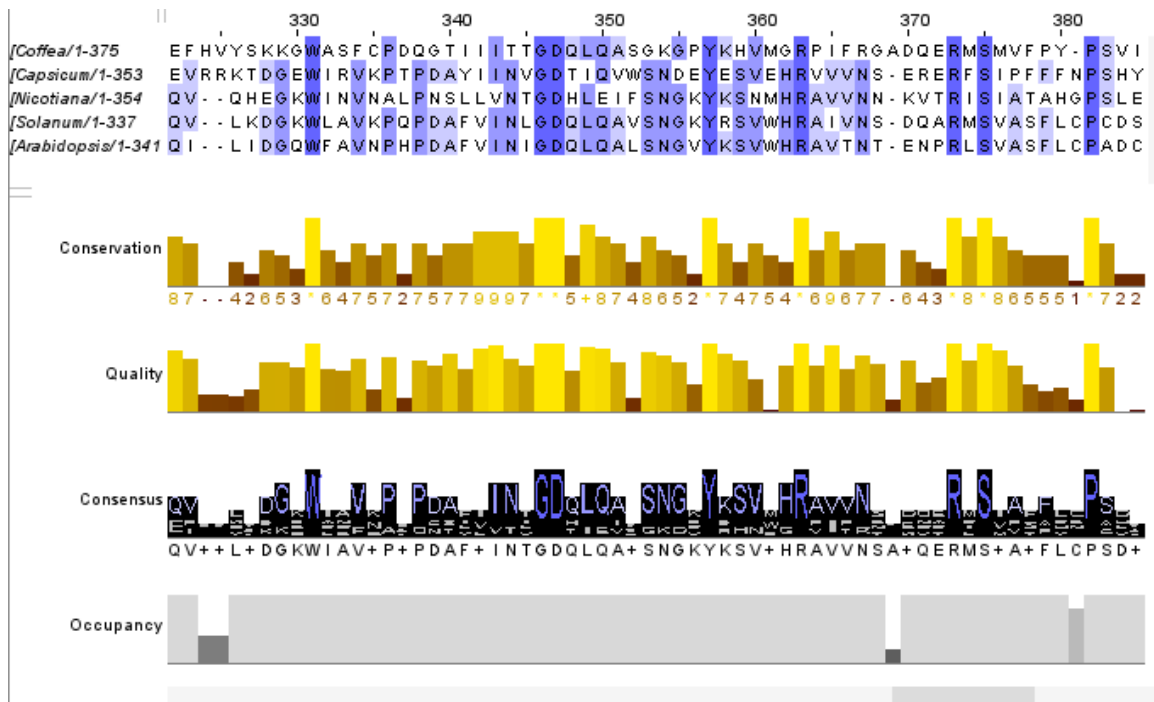


Figura 9 - Visualização do alinhamento local da proteína DMR6-like. Fonte: Autoria própria.

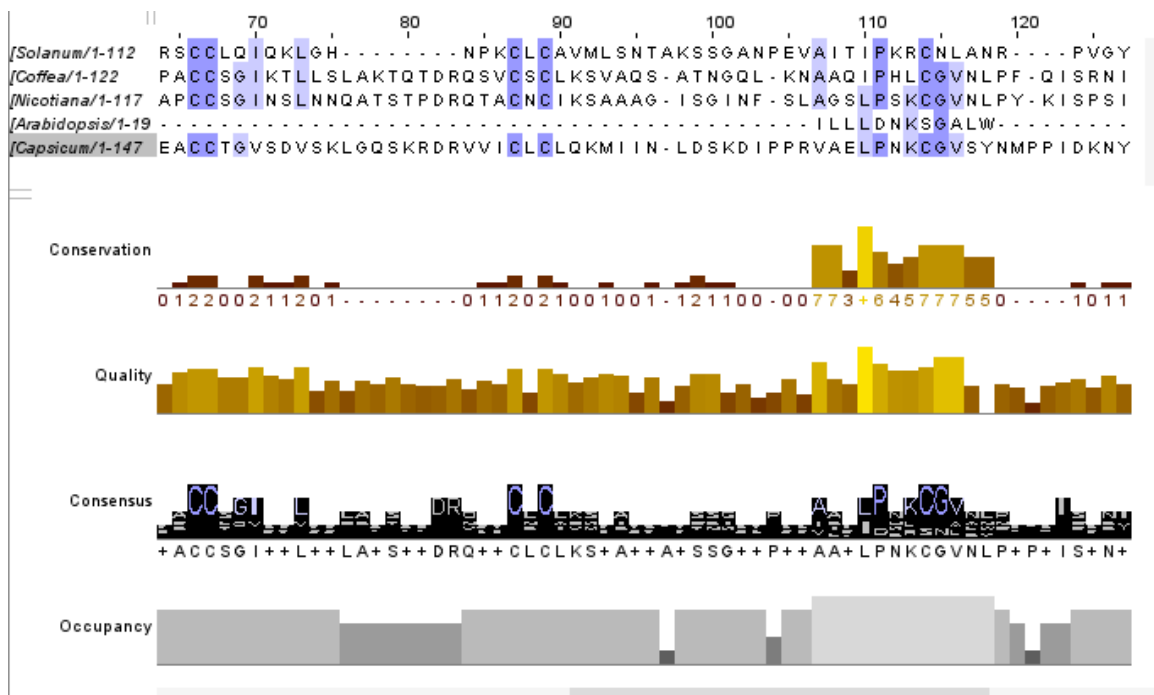


Figura 10 - Visualização do alinhamento local da proteína nsLTP-like. Fonte: Autoria própria.

Entretanto, o padrão observado no gene DMR6-like foi diferente, pois, apesar de a identidade gênica ter atingido até 44,4% em *N. tabacum*, os valores de identidade proteica foram consideravelmente mais baixos, variando apenas de 20,8% a 22,0%. A similaridade entre proteínas também não apresentou valores muito altos, ficando entre 34,0% e 40,4%. Isso sugere

que, embora haja identidade nucleotídica, as substituições de aminoácidos nas regiões codificantes divergem de forma mais acentuada, com muitas delas não sendo conservativas.

O gene non-specific lipid-transfer protein 1-like (nsLTP-like) apresentou uma divergência interessante entre as análises nucleotídica e proteica: os valores de identidade e similaridade proteica em *Arabidopsis thaliana* foram notavelmente mais elevados do que a identidade nucleotídica. Isso pode indicar que, apesar das diferenças na sequência de DNA, as regiões codificantes que geram a proteína podem ser conservadas.

O desenho dos RNAs guias gerados através dos genes MLO-like, DMR6-like e nsLTP-like em cafeeiro gerou sequências que atendem aos parâmetros de determinação de guias com provável eficiência para edição gênica (Tabela 3).

Para o gene MLO-like, seis gRNAs foram avaliados, todos apresentando *scores* MIT superiores a 90, o que indica alta especificidade. O gRNA 90forw se destaca com um *score* de 95 e GC de 48%, estando dentro da faixa ideal para estabilidade e eficiência. De forma semelhante, os gRNAs 237forw (*score* 93, GC 43%), 1355rev (*score* 92, GC 48%), 1426rev (*score* 92, GC 52%) e 456rev (*score* 92, GC 52%) também apresentam bons parâmetros.

O gene DMR6-like apresenta cinco gRNAs com *scores* entre 95 e 99, sendo os mais altos entre os avaliados, com destaque para a sequência 1210rev, com *score* MIT de 99 e GC de 43%, indicando altíssima especificidade, embora o conteúdo GC um pouco mais baixo sugira menor estabilidade. Os guias 1203rev (*score* 98, GC 43%) e 1202rev (*score* 96, GC 48%) possuem sequências bastante semelhantes, com apenas pequenas variações, e ambos apresentam valores altos de especificidade. Já as sequências 1248forw e 97forw, ambas com *score* 96 e 95 respectivamente, também oferecem GC adequados (43% e 48%), reforçando o gene DMR6-like como um alvo promissor para edição com CRISPR/Cas9.

Por sua vez, os RNAs guias para o gene nsLTP-like apresentam *scores* mais baixos quando comparados com os anteriores, sendo as sequências 379rev (*score* 89, GC 48%) e 385rev (*score* 87, GC 48%) as com maior favoráveis entre as avaliadas para o gene. O gRNA 380rev, com *score* MIT de 82 e conteúdo GC de 43%, apresenta o desempenho mais fraco dentre todos os avaliados, o que pode comprometer sua confiabilidade em sistemas CRISPR.

Tabela 3 - RNAs guias propostos para a edição dos genes MLO-like, DMR6-like e nsLTP-like em *Coffea arabica*.

Gene	ID gRNA	Sequência	Score MIT	%CG
MLO-like	90forw	AGGTTGGACTGCACTAGTATTGG	95	48
	237forw	CAACTGAACACATAAGAAGACTGGG	93	43
	1355rev	CCGAACAGAAGAACAACAAGTAGG	92	48
	1426rev	CTCTACTAGGCTTCCCTACTAGG	92	52
	456rev	TTCCGCACCAAATAGTGGCTGG	92	52
DMR6-like	1210rev	AGCTAGTTACGGAGTTACTATGG	99	43
	1203rev	TACGGAGTTACTATGGTCTATGG	98	43
	1202rev	ACGGAGTTACTATGGTCTATGGG	96	48
	1248forw	TAGCTCTTTTAAAGTACGCCTGG	96	43
	97forw	GAAAACCTCATGAGTACACCGG	95	48
LTP-like	379rev	TGTTGCGAGAAATCTGGAATGGG	89	48
	385rev	AGTCGATGTTGCGAGAAATCTGG	87	48
	380rev	ATGTTGCGAGAAATCTGGAATGG	82	43

Fonte: Autoria própria.

5. CONSIDERAÇÕES FINAIS

Este trabalho possibilitou a identificação e análise de genes candidatos à suscetibilidade à ferrugem-do-cafeeiro (*Hemileia vastatrix*) em *Coffea arabica*, por meio de abordagens comparativas com espécies modelo e ferramentas bioinformáticas. Os genes das famílias MLO-like, DMR6-like e nsLTP-like apresentaram níveis distintos de conservação gênica e proteica, com destaque para *Nicotiana tabacum* como espécie com maior similaridade funcional em todos os casos. A análise dos RNAs gerados a partir desses genes demonstrou alta especificidade, especialmente para os genes MLO-like e DMR6-like, reforçando seu potencial como alvos viáveis para edição por CRISPR/Cas9. É fundamental que esses resultados sirvam de base para futuras validações experimentais *in vitro* e *in vivo*, com o intuito de confirmar a eficácia dos gRNAs projetados e elucidar a funcionalidade dos S-genes identificados no contexto da resistência à ferrugem-do-cafeeiro. A abordagem apresentada demonstra a viabilidade do uso de CRISPR para edição de genes de suscetibilidade em *Coffea arabica*, preenchendo uma lacuna até então não conhecidamente abordada na literatura. A originalidade se deve aos desenhos de gRNAs especificamente para a edição de genes associados à infecção por patógeno, e a ausência de trabalhos publicados com esse foco reforça a contribuição desse estudo para o avanço da edição gênica no cafeeiro.

Novas abordagens, como a identificação de genes diferencialmente expressos durante a infecção por *Hemileia vastatrix*, têm revelado alvos promissores para edição gênica em *Coffea arabica*, como os genes GSTs, GPx, SOD e ALDOA, cuja inativação por CRISPR/Cas9 pode contribuir para resistência à ferrugem (Santos et al., 2025). Além disso, estudos de associação genômica ampla (GWAS - *Genome-Wide Association Study*) permitiram mapear SNPs associados a características agronômicas e resistência, viabilizando a seleção assistida por marcadores e otimizando programas de melhoramento (Silva et al., 2024). Esses estudos abrem novas possibilidades para auxiliar programas de melhoramento genético do cafeeiro, permitindo a integração entre ferramentas de bioinformática, edição genômica e seleção assistida por marcadores. A validação funcional dos genes identificados, aliada ao avanço de tecnologias como CRISPR, pode acelerar o desenvolvimento de cultivares com resistência durável à ferrugem-do-cafeeiro e características agronômicas superiores, promovendo ganhos significativos em produtividade e estabilidade da produção.

6. REFERÊNCIAS

ABDELRAHMAN, M. et al. Multiplex genome-editing technologies for revolutionizing plant biology and crop improvement. **Frontiers in Plant Science**, v. 12, 2021.

ALTSCHUL, S. **Substitution Matrices**. 2008.

ALVARENGA, S. M. et al. Marcadores moleculares derivados de sequências expressas do genoma café potencialmente envolvidas na resistência à ferrugem. **Pesquisa Agropecuária Brasileira**, v. 46, n. 8, p. 890–898, 2011.

ANDRADE, F.T. et al. Obtenção de ESTs potencialmente relacionadas à resistência do cafeeiro à ferrugem por meio de análise in silico. In: **SIMPÓSIO DE PESQUISA DOS CAFÉS DO BRASIL, 5.**, 2007, Águas de Lindóia, SP. Anais. Brasília, DF: Embrapa Café, 2007.

ALLEN, F. et al. Predicting the mutations generated by repair of Cas9-induced double-strand breaks. **Nature Biotechnology**, v. 37, n. 1, p. 64–72, 2019.

ANZALONE, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. **Nature**, v. 576, p. 149–157, 2019.

APPIANO, M. et al. Identification of candidate MLO powdery mildew susceptibility genes in cultivated Solanaceae and functional characterization of tobacco NtMLO1. **Transgenic Research**, v. 24, 2015.

ASSOCIAÇÃO BRASILEIRA DA INDÚSTRIA DE CAFÉ. **Indicadores da indústria de café – 2023**.

BENSON, D. A. et al. GenBank. **Nucleic Acids Research**, v. 41, n. D1, p. D36–D42, 2013.

BREITLER, J. C. et al. CRISPR/Cas9-mediated efficient targeted mutagenesis has the potential to accelerate the domestication of *Coffea canephora*. **Plant Cell, Tissue and Organ Culture**, v. 134, p. 383–394, 2018.

CAMPOS, M. A. et al. ESTs de *Coffea arabica* do tipo genes R classe 3 identificadas no CafEST. In: **SIMPÓSIO DE PESQUISA DOS CAFÉS DO BRASIL**, 5., 2007, Águas de Lindóia. *Anais...* Brasília, DF: Embrapa-Café, 2007.

CASARIN, T. et al. Multiplex CRISPR/Cas9-mediated knockout of the phytoene desaturase gene in *Coffea canephora*. **Scientific Reports**, v. 12, p. 17270, 2022.

CHO, S. W. et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. **Genome Research**, v. 24, n. 1, p. 132-141, 2014.

CHOI, S. H. et al. Development of late-bolting plants by CRISPR/Cas9-mediated genome editing from mesophyll protoplasts of lettuce. **Plant Cell Reports**, v. 41, p. 1627–1630, 2022

COMPANHIA NACIONAL DE ABASTECIMENTO. **Acompanhamento da safra brasileira - Café**. 2024.

CONCORDET, J.; HAEUSSLER, M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. **Nucleic Acids Research**, v. 46, 2018.

CONSELHO DOS EXPORTADORES DE CAFÉ DO BRASIL. **Exportações brasileiras de café – dados estatísticos**.

CUI, L. et al. Improving Pathogen Resistance by Exploiting Plant Susceptibility Genes in Coffee (*Coffea* spp.). **Agronomy**, v. 10, n. 12, p. 1928, 2020.

DÉCHAMP, E. et al. Coffee (*Coffea arabica* L.). In: WANG, K. (Ed.). **Agrobacterium Protocols**. Methods in Molecular Biology, v. 1224. New York: Springer, 2015.

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

DOENCH, J. G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of. **Nature biotechnology**, v. 34, 2016.

FARRAR, M. S. Optimizing Smith-Waterman for the Cell Broadband Engine. **Bioinformatics**, v. 23, p. 156–161, 2008.

FENG, Z. et al. Efficient genome editing in plants using a CRISPR/Cas system. **Cell Research**, v. 23, p. 1229-1232, 2013.

FERNANDÉZ, D. et al. 454-pyrosequencing of *Coffea arabica* leaves infected by the rust fungus *Hemileia*. **Molecular plant pathology**, v. 13, 2012.

FERRÃO, L. F. V. et al. New EST–SSR markers of *Coffea arabica*: transferability and application to studies of molecular characterization and genetic mapping. **Molecular Breeding**, v. 35, n. 1, p. 31, 2015.

FERREIRA, P.; CHOUPINA, A. B. CRISPR/Cas9: a simple, inexpensive and effective technique for gene editing. **Molecular Biology Reports**, v. 49, n. 7, p. 7079-7086, 2022.

FIGUEIRÊDO JUNIOR, M. A. C. **Comparação Paralela de Sequências Biológicas em Múltiplas GPUs com Descarte de Blocos e Estratégias de Distribuição de Carga**. 2021. 172 p. Tese (Doutorado) – Universidade de Brasília, Brasília, 2021.

FISTER, A. S. et al. Transient expression of CRISPR/Cas9 machinery targeting *TcNPR3* enhances defense response in *Theobroma cacao*. **Frontiers in Plant Science**, v. 9, p. 268, 2018.

FLOREZ VARON, J. C. **Transcriptoma do cafeeiro (*Coffea arabica* L.) durante a interação com *Hemileia vastatrix* Berk. & Br.** 2017. 56 f. Tese (Doutorado em Genética e Melhoramento) - Universidade Federal de Viçosa, Viçosa. 2017.

GALLUCCI, F. P. **O uso do sistema CRISPR-Cas9 para silenciamento da MMP9 no tratamento do câncer de bexiga**. 2023. Tese (Doutorado) – Universidade de São Paulo, São Paulo, 2023.

GARCIA-RUIZ, H; SZUREK, B.; VAN DEN ACKERVEKEN, G. Stop helping pathogens: engineering plant susceptibility genes for durable. **Current opinion in biotechnology**, v. 70, 2021.

GAO, S. et al. LTP3 contributes to disease susceptibility in Arabidopsis by enhancing abscisic. **Molecular plant pathology**, v. 17, 2016.

GUERRA-GUIMARÃES, L. et al. Coffee leaf rust resistance: an overview. In: INGELBRECHT, I. L. W.; SILVA, M. do C. L. da; JANKOWICZ-CIESLAK, J. (Org.). **Mutation breeding in coffee with special reference to leaf rust: protocols**. Berlin; Heidelberg: Springer Berlin Heidelberg, 2023. p. 19–38.

GUYOT, R. et al. Ancestral synteny shared between distantly-related plant species from the asterid (*Coffea canephora* and *Solanum* Sp.) and rosid (*Vitis vinifera*) clades. **BMC Genomics**, v. 13, n. 1, p. 103, 2012.

HAEUSSLER, M. et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. **Genome Biology**, v. 17, n. 1, p. 148, 2016.

HENIKOFF, S.; HENIKOFF, J. G Amino acid substitution matrices from protein blocks. **Proceedings of the National Academy of Sciences of the United States of America**, v. 89, 1992.

HSU, P. D. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. **Nature Biotechnology**, v. 31, p. 827–832, 2013.

ISHINO, Y. et al. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, e identificação do gene product. **Journal of Bacteriology**, v. 169, p. 5429-5433, 1987.

JACOTT, C. N. et al. Unmasking mildew resistance locus O. **Trends in plant science**, v. 26, n. 10, p. 1006-1013, 2021.

Ji, J. et al. Plant SWEET Family of Sugar Transporters: Structure, Evolution and Biological. **Biomolecules**, v. 12, 2022.

JINEK, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. **Science**, v. 337, n. 6096, p. 816–821, 2012.

KANEHISA, M. et al. KEGG: new perspectives on genomes, pathways, diseases and drugs. **Nucleic Acids Research**, v. 45, n. D1, p. D353–D361, 2017.

KHAJEH-SAEED, A.; POOLE, S.; PEROT, B. J. Acceleration of the Smith–Waterman algorithm using single and multiple graphics processors. **Journal of Computational Physics**, v. 229, n. 11, p. 4247–4258, 2010. DOI: 10.1016/j.jcp.2010.02.009.

KONSTANTAKOS, V. et al. CRISPR–Cas9 gRNA efficiency prediction: an overview of predictive tools and the role of deep learning. **Nucleic Acids Research**, v. 50, n. 7, p. 3616–3637, 2022.

LABUN, K. et al. CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. **Nucleic Acids Research**, v. 47, n. W1, p. W171–W174, 2019.

LASHERMES, P. et al. Molecular characterisation and origin of the *Coffea arabica* L. genome. *Molecular & general genetics*: **MGG**, v. 261, 1999.

LESK, A. M. **Introduction to Bioinformatics**. 4. ed. Oxford: Oxford University Press, 2013.

LI, C. et al. Computational tools and resources for CRISPR/Cas genome editing. **Genomics, Proteomics & Bioinformatics**, v. 21, n. 1, p. 108–126, 2023.

LIN, C. et al. Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts. **Theoretical and Applied Genetics**, v. 112, n. 1, p. 114–130, 2005.

LIU, H. et al. CRISPR-P 2.0: an improved CRISPR-Cas9 tool for genome editing in plants. **Molecular Plant**, v. 10, n. 3, p. 530-532, 2017.

MADEIRA, F. et al. The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. **Nucleic Acids Research**, v. 52, 2024.

MAIA, T. et al. The *Hemileia vastatrix* effector HvEC-016 suppresses bacterial blight symptoms in. **The New phytologist**, v. 213, 2017.

MANGHWAR, H. et al. CRISPR/Cas systems in genome editing: methodologies and tools for sgRNA design, off-target evaluation, and strategies to mitigate off-target effects. **Advanced Science**, v. 7, n. 6, p. 1902312, 2020.

MEKBIB, Y. et al. Whole-genome resequencing of *Coffea arabica* L. (Rubiaceae) genotypes identify SNP and unravels distinct groups showing a strong geographical pattern. **BMC Plant Biology**, v. 22, n. 1, p. 69, 2022.

MEDRANO, J.F. et al. De novo whole-genome assembly and annotation of *Coffea arabica* var. Geisha, a high-quality coffee variety from the primary origin of coffee. **G3 Genes|Genomes|Genetics**, v. 15, 2024.

MOLLA, K. A.; KARMAKAR, S.; ISLAM, M. T. Wide Horizons of CRISPR-Cas-Derived Technologies for Basic Biology, Agriculture, and Medicine. In: MOLLA, K. A.; KARMAKAR, S.; ISLAM, M. T. (Ed.). **CRISPR-Cas Methods**. New York: Springer, 2020. p. 1–23.

MOLINARI, H. B. C. et al. (Eds.). **Tecnologia CRISPR na edição genômica de plantas: biotecnologia aplicada à agricultura**. Brasília: Embrapa, 2020.

MOSHARAF, P. M. et al. Genome-wide identification, characterization and phylogenetic analysis of Dicer-like (DCL) gene family in *Coffea arabica*. **Bioinformatics**, v. 15, n. 11, p. 824–831, 2019.

MOUNT, D. W. **Bioinformatics: Sequence and Genome Analysis**. 2. ed. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, 2004.

OZ, M. T. et al. CRISPR/Cas9-mediated multi-allelic gene targeting in sugarcane confers herbicide tolerance. **Frontiers in Genome Editing**, v. 3, p. 673566, 2021.

PARK, J.; BAE, S.; KIM, J. Cas-Designer: a web-based tool for choice of CRISPR-Cas9 target sites. **Bioinformatics**, v. 31, n. 24, p. 4014–4016, 2015.

PEVSNER, J. **Bioinformatics and Functional Genomics**. 3. ed. Hoboken: Wiley-Blackwell, 2015.

PRADO, G. S. et al. CRISPR technology towards genome editing of the perennial and semi-perennial. **Frontiers in plant science**, v. 14, 2023.

PROCTER J. B. et al. Correction to: Alignment of Biological Sequences with Jalview. In: KATO, K. (Org.). **Multiple Sequence Alignment: Methods and Protocols**. New York, NY: Springer US, 2021. p. C1–C1.

POLETTI, I. **Edição genômica mediada por CRISPR/Cas na era do melhoramento genético vegetal de precisão**. 2021. Trabalho de Conclusão de Curso (Bacharelado em Biotecnologia) – Universidade Federal de Pelotas, Pelotas, 2021.

PORTO, B. N et al. Genome sequencing and transcript analysis of *Hemileia vastatrix* reveal expression. **PLoS one**, v. 14, 2019.

PRAMANIK, D. et al. CRISPR/Cas9-mediated generation of pathogen-resistant tomato against Tomato Yellow Leaf Curl Virus and Powdery Mildew. **International Journal of Molecular Sciences**, v. 22, n. 4, p. 1878, 2021.

RICE, P.; LONGDEN, I.; BLEASBY, A. EMBOSS: the European Molecular Biology Open Software Suite. **Trends in genetics: TIG**, v. 16, 2000.

ROJAS-VÁSQUEZ, R.; GATICA-ARIAS, A. Use of genome editing technologies for genetic improvement of crops of tropical origin. **Plant Cell, Tissue and Organ Culture**, v. 140, p. 215–244, 2020.

RÖNSPIES, M. et al. CRISPR-Cas9-mediated chromosome engineering in *Arabidopsis thaliana*. **Nature Protocols**, v. 17, p. 1332-1358, 2022.

SALAZAR-NAVARRO, A. et al. Coffee leaf rust (*Hemileia vastatrix*) disease in coffee plants and perspectives by the disease control. **Phyton-International Journal of Experimental Botany**, v. 93, n. 5, p. 923–949, 2024.

SALOJÄRVI, J. et al. The genome and population genomics of allopolyploid *Coffea arabica* reveal the diversification history of modern coffee cultivars. **Nature Genetics**, v. 56, n. 4, p. 721–731, 2024.

SÁNCHEZ-LEÓN, S. et al. Low-gluten, nontransgenic wheat engineered with CRISPR/Cas9. **Plant Biotechnology Journal**, v. 16, p. 902-910, 2018.

SANTOS, I. R. et al. *Coffea arabica* susceptibility genes identified by proteomic analysis during coffee-*Hemileia vastatrix* interaction. **Physiological and Molecular Plant Pathology**, v. 139, 2025.

SCALABRIN, S. et al. A single polyploidization event at the origin of the tetraploid genome of *Coffea arabica* is responsible for the extremely low genetic variation in wild and cultivated germplasm. **Scientific Reports**, v. 10, n. 1, p. 4642, 2020.

SCHACHTSIEK, J.; STEHLE, F. Nicotine-free, nontransgenic tobacco (*Nicotiana tabacum* L.) edited by CRISPR-Cas9. **Plant Biotechnology Journal**, v. 17, p. 2228-2230, 2019.

SERA, G. H. et al. Resistance to Leaf Rust in Coffee Carrying SH3 Gene and others SH Genes. **Brazilian Archives of Biology and Technology**, v. 50, n. 5, p. 753-757, 2007.

SILVA, M. C. et al. An Overview of the Mechanisms Involved in Coffee-Hemileia vastatrix Interactions: Plant and Pathogen Perspectives. **Agronomy**, v. 12, n. 2, 2022.

SILVA, R. A. et al. Identification of SNP Markers and Candidate Genes Associated with Major Agronomic. **Plants (Basel, Switzerland)**, v. 13, 2024.

SHEN, M. W. et al. Predictable and precise template-free CRISPR editing of pathogenic variants. **Nature**, v. 563, p. 646–651, 2018.

SHI, J. et al. ARGOS8 variants generated by CRISPR-Cas9 improve maize grain yield under field drought stress conditions. **Plant Biotechnology Journal**, v. 15, n. 2, p. 207–216, 2017.

SMITH, T. F.; WATERMAN, M. S. Identification of common molecular subsequences. **Journal of Molecular Biology**, v. 147, n. 1, p. 195–197, 1981.

SOUZA, F. F. et al. **Características das principais variedades de café cultivadas em Rondônia**. Porto Velho: Embrapa Rondônia, 2004.

TALHINHAS, P. et al. The coffee leaf rust pathogen *Hemileia vastatrix*: one and a half centuries around the tropics. **Molecular Plant Pathology**, v. 18, p. 1039-1051, 2017.

THE UNIPROT CONSORTIUM. UniProt: a worldwide hub of protein knowledge. **Nucleic Acids Research**, v. 47, n. D1, p. D506–D515, 2019.

THOMAZELLA, D. P. T. et al. Loss of function of a DMR6 ortholog in tomato confers broad-spectrum disease resistance. **Proceedings of the National Academy of Sciences of the United States of America**, v. 118, n. 27, p. e2026152118, 2021.

TRIPATHI, J. N. et al. CRISPR/Cas9 editing of endogenous banana streak virus in the B genome of *Musa* spp. overcomes a major challenge in banana breeding. **Communications Biology**, v. 2, n. 46, 2019.

TYAGI, S. et al. Engineering disease resistant plants through CRISPR-Cas9 technology. **GM Crops & Food**, v. 12, n. 1, p. 125–144, 2020.

VAN DAMME, M.; HUIBERS, R. P.; ELBERSE, J.; VAN DEN ACKERVEKEN, G. Arabidopsis DMR6 encodes a putative 2OG-Fe(II) oxygenase that is defense-associated but required for susceptibility to downy mildew. **Plant Journal**, v. 54, n. 5, p. 785–793, jun. 2008.

VAN DER VOSSSEN, H.; BERTRAND, B.; CHARRIER, A. Next generation variety development for sustainable production of arabica coffee (*Coffea arabica* L.): a review. **Euphytica**, v. 204, p. 243–256, 2015.

VIEIRA, L. G. et al. Brazilian coffee genome project: an EST-based genomic resource. **Brazilian Journal of Plant Physiology**, 2006, 18(1), 95–108. <https://doi.org/10.1590/S1677-04202006000100008>

VIEIRA JÚNIOR, J. R. et al. **Avaliação da severidade da ferrugem (*Hemileia vastatrix*) em cafeeiros (*Coffea canephora*) cultivados em condições de sombreamento**. Porto Velho: Embrapa Rondônia, 2008.

WANG, P. et al. High efficient multisites genome editing in allotetraploid cotton (*Gossypium*). **Plant biotechnology journal**, v. 16, 2018.

WATERHOUSE, Andrew et al. Jalview: Visualization and Analysis of Molecular Sequences, Alignments, and Structures. **BMC Bioinformatics**, v. 6, n. 3, p. P28, 2005.

WWPDB CONSORTIUM. Protein Data Bank: the single global archive for 3D macromolecular structure data. **Nucleic Acids Research**, v. 47, n. D1, p. D520–D528, 2019.

XIE, X. et al. CRISPR-GE: a convenient software toolkit for CRISPR-based genome editing. **Molecular Plant**, v. 10, n. 9, p. 1246-1249, 2017.

XUEBO, W. et al. CRISPR/Cas9-mediated targeted mutagenesis of two homoeoalleles in tobacco confers resistance to powdery mildew. **Euphytica**, v. 219, n. 6, p. 67, 2023.

YASCHENKO, A. E.; ALONSO, J. M.; STEPANOVA, A. N. Arabidopsis as a model for translational research. **Plant Cell**, v. 37, n. 5, 2025.

ZAIDI, S. S.; MUKHTAR, M. S.; MANSOOR, S. Genome Editing: Targeting Susceptibility Genes for Plant Disease Resistance. **Trends in Biotechnology**, v. 36, n. 9, p. 898–906, 1 set. 2018.

ZHANG, Y. et al. CRISPR/Cas9-mediated targeted mutagenesis of GmUGT enhanced soybean resistance against leaf-chewing insects through flavonoids biosynthesis. **Frontiers in Plant Science**, v. 13, p. 802716, 2022.

ZHANG, Y. et al. S5H/DMR6 encodes a salicylic acid 5-hydroxylase that fine-tunes salicylic acid homeostasis. **Plant Physiology**, v. 175, n. 3, p. 1082–1093, 2017.

ZHANG, D. et al. CRISPR/Cas: A powerful tool for gene function study and crop improvement. **Journal of Advanced Research**, v. 29, p. 207–221, 2021.

ZHENG, Z. et al. Loss of function in Mlo orthologs reduces susceptibility of pepper and tomato to . **PloS one**, v. 8, 2013

APÊNDICE A – CÓDIGO EM PYTHON PARA MINERAÇÃO DOS RNAs GUIA

```
import pandas as pd

arquivo = 'guias_DMR6.tsv'

df = pd.read_csv(arquivo, sep='\t')

def calcular_gc(seq):
    gc = seq.count('G') + seq.count('C')
    return (gc / len(seq)) * 100

df['GC%'] = df['targetSeq'].apply(calcular_gc)

gRNAs_filtrados = df[
    (df['mitSpecScore'] >= 80) &
    (df['GC%'] >= 40) &
    (df['GC%'] <= 60)
]

colunas_para_mostrar = ['#guideId', 'targetSeq', 'mitSpecScore',
'GC%']
gRNAs_filtrados[colunas_para_mostrar]
gRNAs_filtrados[colunas_para_mostrar].to_csv('gRNAs_filtrados.csv', index=False)

files.download('gRNAs_filtrados.csv')
```