



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE BIOCÊNCIAS
DEPARTAMENTO DE GENÉTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA

VINICIUS TORRES GUERRA

**CO-EXPRESSÃO DE GENES LIGADOS EM *VIGNA UNGUICULATA* (L.) WALP
APÓS ESTÍMULO DE DESIDRATAÇÃO RADICULAR**

Recife
2021

VINICIUS TORRES GUERRA

**CO-EXPRESSÃO DE GENES LIGADOS EM *VIGNA UNGUICULATA* (L.) WALP
APÓS ESTÍMULO DE DESIDRATAÇÃO RADICULAR**

Dissertação apresentada ao Programa de Pós-Graduação em Genética da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de mestre em Genética.
Área de concentração: Biologia Molecular

Orientador (a): Éderson Akio Kido

Coorientador (a): Antônio Félix da Costa

Recife

2021

Catálogo na fonte:
Bibliotecária Claudina Queiroz, CRB4/1752

Guerra, Vinicius Torres

Co-expressão de genes ligados em *Vigna unguiculata* (L.) Walp após estímulo de desidratação radicular / Vinicius Torres Guerra - 2021.

108 folhas: il., fig., tab.

Orientador: Éderson Akio Kido

Coorientador: Antônio Félix da Costa

Dissertação (mestrado) – Universidade Federal de Pernambuco. Centro de Biociências. Programa de Pós-Graduação em Genética. Recife, 2021.

Inclui referências e apêndices.

1. Feijão-caupi 2. Transcriptômica 3. Bioinformática

I. Kido, Éderson Akio (Orientador) II. Costa, Antônio Félix da (Coorientador) III. Título

583.74 CDD (22.ed.)

UFPE/CB-2021-355

VINICIUS TORRES GUERRA

**CO-EXPRESSÃO DE GENES LIGADOS EM *VIGNA UNGUICULATA* (L.) WALP
APÓS ESTÍMULO DE DESIDRATAÇÃO RADICULAR**

Dissertação apresentada ao Programa de Pós Graduação em Genética da Universidade Federal de Pernambuco, Centro Acadêmico de Biociências, como requisito para a obtenção do título de Mestre em Genética. Área de concentração: Biologia Molecular

Aprovado em: 27/08/2021.

BANCA EXAMINADORA

Prof. Dr. Éderson Akio Kido (Orientador)
Universidade Federal de Pernambuco

Prof^a. Dr^a. Flávia Figueira Aburjaile (Examinador Externo)
Universidade Federal de Minas Gerais

Prof^o. Dr. Maurisrael De Moura Rocha (Examinador Externo)
EMBRAPA Meio-Norte

Prof^o. Dr. Tercilio Calsa Júnior (Examinador Interno)
Universidade Federal de Pernambuco

AGRADECIMENTOS

Agradeço:

Ao **Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)** pelo financiamento concedido.

Ao **Instituto Agrônomo de Pernambuco (IPA)** e a **Universidade Federal de Pernambuco (UFPE)** por proverem toda estrutura necessária.

Ao **Dr. Ederson Akio Kido** pela orientação e paciência ao longo de tantos anos.

Ao **Dr. Antônio Félix da Costa** pela extrema simpatia e solicitude.

À toda equipe do **Laboratório de Genética Molecular de Plantas (LGMP)** por todo auxílio proporcionado.

À todos os componentes da **Rede Nordeste de Biotecnologia (NordEST)**, em especial a **Dr^a. Ana Maria Benko-Iseppon**, coordenadora da rede, pela disponibilização das *unitags* SuperSAGE.

À todos que contribuíram de forma direta ou indireta para a execução do projeto

Grazie a tutti!

RESUMO

Em Ciências Agrárias, identificar alelos associados com tolerância a condições de seca será imprescindível para suprir as demandas futuras por alimentos. A natureza complexa dos *loci* controladores de caráter quantitativo (QTLs) limita a manipulação destes nos programas de melhoramento vegetal. O feijão-caupi (*Vigna unguiculata*) é uma cultura importante socioeconomicamente, principalmente em países pobres, e tem emergido como um modelo para estudos envolvendo o estresse de seca. O presente trabalho aplicou uma nova proposta afim de identificar genes candidatos para estudos associativos em feijão-caupi tendo por base a expressão gênica, a ligação física entre genes, e evidências funcionais disponíveis por *data-mining*. Para tanto, *unitags* HT-SuperSAGE de dois cultivares com respostas contrastantes a seca (tolerante e sensível), cujas raízes foram expostas ao ar (até 150 min.), foram identificadas em transcritos de referência da espécie, e aqueles de co-expressão induzida mapeados em *loci* consecutivos no genoma foram associados a 79 blocos, dos quais 21 (67 genes) apresentaram evidências funcionais, em estudos com estresses em plantas. Daqueles cujas funções já foram demonstradas por genética reversa, vários seriam potenciais marcadores moleculares funcionais, especialmente blocos com componentes das vias de sinalização por fitohormônios (ARIA, WRKY51 e OPR3) ou envolvendo respostas a estresses (CHS17 e PAL2). Espera-se que a caracterização proposta possa ser útil em programas de melhoramento do feijão-caupi e espécies correlatas, a partir da manipulação dos blocos ou dos genes individuais

Palavras-chave: Feijão-caupi; Transcriptômica; Bioinformática; Estresse Abiótico.

ABSTRACT

In Agricultural Sciences, identifying genes/alleles implicated with drought tolerance will be crucial for future food supply. The complex nature of quantitative trait loci (QTLs) limits the progress of crop breeding programs. The cowpea [*Vigna unguiculata* (L.) Walp.] is a socioeconomically relevant crop, mainly in undeveloped countries, and is an emerging model for studies covering plants under drought stress. The present work applied an improved method to identify candidate genes for cowpea associative studies, combining the gene expression profile with the physical linkage of genes supported by reverse genetic studies. Thus, two cowpea cultivars showing contrasting drought responses (tolerant and sensitive) generated unitags HT-SuperSAGE from their roots exposed to air (up to 150 min.), and these unitags were identified in cowpea reference transcripts. Those transcripts showing unitags, and presenting induced expression were mapped in the cowpea genome, allowing the identification of 79 blocks of genes linked. From these, 67 genes (from 21 blocks) showed functional evidence in plants under stress. These genes with available reverse genetics studies demonstrating their functions are potentially functional markers, especially those associated with phytohormones signaling pathways (e.g., block with genes ARIA, WRKY52, and OPR3) or involved in stress responses (e.g., block with genes CHS17 and PAL2). We hope the proposed method could be useful in cowpea breeding programs and also for its related species helping to identify candidate genes for associative studies and biotechnological gene manipulation.

Keywords: Cowpea; Transcriptomics; Bioinformatics; Abiotic Stress.

LISTA DE FIGURAS

Figura 1 –	Produção de feijão-caupi no Mundo entre os anos de 1994 à 2016	17
Figura 2 –	Possíveis respostas e efeitos da exposição a condição de seca em plantas	18
Figura 3 –	Comparação entre características monogênicas e poligênicas	19
Figura 4 –	Protocolo SuperSAGE. Retângulos em cinza indicam os sítios de reconhecimento da enzima <i>EcoP15</i>	24
Figura 5 –	Metodologias de artigos científicos encontrados na busca PubMed contendo as palavras “ <i>cowpea</i> ”, “ <i>drought</i> ” e “ <i>expression</i> ” em seu resumo e/ ou título	31
Figura 6 –	Identificação dos blocos de genes de co-expressão em cromossomos de <i>Vigna unguiculata</i>	37
Figura 7 –	Diagramas de Veen destacando números de: <i>unitags</i> (A), transcritos mRNA RefSeq (B) e <i>loci</i> (C) de <i>Vigna unguiculata</i> exclusivos e compartilhados entre duas análises (BLAST e Bowtie). Em negrito, aqueles exclusivos da estratégia Bowtie	41
Figura 8 –	Classificação das <i>unitags</i> alinhadas em sequências mRNA RefSeq do feijão-caupi	41
Figura 9 –	Quantitativos referentes a modulação da expressão pelas <i>unitags</i> e pelos transcritos contendo as <i>unitags</i> nas respostas dos cultivares sensível SI (Santo Inácio) e tolerante PO (Pingo de Ouro 1-2) de <i>Vigna unguiculata</i>	42
Figura 10 –	Caracterização dos blocos de genes ligados e co-expressos identificados no genoma de <i>Vigna unguiculata</i> , com destaque para: A) tamanho dos blocos; B) número de <i>loci</i> ; C) vetores que comportariam; D) número de blocos por cromossomo	43
Figura 11 –	Grafico de dispersão com base em valores de $\log_2FC$, dos componentes de blocos de genes co-expressos na resposta UR do cultivar tolerante Pingo de Ouro 1-2 (PO), com expressão na resposta sensível do cultivar Santo Inácio (SI), após um estímulo de desidratação radicular (até 125 minutos), sendo $\log_2FC$ a razão entre a frequência normalizada na biblioteca tratamento <i>versus</i>	44

	controle negativo	
Figura– 12	Termos mais frequentes extraídos de artigos científicos relacionando genes componentes de blocos co-expressos em <i>Vigna unguiculata</i> com estresses ambientais	45
Figura– 13	Famílias de fatores de transcrição com mais interações previstas com promotores dos genes componentes dos blocos de <i>Vigna unguiculata</i> (em azul, famílias com maior número de interações)	46
Figura– 14	Número de interações com fatores de transcrição por bloco. A linha vermelha indica o número médio de interações. Em azul, blocos com interações bastante superior a média	46
Figura– 15	Potencial regulação gênica nos blocos: A) avaliação quantitativa dos blocos, cujos genes co-expressos seriam regulados por um ou mais fatores de transcrição (FTs); B) FTs (e respectivas famílias) em destaque em cada grupo, e número de blocos total ou parcialmente regulados	47
Figura– 16	Interações proteína - proteína previstas entre componentes dos blocos de co-expressão de genes ligados em <i>Vigna unguiculata</i> , segundo modelo em <i>A. thaliana</i> , com indicação de relação direta ou indireta com respostas a estresses ambientais, tendo por base o enriquecimentos dos termos de ontologia gênica	48
Figura– 17	Interações dos componentes do bloco C01.B02.+ (WEX e GAE6) de <i>Vigna unguiculata</i> , com proteínas de reparo do DNA (em vermelho) e do metabolismo da galactose (em azul), usando interações proteína-proteína do modelo de <i>Arabidopsis thaliana</i> e ferramenta String	49
Figura– 18	Interação dos componentes do bloco C03.B06.+ (TMN11 e EDR2) de <i>Vigna unguiculata</i> , com proteínas do metabolismo da tiamina (em vermelho), usando interações proteína-proteína do modelo de <i>Arabidopsis thaliana</i> e ferramenta String	50
Figura– 19	Interações dos componentes do bloco C06.B01.- (BAS) de <i>Vigna unguiculata</i> , com proteínas da biossíntese de sesquiterpenoides e triterpenoides (em vermelho) e com a biossíntese da <i>backbone</i> terpenoide (em azul), usando interações proteína-proteína do	50

	modelo de <i>Arabidopsis thaliana</i> e ferramenta String	
Figura– 20	Número de potenciais ortólogos em diferentes leguminosas, associados aos 249 transcritos de genes componentes dos blocos de co-expressão de <i>Vigna unguiculata</i>	51
Figura– 21	Diagrama de Veen mostrando o número de blocos compartilhados ou exclusivos entre <i>V. angularis</i> , <i>V. radiata</i> , <i>P. vulgaris</i> e <i>G. max</i>	52
Figura– 22	Diagramas circulares mostrando a posição cromossômica dos blocos de <i>V.unguiculata</i> com ortólogos identificados em : A) <i>V.angularis</i> , B) <i>V. radiata</i> , C) <i>P.vulgaris</i> e D) <i>G.max</i>	52

LISTA DE TABELAS

Tabela 1 –	QTLs identificados em feijão-caupi relativos a características associadas ao estresse de seca	20
Tabela 2 –	Espécies do gênero <i>Vigna</i> com genoma referência depositados no NCBI	22
Tabela 3 –	Exemplos de estudos empregando a técnica SuperSAGE em diversos grupos. Espécies de plantas destacadas em negrito	26
Tabela 4 –	Potenciais vetores capazes de abrigar os blocos de genes de co-expressão observados em <i>Vigna unguiculata</i>	37
Tabela 5 –	Resultados de diferentes estratégias (BLAST vs Bowtie) relacionando <i>unitags</i> SuperSAGE com transcritos mRNA RefSeq de feijão-caupi	40

SUMÁRIO

1	INTRODUÇÃO	14
2	REVISÃO DE LITERATURA	16
2.1	Feijão-caupi: aspectos gerais da cultura	16
2.2	Natureza quantitativa da resposta ao estresse de seca em plantas	17
2.3	Genômica estrutural de <i>Vigna</i> spp	21
2.4	Genômica Funcional	22
2.4.1	<i>Transcriptômica do feijão-caupi</i>	28
2.5	Bioinformática	31
3	OBJETIVOS	34
3.1	Objetivo geral	34
3.2	Objetivos específicos	34
4	MATERIAIS E MÉTODOS	35
4.1	Material biológico, ensaio de desidratação radicular, origem e análise das <i>unitags</i> SuperSAGE	35
4.2	Cálculo da expressão diferencial de transcritos de <i>V. unguiculata</i> com base em <i>unitags</i> HT-SuperSAGE	36
4.3	Mapeamento <i>in silico</i> de transcritos com <i>unitags</i> SuperSAGE e identificação de blocos de genes de co-expressão em genoma de <i>V. unguiculata</i>	37
4.4	Predição dos fatores de transcrição associados aos genes ligados em blocos co-expressos de <i>V. unguiculata</i>	38
4.5	Caracterização de evidências funcionais dos genes de co-expressão dos blocos de <i>V. unguiculata</i> com respostas a estresses ambientais	38
4.5.1	<i>Evidências da literatura especializada</i>	38
4.5.2	<i>Análise de interação proteína-proteína</i>	38
4.6	Conservação dos blocos e inferência de sintenia em legumes tropicais	39
5	RESULTADOS	40
5.1	Análise de <i>unitags</i> HT-SuperSAGE e de transcritos de <i>V. unguiculata</i>	40
5.2	Os blocos de genes de co-expressão induzida no genoma do feijão-caupi	42
5.3	Predição de FTs e de elementos <i>cis</i> regulatórios associados aos promotores de genes co-expressos nos blocos de <i>V. unguiculata</i>	45

5.4	Análise de interação proteína - proteína	47
5.5	Sintenia dos blocos de <i>V. unguiculata</i> em leguminosas diversas	50
6	DISCUSSÃO	53
6.1	SuperSAGE: a expressão de genes a partir da contagem de <i>unitags</i>	53
6.2	Perspectivas na geração de marcadores funcionais aplicados a Seleção Assistida	55
6.3	Perspectiva funcional dos genes dos blocos de co-expressão induzida	56
6.4	Possíveis interações proteína-proteína	58
6.5	Perspectivas na manipulação por engenharia genética	60
6.6	Perspectivas em genômica comparativa	60
7	CONCLUSÕES	63
	REFERÊNCIAS	64
	APÊNDICE A – COMANDOS PARA ALINHAR AS UNITAGS SUPERSAGE CONTRA O TRANSCRITPOMA REFEREÊNCIA DO FEIJÃO-CAUPI VIA BOWTIE	84
	APÊNDICE B – COMANDOS PARA ALINHAR AS UNITAGS SUPERSAGE CONTRA O TRANSCRITPOMA REFEREÊNCIA DO FEIJÃO-CAUPI VIA BLAST	85
	APÊNDICE C – CÓDIGO CORRESPONDENTE AS ETAPAS DO ALGORITMO PARA IDENTIFICAÇÃO DE BLOCOS CO-EXPRESSOS E CONSECUTIVOS NO GENOMA DE <i>V. UNGUICULATA</i>	86
	APÊNDICE D – COMANDOS PARA ALINHAR AS PROTEÍNAS CORRESPONDENTES AOS BLOCOS MAPEADOS CONTRA SEQUÊNCIAS DE <i>A. THALIANA</i> DEPOSITADAS NO STRING VIA BLAST	91
	APÊNDICE E – COMANDO UTILIZADO PARA ANÁLISE COM A FERRAMENTA GET_HOMOLOGUES	92
	APÊNDICE F – DETALHES ACERCA DOS COMPONENTES DOS BLOCOS DE GENES CONSECUTIVOS E CO-EXPRESSOS IDENTIFICADOS NO GENOMA DE <i>V. UNGUICULATA</i>	93
	APÊNDICE G – REPRESENTAÇÃO DOS BLOCOS QUE APRESENTARAM TODOS OS COMPONENTES COM	100

**RELAÇÃO A ESTRESSES. EM LARANJA BLOCOS COM
TERMO GENÉTICA REVERSA E EM AZUL COMPONENTES
COM ESTUDOS REFERENTES A TRANSCRIPTÔMICA E
SECA**

**APÊNDICE H – RELAÇÃO DOS CROMOSSOMOS ENTRE
LEGUMINOSAS CORRESPONDENTES AOS ORTÓLOGOS
DOS GENES EM BLOCOS IDENTIFICADOS**

104

1 INTRODUÇÃO

A seca é um dos principais limitantes da produção em plantas. Estima-se que na última década as perdas econômicas em decorrência desse estresse tenham sido de \$30bi. Além disso, modelos climáticos apontam para escassez de recursos hídricos no médio-longo prazo. Nesse contexto, produzir novos cultivares com performance aceitável nessas condições é uma prioridade para garantir saúde e segurança alimentar global. Entretanto, a dificuldade de manipulação de características relacionadas a tolerância ao déficit hídrico imposta pela natureza molecular quantitativa ou QTLs (*loci* de traço quantitativo) requer novas estratégias para identificação de alelos com potencial biotecnológico que possam ser aplicados a programas de melhoramento vegetal.

Vigna unguiculata (L.) Walp., popularmente conhecida como feijão-caupi, feijão macassar ou feijão-da-corda, é uma leguminosa de grão altamente nutritiva e essencial na dieta de milhares de pessoas, especialmente em países emergentes. Inclusive, representa a principal fonte de proteína vegetal consumida nas regiões Norte e Nordeste do Brasil. Sua resiliência natural a seca e altas temperaturas aliada ao seu menor tamanho de genoma com relação a outras plantas fazem da espécie um modelo adequado para estudos voltados a esses estresses. Apesar de dispor de QTLs relacionados a seca identificados pouco é relatado na literatura a nível de gene para espécie.

O mapeamento associativo por gene candidato é uma forma direta e menos custosa para descoberta de genes relacionados com a variação fenotípica. Por propiciar uma maior resolução de mapeamento em comparação aos métodos tradicionais de análise de QTL (por *linkage*), o feijão-caupi pode se beneficiar diretamente de seu uso em maior escala. Entretanto sua aplicação é limitada pelo conhecimento prévio da característica em questão, o que requer alternativas para superar esse gargalo.

Aqui, foi proposta uma nova maneira de selecionar genes candidatos para estudos associativos tendo por base a expressão em acessos de resposta contrastante (tolerante vs sensível) a seca após estímulo de desidratação radicular ao ar (até 150 min) , evidências funcionais encontradas na literatura e na forte ligação física entre blocos gênicos presentes na montagem referencia do feijão-caupi. Ademais, o cálculo de expressão considerou todas as *units* SuperSAGE (26pb) alinhadas (bowtie) em uma determinada sequência referência, o que difere do observado em estudos dessa natureza para a espécie.

Logo, é esperado que os dados gerados possam auxiliar a desvendar a caixa preta por trás de características quantitativas como resposta a estresse de seca para o feijão-caupi e contribua para dissecar QTLs em componentes individuais fornecendo alvos para aplicação em programas de melhoramento da espécie e outras relacionadas.

2 REVISÃO DE LITERATURA

2.1 Feijão-caupi: aspectos gerais da cultura

O feijão-caupi, feijão macassar ou feijão-da-corda [*Vigna unguiculata* (L.) Walp.] é uma leguminosa de origem africana com grande importância socioeconômica em países emergentes (Boukar *et al.*, 2019). Taxonomicamente, pertence à família Fabaceae, subfamília Faboidae, tribo Phaseolae, subtribo Phaseolinae, Seção Catiang, e subespécie unguiculata, a qual é subdividida em cinco cultigrupos: Unguiculata, Sesquipedalis, Biflora, Textilis e Melanophthalmus (Pasquet and Padulosi, 2013). Se trata de um organismo diploide ($2n = 22$) com tamanho do genoma nuclear estimado em 658 Mb (Lonardi *et al.*, 2019). Sua reprodução é prioritariamente autógama, com baixo índice de fertilização cruzada, devido a ocorrência de cleistogamia em suas flores (Ehlers and Hall, 1997).

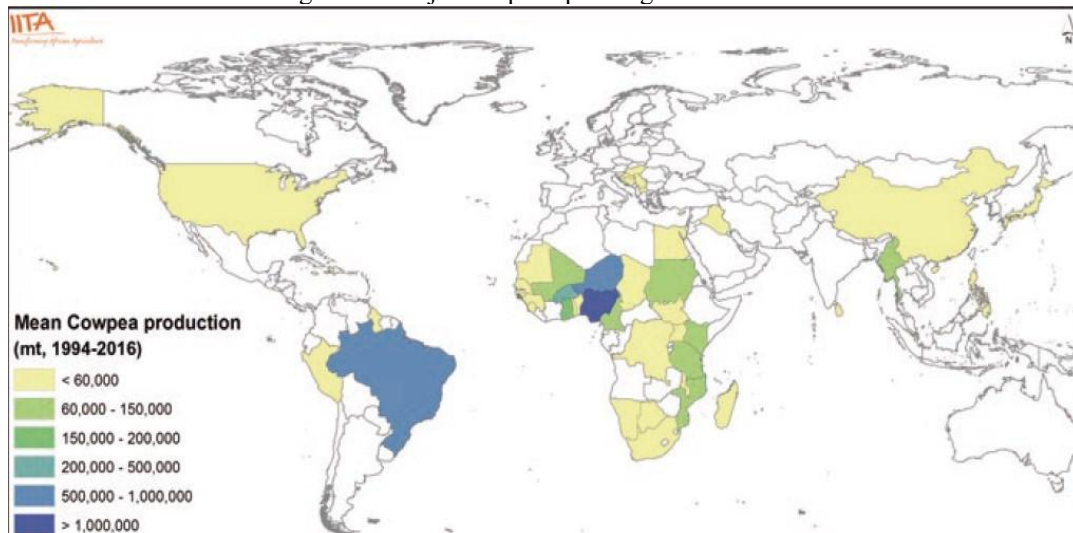
Também, a espécie possui resiliência natural a seca e altas temperaturas além de conseguir performar em diversos tipos de solo, com gradações variando de inférteis a arenosos e alcalinos a ácidos (Hall, 2004; Ehlers and Hall, 1997). Por isso, é bem adaptada a regiões semiáridas dos trópicos, onde a maioria das culturas vegetais não apresentam rendimento satisfatório, sobretudo devido as condições edafoclimáticas severas. Desta forma, o feijão-caupi representa um componente valioso para o sistema de produção tradicional nessas áreas (Singh *et al.*, 2003).

Com relação a suas aplicações, o feijão-caupi além de servir como fonte de alimentação humana, forragem animal e atuar como gerador de renda, também contribui para sustentabilidade dos plantios e do ambiente ao atuar como fixador do nitrogênio atmosférico, justificando a alcunha de legume “multifuncional” (Boukar *et al.*, 2018; Timko and Singh, 2008).

O continente africano concentra mais de 80% da produção mundial de feijão-caupi (Figura 1). Por sua vez, o Brasil foi o terceiro maior produtor com 791.000 toneladas, no ano de 2016, atrás apenas de Nigéria (3.409.992) e Níger (1.959.082) [dados da Companhia Nacional de Abastecimento (CONAB) e Organização das Nações Unidas para Alimentação e Agricultura (FAOSTAT)]. No Brasil, o cultivo, historicamente, era voltado as regiões Norte e Nordeste, locais de maior consumo. Atualmente, há uma tendência de migração para o Centro-Oeste onde a produção é altamente mecanizada e em larga escala, contrastando com o baixo uso tecnológico empregado nas regiões tradicionais. Portanto, o mercado nacional de

feijão-caupi no Brasil é dinâmico e sofreu momentos de reformulação (Freire Filho *et al.*, 2011).

Figura 1 – Feijão-caupi: aspectos gerais da cultura



Fonte: Boukar *et al.*, (2019, p.206)

Estima-se que a desnutrição alimentar contribua para mais da metade das mortes infantis ao redor do mundo (Bain *et al.*, 2013). O grão seco do feijão-caupi representa uma fonte protéica de alta qualidade, com teor superior a cereais, tubérculos e a alguns alimentos de origem animal. Tudo isso a um preço acessível, por isso é muitas vezes chamada de “refeição de pobre” (Jayathilake *et al.*, 2018; Timko and Singh, 2008). Adicionalmente, seu perfil nutricional é rico em carboidratos, minerais, aminoácidos essenciais, fibras dietéticas e vitaminas além do baixo índice de gordura (Jayathilake *et al.*, 2018). Assim, o feijão-caupi é uma cultura estratégica para promoção de saúde e segurança alimentar, especialmente em países pobres (Silva *et al.*, 2018).

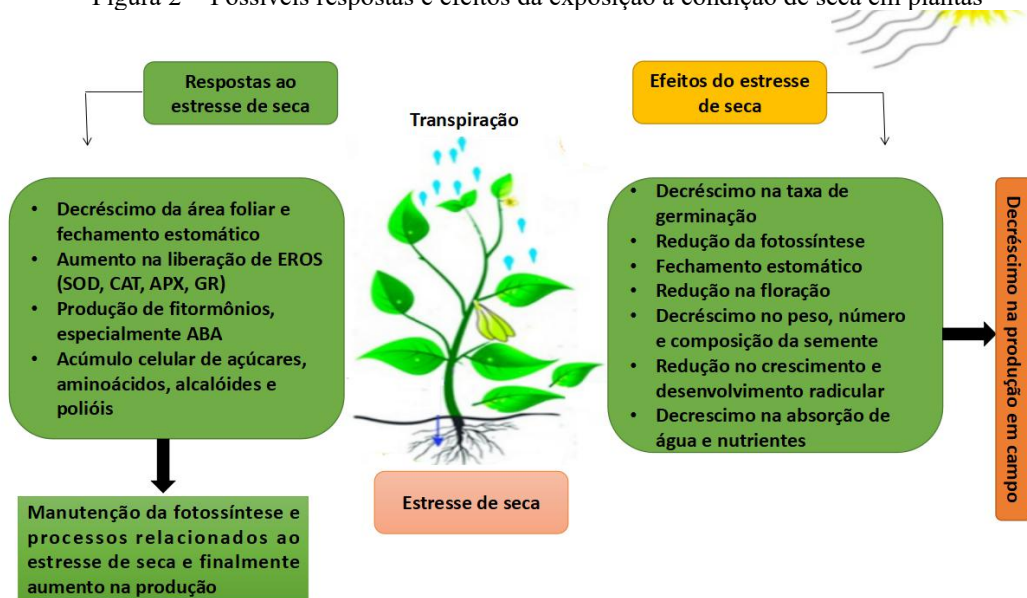
Por todos esses fatores apresentados, há uma expectativa de que o feijão-caupi possa emergir como o principal legume cultivado nas próximas décadas (Singh, 2014). Entretanto, apesar dessa relevância, a produção ainda é baixa em diferentes locais do Mundo sobretudo devido a influência de diversos fatores bióticos e abióticos (Sindhu *et al.*, 2019).

2.2 Natureza quantitativa da resposta ao estresse em plantas

Quando expostas a condição de seca, ou seja a baixa disponibilidade de água no solo e /ou no ar, as plantas respondem com modificações morfofisiológicas e bioquímicas ao longo de

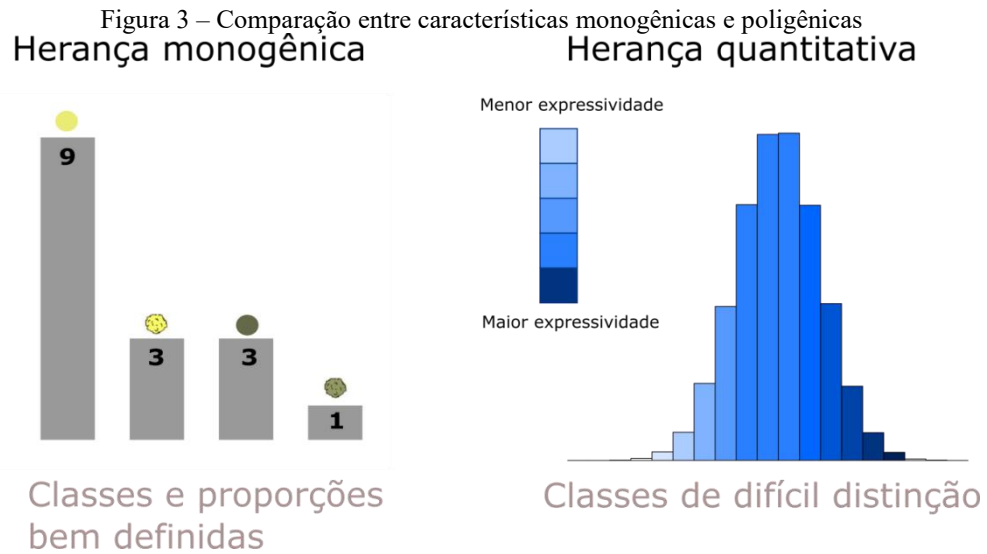
todo organismo (Fang and Xiong, 2015; Farooq *et al.*, 2012). Esses mecanismos, apesar de essenciais para sua sobrevivência, quase sempre levam a perdas agrônomicas consideráveis pelo decréscimo na produção (Figura 2).

Figura 2 – Possíveis respostas e efeitos da exposição a condição de seca em plantas



Fonte: O autor (2021)

Para ser considerada tolerante a planta necessita manter ou restabelecer sua homeostase celular próxima do normal, evitando danos a estruturas como proteínas e membranas durante e após exposição ao estresse (Wang; Vinocur; Altman, 2003). Assim como grande parte das características de interesse agrônomico (produção, altura entre outros), a tolerância ao déficit hídrico depende da ação cumulativa de múltiplas regiões cromossômicas chamadas *Quantitative Trait Loci* (QTL). A interação de alelos em diferentes QTLs entre si e com o ambiente leva a um padrão de herança complexa ou quantitativa (Tanksley, 1993), o que difere da proporção mendeliana de fenótipos fixa e esperada para características regidas por um ou poucos genes (Figura 3). Consequentemente, a performance fenotípica reflete parcialmente o valor genético do indivíduo (Semagn; Bjornstad; Xu, 2010). Como a análise de QTLs consiste em detectar associação entre o fenótipo vs genótipo de marcadores moleculares, identificá-los no genoma pode ser comparado a encontrar uma agulha no palheiro (Collard *et al.*, 2005). Ainda, dissecá-los a nível de genes individuais foi considerado como o maior desafio enfrentado pelos geneticistas no século XX (Luo; Wu; Kearsey, 2002). Logo, a manipulação dessas características por programas de melhoramento tradicionais baseado na seleção fenotípica é dispendiosa.



Fonte: O autor (2021)

No geral, as etapas para identificação ou mapeamento de QTLs consistem em : (i) selecionar ou desenvolver a população apropriada; (ii) fenotipar a população para a característica de interesse em condições controladas; (iii) definir o tipo de marcador molecular e o método para genotipagem gerando dados para um número adequado de marcadores polimórficos e (iv) identificar marcadores ligados a característica de interesse usando programas estatísticos (Semagn; Bjornstad; Xu, 2010). Assim, é possível determinar a posição indireta de um QTL a partir da co-herança entre ele e marcadores molecular, a qual ocorre devido a menor probabilidade de haver eventos de recombinação entre marcas próximas no genoma. Ademais, a influência do marcador pode ser observada pela variação estatisticamente significativa na performance fenotípica entre grupo de indivíduos com um determinado genótipo para marcadores polimórficos. Com isso, é possível testar as hipóteses que: não há QTL presente ou o QTL está presente mas não ligado a um marcador (H_0) ou QTL é presente e está ligado a marcadores (H_a) (Semagn; Bjornstad; Xu, 2010; Collard *et al.*, 2005).

Dentro do contexto anterior, a análise por *linkage* e o mapeamento associativo são as duas ferramentas mais comuns para dissecar traços quantitativos. A diferença entre elas consiste nas oportunidades dadas para que ocorram eventos de recombinação. Ao passo que que na primeira elas são limitadas devido a necessidade do desenvolvimento de mapas genéticos a partir de cruzamentos biparentais, a segunda faz uso de toda meiose acumulada historicamente em coleções de cultivares, sem a necessidade de elaborar cruzamentos e por isso geralmente alcança uma maior resolução de mapa (Sorkher *et al.*, 2008; Zhu *et al.*, 2008). Ainda, a depender do foco do estudo, o mapeamento associativo se enquadra em duas

categorias: (i) por gene candidato (GC), dirigido por hipóteses prévias o qual testa polimorfismos em genes selecionados com provável influência no controle da variação fenotípica para característica complexas e (ii) análise de associação genômica ampla (GWAS), aquele no qual a busca por variantes influenciadoras é realizada ao longo de todo o genoma sem a elaboração de hipóteses prévias (Zhu *et al.*, 2008). GCs podem ser candidatos posicionais (presentes em intervalos de confiança de QTLs já identificados), candidatos funcionais (aqueles nos quais estudos em expressão gênica, *knock-out* ou transformação genética, por exemplo, fornecem pistas sobre a relação do gene com a variação fenotípica observada) ou candidatos trazidos de outras espécies por genômica comparativa (Zhu and Zao, 2007).

Apesar de já existirem QTLs voltado a tolerância a seca mapeados em feijão-caupi (Tabela 1), no geral esses não alcançaram resolução suficiente para identificação de genes individuais, como observado em estudos recentes em outras espécies (Wang *et al.*, 2019). Além disso, genes candidatos na espécie são provenientes basicamente de estudos de sintenia com outras mais bem estudadas (Huynh *et al.*, 2015; Pottorff *et al.*, 2014; Pottorff *et al.*, 2012; Muchero *et al.*, 2011).

Tabela 1 – QTLs identificados em feijão-caupi relativos a características associadas ao estresse de seca

Característica	Marcador	Genes identificados	Refs
Arquitetura da raiz	SNPs	-	Burridge <i>et al.</i> , 2017
Senescencia e maturidade induzida por seca	AFLP	-	Muchero <i>et al.</i> 2009
Senescencia tardia, biomassa e produtividade de grão	SNPs	-	Muchero <i>et al.</i> 2013
Tolerancia ao calor	SNPs	-	Lucas <i>et al.</i> , 2013

Traduzido e adaptado de Chamarthi *et al.*, 2019

Uma vez que genes ou marcadores associados a QTLs forem identificados eles podem ser aplicados a programas de melhoramento na forma de Seleção Assistida por Marcadores (SAM) ou como insertos para abordagens de transgênia, ambos visando o lançamento de novos cultivares tolerantes (Dita *et al.*, 2006; Morgante and Salamini, 2003). A SAM consiste em identificar indiretamente um determinado fenótipo com base no genótipo de um marcador.

Sua aplicação promove ganhos consideráveis quando comparada com a seleção fenotípica tradicional ao permitir selecionar genótipos em estágios iniciais do desenvolvimento além de características com baixa herdabilidade, substituindo ensaios complexos de campo complexos por testes moleculares (Ribaut and Hoisington, 1998). Dentre os tipos de marcadores molecular, o marcador funcional (MF) é o mais desejável para SAM (Gupta and Rustgi, 2004). Seu diferencial é a presença de polimorfismo dentro de genes caracterizados funcionalmente, o que leva a uma associação motivo funcional - marcador é perfeita, ou seja, a ligação genética entre eles não é quebrada por recombinação (Salgotra; Gupta; Stewart, 2014; Andersen and Lübberstedt, 2003).

Tendo em vista as limitações que a natureza quantitativa do caráter impõem a programas de melhoramento tradicional, complementá-los com métodos biotecnológicos permite ao melhorista superar barreiras biológicas naturais (Kumar, 1999). Dados gerados pelas ômicas fornecem a matéria prima para manipular a tolerância a estresses abióticos em nível molecular (Shah *et al.*, 2018).

2.3 Genômica estrutural de *Vigna* spp.

A genômica estrutural representa o ponto de partida da análise de genomas estando associada a determinação da sequência nucleotídica e mapeamento físico e /ou genético completa de um organismo. (Hieter and Boguski, 1997). Uma vez que esses mapas estejam disponíveis é possível inferir sobre a localização e organização de genes.

Durante muito tempo, o feijão-caupi foi considerado uma cultura órfã, ou seja, negligenciada nos campos da pesquisa e indústria, devido a sua baixa influência no mercado global (Boukar *et al.*, 2019). Entretanto, com uma maior acessibilidade a tecnologias genômicas em larga escala, principalmente após advento do sequenciamento de nova geração (NGS), progressos recentes em termos de geração de recursos genômicos foram alcançados (Boukar *et al.*, 2016; Boukar *et al.*, 2019). Parte deles derivam da cultivar IT97K-499-35, com os principais resultados descritos abaixo:

- Timko *et al.* (2008) utilizaram a tecnologia de Filtragem por Metilação (FM) para sequenciar seletivamente regiões do genoma rica em genes, a partir do enriquecimento das porções hipometiladas, evitando assim o trabalhoso sequenciamento do genoma completo.

- Muñoz-Amatriaín *et al.* (2017) realizaram o primeiro sequenciamento de genoma completo (WGS), e produziram uma plataforma de genotipagem para 51.128 Polimorfismos de Nucleotídeos Simples (SNPs) (*The Cowpea iSelect Consortium Array*), além de um mapa genético consenso. No entanto, a montagem gerada não foi representativa de todo o genoma, visto que se encontrava extremamente fragmentada, devido à alta repetitividade do genoma e tamanho das *reads*.
- Lonardi *et al.* (2019) produziram a mais recente montagem genômica até o momento, utilizando a tecnologia de Sequenciamento em Tempo Real de Molécula Única (SMRT) combinado com mapeamento ótico e genético. Diferente da anterior, esta abordagem apresentou a integridade necessária para efetuar anotações precisas, identificação de genes candidatos e comparação robusta entre genomas.

Dentro do gênero *Vigna*, outras três espécies dispõem de genoma referência depositados na base de dados do Centro Nacional de Informações Biotecnológicas (NCBI) (Tabela 2). Espécies de plantas filogeneticamente relacionadas apresentam alto grau de sintenia, ou seja conservação do conteúdo gênico. Comparar a disposição de genes alvos entre elas permite verificar a possibilidade de transferir o conhecimento gerado de uma para as demais (Mudge *et al.*, 2005).

Tabela 2 – Espécies do gênero *Vigna* com genoma referência depositados no NCBI

Espécie	Montagem	Tam (Mpb)	Gen es	Cobertura	Tecnologia de Sequenciamento	Ano
<i>V. mungo</i>	ASM134271 9v1	498,8	-	267x	10x genomics sequencing technology	2020
<i>V. unguiculata</i>	ASM411807 v1	518,9	3438 3	91x	PacBio; Bionano	2019
<i>V. angularis</i>	Vigan 1.1	466,7	2952 3	167,7	Illumina HiSeq 2000	2015
<i>V. radiata</i>	Vradiata_ver 6	463,1	3000 3	300x	Illumina HiSeq	2015

Fonte: O autor (2021)

2.4 Genômica Funcional

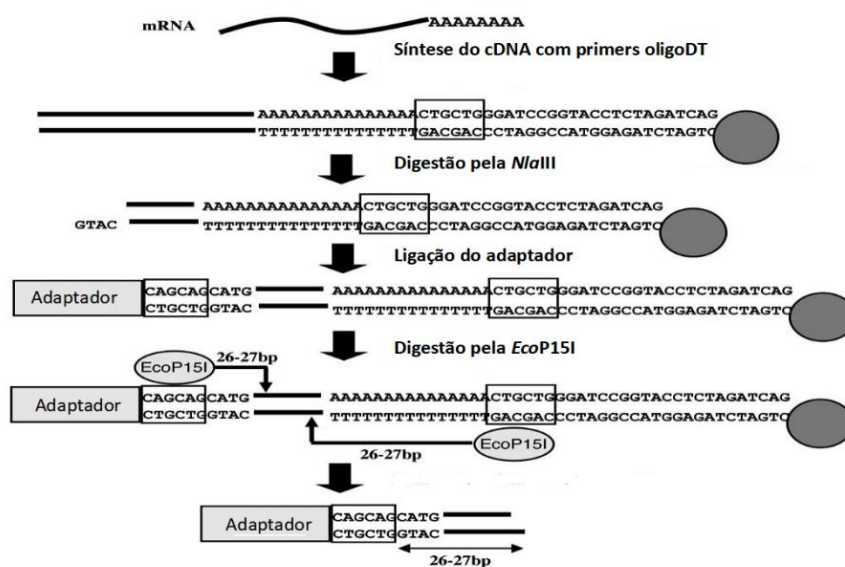
Dados fornecidos pela genômica estrutural por si só não são suficientes para compreender processos que regem os seres vivos (Gubb and Matthiesen, 2009). Para tal, a etapa posterior na análise de genomas consiste em desvendar a função dos genes, bem como sua regulação e padrões de expressão, campo de atuação da genômica funcional (Gubb and Matthiesen, 2009; Hieter and Boguski, 1997). A transcriptômica representa um ramo da genômica funcional que estuda os transcritos de ácidos ribonucleicos (RNAs) de um organismo, sejam eles intermediários na síntese de proteínas [RNAs mensageiro (mRNAs)] ou outros tipos com diversas funções adicionais (Lowe *et al.*, 2017; Gubb and Matthiesen, 2009). Mensurar o nível de expressão de genes em diferentes condições pode fornecer informações de como eles são regulados revelando detalhes sobre a biologia do organismo (Lowe *et al.*, 2017).

As ferramentas recentes para análise global de expressão gênica podem ser separadas em duas grandes classes (Rani and Sharma, 2017); os métodos baseados em sequenciamento, como a Análise Serial da Expressão Gênica (SAGE) (Velculescu *et al.*, 1995) e o Sequenciamento de RNA (RNA-seq) (Wang; Gerstein; Snyder, 2009) e aqueles fundamentados em hibridização, caso dos microarranjos (Schena, 1996). Esse último, no entanto, apresenta diversas limitações como arquitetura fechada e baixa reprodutibilidade entre diferentes fabricantes (Shendure, 2008). Métodos de sequenciamento com arquitetura aberta, expressão digital, alto rendimento e custo cada vez mais acessível, sobretudo devido ao advento das plataformas NGS (Mutz *et al.*, 2013), representam o estado da arte desde 2012 (Lowe *et al.*, 2017). Esses, por sua vez, são baseados em: (i) curtos fragmentos obtidos de posições definidas (*tags*) e (ii) na estratégia *shotgun* de transcritos completos (de Klerk; Den Dunnen; 't Hoen, 2014).

Nas variantes da SAGE *tags* de diferentes tamanhos [13 pb (protocolo original, Velculescu *et al.*, 1995); 21 pb (LongSAGE, Saha *et al.*, 2002); 26 pb (SuperSAGE, (Matsmura *et al.*, 2005)] são geradas a partir da clivagem sucessiva do DNA complementar (cDNA) por duas enzimas de restrições. No caso da SuperSAGE, as enzimas empregadas são a *Nla*III (ancoragem) e a *Eco*P15 (*tagging enzyme*). A primeira reconhece e cliva em todo 5'-CATG-3 e a segunda efetua o corte a uma distância de 26 pb dos seus sítios de restrições, incorporados ao primer oligoDT na síntese do cDNA e a um adaptador ligado ao CATG exposto (retângulos na Figura 4). As *tags* são sequenciadas e contadas, de tal forma que a frequência desses curtos fragmentos indica a abundância do transcrito correspondente na amostra. Ademais, a anotação das *tags* contra bancos de dados permite correlaciona-las com

processos funcionais, sendo assim possível descrever o perfil de expressão gênica de um organismo em diversas condições (Kido *et al.*, 2013). Também é possível associar a metodologia com NGS. A chamada *Deep* SuperSAGE ou HT (Alto rendimento) SuperSAGE promove um sequenciamento massivo e paralelo de *tags* ampliando a sensibilidade da técnica e a capacidade *multiplex*, além de reduzir o tempo e custo do experimento (Matsumura *et al.*, 2012). Desta forma, as trabalhosas etapas de formação e clonagem das *ditags* foram evitadas. Ao invés disso, adaptadores são ligados em extremidades das *tag* permitindo também que esses fragmentos *tag*-adaptador sejam amplificados pela reação em cadeia da polimerase (PCR) (Kido *et al.*, 2013; Matsumura *et al.*, 2010)

Figura 4 – Protocolo SuperSAGE. Retângulos em cinza indicam os sítios de reconhecimento da enzima *EcoP15I*.



Fonte: Traduzido e adaptado de Matsumura *et al.*, 2008

Estudos empregando SuperSAGE variam conforme a estratégia utilizada em diferentes espécies e para diversos fins (Tabela 3). Naqueles envolvendo plantas, nota-se uma predominância no uso da contagem de *tags* atribuídas a *unitags* (*tags* diferentes) ao invés de transcritos ou genes, como visto em outros grupos. Dessa forma, a contribuição de outras *unitags* geradas de um mesmo gene ou transcrito, provavelmente por digestão parcial da enzima *NlaIII* durante a preparação das bibliotecas ou sítios alternativos de poliadenilação e/ou *splicing* é perdida (van Aerle *et al.*, 2013; Hegedus *et al.*, 2009; Gowda *et al.*, 2004). Também observa-se uma tendência em não aplicar métricas para minimizar falsos positivos

(erro tipo I) como a taxa de descoberta falsa (FDR). Esses fatores, em conjunto, podem contribuir para o insucesso em futuras validações *in vivo*. Além disso, há uma predominância no uso do teste audic-claviere, no qual a probabilidade das *reads* ou *tags* terem sido originária de uma determinada região do genoma obedece a distribuição de Poisson. Ainda, alinhadores com funcionamento e propostas diferentes como a Ferramenta de Busca de Alinhamento Local Básico (BLAST) e bowtie são utilizados para anotação das *tags*.

Tabela 3 – Exemplos de estudos empregando a técnica SuperSAGE em diversos grupos. Espécies de plantas destacadas em negrito.

Espécie	Escopo	Anotação	FDR	Teste	Estratégia	Ref
Lagarta da Maça	Identificar genes relacionados com resistência a Vip3A	short-BLAST	N	Audic	Tag	Ayra-Pardo <i>et al.</i> , 2019
Inhame-da-guiné-branco	Identificar genes candidatos a floração	BLAST	S	EdgeR	Tag	Girma <i>et al.</i> , 2019
Robalo europeu	Identificar perfis de expressão gênica entre indivíduos com altas e baixas taxas específicas de crescimento	Galaxy	S	Fisher	Tag	Louro <i>et al.</i> , 2019
Lentilha	Identificar genes resposivos a geada	BLAST	N	Audic	Tag	Barrios <i>et al.</i> , 2017
Nicotiana	Análise do transcriptoma pós inoculação	BLAST	-	-	Tag	Portieles <i>et al.</i> , 2017
Humana	Investigar padrões de expressão diferencial de células estromais	Bowtie	N	DegSeq	Gene	Casado- Díaz <i>et al.</i> , 2016
Ambrosia	Análise do transcriptoma do polen frente a estresses	BLAST	N	Audic	Tag	El Kelish <i>et al.</i> , 2014

	de seca e CO ₂					
Batata	Identificar genes para resistencia contra requeima da batata	Burrows-Wheeler	N	Audic	Transcrito	Draffehn <i>et al.</i> , 2013
Peixe-zebra	Avaliar efeitos da exposição a prata	Bowtie	S	Audic	Gene	van Aerle <i>et al.</i> , 2013
Cana-de-açúcar	Identificar genes responsivos a estresse de seca	BLAST	N	Audic	Tag	Kido <i>et al.</i> , 2012
Fungo da brusone do arroz	Avaliar mudanças na expressão gênica associadas com infecção	Bowtie	S	DeSEQ	Transcrito	Soanes <i>et al.</i> , 2012
Tabaco coio	Identificar genes responsivos em resposta a elicitação do ácido N-linoleico-glutâmico	BLAST	N	Audic	Tag	Gilardoni <i>et al.</i> , 2010
Grão-de-bico	Identificar genes responsivos a um estímulo de dissecação radicular	BLAST	N	Audic	Tag	Molina <i>et al.</i> , 2008

2.4.1 Transcriptômica do feijão-caupi

Estudos em transcriptômica vêm sendo realizados para identificação de genes, vias metabólicas ou processos relacionados com tolerância a diversos estresses abióticos. Genes responsivos a seca em feijão-caupi são deduzidos a partir de candidatos observados em espécies relacionadas ou pela análise da expressão diferencial de mRNAs em plantas estressadas *versus* não estressadas. Diversos destes genes se encontram descritos (Carvalho *et al.*, 2017; Agbicodo *et al.*, 2009). Além desses, iniciativas locais também contribuem para auxiliar a desvendar funcionalmente o genoma do feijão-caupi. A Rede Nordeste de Biotecnologia (NordEST), a qual conta com a colaboração da Universidade Federal de Pernambuco (UFPE) e a Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), gerou milhões de sequências, dentre Etiquetas de Sequências Expressas (ESTs), *tags* HT-SuperSAGE e RNA-seq, a partir de plantas submetidas a diversos estresses bióticos [Vírus do mosaico do feijão-caupi transmitido por afídeos (CABMV) e Vírus do Mosaico Severo do Feijão-caupi (CPSMV)] e abióticos (desidratação radicular e salinidade) (Kido *et al.*, 2011).

Quando comparado a outras leguminosas de importância econômica, o feijão-caupi ainda é uma cultura com estudos de expressão gênica limitados. Da busca avançada (Abril de 2019) na base de dados *PubMed* por artigos de pesquisa contendo em seu título ou resumo os termos “*cowpea*”, “*drought*” e “*expression*”, 30 artigos são encontrados. Após eliminação daqueles que não se referem a estresse de seca, uma vez que o feijão-caupi muitas vezes é descrito como uma cultura “*drought tolerant*” ou “*adapted to drought*”, dos estudos de revisão e aqueles com foco em mais de uma espécie, restaram 21 artigos.

Desses, a maioria remete ao isolamento e caracterização de poucos clones de cDNA induzidos após um estímulo simulando estresse de seca, normalmente supressão da rega, dissecação rápida ou desidratação radicular. Alvos estudados incluem: *VuC1(cystatin)* (Diop *et al.*, 2004), *VuPLD1 (Phospholipase D)* (El Maarouf *et al.*, 1999), *APX (Ascorbate Peroxidase)* (D’Arcy-Lameta *et al.*, 2006), *GR (Glutathione Reductase)* (Contour-Ansel *et al.*, 2006), *VuAP1 (Aspartic Protease)* (De Carvalho *et al.*, 2001), *VuAox (Alternative Oxidase)* (Costa *et al.*, 2007) e FTs (*VuDREB2A*) (Sadhukhan *et al.*, 2014a; Sadhukhan *et al.*, 2014b), *VuPAPa* e *VuPAP (phosphatidic acid phosphatase)* (França *et al.*, 2008; Marcel *et al.*, 2000), *VuCRPD (cowpea clones responsive to dehydration)* (Iuchi *et al.*, 2000; Iuchi *et al.*, 1996),

Vupat1 (*patatin like*) (Matos *et al.*, 2001). Além desses estudos envolvendo regulação por microRNAs (miRNAs) também foram encontrados (Barrera-Figueroa *et al.*, 2011; Shui; Chen and Li., 2013).

Mais recentemente, Yao *et al.* (2016) desenvolveram o *Vigna unguiculata* Gene Expression Atlas (VuGEA) a partir do RNA-seq (montagem *de novo*) de 27 bibliotecas obtidas de tecidos (folhas, caule, raiz e flores) do genótipo IT97K-499-35 cultivado em condições controladas. Nele, estão presentes quase 35 mil transcritos únicos com suas análises de expressão diferencial. Adicionalmente, foi demonstrado que respostas a dissecação da semente e exposição a seca possuem mecanismos comuns, como, no caso da pesquisa, a regulação por um mesmo miRNA.

Amorim *et al.* (2018) reportaram uma avaliação de potenciais genes de referência para PCR em tempo real (qPCR) em raízes de genótipos brasileiros (Pingo de Ouro e Santo Inácio) expostos a desidratação radicular ao ar (de até 150 minutos). A partir dos métodos geNorm, NormFinder, BestKeeper, e ΔCt , todos os candidatos [*VuACT* (*actin*), UE21D (ubiquitin-conjugating enzyme E2 variant 1D), e UNK (*Phaseolus vulgaris* unknown gene)] se mostraram viáveis, sendo *VuACT* e UE21D os mais estáveis segundo as quatro análises. Comparando a expressão desses genes entre diferentes metodologias (HT-SuperSAGE vs qPCR), foi observado uma concordância em 56% dos resultados evidenciando a importância da validação em estudos de transcriptômica. Em resumo, o estudo proporciona ferramentas que possibilitam uma maior precisão em ensaios de qPCR para o feijão-caupi nas referidas condições.

Li *et al.* (2019) identificaram 578 genes órfãos (restrito a espécie ou grupos taxonômico), a maioria deles não codificantes, diferencialmente expressos a partir de análise *microarray*, realizada em outro momento, em raízes e folhas de plantas (cultivar ‘B128’) estressadas e não estressadas a estímulo de seca. Após superexpressão daquele com maior valor de *Fold Change* (FC) (UP12_8740) nas raízes, foi observado um aumento na tolerância dos transgênicos a estresse de seca (supressão de rega) e osmótico [tratamento com PEG (polietilenoglicol) 6% ou 10%]. Este estudo reforçou o potencial dos genes órfãos frente a identificação de novos genes envolvidos em adaptações ambientais.

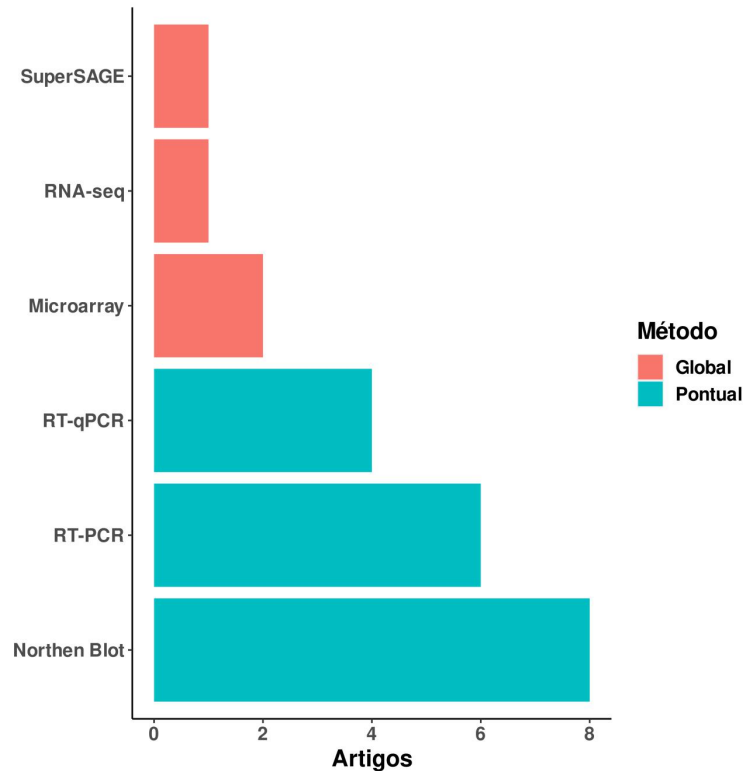
Carvalho *et al.* (2019) realizaram uma ampla análise da tolerância a seca em quatro genótipos [CB 46 (*California blackeye* 46), moderadamente suscetível, *Bambey* 21, altamente suscetível e as variedades comerciais Cp5051 e Vg50] englobando a avaliação de parâmetros

fisiológicos (condutância estomática, absorção de CO₂, eficiência do uso da água), bioquímicos (prolina, malondialdeído, peróxido, antocianinas e enzimas antioxidantes) e moleculares (expressão gênica) a partir de folhas expostas a três regimes de hidratação (seca severa, moderada e controle). Resumidamente, o estudo visou demonstrar se é possível discriminar genótipos de feijão-caupi com base nesses aspectos. Dos cinco genes previamente descritos como responsivos a seca [*VuCPRD14*, *VuCPRD22*, *VuCPRD65* (*Cowpea Responsive to Dehydration*), *VuHsp17.7* (*Small Heat Shock Protein*) e *VuNced1* (*9-cis-epoxycarotenoid dioxygenase*)] tiveram seus perfis de expressão avaliados por PCR semi-quantitativa. Esses genes apresentaram uma relação de proporcionalidade direta entre aumento da expressão e intensidade do estresse, muitos deles também apresentaram baixa expressão em plantas não estressadas. *VuCPRD14* e *VuHsp17.7* também foram avaliados por RT-qPCR se mostrando marcadores adequados para tolerância a seca.

De forma similar, Zegaoui *et al.* (2017) avaliaram as respostas fisiológicas e molecular de duas cultivares argelinas adaptadas a diferentes climas: Menia , quente e árido, e Bejaia , sub húmido, ambos sob desidratação progressiva (supressão da rega). A expressão dos genes *VuDREB2A*, *VuNCED*, *VuAOX*, *VuP5CS*, *VuProDH* e *VuAOX1* foram analisadas por PCR semi-quantitativa. A associação entre os resultados demonstraram que as culturas responderam de forma distinta ao estímulo.

Em relação as metodologias utilizadas (Figura 5), é possível observar uma predominância no uso daquelas voltadas a análise de expressão pontual, envolvendo um ou poucos genes isolados. Isso demonstra que o feijão-caupi ainda carece de estudos globais de seu perfil transcriptômico por meio de metodologias com arquiteturas abertas em larga escala, o que pode ser fundamental para caracterização da resposta molecular frente ao estresse de seca, por exemplo. Também não foi encontrado na literatura nenhuma abordagem envolvendo a associação entre o mapeamento *in silico* em seu genoma, com genes dispondo de dados de expressão gênica, reflexo, talvez, do depósito recente (2019) do genoma de referência em banco público de dados.

Figura 5 – Metodologias de artigos científicos encontrados na busca PubMed contendo as palavras “cowpea”, “drought” e “expression” em seu resumo e/ ou título realizada em Abril de 2019



Fonte: O autor (2021)

2.5 Bioinformática

Bioinformática é um campo fundamentalmente dirigido por dados, prioritariamente computacional e interdisciplinar, o qual engloba aspectos das ciências da computação, matemática e estatística, a fim de armazenar, manipular, analisar e interpretar dados biológicos (Attwood *et al.*, 2019). Recentemente, a magnitude dos dados produzidos apenas pela comunidade científica ultrapassou a casa dos *exabytes* (10^{18}) (Gauthier *et al.*, 2019). A tempestade perfeita formada entre esses dados gerados e avanços computacionais recentes abre novas oportunidades para resolução de problemas em escala antes inimagináveis. Nesse contexto, as ômicas são diretamente beneficiadas. Por exemplo, a associação entre várias ômicas (multiômicas) junto a algoritmos de redes neurais profundas são capazes de propor modelos que predizem o prognóstico em diversos tipos de câncer (Huang *et al.*, 2019; Chaudhary *et al.*, 2018).

Além disso, diversas ferramentas que também fazem interface entre as áreas citadas surgiram. Embora nem todas tenham sido desenvolvidas exclusivamente para aplicação em

estudos ômicos, ainda assim auxiliam na tomada de decisões quando avaliado dados brutos dessa natureza. Pode-se citar alinhadores de sequências que atuam com propósitos e formas diferentes como: o *Basic Local Alignment Search Tool* (BLAST) que busca alinhar por similaridade local sendo importante para análises estruturais e comparativas entre diferentes grupos (Altschul *et al.*, 1990); o bowtie voltado para o alinhamento de curtas sequências (*reads*) contra genomas grandes por indexação, frequentemente utilizado em montagens de transcriptomas por referência (Langmead, 2010; Langmead and Salzberg, 2012) e o diamond otimizado para alinhamentos envolvendo sequências de proteínas, sendo uma alternativa mais veloz ao BLASTp e BLASTx que aumenta sua performance conforme maior for o número de proteomas a serem alinhados (Buchfink, Xie; Huson, 2015).

Outras ferramentas possuem o enfoque mais comparativo entre diferentes espécies como por exemplo o OrthoVenn (Xu *et al.*, 2019) e GET_HOMOLOGUES (Contreras-Moreira and Vinuesa 2013). Apesar de ambas apresentarem o mesmo objetivo central (clusterizar sequências em possíveis ortólogos), elas diferem quanto a sua usabilidade. Enquanto a primeira é mais *user friendly*, disponibilizada online com interface gráfica, a segunda é executada por linhas de comando, dando um maior controle e flexibilidade quanto aos parâmetros utilizados. Essa situação ilustra um cenário comum na bioinformática, onde a escolha da ferramenta a ser utilizada depende do *trade off* entre facilidade de uso e liberdade de customização, mostrando que há alternativas para todo tipo de usuário.

Também há linguagens de programação amplamente utilizadas dentro da bioinformática como R, python e perl as quais permitem uma eficiente manipulação de dados podendo automatizar análises por meio de *pipelines* e pacotes pré-disponibilizados em repositórios, como por exemplo o Deseq (Love; Huber; Anders, 2014) e EdgeR (Robinson; McCarthy; Smyth, 2009) que realizam cálculo de expressão diferencial em ensaios de sequenciamento de alto rendimento.

Outro atributo da bioinformática são os bancos de dados. Sua definição, segundo a oracle: “é uma coleção organizada de informações ou dados - estruturadas, normalmente armazenadas eletronicamente em um sistema de computador”. Assim, simples ações como coleta e consulta a esses bancos de dados fornece ao pesquisador o poder de recuperar rapidamente um grande número de dados além de ser possível depositar os resultados do seu próprio experimento, contribuindo para integrar o conhecimento entre a comunidade científica. O principal repositório de biomoléculas disponível online, o banco americano de dados de sequências de DNA e proteínas, Genbank, possui cerca de 216.214.215 sequências

depositadas (fevereiro de 2020), em sua base de dados pública (<<https://www.ncbi.nlm.nih.gov/genbank/statistics/>>). Outros bancos biológicos incluem o europeu EMB (Laboratório Europeu de Biologia Molecular) e o japonês DDBJ (Banco de Dados de DNA do Japão). Também existem bancos de dados funcionais voltados a estudos de mecanismos mais específicos como o portal PlantRegMap direcionados a fatores de transcrição, elementos regulatórios e interações entre eles visando gerar *insights* sobre a regulação transcricional em plantas (Tian *et al.*, 2020). Outro exemplo é o string, um banco de interações proteína - proteína conhecidas e preditas. Atualmente o banco cobre 24.584.628 proteínas de 5.090 organismos (Szkarczyk *et al.*, 2019).

Todo esse ambiente de integração entre ciências ômicas e ferramentas computacionais proporcionado pela bioinformática faz com que essas sejam cada vez mais direcionadas a dados e voltadas a descoberta de novos fenômenos como previsto em momentos anteriores (Lee Kraus, 2005). Ao mesmo tempo que essas ferramentas contribuem para avanços na área, elas também levam a uma democratização do conhecimento gerado, tudo isso de uma forma cada vez mais acessível para o usuário final.

3 OBJETIVOS

3.1 Objetivo geral

- Identificação da expressão diferencial, em acessos de *V. unguiculata* após desidratação radicular, de genes ligados e sintênicos em genomas de *Vigna* spp

3.2 Objetivos específicos

- Identificar e anotar transcritos Diferencialmente Expressos (DE) nas bibliotecas transcriptômicas de raízes de acessos (tolerante e sensível) de feijão-caupi após estímulo de desidratação radicular.
- Mapear *in silico* os transcritos identificados, em genomas de *V. unguiculata*, *V. angularis* e *V. radiata*.
- Identificar grupo de genes DE, ligados e sintênicos, de co-expressão *in silico* similar ou contrastante nas respostas dos acessos, em relação aos respectivos controles negativos

4 MATERIAIS E MÉTODOS

4.1 Material Biológico, ensaio de desidratação radicular, origem e análise das unitags SuperSAGE

Cultivares de *V. unguiculata* considerados tolerante [Pingo de Ouro 1-2 (PO)] e sensível [Santo Inácio (SI)] ao déficit hídrico (Nascimento *et al.*, 2011) foram mantidos em solução hidropônica (estação EMBRAPA-soja, Londrina, PR), conforme Kulcheski *et al.* (2011). De forma resumida, sementes foram pré-germinadas (no escuro, em casa de vegetação: $25^{\circ}\text{C} \pm 2^{\circ}\text{C}$ e $60\% \pm 5\%$ de umidade relativa). As mudas foram transplantadas para suportes de poliestireno (30L), deixando-se as raízes submersas em solução nutritiva (pH 6,6). Após 15 dias (estádio V2; primeira folha trifoliada desenvolvida), parte das plantas foram submetidas ao tratamento que consistiu na exposição das raízes ao ar (desidratação radicular) e parte não (controle negativo). Foram determinados seis tempos de exposição [T0 (controle negativo), T1 (25 min.), T2 (50 min.), T3 (75 min.), T4 (100 min.), T5 (125 min.) e T6 (150 min.)], com coletas das raízes após cada período, sendo estas imediatamente congeladas (nitrogênio líquido) e armazenadas (-80°C) até extração dos RNAs totais, com o kit *SV Total RNA Isolation System* (Promega). Amostras foram quantificadas em espectrofotômetro NanoDrop ND-1000 UV-Vis (*Thermo Fisher Scientific*), e avaliadas em eletroforese (gel de agarose a 1% para RNA). RNAs de boa qualidade foram transcritos reversamente em cDNAs (kit *ImProm-IITM Reverse Transcription System*). A partir deles,, tags HT-SuperSAGE de 26pb foram geradas e sequenciadas (*Illumina Genome Analyzer II*) por intermédio da empresa GenXPro (Frankfurt, Alemanha). Duas bibliotecas HT-SuperSAGE (controle e tratamento) para cada cultivar (tolerante e sensível) foram sintetizadas, conforme descrição abaixo:

- SI_T0: controle negativo do cultivar sensível Santo Inácio.
- PO_T0: controle negativo do cultivar tolerante Pingo de Ouro 1-2.
- SI_T1-6: *Bulk* contendo 1/6 de cada mRNAs representativos de cada tempo (de 25 a 150 min.) referente ao cultivar sensível Santo Inácio.
- PO_T1-6: *Bulk* contendo 1/6 de cada mRNAs representativos de cada tempo (de 25 a 150 min.) referente ao cultivar tolerante Pingo de Ouro.

Tags agrupadas em *unitags* (tags diferentes) foram normalizadas em cada biblioteca para *reads* por milhão (RPM). *Unitags singlets* foram desconsideradas nas análises. Da comparação das *unitags* normalizadas na biblioteca estimulada em relação ao controle negativo do respectivo acesso se identificou aquelas *unitags* diferencialmente expressas (DE),

usando $p < 0,05$ em testes seguindo distribuição de Poisson [teste *Audic Claverie*, Audic and Claverie, 1997)], conforme implementado no *software DiscoverySpace* (Robertson *et al.*, 2007). Valores de FC consistiram da razão (R) das *unitags* normalizadas na biblioteca estimulada (numerador) em relação ao controle negativo (denominador). Frequências igual a 0 no denominador foram consideradas 1 (evitando a divisão por zero). Quando $R < 1$, o valor de FC considerou $-1/R$; valores negativos de FC indicaram *unitags downregulated* (DR) ou reprimidas, e valores positivos, *unitags upregulated* (UR) ou induzidas.

4.2 Cálculo da expressão diferencial de transcritos de *V. unguiculata* com base em *unitags* HT-SuperSAGE

Unitags foram alinhadas via Bowtie 1.2.3 (Langmead *et al.*, 2009) e NCBI-blast+2.6.0-1 contra o transcriptoma RefSeq de *V. unguiculata* (assembly ASM411807v1, GCA_004118075.1 (Lonardi *et al.*, 2019), recuperado em agosto de 2019) (ver comandos em **APÊNDICE A** para alinhamentos com o bowtie e **APÊNDICE B** para alinhamentos BLASTn). *Unitags* associados a ncRNA ou miscRNA foram desconsideradas das análises posteriores. Alinhamentos aceitos apresentaram: (i) presença do 5'CATG (que corresponde ao sítio da enzima *NlaIII*, conforme protocolo SuperSAGE); (ii) máximo de um erro em 26 pb do alinhamento; (iii) *unitag* presente na fita codificadora do transcrito; (iv) $e\text{-value} < 1e^{-4}$ para alinhamentos BLAST. *Unitags* presentes em mesmo transcrito tiveram as abundâncias somadas e atribuídas ao transcrito. As contagens de *tags*/transcrito foram normalizadas pelo total de *tags* sequenciadas por biblioteca dividido por um milhão e posteriormente convertidas para um milhão de transcritos.

Testes de hipótese verificaram a probabilidade ($P\text{-value}$) de rejeição da H_0 (transcrito não DE), a partir da comparação pareada de *reads counts* dos transcritos considerando cada contraste (SI_T0 *versus* SI_T1-T6 e PO_T0 *versus* PO_T1-6), realizada com a ferramenta 2 da plataforma ACDtool (Claverie and Ta, 2019). Transcritos com frequências nulas foram removidos, para não influenciar no número total de testes. As estimativas de $P\text{-values}$ foram ajustadas para correção de erro tipo I [*False Discovery Rate* (FDR), Benjamini and Hochberg, 1995] por meio da função *p.adjust* em linguagem R. Valores de FC dos transcritos correspondeu a razão entre as contagens normalizadas nas bibliotecas tratamento e controle negativo, similar ao aplicado às *unitags*. Transcritos DE apresentaram $P\text{-value} < 0,05$ e $FDR < 0,01$ nas análises, sendo aqueles com $\log_2FC > 1$ e $\log_2FC < -1$ declarados induzidos e

reprimidos após aplicação do estímulo, respectivamente. O processo (com exceção do teste estatístico) foi automatizado em linguagem R, estando disponível para análises futuras.

4.3 Mapeamento *in silico* de transcritos com unitags SuperSAGE e identificação de blocos de genes de co-expressão em genoma de *V. unguiculata*

A posição dos transcritos com *unitags* nos 11 cromossomos de *V. unguiculata* foi recuperada via *File Transfer Protocol* (FTP), das anotações RefSeq NCBI (arquivo *feature_table.txt*) da montagem genômica de referência (GCF_004118075.1). O mapeamento e identificação de blocos de genes consecutivos declarados UR em PO e expressos em SI também foi automatizado por meio da linguagem R. O algoritmo com as etapas e os códigos correspondentes a cada uma delas encontra-se no **APÊNDICE C**. Blocos foram nomeados conforme representado na Figura 6, ressaltando o cromossomo e os blocos consecutivos, em ordem crescente da numeração das bases, identificados em uma das fitas (+ ou -). Os nomes dos genes, representados pelas siglas foram oriundos da base de dados Uniprot, a partir das anotações RefSeq. Os blocos identificados foram agrupados de acordo com o tamanho (diferença entre a posição do último *locus* do bloco em relação ao primeiro) e consideradas nas análises de potenciais vetores que pudessem abrigá-los, conforme abordados por Nierman and Feldbyum (2001) (Tabela 4).

Figura 6 – Identificação dos blocos de genes de co-expressão em cromossomos de *Vigna unguiculata*

C02.b03.+
Cromossomo. Número do bloco. fita

Fonte: O autor (2021)

Tabela 4 – Potenciais vetores capazes de abrigar os blocos de genes de co-expressão observados em *Vigna unguiculata*

Vetor	Intervalo em kpb
Plasmídeo	0-15
Fago λ	15-35
Cosmídeo	35-45
Fago P1	45-90
Bacteriofágo	>90

Fonte: O autor (2021)

4.4 Predição dos fatores de transcrição associados aos genes ligados em blocos de co-expressão de *V. unguiculata*

Interações entre potenciais Fatores de Transcrição (FTs) e os genes de co-expressão ligados em blocos foi realizada a partir da ferramenta “*Regulation Prediction*” do portal PlantRegMap (Tian *et al.*, 2020), tendo como os alvos sequências de *V. angularis* similares aos transcritos dos genes identificados nos blocos. A predição dos sítios de ligação dos FTs considerou aqueles de $p \leq 1e^{-5}$ com uma posição relativa entre -500 a +100 bp do sítio inicial da transcrição (TSS) de cada gene.

4.5 Caracterização de evidências funcionais dos genes de co-expressão dos bLocos de *V. unguiculata* com respostas a estresses ambientais

4.5.1 Evidências da literatura especializada

A coleta na literatura especializada de evidências funcionais entre os genes componentes dos blocos com respostas a estresses ambientais usou a estratégia de busca por palavras-chave, incluindo a anotação RefSeq e *stress*, na ferramenta do portal *Google Scholar*. Artigos observados foram classificados, de acordo com a natureza dos estudos, por termos técnicos que os resumem, tais como abordagens ômicas (transcriptômica e proteômica), genética reversa (superexpressão de genes ou mutagenese por perda de função), envolvimento com fitormônios, bem como em relação aos estresses associados (bióticos, abióticos ou induzidos de natureza química).

4.5.2 Análise de interação proteína-proteína

Nessa análise interatômica, proteínas correspondentes aos transcritos RefSeq dos genes coexpressos dos blocos identificados foram alinhadas localmente via NCBI-blast + 2.6.0-1 (ver comandos em **APÊNDICE D**) contra sequências de *A. thaliana* oriundas da base de dados String [versão 11.0b, março de 2021, <https://string-db.org/>] (Szklarczyk et al., 2019)]. Arquivos FASTA serviram de *input* na aba *Multiple sequence* (organismo *A. thaliana*); em um primeiro momento, com um único arquivo contendo todas as proteínas, e posteriormente,

um arquivo por bloco. A pontuação mínima de interação exigida foi o mais restrigente disponível (0,900).

4.6 Conservação dos blocos e inferência de sintenia em legumes tropicais

Para inferir sobre o grau de sintenia e conservação dos blocos, proteomas de *Vigna. angularis* (assembly Vigan1.1, GCA_001190045.1), *Vigna. radiata* (assembly Vradiata_ver6, GCA_000741045.2), *Phaseolus. vulgaris* (assembly PhaVulg1_0, GCA_000499845.1) e *Glycine. max* (assembly Glycine_max_v4.0, GCA_000004515.5) foram recuperados das bases de dados do NCBI (abril de 2021). Cada proteoma foi confrontando separadamente (Diamond, $e\text{-value} < 1^{-10}$) contra o arquivo FASTA das proteínas correspondentes aos genes pertencentes aos blocos identificados em *V. unguiculata*, afim de identificar possíveis *clusters* ortólogos por meio do algoritmo *bidirectional best-hits* (BDBH), da ferramenta GET_HOMOLOGUES (Contreras-Moreira and Vinuesa_2013), cujos comandos encontram-se no **APÊNDICE E**. blocos compartilhados ou exclusivos nas espécies estudadas foram visualizados a partir da ferramenta *online* InteractiVenn (Heberle *et al.*, 2015) e as posições cromossômicas dos blocos em *V. unguiculata*, e de seus respectivos ortólogos de cada espécie estudada (exceto aqueles em *scaffolds*) foram ilustradas em digramas circulares gerados com a ferramenta Circos (Krzywinski *et al.*, 2009).

5 RESULTADOS

5.1 Análise de *unitags* HT-SuperSAGE e de transcritos de *V. unguiculata*

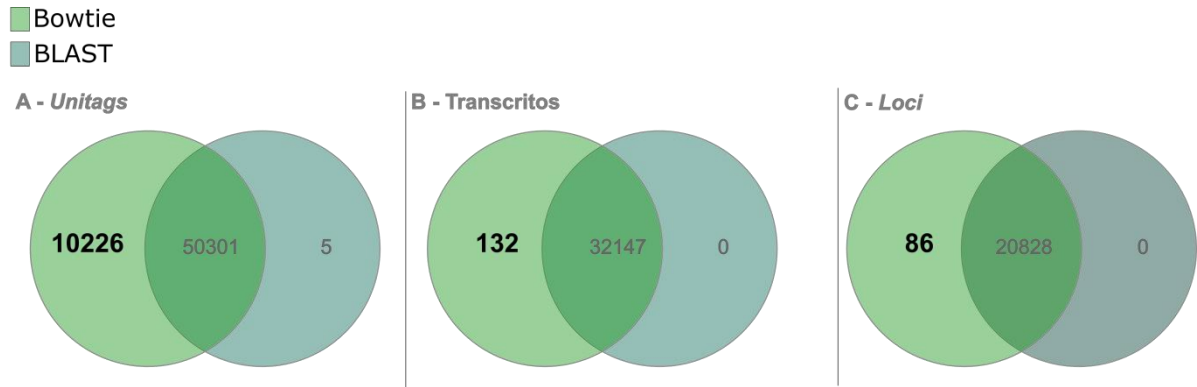
As *tags* HT-SuperSAGE sequenciadas das quatro bibliotecas (14.636.375 do cultivar sensível SI e 11.630.375 do cultivar tolerante PO), após exclusão dos *singletons*, resultaram em 149.305 *unitags* (*tags* distintas) disponíveis ao estudo. A análise BLASTn entre essas *unitags* e os 41.089 transcritos mRNA de referência (GCF_004118075.1), correspondentes a 28.230 *loci*, proporcionaram 96.822 alinhamentos aceitáveis, e envolveram 50.306 *unitags* (33,7% do total) e 32.147 transcritos distintos (78% dos previstos), abrangendo 20.828 *loci* (73,8% dos previstos). Um total de 72.711 *unitags* não apresentaram similaridades significativas com os transcritos (Tabela 5). Alternativamente, por meio da estratégia com a ferramenta Bowtie, 102.810 alinhamentos aceitáveis mapearam 60.527 *unitags* (40%) em 32.279 (78,5%) transcritos de 20.914 *loci* (74%), com 88.778 *unitags* não apresentando similaridades relevantes com as sequências (Tabela 5). Assim, a análise com Bowtie envolveu maior número de *unitags* (10.2226, Figura 7A), transcritos (132, Figura 7B) e *loci* (86, Figura 7C), além de ter detectado inclusive aqueles observados na análise BLAST (Figura 7).

Tabela 5 – Resultados de diferentes estratégias (BLASTvs Bowtie) relacionando *unitags* SuperSAGE com transcritos mRNA RefSeq de feijão-caupi

	BLAST	Bowtie	Disponíveis
<i>Unitags</i>	50.306 (33,7%)	60.527 (40%)	149.305
Transcritos	32.147 (78%)	32.279 (78,5%)	41.089
Loci	20.828 (73,8%)	20.914 (74%)	28.230

Fonte: O autor (2021)

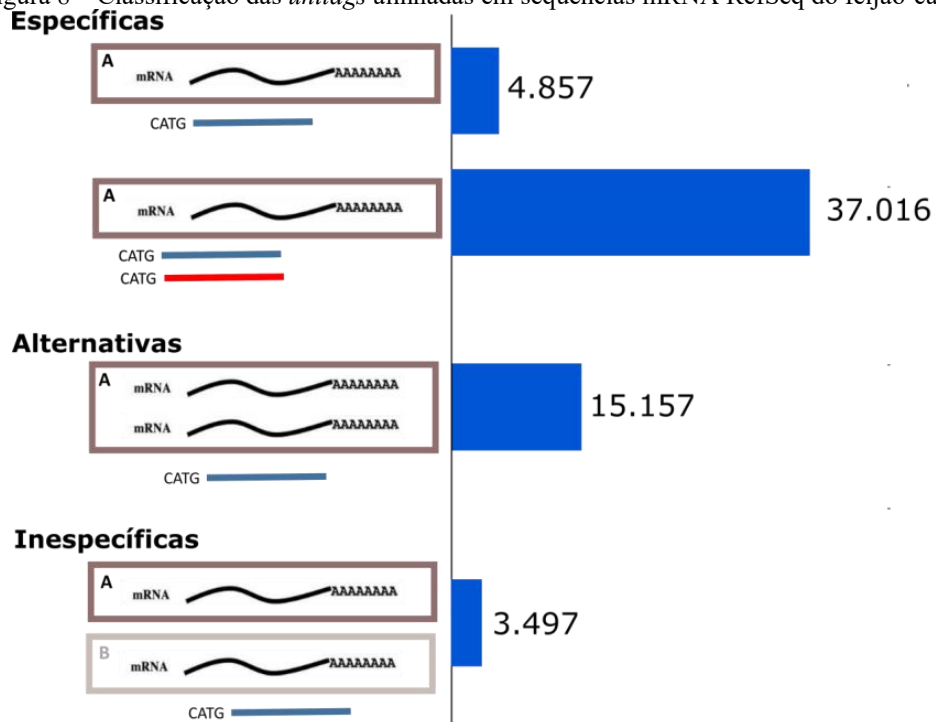
Figura 7 – Diagramas de Venn destacando números de: *unitags* (A), transcritos mRNA RefSeq (B) e *loci* (C) de *Vigna unguiculata* exclusivos e compartilhados entre duas análises (BLASTe Bowtie). Em negrito, aqueles exclusivos da estratégia Bowtie



Fonte: O autor (2021)

Considerando as 60.527 unitags mapeadas pelo Bowtie em sequências mRNA RefSeq de feijão-caupi, 41.873 mapearam em sequências únicas, mas destas somente 4.857 apresentaram relação unitag - transcrito de 1:1. As 37.016 restantes apresentaram a mesma relação >2:1 (duas ou mais unitags associadas a um mesmo transcrito). Já 15.157 outras unitags apresentaram a relação unitag - transcrito de 1:>2 (unitags que alinharam com dois ou mais transcritos alternativos, de mesmo locus). Por fim, 3.497 não foram específicas a nível de gene ou transcrito (Figura 8).

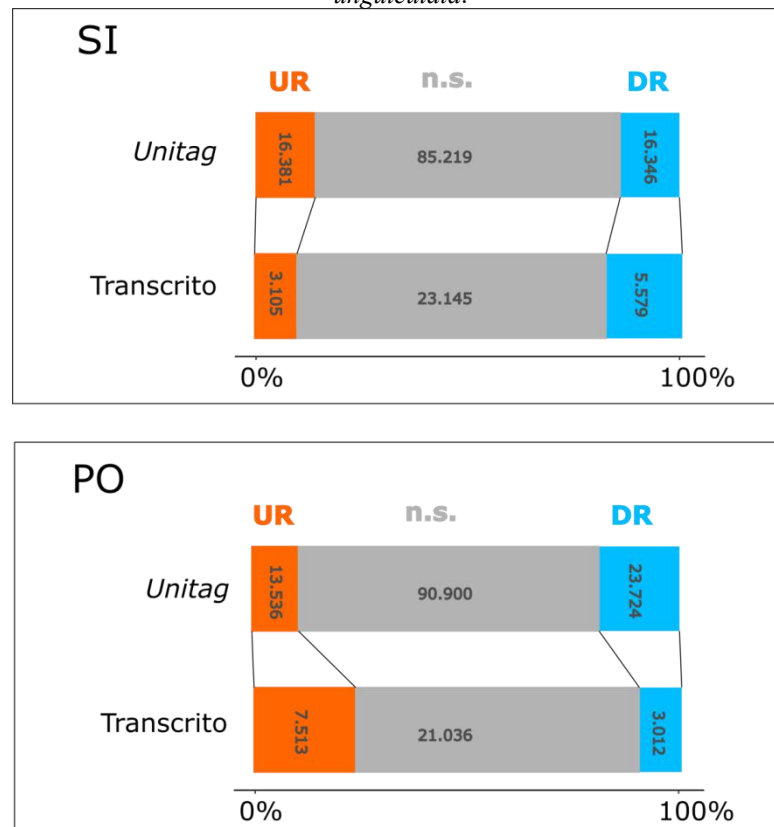
Figura 8 – Classificação das *unitags* alinhadas em sequências mRNA RefSeq do feijão-caupi



Fonte: O autor (2021)

Considerando as respostas dos cultivares ao estímulo aplicado, tendo por base as unitags, em SI a resposta foi uniforme entre repressão e indução. Já em PO, foram quase duas vezes mais unitags DR do que UR. A resposta predominante para ambos os cultivares evidenciou uma maioria de unitags que não modularam significativamente a expressão ($p < 0,05$). Entretanto, ao se avaliar a expressão diferencial dos transcritos contendo unitags SuperSAGE, o cultivar sensível SI reprimiu mais genes do que induziu, ao passo que o cultivar tolerante PO induziu mais do que reprimiu (Figura 9).

Figura 9 – Quantitativos referentes a modulação da expressão pelas *unitags* e pelos transcritos contendo as *unitags* nas respostas dos cultivares sensível SI (Santo Inácio) e tolerante PO (Pingo de Ouro 1-2) de *V. unguiculata*.



Fonte: O autor (2021)

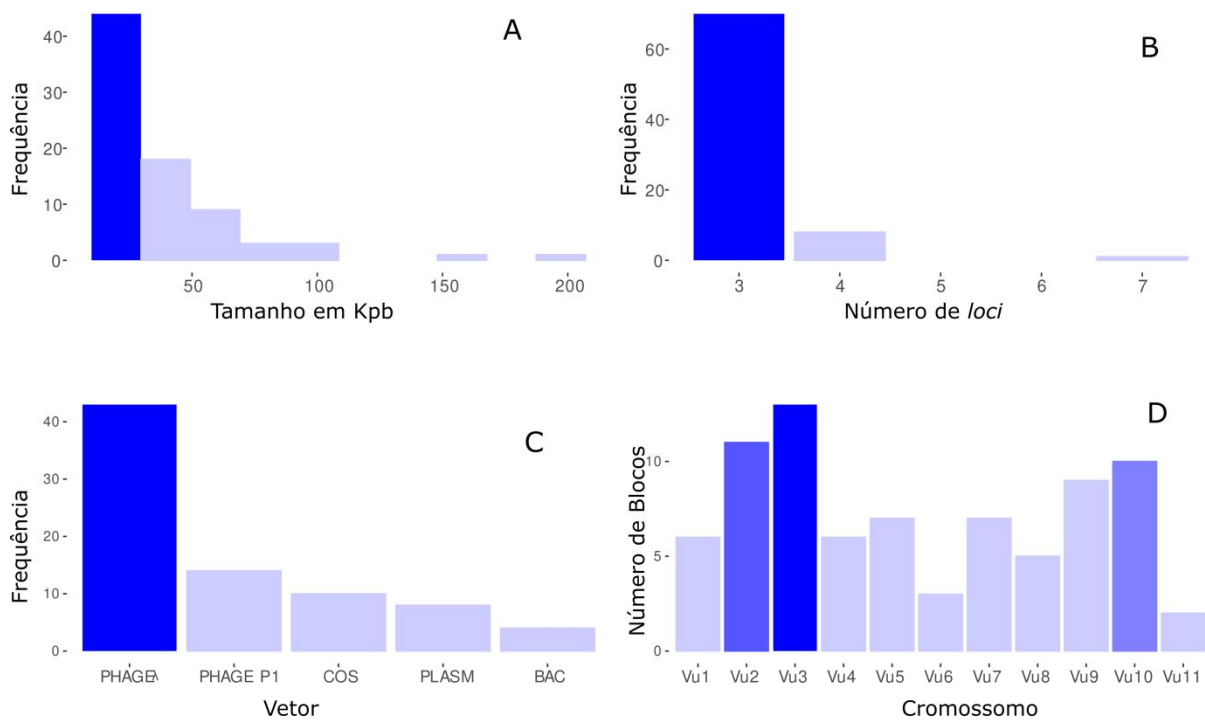
5.2 Os blocos de genes de co-expressão induzida no genoma do feijão-caupi

No total, 79 blocos de genes co-expressos foram identificados nos 11 cromossomos do feijão-caupi (44 na fita *forward* e 35 na *reverse*), abrangendo 249 *loci* distintos, cujos os genes foram DE (UR) em PO, após o estímulo aplicado. Os blocos variaram de 10,87 a 188,39 kpb (média de 37,23 kpb), sendo a maioria deles com até 50 kpb (Figura 10A). Esta maioria apresentou três ou quatro *loci*, com um bloco tendo apresentado sete *loci*, todos genes

chalcone synthase (CHS17) (Figura 10B). A maioria dos blocos comportaria em vetores de transformação do tipo fago λ (Figura 10C), no caso do interesse na introdução conjunta deles. O cromossomo com o maior número de blocos foi o *Vu3* (13), seguido por *Vu2* (11) e *Vu10* (10) (Figura 10D).

Em relação a expressão, os transcritos UR da resposta de PO apresentaram expressão UR ou n.s., com maior frequência do que DR, na resposta de SI (Figura 11), tendendo aqueles de maiores modulação de FC por PO também por SI, em uma correlação positiva. A anotação desses genes estão no **APÊNDICE F**.

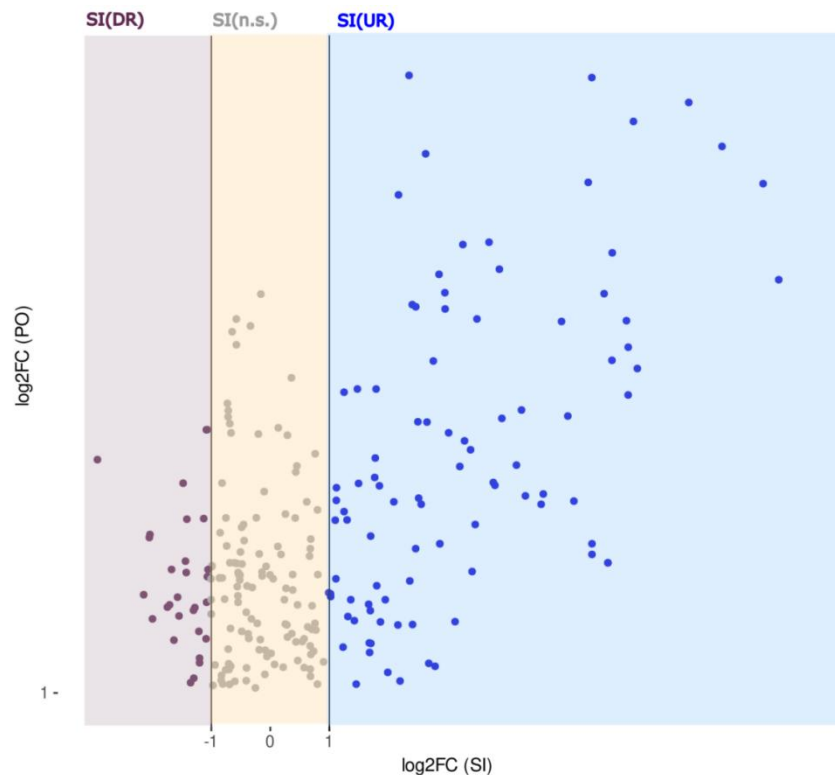
Figura 10 – Caracterização dos blocos de genes ligados e co-expressos identificados no genoma de *Vigna unguiculata*, com destaque para: A) tamanho dos blocos; B) número de *loci*; C) vetores que comportariam; D) número de blocos por cromossomo.



Fonte: O autor (2021)

Figura 11 – Gráfico de dispersão com base em valores de $\log_2FC$, dos componentes de blocos de genes co-expressos na resposta UR do cultivar tolerante Pingo de Ouro 1-2 (PO),

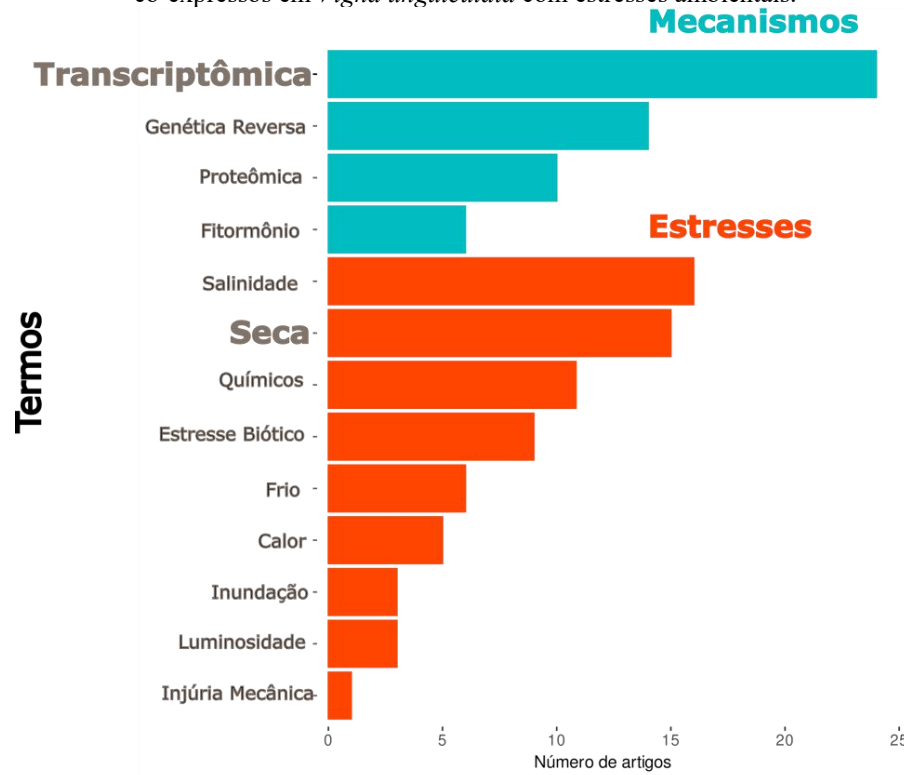
com expressão na resposta sensível do cultivar Santo Inácio (SI), após um estímulo de desidratação radicular (até 125 minutos), sendo $\log_2FC$ a razão entre a frequência normalizada na biblioteca tratamento *versus* controle negativo.



Fonte: O autor (2021)

A relação funcional, com base na literatura indicou 21 blocos (67 genes), em que todos os componentes (exceto aqueles não caracterizados) estiveram associados com estudos reportando estresses em plantas, tendo os termos transcriptômica e estresse de seca aparecidos frequentemente nos artigos (Figura 12). Esses incluem: *2-hydroxyisoflavanone dehydratase-like* (HIDH), *serine/threonine-protein kinase AFC2* (AFC2), *dihydroxy-acid dehydratase* (DHAD), *putative F-box protein PP2-B12* (PP2B12), *F-box protein PP2-B10* (PP2B10), *patatin-like protein 2* (PLP2), *cytochrome P450 CYP82D47* (CYP82D47), *receptor-like protein kinase HSL1* (HSL1), *subtilisin-like protease SBT1.6* (SBT1.6) e *cytochrome P450 94A1-like* (CYP94A1). Ademais, blocos em que todos os componentes, desconsiderando os componentes não-caracterizados, impactaram o fenótipo da planta, conforme os estudos de genética reversa, foram: C02.B03.+ [*ARM REPEAT PROTEIN INTERACTING WITH ABF2-like* (ARIA) e *hydroxyisoflavanone dehydratase-like* (HIDH)]; C03.B02.+ [*probabe WRKY transcription factor 51* (WRKY51) e *small GTPase LIP1* (LIP1)]; C03.B07.+ [dois loci *12-oxophytodienoate reductase 3-like* (OPR3)]; C08.B01.+ [três loci *patatin-like protein 2* (PLP2)]. Referências associadas aos genes dos blocos citados encontram-se no **APÊNDICE G**

Figura 12 – Termos mais frequentes extraídos de artigos científicos relacionando genes componentes de blocos co-expressos em *Vigna unguiculata* com estresses ambientais.

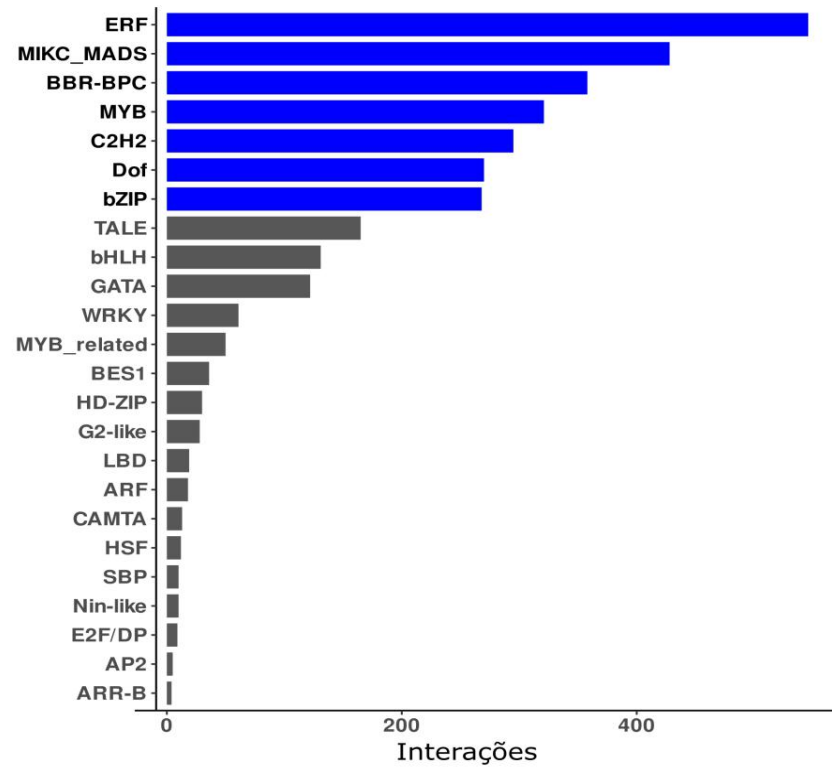


Fonte: O autor (2021)

5.3 Predição de FTs e de elementos *cis* regulatórios associados aos promotores de genes co-expressos nos blocos de *Vigna unguiculata*

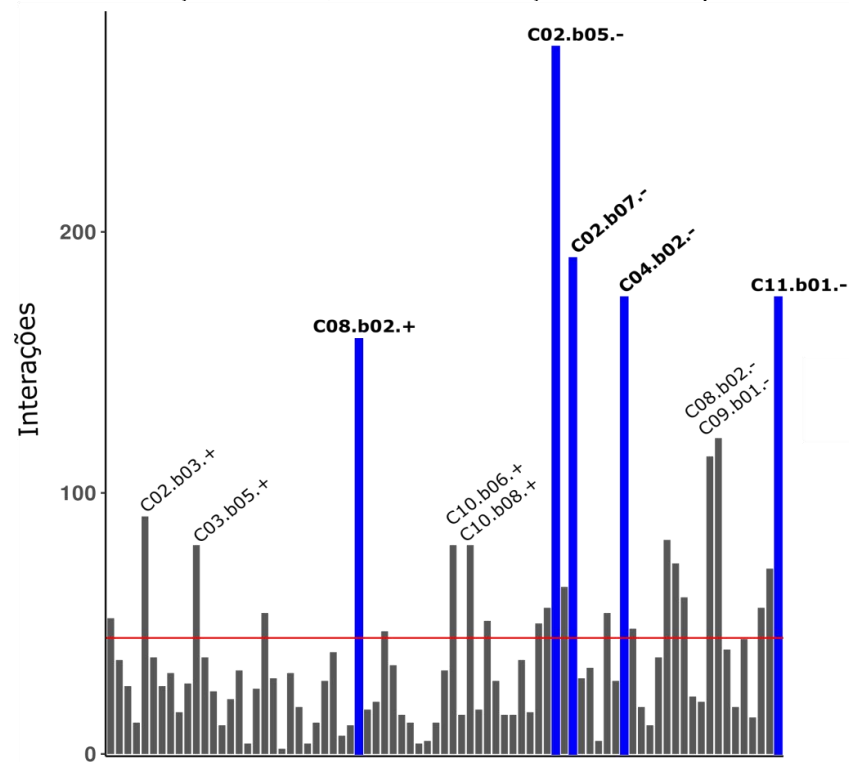
A busca por FTs interagindo com promotores dos 249 genes co-expressos dos blocos de *Vigna unguiculata* previu 1.847 interações entre FTs (259) e genes (221). Famílias com mais interações foram: ERF (546), MIKC_MADS (428), BBR-BPC (358), MYB (321), C2H2 (295), Dof (270) e bZIP (268) (em destaque na Figura 13). Os blocos potencialmente mais regulados pelos FTs, com base no número de interações estão representados na Figura 14.

Figura 13 – Famílias de fatores de transcrição com mais interações previstas com promotores dos genes componentes dos blocos de *Vigna unguiculata* (em azul, famílias com maior número de interações)



Fonte: O autor (2021)

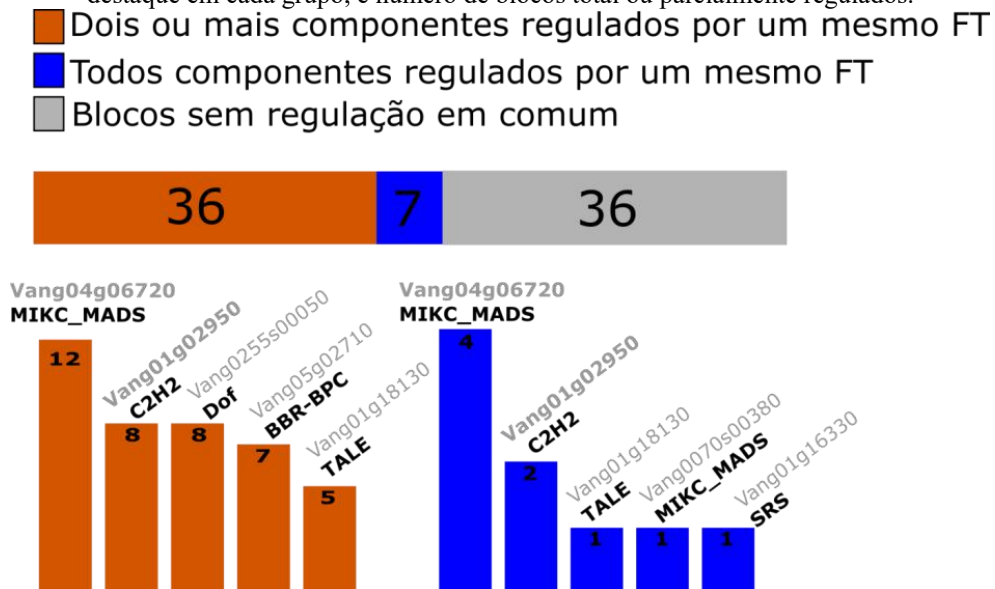
Figura 14 – Número de interações com fatores de transcrição por bloco. A linha vermelha indica o número médio de interações. Em azul, blocos com interações bastante superior a média



Fonte: O autor (2021)

Individualmente, em sete blocos (C01.b02+; C02.b03+; C09.b03+; C06.b01-; C08.b02-; C09.b03-; C11.b01-) todos os genes do bloco seriam reconhecidos por um mesmo FT, em 32 outros blocos, parte dos genes seriam reconhecidos por um mesmo FT, e em outros 32 blocos, genes não seriam reconhecidos por FTs em comum (Figura 15). Dentre os FTs com regulação prevista em comum (parcial ou total entre os componentes de um mesmo bloco), os FTs *Vang04g06720* e *Vang01g02950* (famílias MIKC_MADS e C2H2, respectivamente) se destacaram por sua maior capacidade de reconhecer elementos *cis* regulatórios de promotores dos genes identificados em blocos, cujos transcritos foram UR na resposta do cultivar PO (Figura 15).

Figura 15 – Potencial regulação gênica nos blocos: A) avaliação quantitativa dos blocos, cujos genes co-expressos seriam regulados por um ou mais fatores de transcrição (FTs); B) FTs (e respectivas famílias) em destaque em cada grupo, e número de blocos total ou parcialmente regulados.



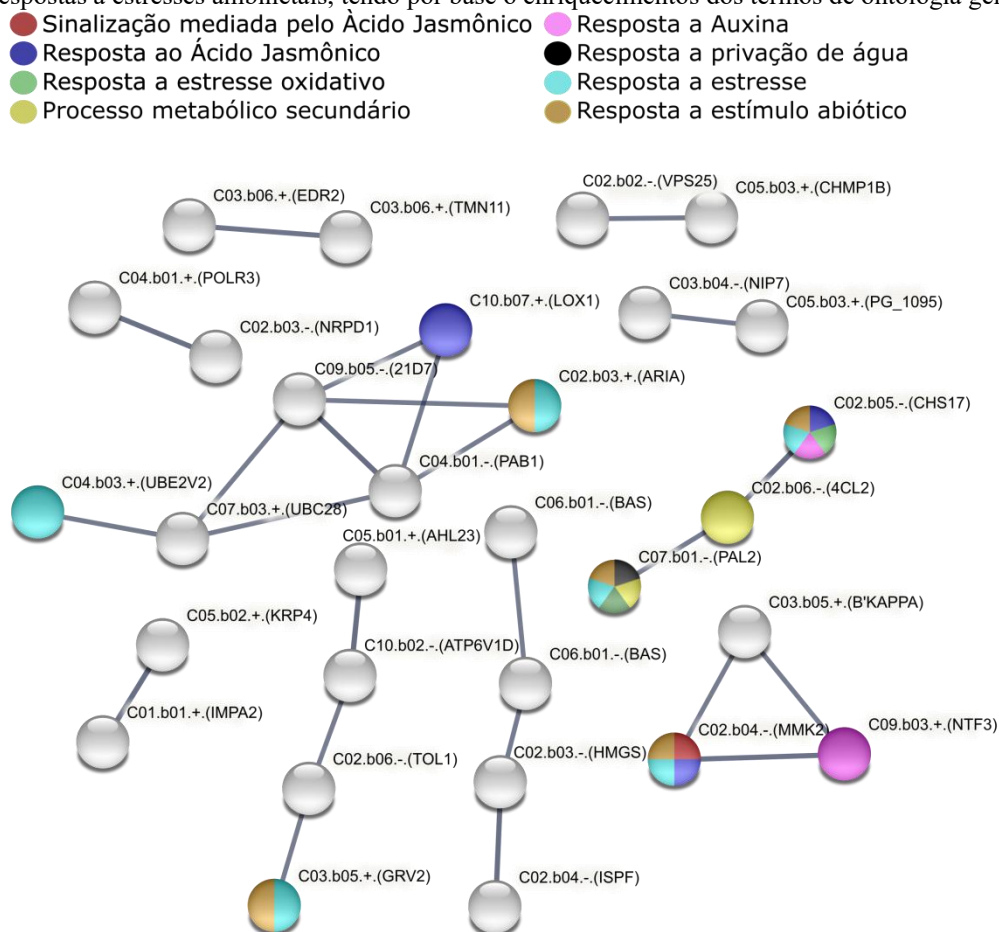
Fonte: O autor (2021)

5.4 Análise de interação proteína - proteína

Para a análise PPI, dos 249 transcritos co-expressos agrupados em blocos nos cromossomos de *V. unguiculata*, 229 foram associadas com proteínas de *A. thaliana* da base de dados *String*, totalizando 174 proteínas distintas. Dessas, 30 mostraram interações, com nível de confiança relevante entre elas, sendo sete [UBE2V2 (UBIQUITIN-CONJUGATING ENZYME E2V2), LOX1 (LINOLEATE 9S-LIPOXYGENASE 1), ARIA), CHS17 (CHALCONE SYNTHASE 17), 4CL2 (4-COUMARATE--COA LIGASE 2) e PAL2

(PHENYLALANINE AMMONIA-LYASE CLASS 2)] associadas direta ou indiretamente com respostas a estresses abióticos, conforme termos GO enriquecidos (Figura 16).

Figura 16 – Interações proteína - proteína previstas entre componentes dos blocos de co-expressão de genes ligados em *Vigna unguiculata*, segundo modelo em *A. thaliana*, com indicação de relação direta ou indireta com respostas a estresses abióticos, tendo por base o enriquecimento dos termos de ontologia gênica



Fonte: O autor (2021)

Interações envolvendo CHS17, 4CL2 e PAL2, se referem a biossíntese de metabólitos secundários, sendo 4CL2 fazendo a ligação entre a produção de flavonoides (interação com CHS17) e biossíntese de fenilpropanoides e metabolismo da fenilalanina (interação com PAL2). Evidências funcionais apontam para a co-expressão de 4CL2 com CHS17 (0,734) e também com PAL2 (0,145), além de co-menção em artigos relacionando-as com a produção de lignina e antocianinas.

Por sua vez, a interação entre MMK2 (MITOGEN-ACTIVATED PROTEIN KINASE HOMOLOG MMK2) e NTF3 (MITOGEN-ACTIVATED PROTEIN KINASE HOMOLOG NTF3), demonstrada experimentalmente (*two hybrid assay*, MI:0018) na via de sinalização

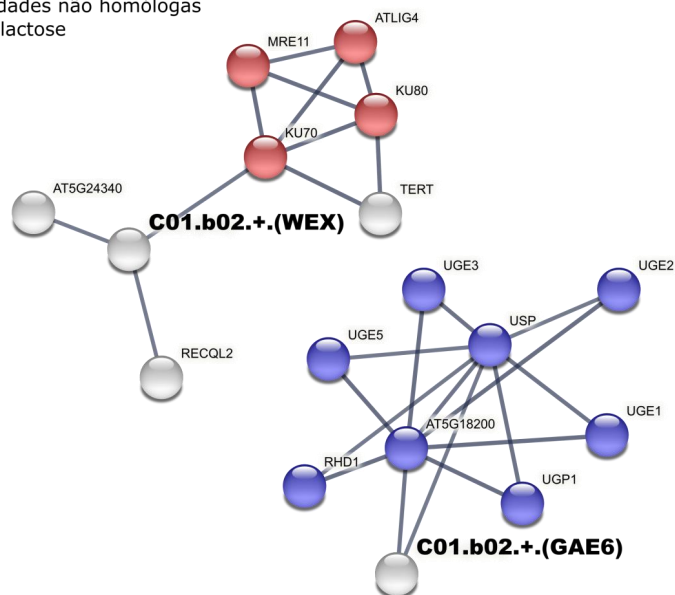
MAPK, apresentou em comum a fosfatase B'KAPPA (SERINE/THREONINE PROTEIN PHOSPHATASE 2A 57 KDA REGULATORY SUBUNIT B' KAPPA).

Ainda, as interações de LOX1, ARIA e UBE2V2 se basearam em associações em bancos curados (*Reactome*) em processos de catálise, complexo proteassoma 26S e exocitose de proteínas do lúmen do grânulo secretor, além da co-expressão de UBC28 (UBIQUITIN-CONJUGATING ENZYME E2 28) e UBEVE2 (0,045), que pertencem ao complexo Ubiquitina: enzimas de conjugação E2.

Componentes de três blocos (C01.B02.+; C03.B06.+; e C06.B01.-) apresentaram interações, com relevante nível de confiança, com proteínas externas àquelas 174 de *Arabidopsis* declaradas similares às de *V. unguiculata*. No bloco C01.B02.+, a proteína WEX (WERNER SYNDROME-LIKE EXONUCLEASE) interagiu com proteínas voltadas ao reparo do DNA (junção de extremidades não homologas), por meio do heterodímero KU (KU70/KU80), em vermelho na Figura 17. Por sua vez, GAE6 (UDP-GLUCURONATE 4-EPIMERASE 6) interagiu com proteínas relacionadas ao metabolismo de galactose (em azul na Figura 17). No bloco C03.B06.+, a proteína EDR2 (PROTEIN ENHANCED DISEASE RESISTANCE 2) e TMN11 (TRANSMEMBRANE 9 SUPERFAMILY MEMBER 11) interagiram com proteínas voltadas a biossíntese de tiamina (Figura 18). No bloco C06.B01.-, as duas BAS (BETA-AMYRIN SYNTHASE-LIKE) interagiram com proteínas das vias de biossíntese do sesquiterpenoide e backbone terpenoid (Figura 19).

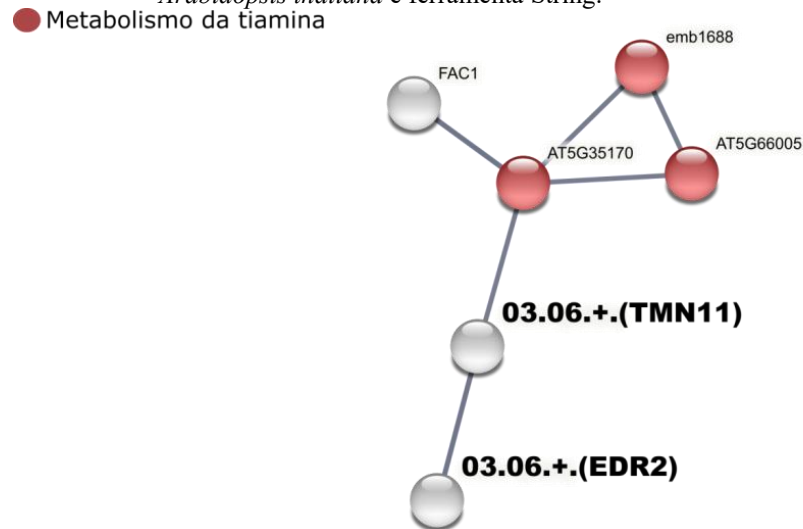
Figura 17 – Interações dos componentes do bloco C01.B02.+ (WEX e GAE6) de *Vigna unguiculata*, com proteínas de reparo do DNA (em vermelho) e do metabolismo da galactose (em azul), usando interações proteína-proteína do modelo de *Arabidopsis thaliana* e ferramenta String

● Junção de extremidades não homólogas
● Metabolismo da galactose



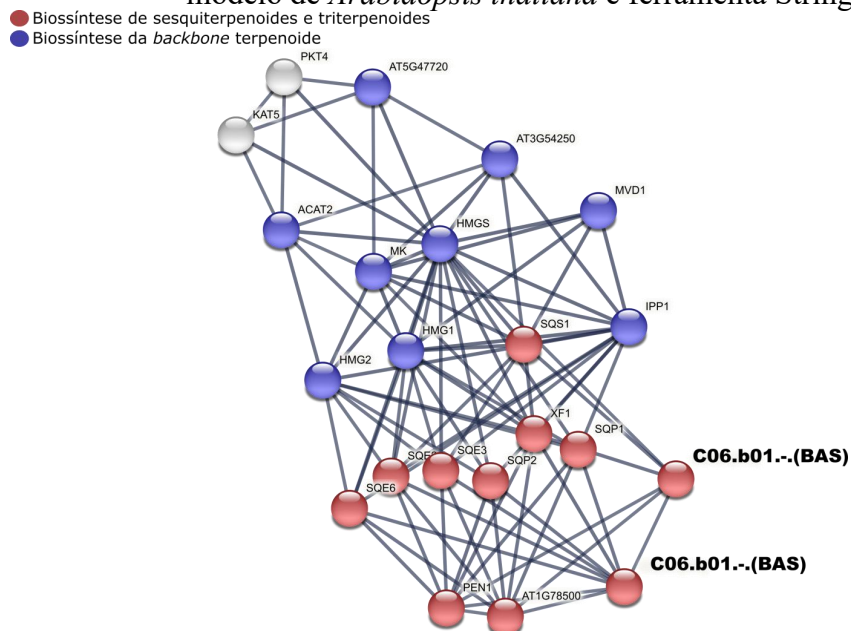
Fonte: O autor(2021)

Figura 18 – Interação dos componentes do bloco C03.B06.+ (TMN11 e EDR2) de *Vigna unguiculata*, com proteínas do metabolismo da tiamina (em vermelho), usando interações proteína-proteína do modelo de *Arabidopsis thaliana* e ferramenta String.



Fonte: O autor (2021)

Figura 19 – Interações dos componentes do bloco C06.B01.- (BAS) de *Vigna unguiculata*, com proteínas da biossíntese de sesquiterpenoides e triterpenoides (em vermelho) e com a biossíntese da *backbone* terpenoide (em azul), usando interações proteína-proteína do modelo de *Arabidopsis thaliana* e ferramenta String

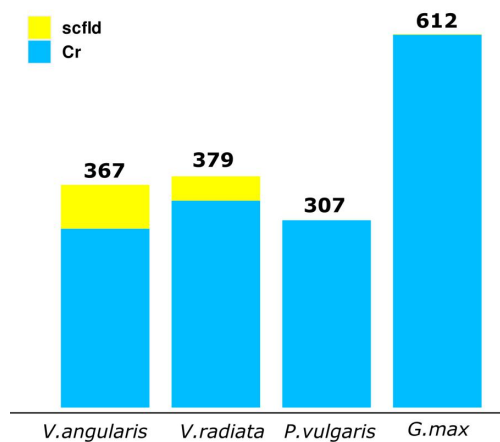


Fonte: O autor (2021)

5.5 Sintenia dos blocos de *Vigna unguiculata* em leguminosas diversas

Dos 249 transcritos de referência, de 249 *loci* de genes ligados nos 79 blocos de co-expressão de *V. unguiculata*, *G. max* foi a espécie com o maior número de ortólogos identificados (612), seguidos por *V. radiata* (379), *V. angularis* (367) e *P. vulgaris* (307), contudo, deve ser ressaltado que as espécies relacionadas de *Vigna*, ao contrário das demais, apresentaram uma parcela expressa de sequências mapeadas em *scaffolds* (Figura 20). Em relação aos 249 transcritos de *V. unguiculata*, relativos aos genes dos blocos identificados, aqueles representados nos genomas de *V. radiata* (216), *V. angularis* (209), *P. vulgaris* (207) e *G. max* (207) corresponderam a cerca de 83,1 - 86,7%. Dos blocos, 35 estiveram conservados nas quatro leguminosas, e aqueles mapeados em ao menos uma das espécies correspondeu a 65 dos 79 blocos (Figura 21). A lista dos blocos mapeados nas leguminosas é dada no **APÊNDICE H** Por fim, a relação de macrossintenia entre os cromossomos de *V. unguiculata* com cada uma das espécies consideradas e representado no gráfico da Figura 22.

Figura 20 – Número de potenciais ortólogos em diferentes leguminosas, associados aos 249 transcritos de genes componentes dos 79 blocos de co-expressão de *Vigna unguiculata*



Fonte: O autor (2021)

6 DISCUSSÃO

Dos fatores ambientais que afetam a produção das plantas, a seca é um dos mais agressivos. Desenvolver variedades tolerantes a esse estresse ambiental é fundamental para uma agricultura mais sustentável, tendo em vista o crescimento populacional e cenários que apontam para uma escassez de água no médio-longo prazo (Gupta *et al.*, 2020). A tarefa é um desafio, sobretudo devido a complexidade do caráter, o que limita os progressos dos programas de melhoramento (Semagn; Bjornstad; Xu, 2010). Logo, introduzir abordagens ômicas, aprimorando a seleção de genes que possam ser explorados na seleção assistida ou na geração de transgênicos, em complemento aos métodos clássicos, auxiliam os melhoristas e as probabilidades de sucesso dos programas de melhoramento vegetal (Dita *et al.*, 2006).

Os princípios adotados na estratégia do presente projeto consideraram: (i) a atividade transcricional de genes como indicativo de uma possível relação funcional entre eles, e evidências nas implicações ao estímulo aplicado (Abou-Elwafa, 2008); (ii) análises com várias marcas, atuantes como um haplótipo, promovem uma maior eficiência na identificação de potenciais *locos* controladores de caracteres quantitativos - QTLs (Boleckova *et al.*, 2012); (iii) quanto menor a distância entre os genes considerados, menores as chances de haver recombinação entre eles (Collard *et al.*, 2005), e (iv) marcadores funcionais são desejáveis para a seleção assistida por marcadores (SAM) por apresentarem associação característica-marcador perfeita (Gupta and Rustgi, 2004). Com base nisso, buscou-se os genes co-expressos e co-induzidos, nas respostas ao estímulo de desidratação radicular (de até 150 min.), de dois cultivares de *V. unguiculata*, um considerado tolerante (cultivar Pingo de Ouro, PO) e outro sensível (cultivar Santo Inácio, SI) à seca, e que estivessem ligados de forma consecutiva nos cromossomos do genoma de referência da espécie. A expressão gênica considerou transcritos portadores de *unitags* HT-SuperSAGE anteriormente obtidas e disponibilizadas (Amorin *et al.*, 2018; Kido *et al.*, 2011; Kido *et al.*, 2010), sendo aqueles diferencialmente expressos e de expressão UR na resposta tolerante de PO, possivelmente oriundos da ativação de genes de diferentes via de sinalização e/ou dos complexos efetadores importantes na resposta desejável de tolerância.

6.1 SuperSAGE: a expressão de genes a partir da contagem de *unitags*

Visando analisar funcionalmente os genomas, métodos transcriptômicos baseados em *tags* foram amplamente explorados em diversas espécies vegetais. Em relação a técnica SuperSAGE ou similares, diferentes estratégias são empregadas para a identificação *in silico* do transcrito que originou a *tag*. Duas ferramentas utilizadas com essa finalidade correspondem ao BLAST (Girma *et al.*, 2019; Barrios *et al.*, 2017) e ao Bowtie (van Aerle *et al.*, 2013; Soanes *et al.*, 2012). No presente estudo, a metodologia com Bowtie incorporou as mesmas *unitags* (*unique tags*), transcritos e *loci* que a estratégia com BLAST, porém indo além, envolvendo também *unitags* complexas (com repetições em *tandem* ou de uma mesma base) que não foram detectadas pelo BLAST. Talvez, devido a sua natureza de alinhador local, ao passo que o Bowtie é otimizado para mapear pequenas *reads* em genomas por indexação, atribuindo mais eficientemente essa função. Entretanto, independente do método usado, menos da metade das *unitags* já representou quase a totalidade do transcriptoma RefSeq. Essa diferença é muito discrepante para ser atribuída a possíveis artefatos de sequenciamento e/ou variações intraespecíficas entre o acesso africano IT97K-499-35, e os cultivares brasileiros PO e SI. Logo é possível que a atual montagem do genoma de referência no NCBI, em relação aos transcritos, não refleta o potencial transcriptômico do feijão-caupi, e novos genes/transcritos podem ser descobertos.

O protocolo SuperSAGE prevê a geração de uma única e exclusiva *tag* para cada cópia de cDNA, sendo esta a mais próxima da cauda poli A (Matsmura *et al.*, 2008), em uma relação *unitag* - transcrito de 1:1. Na realidade, esse resultado esperado ocorre em uma pequena parcela das *unitags* alinhadas. Possíveis explicações incluem: digestão parcial dos cDNAs pela enzima *NlaIII*, ocasionado pelo reconhecimento desigual dos sítios CATGs em cópias de um mesmo cDNA, ou devido a sítios alternativos de poliadenilação no *splicing* alternativo (Van Aerle *et al.*, 2013; Hegedus *et al.*, 2009; Gowda *et al.*, 2004). De modo geral, a contribuição das *unitags* que alinham de forma apropriada em mais de um transcrito, sejam eles de mesmo *loci* ou de *loci* diferentes, são normalmente desconsideradas da expressão do transcrito e tratadas individualmente, como atestam estudos anteriores com as *tags* SuperSAGE (Amorin *et al.*, 2018; Kido *et al.*, 2011; Kido *et al.*, 2010). No presente caso, as frequências das *unitags* alinhadas em um mesmo transcrito contribuíram para se estimar a abundância do próprio transcrito, sem efetuar a escolha de uma *unitag* para representar o transcrito, seja ela a mais próxima da cauda poli A ou a mais abundante. Entretanto, isso acarretaria uma super representação daquelas *unitags* alinhadas em mais de um *locus*, sendo esta minimizada ao se normalizar as bibliotecas pelo total de *tags* sequenciadas, ao invés das alinhadas. Assim, essas *unitags* contribuiriam com $1/n$ da contagem total, sendo n o número

de transcritos alinhados com a mesma *unitag*. É importante ressaltar que essas *unitags* podem atuar como marcadores a nível de isoformas ou família gênica. Para a expressão diferencial, o teste estatístico escolhido consistiu de uma atualização do método Audic, sendo agora mais adequado para a grande massa de dados gerados por NGS, e casos sem réplicas, conforme formulação matemática revisada (Claverie and Ta, 2019). Também, a distribuição de Poisson proposta parece ser ideal para a modelagem dos dados obtidos, tendo em vista que o número de *tags* geradas de um determinado gene ou transcrito é uma variável aleatória discreta, que se encontra dentro da premissa da distribuição (número de eventos em um intervalo específico), sendo utilizada em muitos estudos com SuperSAGE (Ayra-Pardo *et al.*, 2019; Barrios *et al.*, 2017; El Kelish *et al.*, 2014). Ainda, o fato de pacotes especializados para cálculo de expressão diferencial, como o Deseq e EdgeR, não terem apresentado resultados promissores (dados não mostrados), possivelmente devido a falta de réplicas, reforçam o uso da metodologia proposta. Por fim, valores de p foram corrigidos para erros do tipo I, normalmente não aplicados nos estudos iniciais com SuperSAGE.

Desta forma, avaliar a expressão gênica pelas *unitags* individualmente ou com contagens atribuídas para o transcrito revelaram cenários distintos. Individualmente, as *unitags*, apesar de informativas por si só, não evidenciaram o contraste do cultivar tolerante PO em relação ao sensível SI após o estímulo aplicado. Mas, ao se avaliar os transcritos contendo as *unitags*, houve maior indução de genes por PO do que por SI, denotando haver em SI gargalos genéticos reprimindo a expressão de genes supostamente importantes na tolerância. Logo, essas respostas estariam de acordo com informações prévias do melhoramento, que reportam PO como tolerante e SI como sensível à seca (Rodrigues *et al.*, 2018), reforçando a nova metodologia e a viabilidade de caracterizar funcionalmente os cultivares, bem como a prospecção por genes associados a tolerância.

6.2 Perspectivas na geração de marcadores funcionais aplicados a Seleção Assistida

Uma vez que marcadores moleculares são associados a características de interesse, há potencial de aplicá-los no melhoramento, na seleção indireta de cultivares para melhor direcionar os cruzamentos e otimizar as buscas pelos fenótipos desejados (Morgante and Salimini, 2003; Collard *et al.*, 2005). Os métodos tradicionais procuram detectar a correlação fenótipo-marcador molecular usando populações biparentais (métodos baseados em *linkage*) ou em populações naturais (estudos associativos) (Semagn; Bjornstad; Xu, 2010). A segunda opção apresenta maior resolução ao explorar eventos de recombinação acumulados ao longo

da história evolutiva de um conjunto de indivíduos, não se limitando aos *loci* que diferem os parentais (Sorkheh *et al.*, 2008).

Estudos de QTL no feijão-caupi não chegam a identificar genes individuais (Chamarthi *et al.*, 2019), logo, encorajar estudos associativos ajudam a superar essa limitação de uma forma menos custosa e trabalhosa do que gerar mapas genéticos de alta resolução, os quais dependem de cruzamentos complexos e fenotipagem robusta (Wang *et al.*, 2019). Esses estudos podem focar em regiões cromossômicas abrangendo inúmeros genes (associação de amplo genoma) ou ser direto no gene candidato que afeta o fenótipo (Zhu and Zao, 2007). Embora mais poderoso e econômico, estudos associativos por genes candidatos dependem de hipóteses prévias, baseadas no conhecimento molecular da característica em questão (Zhu and Zao, 2007). Assim, é de grande importância desenvolver abordagens alternativas que eliminem a limitação imposta por esses gargalos de informações (Zhu and Zao, 2007; Abou-Elwafa, 2018).

Nesse sentido, estratégias podem ser somadas, como pensar em marcadores funcionais, aqueles nos quais polimorfismos estão dentro de sequências funcionalmente caracterizadas, o que elimina a perda da ligação entre elas por recombinação (Gupta and Rustgi, 2004; Andersen and Lübberstedt, 2003), em conjunto com o uso de marcadores haplótipos, que são mais precisos em análises de QTL do que aquelas de marcas simples (Boleckova *et al.*, 2012). Logo, identificar marcadores potencialmente funcionais e fortemente ligados no genoma pode ser uma alternativa interessante em estudos desta natureza. Desta forma, reconhecer um *pool* de genes candidatos, baseados na forte ligação física, maximizando o estado de desequilíbrio de ligação (associação não aleatória entre alelos), com base na expressão diferencial ao estímulo abiótico, e na anotação funcional favorável, além de fugir da tendência observada, é uma nova maneira de propor genes candidatos ao mapeamento por associação, não se restringindo àqueles normalmente disponíveis ou anônimos. A ligação dos genes de co-expressão induzida garantiria maior co-herança dos alelos marcadores, minimizando a perda destes por recombinação meiótica, sendo interessantes para a aplicação da SAM.

6.3 Perspectiva funcional dos genes dos blocos de co-expressão induzida

Afim de propor os componentes dos blocos de co-expressão induzida como potenciais marcadores funcionais suas relações com estresses foram determinadas a partir de estudos disponíveis na literatura. Assim, dos artigos reportando esses genes, a maior parte deles remete a abordagens ômicas, com enfoque descritivo, onde são geradas listas de transcritos

ou proteínas que mudam sua abundância entre diferentes condições. No entanto, esses não são tão informativos em termos de função gênica. Por outro lado, outros apresentaram o termo “genética reversa”. Caso do bloco C02.b03.+, cuja superexpressão de *ARIA* levaria a um aumento na sensibilidade ao ácido abscísico (ABA), e o envolvimento deste nas respostas a estresses dependente do hormônio (Kim *et al.*, 2004), conferindo maior tolerância ao estresse abiótico (salino), enquanto a superexpressão de *HIDH* possibilitaria um acúmulo de isoflavonas (Shimamura *et al.*, 2007), metabólitos secundários (flavonoides) com atuação direta na redução de danos oxidativos (Amic *et al.*, 2007). No caso do bloco C03.b02.+, a superexpressão de *WRKY51* levaria a uma maior formação de raízes laterais, importantes na absorção de água e nutrientes, e na regulação negativa da biossíntese de etileno, ETH (Hu *et al.*, 2018), ao passo que o outro gene ligado (*LIP1*) mostrou ser importante na resposta salina, pois mutantes com perda de função mostraram-se hipersensíveis ao estresse (Terecskei *et al.*, 2013). No bloco C03.b07.+, a atividade antioxidante dos dois *loci* *OPR3* se ressaltaria, pois mutantes *OPR3* mostraram-se reprimidos nesta atividade, quando expostos a O₃ em estresse sinalizado por jasmonato, JA (Sasaki-Sekimoto *et al.*, 2005). Logo, blocos com componentes assim associados são promissores pois possuem informações de impactos definidos em fenótipos favoráveis, situação normalmente não observada em estudos somente com abordagens ômicas. A busca de polimorfismos nesses genes é passo futuro, tendo em vista a definição de MFs, como polimorfismos em sequências funcionalmente relevantes para a característica de interesse (Andersen and Lübberstedt, 2003).

Ainda na caracterização dos genes dos blocos co-expressos induzidos, a interação com fitormônios, tipo ABA, ETH e JA, já comentados anteriormente para *ARIA* (C02.b03.+), *WRKY51* (C03.b02.+) e *OPR3* (C03.b07.+) se somam aos exemplos de acúmulo de citocininas (CK) e de auxinas (AUX), observados nas raízes de tabaco superexpressando *ZOG1* (C07.b02.+), que após supressão da rega evidenciaram ações destes fitormônios no ambiente radicular induzido pela seca (Havlová *et al.*, 2008), bem como na ação prolil peptidil cis-trans isomerase de CYP21-1 (bloco C09.b04.+), que desencadeiaria uma sinalização via ABA, com reflexos na germinação e crescimento de mudas, em resposta a estresses abióticos (100mM NaCl, 200mM manitol, 4°C e 37°C) (Jung *et al.*, 2020). Visto serem os fitormônios centrais em coordenar a transdução de sinais durante respostas aos estresses abióticos, suas interações são importantes, pois podem interagir com vários genes/produtos favoráveis (Wani *et al.*, 2016). Assim, os blocos e os genes acima mencionados, com funções comprovadas por ensaios de genética reversa, são bons candidatos para exploração nos programas de melhoramento de plantas.

Com relação aos estresses relacionados aos genes componentes dos blocos de co-expressão induzida, muitos, artigos denotaram a predominância de fatores ambientais, como seca, temperatura, salinidade e compostos químicos, os quais são comumente conectados por estresses secundários, como oxidativos e osmóticos (Wang; Vinocur; Altman, 2003). O mesmo ocorre com estresses de natureza biótica, em que uma ação cruzada ou *croostalk* entre sinalizadores, que regulam genes tanto de respostas bióticas como abióticas, leva a um mesmo efetuator (Ben Rejeb *et al.*, 2014; Fujita *et al.*, 2006). Por exemplo, proteínas PR (blocos C03.b03.- e C03.b09.+), que são rapidamente acumuladas em plantas sob ação de patógenos (van Loon; Rep; Pieterse, 2006), também apresentam atividade anticongelamento (Stressmann *et al.*, 2004; Hon *et al.*, 1995) e contribuem na tolerância a seca (Lee *et al.* 2008). Logo, componentes de bloco sem estudos específicos na resposta a seca, podem ainda ser interessantes, em razão de uma potencial sobreposição de tarefas em alguma das etapa da resposta abiótica (percepção e transdução de sinal, controle da transcrição, e resposta efetiva), visando reestabelecer a homeostase celular normal e proteger estruturas sob estresse (Wang; Vinocur; Altman, 2003).

Comparando os dados aqui obtidos de expressão UR do cultivar tolerante PO com aqueles da literatura de plantas expostas ao estresse de seca, ainda que de tratamentos distintos, se observou concordância em oito de dez genes (*HIDH*, *AFC2*, *DHAD*, *PP2B12*, *PP2B10*, *PLP2*, *CYP82D47*, *HSL1*, *SBT1.6* e *CYP94A1*) também induzidos nos estudos transcriptômicos, em respostas a seca de variedades tolerantes de espécies distintas, como batata doce, amendoeira, feijão-azuki, mandioca, algaroba ou o trigo rastejante mongol (*Mongolian wheatgrass*). Os dois casos de discordância envolveram os genes *SBT1.6* e *CYP94A1*, que foram reprimidos em folhas de mandioca e trigo rastejante mongol (plantas tolerantes), após 15 e 20 dias de supressão de rega, respectivamente (Xiao *et al.*, 2019; Ma *et al.*, 2018). A expressiva concordância de expressão com a literatura reforça a metodologia aplicada, e permite hipotetizar a importância destes genes/blocos no caráter de tolerância, de forma até precoce, pois o presente estudo detectou as expressões UR após até 150 min. do estímulo aplicado, ao contrário dos estudos da literatura, todos mais tardios (de 12 horas até 21 dias), em diferentes tecidos (folha, raiz e semente), e tratamentos (PEG, supressão da rega, manitol).

6.4 Possíveis interações proteína-proteína

A co-expressão diferencial, os estudos da literatura, e a ligação física sugerem a atuação de componentes e blocos como possíveis QTLs para tolerância a seca ou outros estresses abióticos. A interação entre alelos de diferentes QTLs em plantas é um dos fatores que torna complexo estudar características regidas por múltiplos genes (Semagn; Bjornstad; Xu, 2010). Afim de simular o funcionamento de um QTL no conjunto de dados, se buscou evidências funcionais que relacionassem as proteínas relativa aos genes dos blocos identificados. A previsão de rede PPI com essas proteínas, apesar de não ter envolvido muitas delas, apresentou mais interações do que o esperado para um conjunto aleatório do mesmo tamanho, reforçando as poucas interações neste seletor grupo. Neste sentido, ao se avaliar os termos GO enriquecidos associados a respostas a estresses abióticos do conjunto interativo CHS17, 4CL2 e PAL2, relacionando a biossíntese de flavonoides com a de fenilpropanoides, se sobressai as atuações mitigadoras de danos oxidativos (Fini *et al.*, 2012; Amic *et al.*, 2007). O conjunto interativo MMK2 e NTF2 é parte da via de sinalização MAPK, importante na percepção de estresse abiótico em plantas (Munnik and Meijer, 2001), e está associado a *crosstalk* na sinalização por AUX e ABA. Outra proteína que também se relaciona com fitormônios e se destacou em termos GO de respostas a estresses abióticos foi ARIA. Logo, a análise de interação proteína-proteína e o enriquecimento de termos GO permitiu relacionar funcionalmente produtos de alguns dos genes presentes nos blocos de co-expressão induzida.

As interações previstas dos componentes de cada bloco individualmente, com outras proteínas da célula, para tentar prever que processos seriam afetados em possíveis eventos de transformação com os blocos de co-expressão induzida, mostrou somente três casos previstos de interações. Isso pode ter decorrido do *input*, em geral, menor do que cinco proteínas, o que aumentaria a chance do conjunto ser considerado randômico, dificultando o encontro das interações, o que não significa que estas não existam ou não se tenha sentido biológico algum. No caso do conjunto interativo TMN11 e EDR2 (bloco C03.b06.+), a interação é suportada por *data-mining* (Shangguan *et al.*, 2013). No caso das duas BAS (bloco C06.b01.-), a interação se dá na biossíntese de terpenos, e no caso da proteína WEX (bloco C01.b02.+), o envolvimento se dá com mecanismos de reparo do DNA, por intermédio do complexo KU70/80, o qual estimula a ação exonuclease, sendo que a superexpressão de *WEX* conferiria maior tolerância aos danos do DNA (West *et al.*, 2002), fator interessante nas respostas a estresses agressivos aos ácidos nucleicos.

6.5 Perspectivas na manipulação por engenharia genética

Uma sugestão do uso dos genes de blocos de co-expressão induzida é usá-los em conjunto, como haplótipos, em estudos associativos (principalmente aqueles de função assinalada, como potenciais marcadores funcionais). Neste sentido, foi importante verificar o tamanho do possível inserto (bloco) e compará-lo quanto a capacidade de abrigo dos potenciais vetores de transformação. Entretanto, há necessidade ainda de se avaliar os sítios de restrição presentes e a possibilidade de colocá-los sob influencia de promotores fortes, em vetores de expressão, afim de superexpressá-los em conjunto, o que minimizaria a necessidade de realizar a pirimidização de genes e de eventos de transgenia.

Outra possibilidade de manipulação decorrente do estudo foi propor FTs como transgenes, para superexpressar potenciais genes de co-expressão induzida dos diferentes blocos. Neste sentido, alguns blocos foram identificados como altamente regulados pela expressiva interação com FTs. Essa regulação fina vai de encontro com o esperado para genes responsáveis a estresses, o que reforça a importancia deles. Foi possível identificar ainda FTs regulando todos os componentes do bloco, o que vai de acordo com a co-expressão. Uma porção relevante dos blocos apresentou parte dos componentes reconhecidos por FTs em comum. Mesmo nesses casos a estratégia de superexpressão indireta por FT poderá ser viável, principalmente se envolver *Vang01g02950* (família FT C2H2) ou *Vang04g06720* (família FT MIKC_MADS), que se destacaram pelas interações previstas em vários blocos.

Também promissores são membros das famílias/subfamílias ERF, bZIP e MYB que representam FTs com atuação reconhecida frente a resposta ao estresse de seca (*Drought Stress Gene Database*; Alter *et al.*, 2015), fazem parte das famílias com mais interações previstas.

Também não se descarta a transformação de genes individuais, afim de determinar experimentalmente sua função ou no caso daqueles que se mostraram relevantes nas análises realizadas, como *ARIA* e *CHS*, por exemplo.

6.6 Perspectivas em genômica comparativa

A expressiva conservação da organização gênica em plantas possibilita extrapolar a informação gerada de espécies mais bem estudadas para culturas consideradas orfãs ou mesmo unir conhecimentos de pesquisas independentes, em espécies sexualmente incompatíveis. Assim, a identificação de genes controladores de características quantitativas em uma espécie faz com que eles sejam também candidatos em outras filogeneticamente relacionadas (Gale and Devos, 1998).

O relevante grau de compartilhamento dos blocos identificados entre as quatro espécies de leguminosas foi de acordo com o esperado, permitindo os programas de melhoramentos das referidas espécies usufruir das informações geradas. Interessante foi o caso de *ARIA*, membro do bloco C03.b02.+, que se mostrou um componente relevante nas análises, não ter apresentado ortólogos identificáveis. Contudo, a relevância e influência de genes orfãos na resposta a seca em feijão-caupi já foi demonstrada (Li *et al.*, 2019). Dos blocos discutidos anteriormente, três foram representados em todas as espécies de leguminosas avaliadas, dos quais C02.b06.- e C07.b01.-, que apresentaram os produtos 4CL2 e PAL2, respectivamente, fazendo parte do conjunto interativo envolvido na biossíntese de metabólitos secundários (conectando a produção de flavonoides com a biossíntese de fenilpropanoides), e C03.b02.+, cujos os produtos gênicos (WRKY e LIP1) foram associados em estudo de genética reversa. Outros blocos compartilhados, que merecem destaques, e apresentaram ortólogos nas espécies relacionadas foram: (i) C09.b04.+ (relação de CYP21-1 com fitormônios), em *V. angularis*, *V. radiata* e *G. max*; (ii) C08.b01.- (estudos de genética reversa de PLP2), com as duas *Vignas spp.*; (iii) C02.b05.- (termos GO enriquecidos de CHS17), em *G.max*; (iv) C07.b02.+ (relação com fitormônios de ZOG1), em *V. angularis*; (v) C03.b07.+ (estudos de de genética reversa de OPR3), em *V. radiata*. Os resultados aqui obtidos reforçam estudos recentes acerca da sintenia entre *V. unguiculata*, *V. radiata* e *P. vulgaris* (Lonardi *et al.*, 2019; Muñoz-Amatriaín *et al.*, 2017), em que regiões de macrossintenia coincidem com a relação dos cromossomos em *V. unguiculata* - *V. radiata*, e *V. unguiculata* - *P. vulgaris*. Curiosamente, isso não foi observado em *V. angularis*, talvez devido a uma mudança na nomenclatura. Com relação a *G. max*, a relação foi de dois *loci* para um na maioria dos cromossomos de *V. unguiculata*, o que se justifica tendo em vista que a soja sofreu dois eventos de duplicação de seu genoma ao longo do tempo evolutivo (Schmutz *et al.*, 2010).

No geral, os blocos de co-expressão induzida permaneceram conservados nas espécies filogeneticamente relacionadas de feijão-caupi, principalmente *V. angularis* e *V. radiata*, bem como naquelas economicamente mais importantes (*P. vulgaris* e *G. max*), abrindo a

possibilidade de se aplicar os conhecimentos aqui gerados em seus respectivos programas de melhoramento, visando minimizar o tempo na geração destas informações e no tempo decorrido para o desenvolvimento de cultivares mais tolerantes as condições adversas de estresse.

Espera-se que fornecer a priori blocos de genes: funcionalmente relacionados com o estresse de seca (conforme dados de expressão e triagem na literatura), passíveis de serem marcadores funcionais, a depender da presença de polimorfismos (que pode ser maximizada avaliando SNPs amplamente presente no genoma) e que ao mesmo tempo estão fortemente ligados no cromossomo garantido a co-herança de seus alelos, além de conservados em espécies de plantas relacionadas evidenciando uma importância para o grupo, possa encorajar seu uso em estudos associativos, facilitando a identificação de variantes que afetam o fenótipo, o que auxiliaria a desvendar a caixa preta por trás da natureza quantitativa aumentando a eficácia da aplicação da seleção assistida por programas de melhoramento de feijão-caupi e outras leguminosas.

7 CONCLUSÕES

Este trabalho, apresentando uma nova estratégia para selecionar genes candidatos aos estudos associativos e ao melhoramento do feijão-caupi, permitiu concluir que:

a) A metodologia proposta para cálculo da expressão diferencial se mostrou mais adequada para explicar os perfis de PO e SI tendo em vista as respostas fenotípicas esperadas;

b) A ligação física entre genes de co-expressão induzida no PO permitiu identificar blocos de genes (79) com três ou mais *loci* consecutivos no genoma do feijão-caupi com tamanho médio de 37kpb e que, em sua maioria, seriam abrigados em vetores do tipo bacteriófagos e cosmídeos;

c) Os genes dos blocos permitiram identificar FTs associados incluindo famílias normalmente citadas em plantas sob estresse, além dos FTs Vang04g06720 e Vang01g02950 (MIKC_MADS e C2H2) propostos como transgenes devido as interações detectadas;

d) Com base em análises funcionais dos genes em blocos (co-expressão, estudos disponíveis na literatura e previsão da interação proteína-proteína) foi possível eleger candidatos a marcadores funcionais que podem ser manipulados de forma individual ou conjuntamente através da biotecnologia vegetal;

e) No geral, as sentenças observadas permitem antever sucessos na transferibilidade e compartilhamento para outras leguminosas (*V. angularis*, *V. radiata*, *P. vulgaris* e *G.max*), visto que 44% dos blocos foram completamente conservados nestas espécies.

REFERÊNCIAS

- Abou-Elwafa SF (2018) Identification of genes associated with drought tolerance in barley. *Biol Plant* 62:299–306. doi:10.1007/s10535-017-0765-0
- Agbicodo EM, Fatokun CA, Muranaka S, Visser RGF and Linden Van Der CG (2009) Breeding drought tolerant cowpea: Constraints, accomplishments, and future prospects. *Euphytica* 167:353–370. doi: 10.1007/s10681-009-9893-8
- Alimohammadi A, Shiran B, Martínez-Gómez P and Ebrahimie E (2013) Identification of water-deficit resistance genes in wild almond *Prunus scoparia* using cDNA-AFLP. *Sci Hortic (Amsterdam)* 159:19–28. doi: 10.1016/j.scienta.2013.04.023
- Alter S, Bader KC, Spannagl M, Wang Y, Bauer E, Schön CC and Mayer KF (2015) DroughtDB: an expert-curated compilation of plant drought stress genes and their homologs in nine species. *Database (Oxford)*. 15:2015-bav046. doi: 10.1093/database/bav046
- Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410. doi: 10.1016/S0022-2836(05)80360-2
- Amic D, Davidovic-Amic D, Beslo D, Rastija V, Lucic B and Trinajstić N (2007) SAR and QSAR of the Antioxidant Activity of Flavonoids. *Curr Med Chem* 14:827–845. doi: 10.2174/092986707780090954
- Amorim LLB, Ferreira-Neto JRC, Bezerra-Neto JP, Pandolfi V, de Araújo FT, da Silva Matos MK, Santos MG, Kido EA and Benko-Iseppon AM (2018) Cowpea and abiotic stresses: identification of reference genes for transcriptional profiling by qPCR. *Plant Methods* 14:88. doi: 10.1186/s13007-018-0354-z
- Andersen JR and Lübberstedt T (2003) Functional markers in plants. *Trends Plant Sci* 8:554–560. doi: 10.1016/j.tplants.2003.09.010
- Arisha MH, Ahmad MQ, Tang W, Liu Y, Yan H, Kou M, Wang X, Zhang Y and Li Q (2020) RNA-sequencing analysis revealed genes associated drought stress responses of different durations in hexaploid sweet potato. *Sci Rep* 10:1–17. doi: 10.1038/s41598-020-69232-3
- Arisha MH, Taha TF and Elakkad HA (2018) Expression profile of redox active secondary metabolites related genes obtained by transcriptome sequencing under salt and drought stress in sweet potato. *The Future Journal of Biology*. 2:1-10. doi: 10.37229/fsa.fjb.2018.04.15
- Attwood TK, backford S, Brazas MD, Davies A and Schneider MV (2019) A global perspective on evolving bioinformatics and data science training needs. *Brief Bioinform* 20:398–404. doi: 10.1093/bib/bbx100
- Audic S and Claverie JM (1997) The Significance of Digital Gene Expression Profiles. *Genome Research* 7:986-995. doi: 10.1101/gr.7.10.986

- Ayra-Pardo C, Ochagavía ME, Raymond B, Gulzar A, Rodríguez-Cabrera L, Rodríguez de la Noval C, Morán Bertot I, Terauchi R, Yoshida K, Matsumura H et al. (2019) HT-SuperSAGE of the gut tissue of a Vip3Aa-resistant *Heliothis virescens* (Lepidoptera: Noctuidae) strain provides insights into the basis of resistance. *Insect Sci* 26:479–498. doi: 10.1111/1744-7917.12535
- Bain LE, Awah PK, Geraldine N, Kindong NP, Sigal Y, Bernard N and Tanjeko AT (2013) Malnutrition in Sub - Saharan Africa: Burden, causes and prospects. *Pan Afr Med J* 15:1–9. doi: 10.11604/pamj.2013.15.120.2535
- Barrera-Figueroa BE, Gao L, Diop NN, Wu Z, Ehlers JD, Roberts PA, Close TJ, Zhu JK and Liu R (2011) Identification and comparative analysis of drought-associated microRNAs in two cowpea genotypes. *BMC Plant Biol*. doi: 10.1186/1471-2229-11-127
- Barrios A, Caminero C, García P, Krezdorn N, Hoffmeier K, Winter P and Pérez de la Vega M (2017) Deep Super-SAGE transcriptomic analysis of cold acclimation in lentil (*Lens culinaris* Medik.). *BMC Plant Biol* 17:1–15. doi: 10.1186/s12870-017-1057-8
- Ben Rejeb I, Pastor V and Mauch-Mani B (2014) Plant responses to simultaneous biotic and abiotic stress: Molecular mechanisms. *Plants* 3:458–475. doi: 10.3390/plants3040458
- Benjamini Y and Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B* 57:289–300.
- Berni R, Mandlik R, Hausman JF and Guerriero G (2021) Silicon-induced mitigatory effects in salt-stressed hemp leaves. *Physiol Plant* 171:476–482. doi: 10.1111/ppl.13097
- Bidabadi MS, Shiran B, Fallahi H, Rafiei F, Esmaeili F and Ebrahimie E (2015) Identification of differential expressed transcripts of almond (*Prunus dulcis* ‘Sefied’) in response to water-deficit stress by cDNA-AFLP. *Journal of Forest Research* 20:403–410. doi: 10.1007/s10310-015-0494-1
- Boleckova J, Christensen OF, Sørensen P and Sahana G (2012) Strategies for haplotype-based association mapping in a complex pedigreed population. *Czech J Anim Sci* 57:1–9. doi: 10.17221/5478-cjas
- Boukar O, Togola A, Chamarthi S, Belko N, Ishikawa H, Suzuki K and Fatokoun C (2019) Cowpea [*Vigna unguiculata* (L.) Walp.] Breeding. In: Al-Khayri JM, Jain SM and Johnson DV (eds) *Advances in Plant Breeding Strategies: Legumes Volume 7*. 1nd edition. Springer, [S.l.], pp 201–245. doi: 10.1007/978-3-030-23400-3_6
- Boukar O, Belko N, Chamarthi S, Togola A, Batieno J, Owusu E, Haruna M, Diallo S, Umar ML, Olufajo O et al. (2018) Cowpea (*Vigna unguiculata*): Genetics, genomics and breeding. *Plant Breed* 138:415–424. doi: 10.1111/pbr.12589
- Boukar O, Fatokun CA, Huynh b, Roberts PA and Close TJ (2016) Genomic tools in cowpea breeding programs: Status and perspectives. *Front Plant Sci* 7:1–13. doi: 10.3389/fpls.2016.00757

Buchfink B, Xie C and Huson DH (2014) Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12:59–60. doi: 10.1038/nmeth.3176

BurrIDGE JD, Schneider HM, Huynh b, Roberts PA, Bucksch A and Lynch JP (2017) Genome-wide association mapping and agronomic impact of cowpea root architecture. *Theor Appl Genet* 130:419–431. doi: 10.1007/s00122-016-2823-y

Carvalho M, Castro I, Moutinho-Pereira J, Correia C, Egea-Cortines M, Matos M, Rosa E, Carnide V and Lino-Neto T (2019) Evaluating stress responses in cowpea under drought stress. *J Plant Physiol* 241:153001. doi: 10.1016/j.jplph.2019.153001

Carvalho M, Lino-Neto T, Rosa E and Carnide V (2017) Cowpea: a legume crop for a challenging environment. *J Sci Food Agric* 97:4273–4284. doi: 10.1002/jsfa.8250

Casado-Díaz A, Anter J, Müller S, Winter P, Quesada-Gómez JM and Dorado G (2017) Transcriptomic Analyses of Adipocyte Differentiation From Human Mesenchymal Stromal-Cells (MSC). *J Cell Physiol* 232:771–784. doi: 10.1002/jcp.25472

Chakrabarty D, Trivedi PK, Misra P, Tiwari M, Shri M, Shukla D, Kumar S, Rai A, Pandey A, Nigam D et al. (2009) Comparative transcriptome analysis of arsenate and arsenite stresses in rice seedlings. *Chemosphere* 74:688–702. doi: 10.1016/j.chemosphere.2008.09.082

Chamarthi SK, Belko N, Togola A, Fatokun CA and Boukar O (2019) Genomics-Assisted Breeding for Drought Tolerance in Cowpea. In: Rajpal VR, Sehgal D, Kumar A and Raina SN (eds) *Genomics Assisted Breeding of Crops for Abiotic Stress Tolerance Volume 2*. 1st edition. Springer, Cham, pp187-209

Chandran D, Tai YC, Hather G, Dewdney J, Denoux C, Burgess DG, Ausubel FM, Speed TP and Wildermuth MC (2009) Temporal global expression data reveal known and novel salicylate-impacted processes and regulators mediating powdery mildew growth and reproduction on arabidopsis. *Plant Physiol* 149:1435–1451. doi: 10.1104/pp.108.132985

Chaudhary K, Poirion OB, Lu L and Garmire LX (2018) Deep learning-based multi-omics integration robustly predicts survival in liver cancer. *Clin Cancer Res* 24:1248–1259. doi: 10.1158/1078-0432.CCR-17-0853

Claverie JM and Ta TN (2019) ACDtool: A web-server for the generic analysis of large data sets of counts. *Bioinformatics* 35:170–171. doi: 10.1093/bioinformatics/bty640

Collard BCY, Jahufer MZZ, Brouwer JB and Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica* 142:169–196. doi: 10.1007/s10681-005-1681-5

Contour-Ansel D, Torres-Franklin ML, Cruz De Carvalho MH, D’Arcy-Lameta A and Zuily-Fodil Y (2006) Glutathione reductase in leaves of cowpea: Cloning of two cDNAs, expression

and enzymatic activity under progressive drought stress, desiccation and abscisic acid treatment. *Ann Bot* 98:1279–1287. doi: 10.1093/aob/mcl217

Contreras-Moreira B and Vinuesa P (2013) GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl Environ Microbiol* 79:7696–7701. doi: 10.1128/AEM.02411-13

Costa JH, Jolivet Y, Hasenfratz-Sauder MP, Orellano EG, da Guia Silva Lima M, Dizengremel P and Fernandes de Melo D (2007) Alternative oxidase regulation in roots of *Vigna unguiculata* cultivars differing in drought/salt tolerance. *J Plant Physiol* 164:718–727. doi: 10.1016/j.jplph.2006.04.001

D’Arcy-Lameta A, Ferrari-Iliou R, Contour-Ansel D, Pham-Thi AT and Zuily-Fodil Y (2006) Isolation and characterization of four ascorbate peroxidase cDNAs responsive to water deficit in cowpea leaves. *Ann Bot* 97:133–140. doi: 10.1093/aob/mcj010

Damaris RN, Li M, Liu Y, Chen X, Murage H and Yang P (2016) A proteomic analysis of salt stress response in seedlings of two African rice cultivars. *Biochim Biophys Acta - Proteins Proteomics* 1864:1570–1578. doi: 10.1016/j.bbapap.2016.08.011

De Carvalho MHC, D’Arcy-Lameta A, Roy-Macauley H, Gareil M, El Maarouf H, Pham-Thi AT and Zuily-Fodil Y (2001) Aspartic protease in leaves of common bean (*Phaseolus vulgaris* L.) and cowpea (*Vigna unguiculata* L. Walp): Enzymatic activity, gene expression and relation to drought susceptibility. *FEBS Lett* 492:242–246. doi: 10.1016/S0014-5793(01)02259-1

de Klerk E, Den Dunnen JT and ’t Hoen PAC (2014) RNA sequencing: From tag-based profiling to resolving complete transcript structure. *Cell Mol Life Sci* 71:3537–3551. doi: 10.1007/s00018-014-1637-9

Díaz M, Polanco V, Ramírez I and Peña-Cortés H (2012) Molecular cloning and expression analysis of 12-oxophytodienoate reductase cDNA by wounding in *Solanum tuberosum*. *Electron J Biotechnol* 15:1–7. doi: 10.2225/vol15-issue1-fulltext-3

Diop NN, Kidrič M, Repellin A, Gareil M, D’Arcy-Lameta A, Pham Thi AT and Zuily-Fodil Y (2004) A multicystatin is induced by drought-stress in cowpea (*Vigna unguiculata* (L.) Walp.) leaves. *FEBS Lett* 577:545–550. doi: 10.1016/j.febslet.2004.10.014

Dita MA, Rispail N, Prats E, Rubiales D and Singh KB (2006) Biotechnology approaches to overcome biotic and abiotic stress constraints in legumes. *Euphytica* 147:1–24. doi: 10.1007/s10681-006-6156-9

Draffehn AM, Li L, Krezdorn N, Ding J, Lübeck J, Strahwald J, Muktar MS, Walkemeier B, Rotter B and Gebhardt C (2013) Comparative transcript profiling by SuperSAGE identifies novel candidate genes for controlling potato quantitative resistance to late blight not compromised by late maturity. *Front Plant Sci* 4:1–21. doi: 10.3389/fpls.2013.00423

Du J, Guo S, Sun J and Shu S (2018) Proteomic and physiological analyses reveal the role of exogenous spermidine on cucumber roots in response to $\text{Ca}(\text{NO}_3)_2$ stress. *Plant Mol Biol* 97:1–21. doi: 10.1007/s11103-018-0721-1

Ehlers JD and Hall AE (1997) Cowpea (*Vigna unguiculata* L. Walp.). *Field Crops Research* 53:187–204. doi: 10.1201/9781351072533

Eichmeier A, Kiss T, Penazova E, Pecenka J, Berraf-Tebbal A, Baranek M, Pokluda R, Cechova J, Gramaje D and Grzebelus D (2019) MicroRNAs in *Vitis vinifera* cv. Chardonnay Are Differentially Expressed in Response to Diaporthe Species. *Genes (Basel)*. doi:10.3390/genes10110905

El Kelish A, Zhao F, Heller W, Durner J, Winkler JB, Behrendt H, Traidl-Hoffmann C, Horres R, Pfeifer M, Frank U et al. (2014) Ragweed (*Ambrosia artemisiifolia*) pollen allergenicity: SuperSAGE transcriptomic analysis upon elevated CO_2 and drought stress. *BMC Plant Biol* 14:1–16. doi: 10.1186/1471-2229-14-176

El Maarouf H, Zuily-Fodil Y, Gareil M, D'Arcy-Lameta A and Pham-Thi AT (1999) Enzymatic activity and gene expression under water stress of phospholipase D in two cultivars of *Vigna unguiculata* L. Walp. differing in drought tolerance. *Plant Mol Biol* 39:1257–1265. doi: 10.1023/A:1006165919928

Fang Y and Xiong L (2015) General mechanisms of drought response and their application in drought resistance improvement in plants. *Cell Mol Life Sci* 372:673–689. doi: 10.1007/s00018-014-1767-0

Farooq M, Hussain M, Wahid A and Siddique K (2012) Drought stress in plants: An overview. In: Aroca R (eds) *Plant Responses to Drought Stress*. Springer, New York, pp 1–33. doi: 10.1007/978-3-642-32653-0_1

Fini A, Guidi L, Ferrini F, Brunetti C, Di Ferdinando M, Biricolti S, Pollastri S, Calamai L and Tattini M (2012) Drought stress has contrasting effects on antioxidant enzymes activity and phenylpropanoid biosynthesis in *Fraxinus ornus* leaves: An excess light stress affair? *J Plant Physiol* 169:929–939. doi: 10.1016/j.jplph.2012.02.014

Freire Filho FR, Ribeiro VQ, Rocha MM, Silva KJD, Nogueira MS and Rodrigues EV (2011) *Feijão-caupi no Brasil: produção, melhoramento genético, avanços e desafios*. 1st edition. Embrapa Meio-Norte, Teresina, 84p

França MGC, Matos AR, D'arcy-Lameta A, Passaquet C, Lichtlé C, Zuily-Fodil Y and Pham-Thi AT (2008) Cloning and characterization of drought-stimulated phosphatidic acid phosphatase genes from *Vigna unguiculata*. *Plant Physiol Biochem* 46:1093–1100. doi: 10.1016/j.plaphy.2008.07.004

Fujita M, Fujita Y, Noutoshi Y, Takahashi F, Narusaka Y, Yamaguchi-Shinozaki K and Shinozaki K (2006) Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. *Curr Opin Plant Biol* 9:436–442. doi: 10.1016/j.pbi.2006.05.014

- Fujiwara K, Cabanos C, Toyota K, Kobayashi Y and Maruyama N (2014) Differential expression and elution behavior of basic 7S globulin among cultivars under hot water treatment of soybean seeds. *J Biosci Bioeng* 117:742–748. doi: 10.1016/j.jbiosc.2013.11.004
- Gai Z, Wang Y, Ding Y, Qian W, Qiu C, Xie H, Sun L, Jiang Z, Ma Q, Wang L et al. (2020) Exogenous abscisic acid induces the lipid and flavonoid metabolism of tea plants under drought stress. *Sci Rep* 10:1–13. doi: 10.1038/s41598-020-69080-1
- Gale MD and Devos KM (1998) Plant comparative genetics after 10 years. *Science* (80-) 282:656–659. doi: 10.1126/science.282.5389.656
- Gao QM, Venugopal S, Navarre D and Kachroo A (2011) Low oleic acid-derived repression of jasmonic acid-inducible defense responses requires the WRKY50 and WRKY51 proteins. *Plant Physiol* 155:464–476. doi: 10.1104/pp.110.166876
- Gauthier J, Vincent AT, Charette SJ and Derome N (2019) A brief history of bioinformatics. *Brief Bioinform* 20:1981–1996. doi: 10.1093/bib/bby063
- George S, Manoharan D, Li J, Britton M and Parida A (2017) Transcriptomic responses to drought and salt stress in desert tree *Prosopis juliflora*. *Plant Gene* 12:114–122. doi: 10.1016/j.plgene.2017.09.004
- Gilardoni PA, Schuck S, Jüngling R, Rotter B, Baldwin IT and Bonaventure G (2010) SuperSAGE analysis of the *Nicotiana attenuata* transcriptome after fatty acid-amino acid elicitation (FAC): Identification of early mediators of insect responses. *BMC Plant Biol* 10:66. doi: 10.1186/1471-2229-10-66
- Girma G, Natsume S, Carluccio AV, Takagi H, Matsumura H, Uemura A, Muranaka S, Takagi H, Stavolone L, Gedil M et al. (2019) Identification of candidate flowering and sex genes in white Guinea yam (*D. Rotundata* Poir.) by SuperSAGE transcriptome profiling. *PLoS One* 14:1–18. doi: 10.1371/journal.pone.0216912
- Gowda M, Jantasuriyarat C, Dean RA and Wang GL (2004) Robust-LongSAGE (RL-SAGE): A substantially improved LongSAGE method for gene discovery and transcriptome analysis. *Plant Physiol* 134:890–897. doi: 10.1104/pp.103.034496
- Gubb E and Matthiesen R (2009) Introduction to omics. In: Matthiesen R (ed) *Bioinformatics Methods in Clinical Research. Methods in Molecular Biology (Methods and Protocols)*, vol 593. 1st edition, Humana Press, SI , pp 1-23. doi: 10.1007/978-1-60327-194-3_1
- Guerrero-González ML, Rodríguez-Kessler M, Rodríguez-Guerra R, González-Chavira M, Simpson J, Sanchez F and Jiménez-Bremont JF (2011) Differential expression of *Phaseolus vulgaris* genes induced during the interaction with *Rhizoctonia solani*. *Plant Cell Rep* 30:1465–1473. doi: 10.1007/s00299-011-1055-5

- Guo R, Wang X, Han X, Li W, Liu T, Chen B, Chen X and Wang-Pruski G (2019) Comparative transcriptome analyses revealed different heat stress responses in high- and low-GS *Brassica alboglabra* sprouts. *BMC Genomics* 20:1–17. doi: 10.1186/s12864-019-5652-y
- Gupta A, Rico-Medina A and Caño-Delgado AI (2020) The physiology of plant responses to drought. *Science* (80-) 368:266–269. doi: 10.1126/science.aaz7614
- Gupta PK and Rustgi S (2004) Molecular markers from the transcribed/expressed region of the genome in higher plants. *Funct Integr Genomics* 4:139–162. doi: 10.1007/s10142-004-0107-0
- Haisel D, Vaňoková R, Synková H and Pospíšilová J (2008) The impact of trans-zeatin O-glucosyltransferase gene over-expression in tobacco on pigment content and gas exchange. *Biol Plant* 52:49–58. doi: 10.1007/s10535-008-0007-6
- Hall AE (2004) Breeding for adaptation to drought and heat in cowpea. *Eur J Agron* 21:447–454. doi: 10.1016/j.eja.2004.07.005
- Havlová M, Dobrev PI, Motyka V, Štorchová H, Libus J, Dobrá J, Malbeck J, Gaudinová A and Vanková R (2008) The role of cytokinins in responses to water deficit in tobacco plants over-expressing trans-zeatin O-glucosyltransferase gene under 35S or SAG12 promoters. *Plant, Cell Environ* 31:341–353. doi: 10.1111/j.1365-3040.2007.01766.x
- He W Di, Gao J, Dou TX, Shao XH, Bi FC, Sheng O, Deng GM, Li CY, Hu CH, Liu JH et al. (2018) Early cold-induced peroxidases and aquaporins are associated with high cold tolerance in *Dajiao* (*Musa* spp. ‘*Dajiao*’). *Front Plant Sci* 9:1–18. doi: 10.3389/fpls.2018.00282
- Heberle H, Meirelles VG, da Silva FR, Telles GP and Minghim R (2015) InteractiVenn: A web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* 16:169. doi: 10.1186/s12859-015-0611-3
- Hegedus Z, Zakrzewska A, Ágoston VC, Ordas A, Rácz P, Mink M, Spaink HP and Meijer AH (2009) Deep sequencing of the zebrafish transcriptome response to mycobacterium infection. *Mol Immunol* 46:2918–2930. doi: 10.1016/j.molimm.2009.07.002
- Hieter P and Boguski M (1997) Functional Genomics : It s All How You Read It Predictive Genetic Testing : From Basic Research to Clinical Practice. October 278:601–602. Hodgkin J (2001) Genome Size. *Encycl Genet* 865. doi: 10.1006/rwgn.2001.0557
- Hon WC, Griffith M, Mlynarz A, Kwok YC and Yang DS (1995). Antifreeze proteins in winter rye are similar to pathogenesis-related proteins. *Plant Physiol.* 109:879-889. doi:10.1104/pp.109.3.879
- Hu Z, Wang R, Zheng M, Liu X, Meng F, Wu H, Yao Y, Xin M, Peng H, Ni Z et al. (2018) TaWRKY51 promotes lateral root formation through negative regulation of ethylene biosynthesis in wheat (*Triticum aestivum* L.). *Plant J* 96:372–388. doi: 10.1111/tpj.14038

Huang W, Wu F, Huang H, Huang W, Deng C, Yiang L, Huang Z and Chen L (2020) Excess copper-induced alterations of protein profiles. *Plants* 9:291. doi: 10.3390/plants9030291

Huang Z, Zhan X, Xiang S, Johnson TS, Helm B, Yu CY, Zhang J, Salama P, Rizkalla M, Han Z et al. (2019) Salmon: Survival analysis learning with multi-omics neural networks on breast cancer. *Front Genet* 10:1–13. doi: 10.3389/fgene.2019.00166

Huynh b, Ehlers JD, Ndeve A, Wanamaker S, Lucas MR, Close TJ and Roberts PA (2015) Genetic mapping and legume synteny of aphid resistance in African cowpea (*Vigna unguiculata* L. Walp.) grown in California. *Mol Breed*. doi: 10.1007/s11032-015-0254-0

Ismail A, Riemann M and Nick P (2012) The jasmonate pathway mediates salt tolerance in grapevines. *J Exp Bot* 63:2127–2139. doi: 10.1093/jxb/err426

Iuchi S, Kobayashi M, Yamaguchi-Shinozaki K and Shinozaki K (2000) A stress-inducible gene for 9-cis-epoxycarotenoid dioxygenase involved in abscisic acid biosynthesis under water stress in drought-tolerant cowpea. *Plant Physiol* 123:553–562. doi: 10.1104/pp.123.2.553

Iuchi S, Yamaguchi-Shinozaki K, Urao T, Terao T and Shinozaki K (1996) Novel drought-inducible genes in the highly drought-tolerant cowpea: Cloning of cDNAs and analysis of the expression of the corresponding genes. *Plant Cell Physiol* 37:1073–1082. doi: 10.1093/oxfordjournals.pcp.a029056

Jayathilake C, Visvanathan R, Deen A, Bangamuwage R, Jayawardana BC, Nammi S and Liyanage R (2018) Cowpea: an overview on its nutritional facts and health benefits. *J Sci Food Agric* 98:4793–4806. doi: 10.1002/jsfa.9074

Jiang Z, Jin F, Shan X and Li Y (2019) Itraq-based proteomic analysis reveals several strategies to cope with drought stress in maize seedlings. *Int J Mol Sci* 20:1–17. doi: 10.3390/ijms20235956

Jung H, Jo SH, Park HJ, Lee A, Kim HS, Lee HJ and Cho HS (2020) Golgi-localized cyclophilin 21 proteins negatively regulate ABA signalling via the peptidyl prolyl isomerase activity during early seedling development. *Plant Mol Biol* 102:19–38. doi: 10.1007/s11103-019-00928-5

Kido ÉA, Ferreira Neto JRC, Kido SAB, Pandolfi V and Benko-Iseppon AM (2013) DeepSuperSAGE in a friendly bioinformatic approach: Identifying molecular targets responding to abiotic stress in plants. In: Gaur RK and Sharma P (eds) *Molecular Approaches in Plant Abiotic Stress*. 1st edition. CRC Press, SI, pp 108-126 doi: 10.1201/b15538-7.

Kido ÉA, Ferreira Neto JRC, Silva RL, Pandolfi V, Guimares ACR, Veiga DT, Chabregas SM, Crovella S and Benko-Iseppon AM (2012) New insights in the sugarcane transcriptome responding to drought stress as revealed by supersage. *Sci World J*. doi: 10.1100/2012/821062

- Kido ÉA, Kelle ABP, Ferreira Neto JRC, Pandolfi V, Houllou-Kido LM, Crovella S, Molina C, Kahl G and Benko-Iseppon AM (2011) Identification of Plant Protein Kinases in Response to Abiotic and Biotic Stresses Using SuperSAGE. *Curr Protein Pept Sci* 12:643–656. doi: 10.2174/1389203711109070643
- Kido ÉA, Pandolfi V, Houllou-Kido LM, Andrade PP, Marcelino FC, Nepomuceno AL, Abdelnoor RV, Burnquist WL and Benko-Iseppon AM (2010) Plant Antimicrobial Peptides: An Overview of SuperSAGE Transcriptional Profile and a Functional Review. *Curr Protein Pept Sci* 11:220–230. doi: 10.2174/138920310791112110
- Kim S, Choi HI, Ryu HJ, Ji HP, Myoung DK and Soo YK (2004) ARIA, an arabidopsis arm repeat protein interacting with a transcriptional regulator of abscisic acid-responsive gene expression, is a novel abscisic acid signaling component. *Plant Physiol* 136:3639–3648. doi: 10.1104/pp.104.049189
- Komatsu S, Kuji R, Nanjo Y, Hiraga S and Furukawa K (2012) Comprehensive analysis of endoplasmic reticulum-enriched fraction in root tips of soybean under flooding stress using proteomics techniques. *J Proteomics* 77:531–560. doi: 10.1016/j.jprot.2012.09.032
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Jones SJ and Marra MA (2009) Circos- An information aesthetic for comparative genomics. *Genome research* 19:1639-1645. doi: 10.1101/gr.092759.109
- Kulcheski FR, de Oliveira LFV, Molina LG, Almerão MP, Rodrigues FA, Marcolino J, Barbosa JF, Stolf-Moreira R, Nepomuceno AL, Marcelino-Guimarães FC et al. (2011) Identification of novel soybean microRNAs involved in abiotic and biotic stresses. *BMC Genomics*. doi: 10.1186/1471-2164-12-307
- Kumar LS (1999) DNA markers in plant improvement: An overview. *Biotechnol Adv* 17:143–182. doi: 10.1016/S0734-9750(98)00018-4
- La Camera S, Balagué C, Göbel C, Geoffroy P, Legrand M, Feussner I, Roby D and Heitz T (2009) The Arabidopsis patatin-like protein 2 (PLP2) plays an essential role in cell death execution and differentially affects biosynthesis of oxylipins and resistance to pathogens. *Mol Plant-Microbe Interact* 22:469–481. doi: 10.1094/MPMI-22-4-0469
- Langmead B and Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. doi: 10.1038/nmeth.1923
- Langmead B, Trapnell C, Pop M and Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. doi: 10.1186/gb-2009-10-3-r25
- Lee Kraus W (2015) Editorial: Would you like a hypothesis with those data? Omics and the age of discovery science. *Mol Endocrinol* 29:1531–1534. doi: 10.1210/me.2015-1253

- Lee BR, Jung WJ, Lee BH, Avice JC, Ourry A and Kim TH (2008) Kinetics of drought-induced pathogenesis-related proteins and its physiological significance in white clover leaves. *Physiol Plant* 132:329–337. doi: 10.1111/j.1399-3054.2007.01014.x
- Li LQ, Lyu CC, Li JH, Wan CY, Liu L, Xie MQ, Zuo RJ, Ni S, Liu F, Zeng FC et al. (2020) Quantitative proteomic analysis of alligator weed leaves reveals that cationic peroxidase 1 plays vital roles in the potassium deficiency stress response. *Int J Mol Sci*. doi: 10.3390/ijms21072537
- Li G, Wu X, Hu Y, Muñoz-Amatriaín M, Luo J, Zhou W, Wang B, Wang Y, Wu X, Huang L et al. (2019) Orphan genes are involved in drought adaptations and ecoclimatic-oriented selections in domesticated cowpea. *J Exp Bot* 70:3101–3110. doi: 10.1093/jxb/erz145
- Li K, Xu C, Zhang K, Yang A and Zhang J (2007) Proteomic analysis of roots growth and metabolic changes under phosphorus deficit in maize (*Zea mays* L.) plants. *Proteomics* 7:1501–1512. doi: 10.1002/pmic.200600960
- Li R, Sosa JL and Zavala ME (2000) Accumulation of zeatin O-glycosyltransferase in *Phaseolus vulgaris* and *Zea mays* following cold stress. *Plant Growth Regul* 32:295–305. doi: 10.1023/A:1010755901072
- Lonardi S, Muñoz-Amatriaín M, Liang Q, Shu S, Wanamaker SI, Lo S, Tanskanen J, Schulman AH, Zhu T, Luo MC et al. (2019) The genome of cowpea (*Vigna unguiculata* [L.] Walp.). *Plant J* 98:767–782. doi: 10.1111/tpj.14349
- Louro B, Martins RST, Pinto PIS, Reinhardt R, de Koning DJ, Canario AVM and Power DM (2019) SuperSAGE digital expression analysis of differential growth rate in a European sea bass population. *Aquac Fish* 4:17–26. doi: 10.1016/j.aaf.2018.03.001
- Love MI, Huber W and Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:1–21. doi: 10.1186/s13059-014-0550-8
- Lowe R, Shirley N, beackley M, Dolan S and Shafee T (2017) Transcriptomics technologies. *PLoS Comput Biol* 13:1–23. doi: 10.1371/journal.pcbi.1005457
- Lucas MR, Huynh b, Vinholes P da S, Cisse N, Drabo I, Ehlers JD, Roberts PA and Close TJ (2013) Association studies and legume synteny reveal haplotypes determining seed size in *Vigna unguiculata*. *Front Plant Sci* 4:1–9. doi: 10.3389/fpls.2013.00095
- Luo ZW, Wu CI and Kearsey MJ (2002) Precision and high-resolution mapping of quantitative trait loci by use of recurrent selection, backcross or intercross schemes. *Genetics* 161:915–929.
- Ma Y, Yu X, Yu Z, Sun F, Li X and Li XL (2018) RNA-Seq of *Agropyron mongolicum* Keng in response to drought stress. *Grassl Sci* 64:3–15. doi: 10.1111/grs.12176

Maher M, Ahmad H, Nishawy E, Li Y and Luo J (2021) Novel transcriptome study and detection of metabolic variations in UV-B-treated date palm (*Phoenix dactylifera* cv. khalas). *Int J Mol Sci* 22:1–24. doi: 10.3390/ijms22052564

Maheswary V, Khairun HN, Vasanthi S and Qistina ON (2016) Digital gene expression of water-stressed *Phyllanthus urinaria* (dukung anak). *J Trop Agric Food Sci* 44:121–136

Marcel GC, Matos A, d'arcy-Lameta A, Kader JC, Zuily-Fodil Y and Pham-Thi A (2000). Two novel plant cDNAs homologous to animal type-2 phosphatidate phosphatase are expressed in cowpea leaves and are differently regulated by water deficits. *Biochem Soc Trans.* 6:915-917

Matos AR, D'Arcy-Lameta A, França M, Pêtres S, Edelman L, Kader JC, Zuily-Fodil Y and Pham-Thi AT (2001) A novel patatin-like gene stimulated by drought stress encodes a galactolipid acyl hydrolase. *FEBS Lett* 491:188–192. doi: 10.1016/S0014-5793(01)02194-9

Matsumura H, Urasaki N, Yoshida K, Krüger DH, Kahl G and Terauchi R (2012) SuperSAGE: Powerful Serial Analysis of Gene Expression In: Hailing J, Walter G (eds) *RNA Abundance Analysis: Methods and Protocols*. 1nd edition. Humana Press, Totowar, pp 1-17

Matsumura H, Yoshida K, Luo S, Kimura E, Fujibe T, Albertyn Z, Barrero RA, Krüger DH, Kahl G, Schroth GP et al. (2010) High-throughput superSAGE for digital gene expression analysis of multiple samples using next generation sequencing. *PLoS One*. doi: 10.1371/journal.pone.0012010

Matsumura H, Kruger D, Kahl G and Terauchi R (2008) SuperSAGE: A Modern Platform for Genome-Wide Quantitative Transcript Profiling. *Curr Pharm Biotechnol* 9:368–374. doi: 10.2174/138920108785915157

Matsumura H, Ito A, Saitoh H, Winter P, Kahl G, Reuter M, Kruger D and Terauchi RS (2005) Technoreview SuperSAGE. *Cell Microbiol* 7:11-18. doi: 10.1111/j.1462-5822.2004.00478.x

Meng H, Sun M, Jiang Z, Liu Y, Sun Y, Liu D, Jiang C, Ren M, Yuan G, Yu W et al. (2021) Comparative transcriptome analysis reveals resistant and susceptible genes in tobacco cultivars in response to infection by *Phytophthora nicotianae*. *Sci Rep* 11:1–13. doi: 10.1038/s41598-020-80280-7

Molina C, Rotter B, Horres R, Udupa SM, Besser B, Bellarmino L, Baum M, Matsumura H, Terauchi R, Kahl G et al. (2008) SuperSAGE: The drought stress-responsive transcriptome of chickpea roots. *BMC Genomics* 9:1–28. doi: 10.1186/1471-2164-9-553

Moon SJ, Han SY, Kim DY, Kim BG, Yoon IS, Shin D, Kwon H Bin and Byun MO (2014a) Ectopic expression of CaWRKY1, a pepper transcription factor, enhances drought tolerance in transgenic potato plants. *J Plant Biol* 57:198–207. doi: 10.1007/s12374-013-0511-x

Moon JC, Yim WC, Lim SD, Song K and Lee BM (2014b) Differentially expressed genes and in silico analysis in response to ozone (O₃) stress of soybean leaves. *Aust J Crop Sci* 8:276–283.

Morgante M and Salamini F (2003) From plant genomics to breeding practice. *Curr Opin Biotechnol* 14:214–219. doi: 10.1016/S0958-1669(03)00028-4

Muchero W, Roberts PA, Diop NN, Drabo I, Cisse N, Close TJ, Muranaka S, Boukar O and Ehlers JD (2013) Genetic Architecture of Delayed Senescence, Biomass, and Grain Yield under Drought Stress in Cowpea. *PLoS One*: 30;8(7):e70041. doi: 10.1371/journal.pone.0070041

Muchero W, Ehlers JD, Close TJ and Roberts PA (2011) Genic SNP markers and legume synteny underlying QTL for M phaseolina resistance and maturity in cowpea. *BMC Genomics* 12:1–14 doi: 10.1186/1471-2164-12-8.

Muchero W, Ehlers JD, Close TJ and Roberts PA (2009) Mapping QTL for drought stress-induced premature senescence and maturity in cowpea [*Vigna unguiculata* (L.) Walp.]. *Theor Appl Genet* 118:849–863. doi: 10.1007/s00122-008-0944-7

Mudge J, Cannon SB, Kalo P, Oldroyd GED, Roe BA, Town CD and Young ND (2005) Highly syntenic regions in the genomes of soybean, *Medicago truncatula*, and *Arabidopsis thaliana*. *BMC Plant Biol* 5:1–16. doi: 10.1186/1471-2229-5-15

Munnik T and Meijer HJG (2001) Osmotic stress activates distinct lipid and MAPK signalling pathways in plants. *FEBS Lett* 498:172–178. doi: 10.1016/S0014-5793(01)02492-9

Muñoz-Amatriaín M, Mirebrahim H, Xu P, Wanamaker SI, Luo MC, Alhakami H, Alpert M, Atokple I, Batiemo BJ, Boukar O et al. (2017) Genome resources for climate-resilient cowpea, an essential crop for food security. *Plant J* 89:1042–1054. doi: 10.1111/tpj.13404

Murgia I, Tarantino D, Soave C and Morandini P (2011) *Arabidopsis* CYP82C4 expression is dependent on Fe availability and circadian rhythm, and correlates with genes involved in the early Fe deficiency response. *J Plant Physiol* 168:894–902. doi: 10.1016/j.jplph.2010.11.020

Mutz KO, Heilkenbrinker A, Lönne M, Walter JG and Stahl F (2013) Transcriptome analysis using next-generation sequencing. *Curr Opin Biotechnol* 24:22–30. doi: 10.1016/j.copbio.2012.09.004

Nadeem M, Li J, Yahya M, Sher A, Ma C, Wang X and Qiu L (2019) Research Progress and Perspective on Drought Stress in Legumes: A Review. *Int J Mol Sci* 20:2541–2573. doi: 10.3390/ijms20102541

Nascimento SP do, Bastos EA, Araújo ECE, Filho FRF and Silva EM (2011) Tolerância ao déficit hídrico em genótipos de feijão-caupi Tolerance to water deficit of cowpea genotypes. *Rev Bras Eng Agrícola e Ambient* 15:853–860.

- Nasrollahi V, Mirzaie-Asl A, Piri K, Nazeri S and Mehrabi R (2014) The effect of drought stress on the expression of key genes involved in the biosynthesis of triterpenoid saponins in liquorice (*Glycyrrhiza glabra*). *Phytochemistry* 103:32–37. doi: 10.1016/j.phytochem.2014.03.004
- Ni W, Zhu L, Sha R, Tao J, Cai B and Wang S (2017) Comparative iTRAQ proteomic profiling of susceptible and resistant apple cultivars infected by *Alternaria alternata* apple pathotype. *Tree Genet Genomes* 13:1–13. doi: 10.1007/s11295-017-1104-5
- Nierman WC and Feldbyum TV (2001) Genomic Library. In: Brenner S and Miller JH (eds) *Encyclopedia of Genetics*. Academic Press, [S.l.], pp 201-245. doi: 10.1006/rwgn.2001.0559.
- Ning LH, Du W kai, Song HN, Shao HB, Qi WC, Sheteiwy MSA and Yu D yue (2019) Identification of responsive miRNAs involved in combination stresses of phosphate starvation and salt stress in soybean root. *Environ Exp Bot* 167:103823. doi: 10.1016/j.envexpbot.2019.103823
- Park SU, Kim YH, Lee CJ, Kim SE, Lim YH, Yoon UH, Kim HS and Kwak SS (2020) Comparative transcriptome profiling of two sweetpotato cultivars with contrasting flooding stress tolerance levels. *Plant Biotechnol Rep* 14:743–756. doi: 10.1007/s11816-020-00650-5
- Pasquet R.S. and Padulosi S. (2013) Genus *Vigna* and Cowpea (*V. unguiculata* [L.] Walp.) taxonomy: current status and prospects. *Innov Res* 66–87.
- Portieles R, Ochagavia ME, Canales E, Silva Y, Chacón O, Hernández I, López Y, Rodríguez M, Terauchi R, Borroto C et al. (2017) High-throughput SuperSAGE for gene expression analysis of *Nicotiana tabacum*-*Rhizoctonia solani* interaction. *BMC Res Notes* 10:1–7. doi: 10.1186/s13104-017-2934-9
- Pottorff M, Roberts PA, Close TJ, Lonardi S, Wanamaker S and Ehlers JD (2014) Identification of candidate genes and molecular markers for heat-induced brown discoloration of seed coats in cowpea [*Vigna unguiculata* (L.) Walp]. *BMC Genomics* 15:1–11. doi: 10.1186/1471-2164-15-328
- Pottorff M, Ehlers JD, Fatokun C, Roberts PA and Close TJ (2012) Leaf morphology in Cowpea [*Vigna unguiculata* (L.) Walp]: QTL analysis, physical mapping and identifying a candidate gene using synteny with model legume species. *BMC Genomics* 13:1–12. doi: 10.1186/1471-2164-13-234
- Qin C, Ahanger MA, Lin B, Huang Z, Zhou J, Ahmed N, Ai S, Mustafa NSA, Ashraf M and Zhang L (2021) Comparative transcriptome analysis reveals the regulatory effects of acetylcholine on salt tolerance of *Nicotiana benthamiana*. *Phytochemistry* 181:112582. doi: 10.1016/j.phytochem.2020.112582
- Rabuma T, Gupta OP, Yadav M and Chhokar V (2021) Integrative analysis of *Capsicum annuum* L. - *Phytophthora capsici* L. pathosystem reveals molecular cross-talk and activation of host defence response. *bioRxiv*. doi: 10.1101/2021.04.03.438323

- Rani B and Sharma VK (2017) Transcriptome profiling: methods and applications-A review. *Agric Rev* 38:271-281. doi: 10.18805/ag.R-1549
- Ribaut J-M and Hoisington D (1998) Marker-assisted selection: new tools and strategies. *Trends Plant Sci* 3:236-239
- Robertson N, Oveisi-Fordorei M, Zuyderduyn SD, Varhol RJ, Fjell C, Marra M, Jones S and Siddiqui A (2007) DiscoverySpace: An interactive data analysis application. *Genome Biol* 8:1-13. doi: 10.1186/gb-2007-8-1-r6
- Robinson MD, McCarthy DJ and Smyth GK (2009) edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139-140. doi: 10.1093/bioinformatics/btp616
- Sadhukhan A, Panda SK and Sahoo L (2014a) The cowpea RING ubiquitin ligase VuDRIP interacts with transcription factor VuDREB2A for regulating abiotic stress responses. *Plant Physiol Biochem* 83:51-56. doi: 10.1016/j.plaphy.2014.07.007
- Sadhukhan A, Kobayashi Y, Kobayashi Y, Tokizawa M, Yamamoto YY, Iuchi S, Koyama H, Panda SK and Sahoo L (2014b) VuDREB2A, a novel DREB2-type transcription factor in the drought-tolerant legume cowpea, mediates DRE-dependent expression of stress-responsive genes and confers enhanced drought resistance in transgenic Arabidopsis. *Planta* 240:645-664. doi: 10.1007/s00425-014-2111-5
- Saha S, Sparks AB, Rago C, Akmaev V, Wang CJ, Vogelstein B, Kinzler KW and Velculescu VE (2002) Using the transcriptome to annotate the genome. *Nat Biotechnol* 20:508-512. doi: 10.1038/nbt0502-508
- Salgotra RK, Gupta BB and Stewart CN (2014) From genomics to functional markers in the era of next-generation sequencing. *Biotechnol Lett* 36:417-426. doi: 10.1007/s10529-013-1377-1
- Sang Q, Shan X, An Y, Shu S, Sun J and Guo S (2017) Proteomic analysis reveals the positive effect of exogenous spermidine in tomato seedlings' response to high-temperature stress. *Front Plant Sci* 8:1-15. doi: 10.3389/fpls.2017.00120
- Sasaki-Sekimoto Y, Taki N, Obayashi T, Aono M, Matsumoto F, Sakurai N, Suzuki H, Hirai MY, Noji M, Saito K et al. (2005) Coordinated activation of metabolic pathways for antioxidants and defence compounds by jasmonates and their roles in stress tolerance in Arabidopsis. *Plant J* 44:653-668. doi: 10.1111/j.1365-313X.2005.02560.x
- Schena M (1996) Genome analysis with gene expression microarrays. *BioEssays* 18:427-431. doi: 10.1002/bies.950180513
- Shabani L, Ehsanpour AA and Esmaeili A (2010) Assessment of squalene synthase and beta-amyrin synthase gene expression in licorice roots treated with methyl jasmonate and salicylic acid using real-time qPCR. *Russ J Plant Physiol* 57:480-484. doi: 10.1134/S1021443710040047

- Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J et al. (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463:178-183
- Semagn K, Bjornstad A and Xu Y (2010) The genetic dissection of quantitative traits in crops. *Electronic Journal of Biotechnology* 13:1–45. doi: 10.2225/vol13-issue5-fulltext-14
- Shah T, Xu J, Zou X, Cheng Y, Nasir M and Zhang X (2018) Omics approaches for engineering wheat production under abiotic stresses. *Int J Mol Sci* 19:1–16. doi: 10.3390/ijms19082390
- Shah Mohammadi N, Buapet P, Pernice M, Signal B, Kahlke T, Hardke L and Ralph PJ (2019) Transcriptome profiling analysis of the seagrass, *Zostera muelleri* under copper stress. *Mar Pollut Bull* 149:110556. doi: 10.1016/j.marpolbul.2019.110556
- Shangguan L, Han J, Kayesh E, Sun X, Zhang C, Pervaiz T, Wen X and Fang J (2013) Evaluation of Genome Sequencing Quality in Selected Plant Species Using Expressed Sequence Tags. *PLoS One*. doi: 10.1371/journal.pone.0069890
- Shendure J (2008) The beginning of the end for microarrays? *Nat Methods* 5:585–587. doi: 10.1038/nmeth0708-585
- Shimamura M, Akashi T, Sakurai N, Suzuki H, Saito K, Shibata D, Ayabe SI and Aoki T (2007) 2-Hydroxyisoflavanone dehydratase is a critical determinant of isoflavone productivity in hairy root cultures of *Lotus japonicus*. *Plant Cell Physiol* 48:1652–1657. doi: 10.1093/pcp/pcm125
- Shirazi Z, Aalami A, Tohidfar M and Sohani MM (2019) Triterpenoid gene expression and phytochemical content in Iranian licorice under salinity stress. *Protoplasma* 256:827–837. doi: 10.1007/s00709-018-01340-4
- Shui XR, Chen ZW and Li JX (2013) MicroRNA prediction and its function in regulating drought-related genes in cowpea. *Plant Sci* 210:25–35. doi: 10.1016/j.plantsci.2013.05.002
- Simova-Stoilova LP, Romero-Rodríguez MC, Sánchez-Lucas R, Navarro-Cerrillo RM, Alberto Medina-Aunon J and Jorrín-Novo J V. (2015) 2-DE proteomics analysis of drought treated seedlings of *Quercus ilex* supports a root active strategy for metabolic adaptation in response to water shortage. *Front Plant Sci* 6:1–16. doi: 10.3389/fpls.2015.00627
- Sindhu M, Kumar A, Yadav H, Chaudhary D, Jaiwal R and Jaiwal PK (2019) Current advances and future directions in genetic enhancement of a climate resilient food legume crop, cowpea (*Vigna unguiculata* L. Walp.). *Plant Cell Tissue Organ Cult* 139:429–453. doi: 10.1007/s11240-019-01695-3
- Singh BB (2014) *Cowpea: The Food Legume of the 21st Century*. 1st edition. Crop Science Society of America, Madison, 192 pp

Singh BB, Ajeigbe HA, Tarawali SA, Fernandez-Rivera S and Abubakar M (2003) Improving the production and utilization of cowpea as food and fodder. *F Crop Res* 84:169–177. doi: 10.1016/S0378-4290(03)00148-5

Sirhindi G, Sharma P, Arya P, Goel P, Kumar G, Acharya V and Singh AK (2016) Genome-wide characterization and expression profiling of TIFY gene family in pigeonpea (*Cajanus cajan* (L.) Millsp.) under copper stress. *J Plant Biochem Biotechnol* 25:301–310. doi: 10.1007/s13562-015-0342-6

Soanes DM, Chakrabarti A, Paszkiewicz KH, Dawe AL and Talbot NJ (2012) Genome-wide transcriptional profiling of appressorium development by the rice blast fungus *Magnaporthe oryzae*. *PLoS Pathog* 8(2): e1002514. doi: 10.1371/journal.ppat.1002514

Sorkheh K, Malysheva-Otto L V., Wirthensohn MG, Tarkesh-Esfahani S and Martínez-Gómez P (2008) Linkage disequilibrium, genetic association mapping and gene localization in crop plants. *Genet Mol Biol* 31:805–814. doi: 10.1590/S1415-47572008000500001

Stressmann M, Kitao S, Griffith M, Moresoli C, Bravo LA and Marangoni AG (2004) Calcium interacts with antifreeze proteins and chitinase from cold-acclimated winter rye. *Plant Physiol* 135:364–376. doi: 10.1104/pp.103.038158

Szklarczyk D, Gabe AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P et al. (2019) STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 47:D607–D613. doi: 10.1093/nar/gky1131

Tanksley SD (1993) Mapping polygenes. *Annual review of genetics* 27:205-233

Terecskei K, Tóth R, Gyula P, Kevei É, Bindics J, Coupland G, Nagy F and Kozma-Bognár L (2013) The circadian clock-associated small GTPase LIGHT INSENSITIVE PERIOD1 suppresses light-controlled endoreplication and affects tolerance to salt stress in *Arabidopsis*. *Plant Physiol* 161:278–290. doi: 10.1104/pp.112.203356

Tian F, Yang DC, Meng YQ, Jin J and Gao G (2020) PlantRegMap: Charting functional regulatory maps in plants. *Nucleic Acids Res* 48:D1104–D1113. doi: 10.1093/nar/gkz1020

Timko MP, Rushton PJ, Laudeman TW, Bokowiec MT, Chipumuro E, Cheung F, Town CD and Chen X (2008) Sequencing and analysis of the gene-rich space of cowpea. *BMC Genomics* 9:1–20. doi: 10.1186/1471-2164-9-103

Timko MP and Singh BB (2008) Cowpea, a Multifunctional Legume. In: Moore PH and Ming R (eds) *Genomics of Tropical Crop Plants. Plant Genetics and Genomics: Crops and Models*, vol 1. 1nd edition. Springer, New York, pp 227-258. doi: 10.1007/978-0-387-71219-2_10

Tombuloglu G, Tombuloglu H, Sakcali MS and Unver T (2015) High-throughput transcriptome analysis of barley (*Hordeum vulgare*) exposed to excessive boron. *Gene* 557:71–81. doi: 10.1016/j.gene.2014.12.012

Torres NL, Cho K, Shibato J, Hirano M, Kubo A, Masuo Y, Iwahashi H, Jwa NS, Agrawal GK and Rakwal R (2007) Gel-based proteomics reveals potential novel protein markers of ozone stress in leaves of cultivated bean and maize species of Panama. *Electrophoresis* 28:4369–4381. doi: 10.1002/elps.200700219Url

Van Aerle R, Lange A, Moorhouse A, Paszkiewicz K, Ball K, Johnston BD, De-Bastos E, Booth T, Tyler CR and Santos EM (2013) Molecular mechanisms of toxicity of silver nanoparticles in zebrafish embryos. *Environ Sci Technol* 47:8005–8014. doi: 10.1021/es401758d

Van Loon LC, Rep M and Pieterse CMJ (2006) Significance of inducible defense-related proteins in infected plants. *Annu Rev Phytopathol* 44:135–162. doi: 10.1146/annurev.phyto.44.070505.143425

Velculescu VE, Zhang L, Vogelstein B and Kinzler KW (1995) Serial Analysis of Gene Expression. *Science* 270:484–487. doi: 10.1126/science.270.5235.484

Wang H, Zaman QU, Huang W, Mei D, Liu J, Wang W, Ding B, Hao M, Fu L, Cheng H et al. (2019) QTL and Candidate Gene Identification for Silique Length Based on High-Dense Genetic Map in *Brassica napus* L. *Front Plant Sci* 10:1–11. doi: 10.3389/fpls.2019.01579

Wang W, Huang F, Qin Q, Zhao X, Li Z and Fu B (2015) Comparative analysis of DNA methylation changes in two rice genotypes under salt stress and subsequent recovery. *Biochem Biophys Res Commun* 465:790–796. doi: 10.1016/j.bbrc.2015.08.089

Wang Y, Kwon SJ, Wu J, Choi J, Lee YH, Agrawal GK, Tamogami S, Rakwal R, Park SR, Kim BG et al. (2014) Transcriptome analysis of early responsive genes in rice during *Magnaporthe oryzae* infection. *Plant Pathol J* 30:343–354. doi: 10.5423/PPJ.OA.06.2014.0055

Wang Z, Gerstein M and Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10:57–63. doi: 10.1038/nrg2484

Wang W, Vinocur B and Altman A (2003) Plant responses to drought, salinity and extreme temperatures: Towards genetic engineering for stress tolerance. *Planta* 218:1–14. doi: 10.1007/s00425-003-1105-5

Wani SH, Kumar V, Shriram V and Sah SK (2016) Phytohormones and their metabolic engineering for abiotic stress tolerance in crop plants. *Crop J* 4:162–176. doi: 10.1016/j.cj.2016.01.010

West CE, Waterworth WM, Story GW, Sunderland PA, Jiang Q and Bray CM (2002) Disruption of the *Arabidopsis* AtKu80 gene demonstrates an essential role for AtKu80 protein

in efficient repair of DNA double-strand breaks in vivo. *Plant J* 31:517–528. doi: 10.1046/j.1365-3113X.2002.01370.x

Wu Y, Wei W, Pang X, Wang X, Zhang H, Dong B, Xing Y, Li X and Wang M (2014) Comparative transcriptome profiling of a desert evergreen shrub, *Ammopiptanthus mongolicus*, in response to drought and cold stresses. *BMC Genomics* 15:1–16. doi: 10.1186/1471-2164-15-671

Xiao L, Shang XH, Cao S, Xie XY, Zeng WD, Lu LY, Chen SB and Yan HB (2019) Comparative physiology and transcriptome analysis allows for identification of lncRNAs imparting tolerance to drought stress in autotetraploid cassava. *BMC Genomics* 20:1–15. doi: 10.1186/s12864-019-5895-7

Xin Z, Zhang J, Ge L, Lei S, Han J, Zhang X, Li X and Sun X (2017) A putative 12-oxophytodienoate reductase gene *CsOPR3* from *Camellia sinensis*, is involved in wound and herbivore infestation responses. *Gene* 615:18–24. doi: 10.1016/j.gene.2017.03.013

Xin H, Zhu W, Wang L, Xiang Y, Fang L, Li J, Sun X, Wang N, Londo JP and Li S (2013) Genome Wide Transcriptional Profile Analysis of *Vitis amurens* and *Vitis vinifera* in Response to Cold Stress. *PLoS One* 8:1–14. doi: 10.1371/journal.pone.0058740

Xiong J, Sun Y, Yang Q, Tian H, Zhang H, Liu Y and Chen M (2017) Proteomic analysis of early salt stress responsive proteins in alfalfa roots and shoots. *Proteome Sci.* doi: 10.1186/s12953-017-0127-z

Xu L, Dong Z, Fang L, Luo Y, Wei Z, Guo H, Zhang G, Gu YQ, Coleman-Derr D, Xia Q et al. (2019) OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res* 47:W52–W58. doi: 10.1093/nar/gkz333

Xu Y and Huang B (2018) Comparative transcriptomic analysis reveals common molecular factors responsive to heat and drought stress in *Agrostis stolonifera*. *Sci Rep* 8:16–19. doi: 10.1038/s41598-018-33597-3

Yang Y, Li HG, Wang J, Wang HL, He F, Su Y, Zhang Y, Feng CH, Niu M, Li Z, et al. (2020) ABF3 enhances drought tolerance via promoting ABA-induced stomatal closure by directly regulating ADF5 in *Populus euphratica*. *J Exp Bot.* 22:7270-7285. doi: 10.1093/jxb/eraa383

Yang QS, Wu JH, Li CY, Wei YR, Sheng O, Hu CH, Kuang R Bin, Huang YH, Peng XX, McCardle JA et al. (2012) Quantitative proteomic analysis reveals that antioxidation mechanisms contribute to cold tolerance in plantain (*Musa paradisiaca* L.; ABB Group) seedlings. *Mol Cell Proteomics* 11:1853–1869. doi: 10.1074/mcp.M112.022079

Yao S, Jiang C, Huang Z, Torres-Jerez I, Chang J, Zhang H, Udvardi M, Liu R and Verdier J (2016) The *Vigna unguiculata* Gene Expression Atlas (VuGEA) from de novo assembly and

quantification of RNA-seq data provides insights into seed maturation mechanisms. *Plant J* 88:318–327. doi: 10.1111/tpj.13279

Yin X, Hiraga S, Hajika M, Nishimura M and Komatsu S (2017) Transcriptomic analysis reveals the flooding tolerant mechanism in flooding tolerant line and abscisic acid treated soybean. *Plant Mol Biol* 93:479–496. doi: 10.1007/s11103-016-0576-2

Yokotani N, Ichikawa T, Kondou Y, Iwabuchi M, Matsui M, Hirochika H and Oda K (2013) Role of the rice transcription factor JAmyb in abiotic stress response. *J Plant Res* 126:131–139. doi: 10.1007/s10265-012-0501-y

Yu L, Ma J, Niu Z, Bai X, Lei W, Shao X, Chen N, Zhou F and Wan D (2017) Tissue-specific transcriptome analysis reveals multiple responses to salt stress in *Populus euphratica* seedlings. *Genes (Basel)* 8:1–14. doi: 10.3390/genes8120372

Zadražnik T, Hollung K, Egge-Jacobsen W, Meglič V and Šuštar-Vozlič J (2013) Differential proteomic analysis of drought stress response in leaves of common bean (*Phaseolus vulgaris* L.). *J Proteomics* 78:254–272. doi: 10.1016/j.jprot.2012.09.021

Zegaoui Z, Planchais S, Cabassa C, Djebbar R, Belbachir OA and Carol P (2017) Variation in relative water content, proline accumulation and stress gene expression in two cowpea landraces under drought. *J Plant Physiol* 218:26–34. doi: 10.1016/j.jplph.2017.07.009

Zhang P, Qian D, Luo C, Niu Y, Li T, Li C, Xiang Y, Wang X and Niu Y (2021) Arabidopsis ADF5 Acts as a Downstream Target Gene of CBFs in Response to Low-Temperature Stress. *Front Cell Dev Biol* 9:1–12. doi: 10.3389/fcell.2021.635533

Zhang W, Zuo C, Chen Z, Kang Y and Qin S (2019a) RNA sequencing reveals that both abiotic and biotic stress-responsive genes are induced during expression of steroidal glycoalkaloid in potato tuber subjected to light exposure. *Genes (Basel)* 10:1–17. doi: 10.3390/genes10110920

Zhang Y, Wei M, Liu A, Zhou R, Li D, Dossa K, Wang L, Zhang Y, Gong H, Zhang X et al. (2019b) Comparative proteomic analysis of two sesame genotypes with contrasting salinity tolerance in response to salt stress. *J Proteomics* 201:73–83. doi: 10.1016/j.jprot.2019.04.017

Zhou J, Huang J, Tian X, Zheng J and He X (2020) Transcriptome analysis reveals dynamic changes in the salt stress response in *Salix*. *J For Res* 31:1851–1862. doi: 10.1007/s11676-019-00941-w

Zhu Z, Chen H, Xie K, Liu C, Li L, Liu L, Han X, Jiao C, Wan Z and Sha A (2020) Characterization of Drought-Responsive Transcriptome During Seed Germination in Adzuki Bean (*Vigna angularis* L.) by PacBio SMRT and Illumina Sequencing. *Front Genet* 11:1–14. doi: 10.3389/fgene.2020.00996

Zhu FY, Chen MX, Chan WL, Yang F, Tian Y, Song T, Xie LJ, Zhou Y, Xiao S, Zhang J et al. (2018) SWATH-MS quantitative proteomic investigation of nitrogen starvation in

Arabidopsis reveals new aspects of plant nitrogen stress responses. *J Proteomics* 187:161–170. doi: 10.1016/j.jprot.2018.07.014

Zhu C, Gore M, Buckler ES and Yu J (2008) Status and Prospects of Association Mapping in Plants. *Plant Genome* 1:5–20. doi: 10.3835/plantgenome2008.02.0089

Zhu M and Zhao S (2007) Candidate gene identification approach: Progress and challenges. *Int J Biol Sci* 3:420–427. doi: 10.7150/ijbs.3.420

APÊNDICE A – COMANDOS PARA ALINHAR AS *UNITAGS* SUPERSAGE CONTRA O TRANSCRITPOMA REFEREÊNCIA DO FEIJÃO-CAUPI VIA BOWTIE

1) Indexando o transcriptoma: bowtie-build [PATH transcriptome.fas] Index_Name

2) Alinhamento: bowtie -v 1 -a --best --threads 4 -x Index_Name -f [PATH unitags.fas] -t > output.txt

APÊNDICE B – COMANDOS PARA ALINHAR AS *UNITAGS* SUPERSAGE CONTRA O TRANSCRITPOMA REFEREÊNCIA DO FEIJÃO-CAUPI VIA BLAST

- 1) Criando arquivos do banco de dados:** `makeblastdb -in [PATH transcriptome.fas] -dbtype nucl`
- 2) Alinhamento:** `blastn -query [PATH unitags.fas] -db [PATH transcriptome.fas] -evalue 1e-4 -outfmt 6 -out output.txt -num_threads 4 -task blastn-short`

APÊNDICE C – CÓDIGO CORRESPONDENTE AS ETAPAS DO ALGORITMO PARA IDENTIFICAÇÃO DE BLOCOS CO-EXPRESSOS E CONSECUTIVOS NO GENOMA DE *V. UNGUICULATA*

#1) Definir diretório e carregar pacotes

```
setwd("/home/vinicius/MEGA/Projeto/Paper")
library(dplyr)
library(tidyr)
library(readr)
```

#2) Carregar arquivos necessários

#2.1 Dados de expressão

```
df<- read_tsv('Outputs/Seca_expressao.txt')

df<-select(df,c(1,22,23))

knitr::kabe(head(df))
```

Transcrito	reg_SI	reg_PO
XM_028045870.1	ns	UR
XM_028045871.1	UR	ns
XM_028045873.1	DR	ns
XM_028045874.1	UR	ns
XM_028045875.1	ns	ns
XM_028045876.1	ns	ns

#2.2 Features RefSeq

```
feature_RefSeq <- read_tsv("Inputs/GCF_004118075.1_ASM411807v1_feature_tabe.txt")
%>%

distinct(product_accession, .keep_all = T) %>% # Remover redundância por ID
filter(grepl('XM_',product_accession)) %>% #Filtrar apenas sequências de mRNA
select(c(11,15,6,10,8,9)) #Selecionar atributos de interesse

knitr::kabe(head(feature_RefSeq))
```

product_accession	symbol	chromosome	strand	start	end
-------------------	--------	------------	--------	-------	-----

XM_028049264.1	LOC114164547	1	+	1951	3883
XM_028067637.1	LOC114181237	1	+	53969	54595
XM_028075424.1	LOC114187231	1	+	64273	73921
XM_028059640.1	LOC114174903	1	-	80597	82467
XM_028072747.1	LOC114185176	1	+	82697	86465
XM_028080716.1	LOC114191509	1	+	113231	129552

#2.3 Gerar DataFrame final

```
df_final <- feature_RefSeq %>%
```

```
left_join(df, by = c('product_accession' = 'Transcrito')) # Unir colunas entre os
Dataframes anteriores pelo ID do transcrito
```

```
knitr::kabe(head(df_final))
```

product_accession	symbol	chromosome	strand	start	end	reg_SI	reg_PO
XM_028049264.1	LOC114164547	1	+	1951	3883	ns	UR
XM_028067637.1	LOC114181237	1	+	53969	54595	NA	NA
XM_028075424.1	LOC114187231	1	+	64273	73921	ns	ns
XM_028059640.1	LOC114174903	1	-	80597	82467	ns	ns
XM_028072747.1	LOC114185176	1	+	82697	86465	DR	UR
XM_028080716.1	LOC114191509	1	+	113231	129552	ns	ns

3 Mapear Transcritos por fita

#3.1 Fita sense

```
map_plus <- df_final %>%
```

```
distinct(symbol, .keep_all = T) %>% #Remover redundância nos loci (elimina i
soformas)
```

```
group_by(chromosome) %>% #Agrupar operações por cromossomo
```

```
mutate(ordem = seq(start)) %>% #Criar índice com base na posição de início
```

```
filter(strand == "+") %>% #Filtrar apenas aqueles na fita sense
```

```
filter( ( (reg_SI == "UR") | (reg_SI == "DR") | (reg_SI == "ns") ) & (reg_PO
== "UR") ) %>% # Filtrar por expressão (UR_SI ou DR_SI ou n.s._SI) e (UR_PO).
```

```
arrange(chromosome,ordem) #Ordenar por índice por cromossomo
```

```
knitr::kabe(head(map_plus))
```

product_accession	symbol	chromosome	strand	start	end	reg_SI	reg_PO	ordem
XM_028049264.1	LOC114164547	1	+	1951	3883	ns	UR	1

XM_028072747.1	LOC114185176	1	+	82697	86465	DR	UR	5
XM_028084539.1	LOC114194366	1	+	291590	297933	ns	UR	22
XM_028082510.1	LOC114192680	1	+	352759	356891	DR	UR	28
XM_028081201.1	LOC114191802	1	+	375159	378543	ns	UR	30
XM_028054770.1	LOC114169535	1	+	414563	422639	ns	UR	33

#3.2 Fita antisense

```
map_minus <- df_final %>%
```

```
  distinct(symbol, .keep_all = T) %>% #Remover redundância nos loci (elimina i
```

```
  group_by(chromosome) %>% #Agrupar operações por cromossomo
```

```
  mutate(ordem = seq(start)) %>% #Criar índice com base na posição de início
```

```
  filter(strand == "-") %>% #Filtrar apenas aqueles na fita antisense
```

```
  filter( ( reg_SI == "UR" ) | ( reg_SI == "DR" ) | ( reg_SI == "ns" ) ) & ( reg_PO
```

```
== "UR" ) ) %>% # Filtrar por expressão (UR_SI ou DR_SI ou n.s._SI) e (UR_PO).
```

```
  arrange(chromosome,ordem) #Ordenar por índice por cromossomo
```

```
knitr::kable(head(map_minus))
```

product_accession	symbol	chromosome	strand	start	end	reg_SI	reg_PO	ordem
XM_028086254.1	LOC114195797	1	-	401976	404219	ns	UR	32
XM_028070165.1	LOC114183228	1	-	558321	577328	ns	UR	44
XM_028083488.1	LOC114193617	1	-	594688	596590	ns	UR	46
XM_028083478.1	LOC114193606	1	-	602877	604643	ns	UR	47
XM_028086848.1	LOC114196257	1	-	638971	640694	ns	UR	51
XM_028071727.1	LOC114184412	1	-	648668	653880	DR	UR	53

4 Definir função para identificar os blocos

```
consec <- function (ordem, name, n) { #argumentos: ordem dos transcritos no genom
```

```
  a, name representa o ID (XM_) e n o número de loci consecutivos a serem buscado
```

```
s
```

```
  temp <- c() #iniciar vetor temporário
```

```
  result <- c() #iniciar vetor de resultados
```

```
  for (i in seq(ordem)) { #iterar pelos índices do vetor de ordem
```

```
    temp <- append(temp, ordem[ i:(i+(n-1))] ) # Adicionar na lista temporária os g
```

```
    enes do índice atual até os próximos n após esse
```

```

    if (sum(temp[n], -temp[1], na.rm = T) == n-1) { # Verificar se a diferença entre
      o ultimo e o primeiro item do vetor temporario é igual a n-1, ou seja são consecut
      ivos

      result <- append(result, name[i:(i+(n-1))]) # Caso verdadeiro, adicionar os elem
      entos correspondentes no vetor de resultados
    }

    temp<- c() #Caso falso, esvaziar a lista temporária
    next # Ainda, caso falso, seguir para próxima iteração

  }

  return(result) #Retornar resultados
}

```

#5 Aplicar função no conjunto de dados

#5.1 DataFrame da fita sense

```

b6_plus <- consec(map_plus$ordem,map_plus$product_accession,6)
b5_plus <- consec(map_plus$ordem,map_plus$product_accession,5)
b4_plus <- consec(map_plus$ordem,map_plus$product_accession,4)
b3_plus <- consec(map_plus$ordem,map_plus$product_accession,3)

```

#5.2 DataFrame da fita antisense

```

b6_minus <- consec(map_minus$ordem,map_minus$product_accession,6)
b5_minus <- consec(map_minus$ordem,map_minus$product_accession,5)
b4_minus <- consec(map_minus$ordem,map_minus$product_accession,4)
b3_minus <- consec(map_minus$ordem,map_minus$product_accession,3)

```

#6 Formatar resultado final

```

df_blocos_plus <- filter(map_plus, product_accession %in% c(b3_plus,b4_plus,b5_plus,
b6_plus))
df_blocos_plus <- df_blocos_plus %>%
  arrange(chromosome, ordem)

df_blocos_minus <- filter(map_minus, product_accession %in% c(b3_minus,b4_minus,b
5_minus,b6_minus))
df_blocos_minus <- df_blocos_minus %>%
  arrange(chromosome, ordem)

```

```
knitr::kabe(head(df_blocos_plus))
```

product_accession	symbol	chromosome	strand	start	end	reg_SI	reg_PO	ordem
XM_028063227.1	LOC114177719	1	+	38541024	38545935	UR	UR	2032
XM_028082462.1	LOC114192650	1	+	38551011	38552293	ns	UR	2033
XM_028077425.1	LOC114188782	1	+	38555097	38562315	ns	UR	2034

XM_028071823.1	LOC114184514	1	+	41877640	41880906	ns	UR	2416
XM_028071831.1	LOC114184520	1	+	41884179	41887233	ns	UR	2417
XM_028070732.1	LOC114183646	1	+	41892357	41894393	ns	UR	2418

APÊNDICE D – COMANDOS PARA ALINHAR AS PROTEÍNAS CORRESPONDENTES AOS BLOCOS MAPEADOS CONTRA SEQUÊNCIAS DE *A. THALIANA* DEPOSITADAS NO STRING VIA BLAST

Criar arquivos para reconhecimento como database: makeblastdb -dbtype prot -in [PATH Ara_stringdb.fasta]

Alinhamento : blastp -query [PATH blocos_vung.fasta] -db [PATH Ara_stringdb.fasta] -outfmt 6 -out bATS_output.txt -num_threads 4 -evalue 1e-20 -max_target_seqs 1 -max_hsps

APÊNDICE E – COMANDO UTILIZADO PARA ANÁLISE COM A FERRAMENTA GET_HOMOLOGUES

```
get_homologues.pl -d Dir-X -E 1e-10 -t 2 -r Vung_blocos.faa -n 4 -I seq_name.txt
```

-d directory with input FASTA files (.faa / .fna)

-X use diamond instead of blastp

-E max E-value

-t report sequence clusters including at least t taxa

-r reference proteome .faa/.gbk file

-n nb of threads for BLAST/HMMER/MCL in 'local' runmode

-I file with .faa/.gbk files in -d to be included

Unir clusters em um único arquivo com quebra de linha entre eles :

```
for f in *.faa; do (cat "${f}"; echo) >> finalfile.txt; done
```

APÊNDICE F – DETALHES ACERCA DOS COMPONENTES DOS BLOCOS DE GENES CONSECUTIVOS E CO-EXPRESSOS IDENTIFICADOS NO GENOMA DE *V. UNGUICULATA*

bloco_ID	RefSeq_ID	Anotação	Sigla_Uniprot	SI	Tam(Kpb)	Vetor
C01.b01.+	LOC114177719; LOC114192650; LOC114188782	protein SAR DEFICIENT 1-like; dnaJ homolog subfamily B member 6; importin subunit alpha-2-like	SARD1; DNAJB6; IMPA2	UR n.s. n.s.	21,3	Fago λ
C01.b02.+	LOC114184514; LOC114184520; LOC114183646	Werner Syndrome-like exonuclease (2x); UDP-glucuronate 4-epimerase 6	WEX (2X); GAE6	n.s.	16,8	Fago λ
C02.b01.+	LOC114174011; LOC114173965; LOC114172973	protein JINGUBANG; uncharacterized LOC114173965; zinc finger protein 5	JGB; UNCH; ZIC5	UR n.s. UR	53,0	Fago P1
C02.b02.+	LOC114173242; LOC114173853; LOC114173854	probabe serine/threonine-protein kinase PIX13 (2x); dirigent protein 11-like	PIX13 (2X); DIR11	UR	12,5	Plasmídeo
C02.b03.+	LOC114173769; LOC114167418; LOC114173084; LOC114166941	ARM REPEAT PROTEIN INTERACTING WITH ABF2-like; 2-hydroxyisoflavanone dehydratase-like (3x)	ARIA; HIDH (3X)	DR UR (3x)	24,9	Fago λ
C02.b04.+	LOC114173208; LOC114168626; LOC114174121	amino acid transporter AVT1I-like; uncharacterized LOC114174121	AVT1L (2x); UNCH	UR	37,4	Cosmídeo
C03.b01.+	LOC114179948; LOC114177763; LOC114178971	uncharacterized LOC114179948; plant UBX domain-containing protein 2; protein TIFY 6B-like	UNCH; PUX2; TIFY6B	n.s.	15,3	Fago λ
C03.b02.+	LOC114177220; LOC114179161; LOC114177289	probabe WRKY transcription factor 51; small GTPase LIP1; uncharacterized LOC114177289	WRKY51; LIP1; UNCH	UR n.s. n.s.	22,7	Fago λ
C03.b03.+	LOC114179628; LOC114178255; LOC114179359	uncharacterized LOC114179628; heterogeneous nuclear ribonucleoprotein H3; probabe serine/threonine-protein kinase PIX13	UNCH; HNRNPH3; PIX13	n.s. n.s. UR	47,0	Fago P1

C03.b04.+	LOC114177973; LOC114176076; LOC114176669	serine/threonine-protein kinase AFC2; cytochrome P450 71A1-like	AFC2; CYP714A1	n.s. UR n.s.	20,4	Fago λ
C03.b05.+	LOC114176935; LOC114178102; LOC114178076	dnaJ homolog subfamily C GRV2; palmitoyl-acyl carrier protein thioesterase, chloroplastic; serine/threonine protein phosphatase 2A 57 kDa regulatory subunit B' kappa isoform-like	GRV2; FATB; B'KAPPA	n.s. n.s. DR	62,4	Fago P1
C03.b06.+	LOC114179179; LOC114178664; LOC114177569	transmembrane 9 superfamily member 11; uncharacterized LOC114178664; protein ENHANCED DISEASE RESISTANCE 2-like,	TMN11; UNCH; EDR2	n.s.	41,2	Cosmídeo
C03.b07.+	LOC114178749; LOC114177571; LOC114176568	12-oxophytodienoate reductase 3-like; uncharacterized LOC114177571; 12- oxophytodienoate reductase 3-like	OPR3; UNCH; OPR3	UR	20,2	Fago λ
C03.b08.+	LOC114178283; LOC114175817; LOC114178291	probabe polygalacturonase; protein TIFY 10A; uncharacterized LOC114178291	pga; TIFY10A; UNCH	n.s. UR UR	21,8	Fago λ
C03.b09.+	LOC114177279; LOC114177706; LOC114177705	pathogenesis-related protein 2-like; pathogenesis- related protein STH-2-like; uncharacterized LOC114177705	PR2; STH-2; UNCH	n.s. UR UR	14,0	Plasmídeo
C04.b01.+	LOC114182316; LOC114182455; LOC114180296	arginine--tRNA ligase, chloroplastic/mitochondrial- like; DNA-directed RNA polymerase III subunit 2; serine/threonine-protein kinase VPS15	EMB1027; POLR3; VPS15	DR n.s. n.s.	41,5	Cosmídeo
C04.b02.+	LOC114181799; LOC114181797; LOC114181798	TMV resistance protein N-like (3x)	N (3X)	UR n.s. UR	16,7	Fago λ
C04.b03.+	LOC114182126; LOC114182340; LOC114181978	probabe methyltransferase PMT16; protein IQ- DOMAIN 14-like; ubiquitin-conjugating enzyme E2 2	At2g45750; IQD14; UBE2V2	n.s. UR n.s.	39,1	Cosmídeo
C05.b01.+	LOC114183868; LOC114185495; LOC114183107	trafficking protein particle complex subunit 2-like protein; V-type proton ATPase subunit B 1; AT-hook motif nuclear-localized protein 23-like	TRAPPC2L; VHA-B1; AHL23	n.s. n.s. DR	22,1	Fago λ
C05.b02.+	LOC114185393; LOC114185487; LOC114183603	transcriptional corepressor LEUNIG_HOMOLOG- like; E3 ubiquitin-protein ligase PRT1; cyclin- dependent kinase inhibitor 4	LUH; PRT1; KRP4	DR n.s. UR	101,0	Bacteriófago
C05.b03.+	LOC114185693; LOC114183835; LOC114183834	ESCRT-related protein CHMP1B; IAA-amino acid hydrolase ILR1-like 4; uncharacterized RNA methyltransferase PG_1095	CHMP1B; ILL4; PG_1095	n.s.	14,6	Plasmídeo

C06.b01.+	LOC114188285; LOC114186774; LOC114186955	dihydroxy-acid dehydratase, chloroplastic; basic 7S globulin-like; basic 7S globulin 2-like	DHAD; BG (2X);	n.s. UR UR	14,7	Plasmídeo
C06.b02.+	LOC114186754; LOC114187063; LOC114188761	non-specific lipid-transfer protein-like protein At5g64080; chaperone protein dnaJ 20, chloroplastic- like (2X)	At5g64080; ATJ20 (2X)	DR UR n.s.	18,3	Fago λ
C07.b01.+	LOC114190114; LOC114190113; LOC114192398	albumin-2-like (2x); uncharacterized LOC114190113	alb2 (2X); UNCH	UR	27,3	Fago λ
C07.b02.+	LOC114190574; LOC114192658; LOC114189618; LOC114189489	zeatin O-glucosyltransferase-like; calcium- transporting ATPase 4, plasma membrane-type-like	ZOG1 (3X); ACA4	n.s.	34,8	Fago λ
C07.b03.+	LOC114191677; LOC114189948; LOC114190678	exportin-7-like; ubiquitin-conjugating enzyme E2 28- like; gibberellin-regulated protein 9-like	XPO7; UBC28; GASA9	n.s n.s. DR	56,6	Fago P1
C07.b04.+	LOC114192748; LOC114192499; LOC114189818	putative F-box protein PP2-B12; F-box protein PP2- B10-like; putative F-box protein PP2-B12	PP2B12; PP2B10; PP2B12	n.s. UR n.s.	16,1	Fago λ
C07.b05.+	LOC114191845; LOC114191832; LOC114191426	cysteine-rich receptor-like protein kinase 29; cysteine-rich receptor-like protein kinase 10; putative receptor-like protein kinase At4g00960	CRK29; CRK10; At4g00960	UR UR n.s.	19,2	Fago λ
C08.b01.+	LOC114193730; LOC114193620; LOC114193125	patatin-like protein 2 (3X)	PLP2 (3X)	UR	21,6	Fago λ
C08.b02.+	LOC114195461; LOC114194005; LOC114193498	protein EXORDIUM-like; protein EXORDIUM-like 2; phospho-2-dehydro-3-deoxyheptonate aldolase 1, chloroplastic-like	EXL; EXL2; DHS1	UR	75,8	Fago P1
C08.b03.+	LOC114195301; LOC114193231; LOC114193230	anaphase-promoting complex subunit 5; 4- hydroxyphenylpyruvate dioxygenase-like	ANAPC5; HPDL (2X)	n.s. UR UR	23,9	Fago λ
C09.b01.+	LOC114164574; LOC114164702; LOC114162167	probabe 6-phosphogluconolactonase 4, chloroplastic (3X)	PGL4	n.s.	25,4	Fago λ
C09.b02.+	LOC114163756; LOC114164138; LOC114164235	polyadenylate-binding protein RBP47C-like; uroporphyrinogen-III synthase, chloroplastic; uncharacterized LOC114164235	RBP47C; UROS; UNCH	n.s. n.s. UR	26,1	Fago λ
C09.b03.+	LOC114164035; LOC114162116; LOC114163897	mitogen-activated protein kinase homolog NTF3; bZIP transcription factor 44-like; uncharacterized LOC114163897	NTF3; BZIP44; UNCH	UR n.s. DR	18,0	Fago λ
C09.b04.+	LOC114162123; LOC114164656; LOC114164655	peptidyl-prolyl cis-trans isomerase CYP21-1-like; actin-depolymerizing factor 5-like; subtilisin-like	CYP21-1; ADF5; SBT1.6	DR	15,1	Fago λ

protease SBT1.6

C10.b01.+	LOC114166300; LOC114166085; LOC114165085	cytochrome P450 CYP82D47-like (2X); xanthotoxin 5-hydroxylase CYP82C4-like	CYP82D47 (2X); CYP82C4	UR UR n.s.	94,6	Bacteriófago
C10.b02.+	LOC114166175; LOC114167114; LOC114165104	NADPH-dependent aldo-keto reductase, chloroplastic-like (3X)	AKR4C9 (3X)	n.s.	23,1	Fago λ
C10.b03.+	LOC114167253; LOC114167252; LOC114167251	cysteine proteinase inhibitor-like; multicystatin-like (2X)	CYS; mcd (2X)	UR	100,3	Bacteriófago
C10.b04.+	LOC114166256; LOC114167365; LOC114167320	cationic peroxidase 1-like; peroxidase P7-like (2X)	PNC1; PER7 (2X)	n.s. UR UR	27,9	Fago λ
C10.b05.+	LOC114165813; LOC114165568; LOC114165811	(+)-neomenthol dehydrogenase-like; uncharacterized LOC114165568; (+)-neomenthol dehydrogenase-like	SDR1; UNCH; SDR1	n.s.	52,5	Fago P1
C10.b06.+	LOC114166979; LOC114167236; LOC114166814	uncharacterized LOC114166979; uncharacterized LOC114167236; cation/calcium exchanger 4	UNCH; CCX4	n.s.	24,7	Fago λ
C10.b07.+	LOC114166341; LOC114166339; LOC114165654	linoleate 9S-lipoxygenase 1; linoleate 9S- lipoxygenase 1-like; seed linoleate 9S-lipoxygenase- 2-like	LOX1; LOX1; LOX1.2	UR	33,1	Fago λ
C10.b08.+	LOC114167210; LOC114167407; LOC114167642	ubiquitin receptor RAD23b-like; ylmG homolog protein 1-1, chloroplastic-like; uncharacterized membrane protein At1g16860-like	RAD23B; YLMG1-1; At1g16860	DR n.s. n.s.	23,8	Fago λ
C11.b01.+	LOC114168762; LOC114168273; LOC114169972	small nuclear ribonucleoprotein Sm D2-like; MDIS1- interacting receptor like kinase 2-like; uncharacterized LOC114169972	SNRPD2; MIK2; UNCH	UR n.s. UR	21,1	Fago λ
C01.b01.-	LOC114193753; LOC114180702; LOC114173644	cytochrome P450 94A1-like (3X)	CYP94A1 (3X)	UR UR n.s.	40,7	Cosmídeo
C01.b02.-	LOC114174173; LOC114182046; LOC114177142	organelle RRM domain-containing protein 2, mitochondrial-like; uncharacterized LOC114182046; uncharacterized LOC114177142	ORRM2; UNCH (2X)	UR n.s. UR	17,2	Fago λ
C01.b03.-	LOC114184616; LOC114184689; LOC114192772	60S ribosomal protein L21-1; sister-chromatid cohesion protein 3; NPL4-like protein 1	RPL21; SCC3; At3g63000	n.s.	17,7	Fago λ
C01.b04.-	LOC114193563; LOC114191257; LOC114180279	uncharacterized LOC114193563; wound-induced protein 1; uncharacterized LOC114180279	UNCH; WUN1; UNCH	n.s. UR UR	19,0	Fago λ

C02.b01.-	LOC114173930; LOC114171733; LOC114167667	E3 ubiquitin protein ligase RIE1-like; probabe UDP-3-O-acyl-N-acetylglucosamine deacetylase 2, mitochondrial; E3 ubiquitin protein ligase RIE1-like	RIE1 (2X); LPXC2	n.s.	84,3	Fago P1
C02.b02.-	LOC114174686; LOC114174689; LOC114168499	sn1-specific diacylglycerol lipase beta-like (2X); vacuolar protein sorting-associated protein 25	DAGLB (2X); VPS25	UR n.s. n.s.	15,3	Plasmídeo
C02.b03.-	LOC114173870; LOC114173701; LOC114174275	hydroxymethylglutaryl-CoA synthase; DNA-directed RNA polymerase IV subunit 1-like; uncharacterized LOC114174275	HMGS; NRPD1; UNCH	DR n.s. DR	38,4	Cosmídeo
C02.b04.-	LOC114166277; LOC114174640; LOC114173246; LOC114174243	mitogen-activated protein kinase homolog MMK2-like; uncharacterized LOC114174640; 1-aminocyclopropane-1-carboxylate synthase-like; 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, chloroplastic	MMK2; UNCH; ACS1; ISPF	UR n.s. UR n.s.	34,0	Fago λ
C02.b05.-	LOC114173397; LOC114173399; LOC114173396; LOC114173400; LOC114173398; LOC114173395; LOC114173840	chalcone synthase 17-like (6X); ERAD-associated E3 ubiquitin-protein ligase HRD1B	CHS17 (6X); HRD1B	UR(6X) n.s.	32,4	Fago λ
C02.b06.-	LOC114174012; LOC114173797; LOC114173839; LOC114173525	poly(ADP-ribose) glycohydrolase 1-like; 4-coumarate--CoA ligase 2; cellulose synthase-like protein D3; TOM1-like protein 1	PARG1; 4CL2; CSLD3; TOM1L1	n.s. UR n.s. n.s.	32,7	Fago λ
C02.b07.-	LOC114171153; LOC114171158; LOC114169940; LOC114169932	putative 12-oxophytodienoate reductase 11 (4x)	OPR11 (4X)	UR	19,4	Fago λ
C03.b01.-	LOC114177831; LOC114176641; LOC114177489	probabe receptor-like protein kinase At5g24010 (3x)	At5g24010 (3X)	n.s.	21,0	Fago λ
C03.b02.-	LOC114176711; LOC114175837; LOC114177908	5'-nucleotidase domain-containing protein DDB_G0275467; HD domain-containing protein 2; UDP-glucuronate 4-epimerase 1	DDB_G0275467; HDDC2; GAE1	n.s. n.s. UR	30,9	Fago λ
C03.b03.-	LOC114176145; LOC114177169; LOC114177113	uncharacterized LOC114176145; pathogenesis-related protein 2-like (2X)	UNCH; PR2 (2X)	n.s. n.s. DR	17,9	Fago λ
C03.b04.-	LOC114178505; LOC114176230; LOC114175166	flowering-promoting factor 1-like protein 3; 60S ribosome subunit biogenesis protein NIP7 homolog; protein SRC2-like	RAA3; NIP7; SRC2	n.s. n.s. UR	15,0	Plasmídeo
C04.b01.-	LOC114182695; LOC114181851; LOC114180414	proteasome subunit alpha type-2-A; receptor-like protein kinase 5; receptor-like protein kinase HSL1	PAB1; RLK5; HSL1	n.s. n.s. UR	16,6	Fago λ

C04.b02.-	LOC114181454; LOC114182366; LOC114180235	extensin-2-like; RNA-binding protein 1-like; U-box domain-containing protein 28-like	EXT2; RBP1; PUB28	DR n.s UR	39,3	Cosmídeo
C04.b03.-	LOC114182415; LOC114181536; LOC114181522; LOC114181427	heavy metal-associated isoprenylated plant protein 39-like (4X)	HIPP39 (4X)	n.s. UR (3X)	27,0	Fago λ
C05.b01.-	LOC114185690; LOC114184949; LOC114183331	metacaspase-1; phosphatidylinositol 4-phosphate 5- kinase 5-like; elongator complex protein 6	AMC1; PIP5K5; ELP6	n.s. n.s. DR	55,1	Fago P1
C05.b02.-	LOC114182960; LOC114185023; LOC114185022	organic cation/carnitine transporter 4-like; probabe terpene synthase 2; (-)-germacrene D synthase-like	OCT4; TPS2; -	n.s. UR n.s.	71,7	Fago P1
C05.b03.-	LOC114183820; LOC114183623; LOC114185771	probabe LRR receptor-like serine/threonine-protein kinase At1g05700; LRR receptor-like serine/threonine-protein kinase IOS1 (2X)	At1g05700; IOS1 (2X)	UR DR DR	26,0	Fago λ
C05.b04.-	LOC114185155; LOC114185352; LOC114184018	transcription factor MYBS3; NADP-dependent malic enzyme-like; ABC transporter G family member 36	MYBS3; MOD1; ABCG36	UR n.s. n.s.	35,8	Cosmídeo
C06.b01.-	LOC114187124; LOC114187127; LOC114187122; LOC114187125	beta-amyrin synthase-like (4X)	BAS (4X)	n.s.	62,5	Fago P1
C07.b01.-	LOC114192183; LOC114192048; LOC114192381; LOC114191551	adenosine kinase 2-like; isocitrate dehydrogenase [NAD] catalytic subunit 5, mitochondrial; phenylalanine ammonia-lyase class 2; ethylene- insensitive protein 2	ADK2; IDH5; PAL2; EIN2	DR DR UR n. s.	161,8	Bacteriófago
C07.b02.-	LOC114190592; LOC114192126; LOC114190734	GEM-like protein 4 (3X)	GEM4 (3X)	n.s. DR n.s.	10,9	Plasmídeo
C08.b01.-	LOC114194309; LOC114193777; LOC114194772	uncharacterized LOC114194309; gamma- glutamylcyclotransferase 2-3; WAT1-related protein At4g08300-like	GGCT2;3; WAT1	n.s. n.s. DR	58,6	Fago P1
C08.b02.-	LOC114194751; LOC114194103; LOC114195608	CTD nuclear envelope phosphatase 1 homolog; dual specificity tyrosine-phosphorylation-regulated kinase 4; probabe 2-oxoglutarate-dependent dioxygenase At5g05600	CTDNEP1; DYRK4; At5g05600	DR n.s. UR	37,7	Cosmídeo
C09.b01.-	LOC114195845; LOC114162220; LOC114196221	uncharacterized LOC114195845; uncharacterized LOC114162220; ankyrin repeat-containing protein ITN1-like	UNCH (2X); ITN1	n.s. n.s. UR	20,5	Fago λ
C09.b02.-	LOC114164728; LOC114162347; LOC114163080	polygalacturonase inhibitor-like; uncharacterized LOC114162347; probabe alpha,alpha-trehalose-	PGIP1; UNCH; TPS9	n.s n.s. UR	61,2	Cosmídeo

phosphate synthase [UDP-forming] 9						
C09.b03.-	LOC114162305; LOC114162315; LOC114196331	cytochrome P450 71A1-like (3X)	CYP71A1 (3X)	n.s.	188,4	Bacteriófago
C09.b04.-	LOC114196316; LOC114163872; LOC114195852	ribonuclease P protein subunit p25-like protein; RNA-binding protein 39-like; protein LIKE COV 2	RPP25L_1; RBM39; LCV2	n.s. n.s. DR	56,5	Fago P1
C09.b05.-	LOC114163169; LOC114163168; LOC114162222	expansin-A11-like; probabe 26S proteasome non- ATPase regulatory subunit 3; uncharacterized LOC114162222	EXPA11; 21D7; UNCH	DR n.s. DR	13,2	Plasmídeo
C10.b01.-	LOC114165748; LOC114165997; LOC114165394	probabe pyridoxal 5'-phosphate synthase subunit PDX2; E3 ubiquitin-protein ligase AIRP2-like; uncharacterized LOC114165394	PDX2; AIRP2; UNCH	DR n.s. n.s.	49,2	Fago P1
C10.b02.-	LOC114167603; LOC114167214; LOC114167196	metal transporter Nramp6-like; Cb-interacting serine/threonine-protein kinase 5-like; V-type proton ATPase subunit D	NRAMP6; CIPK5; ATP6V1D	DR DR n.s.	24,6	Fago λ
C11.b01.-	LOC114170521; LOC114168757; LOC114168556	transcription factor MYB10-like; transcription factor MYB13-like; transcription factor MYB30-like	MYB10; MYB13; MYB30	UR	29,9	Fago λ

APÊNDICE G – REPRESENTAÇÃO DOS BLOCOS QUE APRESENTARAM TODOS OS COMPONENTES COM RELAÇÃO A ESTRESSES. EM LARANJA BLOCOS COM TERMO GENÉTICA REVERSA E EM AZUL COMPONENTES COM ESTUDOS REFERENTES A TRANSCRIPTÔMICA E SECA

bloco_ID	Nome; sigla	Termos Chave	Referências
C02b.03.+	ARM REPEAT PROTEIN INTERACTING WITH ABF2-like; ARIA	Genética Reversa; Salinidade; Fitormônio	Kim <i>et al.</i> , 2004
	2-hydroxyisoflavanone dehydratase-like; HIDH (3x)	Genética Reversa ^A ; Transcriptômica ^{B,C,D} ; Químicos ^B ; Seca ^{C,D} ; Frio ^C	Shimamura <i>et al.</i> , 2007 ^A ; Chakrabarty <i>et al.</i> , 2009 ^B ; Wu <i>et al.</i> , 2014 ^C ; Arisha <i>et al.</i> , 2020 ^D
C03.b01.+	uncharacterized LOC114179948; UNCH	-	-
	plant UBX domain-containing protein 2; PUX2	Genética Reversa ^A ; Transcriptômica ^{A,B,C} ; Estresse Biótico ^{A,C} ; Salinidade ^B	Chandran <i>et al.</i> , 2009 ^A ; Yu <i>et al.</i> , 2017 ^B ; Meng <i>et al.</i> , 2021 ^C
	protein TIFY 6B-like; TIFY6B	Transcriptômica ^{A,B} ; Salinidade ^A ; Calor ^B	Ismail <i>et al.</i> , 2012 ^A ; Sirhindi <i>et al.</i> , 2016 ^B
C03.b02.+	probabe WRKY transcription factor 51; WRKY51	Transcriptômica ^A ; Fitormônio ^{A,B} ; Estresse Biótico ^A ; Genética Reversa ^B	Gao <i>et al.</i> , 2011 ^A ; Hu <i>et al.</i> , 2018 ^B
	small GTPase LIP1; LIP1	Genética Reversa; Salinidade	Terecskei <i>et al.</i> , 2013
	uncharacterized LOC114177289; UNCH	-	-
C03.b04.+	serine/threonine-protein kinase AFC2; AFC2	Transcriptômica ^{A,B,C} ; Seca ^{A,B} ; Químicos ^C	Alimohammadi <i>et al.</i> , 2013 ^A ; Bidabadi <i>et al.</i> , 2015 ^B ; Tombuloglu <i>et al.</i> , 2015 ^C

	cytochrome P450 71A1-like; CYP714A1 (2x)		
C03.b07.+	12-oxophytodienoate reductase 3-like; OPR3 (2x)	Genética Reversa ^A ; Químicos ^A ; Fitormônio ^{A,B,C} ; Injúria Mecânica ^{B,C} ; Estresse Biótico ^C	Sasaki-Sekimoto <i>et al.</i> , 2005 ^A ; Diaz <i>et al.</i> , 2012 ^B ; Xin <i>et al.</i> , 2017 ^C
	uncharacterized LOC114177571; UNCH	-	-
C04.b02.+	TMV resistance protein N-like; N (3x)	Transcriptômica ^{A,B,C} ; Frio ^A ; Salinidade ^B ; Estresse Biótico ^C ; Luminosidade ^C	Xin <i>et al.</i> , 2013 ^A ; Ning <i>et al.</i> , 2019 ^B ; Zhang <i>et al.</i> , 2019a ^C
C06.b01.+	dihydroxy-acid dehydratase, chloroplastic; DHAD	Proteômica^{A,B}; Químicos^{A,B}; Transcriptômica^C; Seca^C	Li <i>et al.</i>, 2007^A; Zhu <i>et al.</i>, 2018^B; Gai <i>et al.</i>, 2020^C
	basic 7S globulin-like; G7S (2x)	Calor ^A , Salinidade ^B	Fujiwara <i>et al.</i> , 2014 ^A ; Wang <i>et al.</i> , 2015 ^B
C06.b02.+	non-specific lipid-transfer protein-like protein At5g64080; LOC101508788	Transcriptômica ^{A,B} ; Estresse Biótico ^{A,B}	Eichmeier <i>et al.</i> , 2019 ^A ; Rabuma <i>et al.</i> , 2021 ^B
	chaperone protein dnaJ 20; ATJ20 (2x)	Transcriptômica ^{A,B} ; Calor ^A ; Salinidade ^B	Guo <i>et al.</i> , 2019 ^A ; Qin <i>et al.</i> , 2021 ^B
C07.b02.+	zeatin O-glucosyltransferase-like; ZOG1 (3x)	Frio ^A ; Fitormônio ^{B,C} ; Genética Reversa ^{B,C} ; Seca ^{B,C}	Li <i>et al.</i> , 2000 ^A ; Haisel <i>et al.</i> , 2008 ^B ; Havlová <i>et al.</i> , 2008 ^C
	calcium-transporting ATPase 4, plasma membrane-type-like; ACA4	Proteômica ^{A,B} ; Inundação ^A ; Salinidade ^B	Komatsu <i>et al.</i> , 2012 ^A ; Damaris <i>et al.</i> , 2016 ^B
C07.b04.+	putative F-box protein PP2-B12; PP2B12 (2x)	Transcriptômica^{A,B}; Seca^A	Moon <i>et al.</i>, 2014a

	F-box protein PP2-B10-like; PP2B10	Proteômica ^A ; Seca ^{A,B} ;	Jiang <i>et al.</i> , 2019 ^A ; Zhu <i>et al.</i> , 2020 ^B
C08.b01.+	patatin-like protein 2; PLP2 (3x)	Genética Reversa ^A ; Estresse Biótico ^A ; Transcriptômica ^B ; Calor ^B ; Seca ^B	La Camera <i>et al.</i> , 2009 ^A ; Xu <i>et al.</i> , 2018 ^B
C09.b01.+	probabe 6-phosphogluconolactonase 4, chloroplastic; PGL4 (3X)	Proteômica ^{A,B} ; Químicos ^{A,B} ; Transcriptômica ^C ; Luminosidade ^C	Du <i>et al.</i> , 2018 ^A ; Huang <i>et al.</i> , 2020 ^B ; Maher <i>et al.</i> , 2021 ^C
	peptidyl-prolyl cis-trans isomerase CYP21-1-like; CYP21-1	Genética Reversa; Fitormônio	Jung <i>et al.</i> , 2020
C09.b04.+	actin-depolymerizing factor 5-like; ADF5	Salinidade ^A ; Químicos ^A ; Genética Reversa ^{B,C} ; Seca ^B ; Transcriptômica ^C ; Frio ^C	Berni <i>et al.</i> , 2020 ^A ; Yang <i>et al.</i> , 2020 ^B ; Zhang <i>et al.</i> , 2021 ^C
	subtilisin-like protease SBT1.6; SBT1.6	Transcriptômica; Seca	Xiao <i>et al.</i> , 2019
C10.b01.+	cytochrome P450 CYP82D47-like; LOC107935599 (2x)	Transcriptômica ^{A,B} ; Seca ^A ; Salinidade ^A ; Inundação ^B	Arisha; Taha; Elakkad, 2018 ^A ; Park <i>et al.</i> , 2020 ^B
	xanthotoxin 5-hydroxylase CYP82C4-like; CYP82C4	Genética Reversa ^A ; Químicos ^A ; Transcriptômica ^B ; Salinidade ^B	Murgia <i>et al.</i> , 2011 ^A ; Yokotani <i>et al.</i> , 2013 ^B
C10.b04.+	cationic peroxidase 1-like; PNC1	Transcriptômica ^A ; Químicos ^A ; Proteômica ^B ; Genética Reversa ^B	Chakrabarty <i>et al.</i> , 2009 ^A ; Li <i>et al.</i> , 2020 ^B
	peroxidase P7-like; PER7 (2x)	Proteômica ^A ; Frio ^A ; Transcriptômica ^{B,C} ; Químicos ^B ; Salinidade ^C	He <i>et al.</i> , 2018 ^A ; Shah Mohammadi <i>et al.</i> , 2019 ^B ; Zhou <i>et al.</i> , 2020 ^C

C10.b05.+	(+)-neomenthol dehydrogenase-like; SDR1 (2x)	Proteômica ^{A,C} ; Seca ^{A,B} ; Transcriptômica ^B ; Estresse Biótico ^C	Simova-Stoilova <i>et al.</i> , 2015 ^A ; Maheswary <i>et al.</i> , 2016 ^B ; Ni <i>et al.</i> , 2017 ^C
	uncharacterized LOC114165568; UNCH	-	-
C01.b01.-	cytochrome P450 94A1-like; CYP94A1 (3x)	Transcriptômica^{A,C}; Químicos^{A,B}; Seca^C	Chakrabarty <i>et al.</i>, 2009^A; Moon <i>et al.</i>, 2014b^B; Ma <i>et al.</i>, 2017^C
C02.b07.-	putative 12-oxophytodienoate reductase 11; OPR11 (4x)	Proteômica ^{A,C} ; Frio ^A ; Transcriptômica ^B ; Luminosidade ^B ; Salinidade ^C ; Fitôrmônio ^C	Yang <i>et al.</i> , 2012 ^A ; Li <i>et al.</i> , 2020 ^B ; Zhang <i>et al.</i> , 2019b ^C
	uncharacterized LOC114176145; UNCH		
C03.b03.-	pathogenesis-related protein 2-like; PR2 (2x)	Proteômica ^{A,C} ; Composto químico ^A ; Transcriptômica ^B ; Estresse Biótico ^B ; Salinidade ^C	Torres <i>et al.</i> , 2007 ^A ; Guerrero-González <i>et al.</i> , 2011 ^B ; Xiong <i>et al.</i> , 2017 ^C
	proteasome subunit alpha type-2-A; PAB1	Proteômica ^{A,B} ; Seca ^A ; Calor ^B	Zadrazilnik <i>et al.</i> , 2013 ^A ; Sang <i>et al.</i> , 2017 ^B
C04.b01.-	receptor-like protein kinase 5; RLK5	Transcriptômica ^{A,B} ; Químicos ^A ; Estresse Biótico ^B	Chakrabarty <i>et al.</i> , 2009 ^A ; Wang <i>et al.</i> , 2014 ^B
	receptor-like protein kinase HSL1; HSL1	Transcriptômica^{A,B}; Seca^A; Salinidade^A; Inundação^B	George <i>et al.</i>, 2017^A; Yin <i>et al.</i>, 2017^B
C06.b01.-	beta-amyrin synthase-like; BAS (4x)	Fitormônio ^A ; Seca ^B ; Salinidade ^C	Shabani <i>et al.</i> , 2010 ^A ; Nasrollahi <i>et al.</i> , 2014 ^B ; Shirazi <i>et al.</i> , 2019 ^C

APÊNDICE H – RELAÇÃO DOS CROMOSSOMOS ENTRE LEGUMINOSAS CORRESPONDENTES AOS ORTÓLOGOS DOS GENES EM BLOCOS IDENTIFICADOS

bloco	<i>V.unguiculata</i>	<i>V.angularis</i>	<i>V.radiata</i>	<i>P.vulgaris</i>	<i>G.max</i>
C01.b01.+	Vu1	Va3	Vr3	-	-
C01.b02.+	Vu1	-	-	-	-
C01.b01.-	Vu1	-	-	-	Gm3 Gm19
C01.b02.-	Vu1	-	Vr3	Pv1	Gm3 Gm19
C01.b03.-	Vu1	Va3	Vr3	Pv1	Gm3 Gm19
C01.b04.-	Vu1	Va3	-	-	-
C02.b01.+	Vu2	Va6 scfld	scfld	Pv2	Gm1 Gm2
C02.b02.+	Vu2	Va6	Vr7 Vr11	Pv2	-
C02.b03.+	Vu2	-	-	-	-
C02.b04.+	Vu2	-	-	-	-
C02.b01.-	Vu2	-	-	-	-
C02.b02.-	Vu2	-	-	-	-
C02.b03.-	Vu2	Va6	Vr11	Pv2	Gm1 Gm11
C02.b04.-	Vu2	-	Vr2 Vr7 Vr11	-	Gm1 Gm11
C02.b05.-	Vu2	-	-	-	Gm1 Gm11
C02.b06.-	Vu2	Va3 Va6 Va7 Va8 scfld	Vr2 Vr3 Vr6 Vr11	Pv1 Pv2 Pv3 Pv7 Pv8 Pv11	Gm1 Gm2 Gm3 Gm9 Gm11

C02.b07.-	Vu2	-	-	-	-
C03.b01.+	Vu3	Va6 scfld	Vr7	Pv2	Gm5 Gm8
C03.b02.+	Vu3	Va1	Vr1	Pv2	Gm5 Gm8
C03.b03.+	Vu3	Va1	Vr7	Pv3	Gm9 Gm16
C03.b04.+	Vu3	Va1	Vr7	Pv5 Pv3	Gm17 Gm5
C03.b05.+	Vu3	Va1	Vr7	Pv3	Gm17 Gm5
C03.b06.+	Vu3	-	Vr7	Pv3	Gm17 Gm13
C03.b07.+	Vu3	-	Vr7 Vr3 scfld	-	-
C03.b08.+	Vu3	-	-	-	-
C03.b09.+	Vu3	-	-	-	-
C03.b01.-	Vu3	Va1	Vr7	Pv3	Gm5
C03.b02.-	Vu3	Va1 scfld	Vr7	Pv3	Gm2 Gm17
C03.b03.-	Vu3	-	Vr7	Pv3	-
C03.b04.-	Vu3	Va1	Vr7 scfld	Pv3	-
C04.b01.+	Vu4	Va10	Vr1	Pv4	Gm1 Gm7
C04.b02.+	Vu4	-	Vr1 Vr5	Pv4	-
C04.b03.+	Vu4				
C04.b01.-	Vu4	Va3 Va10	Vr1	Pv4	Gm1 Gm8 Gm9 Gm16
C04.b02.-	Vu4	-	-	-	-
C04.b03.-	Vu4	-	-	-	-
C05.b01.+	Vu5	Va3	Vr4	Pv8 Pv7	Gm8 Gm9 Gm10 Gm18 Gm20
C05.b02.+	Vu5	Va3 scfld	Vr4	Pv8	Gm8 Gm18

C05.b03.+	Vu5	Va5	Vr5	Pv5 Pv10	Gm13 Gm15
C05.b01.-	Vu5	Va3	scfld	-	Gm1 Gm8
C05.b02.-	Vu5	Va5 scfld	Vr5 scfld	-	-
C05.b03.-	Vu5	Va5	Vr5	Pv5	Gm13 Gm15
C05.b04.-	Vu5	Va1 Va2 Va3 Va10 scfld	Vr3 Vr5 Vr5 Vr8 Vr10	Pv1 Pv3 Pv5 Pv7 Pv8 Pv9 Pv10	Gm3 Gm13 Gm15 Gm19 Gm2 Gm7 Gm14 Gm5 Gm20 Gm17 Gm10 Gm18 Gm6 Gm8 Gm4
C06.b01.+	Vu6	Va9	Vr10	Pv6	Gm13
C06.b02.+	Vu6	Va9	Vr10	Pv6	Gm8 Gm15
C06.b01.-	Vu6	Va9 scfld	Vr10 scfld	Pv6 Pv10	Gm3 Gm7 Gm8 Gm15
C07.b01.+	Vu7	-	-	-	-
C07.b02.+	Vu7	Va1 Va2 Va4 Va8	-	-	-
C07.b03.+	Vu7	Va2 Va4	Vr5 Vr8	Pv7	Gm4 Gm6 Gm10
C07.b04.+	Vu7	-	-	Pv7	Gm3 Gm10
C07.b05.+	Vu7	-	Vr8	Pv7	-
C07.b01.-	Vu7	Va2 scfld	Vr3 Vr8 scfld	Pv1 Pv7	Gm3 Gm10 Gm13 Gm19 Gm20
C07.b02.-	Vu7	Va2	Vr8	Pv7	-
C08.b01.+	Vu8	Va1 scfld	Vr6	-	-
C08.b02.+	Vu8	Va7	scfld	Pv8	Gm2 Gm14
C08.b03.+	Vu8	-	-	Pv8	Gm14
C08.b01.-	Vu8	Va7	Vr6	Pv8	Gm13 Gm14
C08.b02.-	Vu8	Va7	Vr6	Pv8	Gm2 Gm14

C09.b01.+	Vu9	-	Vr10	Pv9	-
C09.b02.+	Vu9	Va9	Vr5	Pv9	Gm4 Gm6
C09.b03.+	Vu9	Va9	Vr5	Pv9	Gm4 Gm6
C09.b04.+	Vu9	Va4 Va8	Vr2 Vr5	-	Gma4 Gma6 Gm11 Gm12
C09.b01.-	Vu9	-	-	Pv3	-
C09.b02.-	Vu9	Va2 Va4 Va6 Va8 scfld	Vr5 Vr8	Pv1 Pv2 Pv3 Pv9	Gm15 Gm2 Gm4 Gm10 Gm6 Gm18 Gm20 Gm5 Gm1 Gm7 Gm8 Gm12 Gm17
C09.b03.-	Vu9	Va4	Vr5	Pv9	Gm6
C09.b04.-	Vu9	Va4 Va11	scfld	-	Gm6
C09.b05.-	Vu9	Va9	Vr5	Pv9	Gm4 Gm6
C10.b01.+	Vu10	Va11	Vr9	Pv10	Gm16
C10.b02.+	Vu10	-	-	-	-
C10.b03.+	Vu10	-	-	-	-
C10.b04.+	Vu10	-	scfld	Pv10	Gm1 Gm10
C10.b05.+	Vu10	-	-	-	-
C10.b06.+	Vu10	Va11	Vr9	Pv10	Gm7 Gm16
C10.b07.+	Vu10	-	Vr5 Vr9	-	-
C10.b08.+	Vu10	scfld	Vr9	Pv10	-
C11.b01.+	Vu11	-	-	-	Gm10 Gm12
C10.b01.-	Vu10	Va3 Va4 Va5 Va11	-	-	-
C10.b02.-	Vu10	scfld	Vr9	Pv10	Gm7 Gm8
C11.b01.-	-	Va8	Vr2	Pv11	Gm10 Gm12