UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS

Análise de transcritoma e expressão gênica de feijoeiro comum (*Phaseolus vulgaris* L.) submetido ao estresse abiótico.

WENDELL JACINTO PEREIRA

GOIÂNIA-GO
2016

sistema de bibliotecas ufg

PRPG
PRÓ-REITORIA DE PÓS GRADUAÇÃO

UFG

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou *download*, a título de divulgação da produção científica brasileira, a partir desta data.

1. Identificação do material bibliográfico: **[X] Dissertação** **[ ] Tese**

2. Identificação da Tese ou Dissertação:

Nome completo do autor: Wendell Jacinto Pereira

Título do trabalho: Análise de transcritoma e expressão gênica de feijoeiro comum (*Phaseolus vulgaris L.)* submetido ao estresse abiótico

3. Informações de acesso ao documento:

Concorda com a liberação total do documento [X] SIM [ ] NÃO[1]

Havendo concordância com a disponibilização eletrônica, torna-se imprescindível o envio do(s) arquivo(s) em formato digital PDF da tese ou dissertação.

Wendell Jacinto Pereira
Assinatura do(a) autor(a)[2]

Ciente e de acordo:
Wanelle
Assinatura do(a) orientador(a)[2]

Data: 19/08/2019

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. A extensão deste prazo suscita justificativa junto à coordenação do curso. Os dados do documento não serão disponibilizados durante o período de embargo.
Casos de embargo:
- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

[2] A assinatura deve ser escaneada.

sistema de bibliotecas ufg

PRPG
PRÓ-REITORIA DE PÓS-GRADUAÇÃO

UFG

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou *download*, a título de divulgação da produção científica brasileira, a partir desta data.

1. Identificação do material bibliográfico: **[X] Dissertação** **[ ] Tese**

2. Identificação da Tese ou Dissertação:

Nome completo do autor: Wendell Jacinto Pereira

Título do trabalho: Análise de transcritoma e expressão gênica de feijoeiro comum (*Phaseolus vulgaris* L.) submetido ao estresse abiótico.

3. Informações de acesso ao documento:

Concorda com a liberação total do documento [ ] SIM [X] NÃO[1]

Havendo concordância com a disponibilização eletrônica, torna-se imprescindível o envio do(s) arquivo(s) em formato digital PDF da tese ou dissertação.

Wendell Jacinto Pereira.
Assinatura do(a) autor(a)[2]

Ciente e de acordo:

Qwanello
Assinatura do(a) orientador(a)[2]

Data: 02 / 08 / 17

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. A extensão deste prazo suscita justificativa junto à coordenação do curso. Os dados do documento não serão disponibilizados durante o período de embargo.

Casos de embargo:
- Solicitação de registro de patente
- Submissão de artigo em revista científica
- Publicação como capítulo de livro
- Publicação da dissertação/tese em livro

[2] A assinatura deve ser escaneada.

WENDELL JACINTO PEREIRA

Análise de transcritoma e expressão gênica de feijoeiro comum (*Phaseolus vulgaris* L.) submetido ao estresse abiótico.

Dissertação apresentada ao Programa de Pós-Graduação em Ciências Biológicas do Instituto de Ciências Biológicas da Universidade Federal de Goiás, como requisito parcial para a obtenção do título de Mestre em Ciências Biológicas.

Área de Concentração: Bioquímica e Genética.

Orientadora: Profa. Dra. Rosana Pereira Vianello.

GOIÂNIA-GO

2016

Ficha de identificação da obra elaborada pelo autor, através do
Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Pereira, Wendell Jacinto
Análise de transcritoma e expressão gênica de feijoeiro comum (Phaseolus vulgaris L.) submetido ao estresse abiótico. [manuscrito] / Wendell Jacinto Pereira. - 2016.
IX, 81 f.

Orientador: Profa. Dra. Rosana Pereira Vianello.
Dissertação (Mestrado) - Universidade Federal de Goiás, Instituto de Ciências Biológicas (ICB), Programa de Pós-Graduação em Biologia, Goiânia, 2016.
Bibliografia.
Inclui siglas, abreviaturas, gráfico.

1. Estresse abiótico. 2. Transcritoma. 3. Genômica funcional. 4. RNA Seq. 5. SNPs. I. Vianello, Rosana Pereira, orient. II. Título.

CDU 577.2

SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS

ATA DA SESSÃO PÚBLICA DE DEFESA DE DISSERTAÇÃO DE Nº 443

Aos oito dias do mês de março do ano de dois mil e dezesseis, às nove horas, na Embrapa Arroz e Feijão, reuniram-se os componentes da banca examinadora: Profa. Dra. Rosana Pereira Vianello, Prof. Dr. Claudio Brondani e Profa. Dra. Anna Cristina Lanna para, em sessão pública presidida pela primeira examinadora citada, procederem à avaliação da defesa de dissertação intitulada "Análise do transcriptoma e expressão gênica de feijoeiro comum (*Phaseolus vulgaris*) submetido ao estresse abiótico", em nível de mestrado, área de concentração em Bioquímica e Genética, de autoria de **Wendell Jacinto Pereira**, discente do Programa de Pós-Graduação em Ciências Biológicas da Universidade Federal de Goiás. A sessão foi aberta pela presidenta, que fez a apresentação formal dos membros da banca. A palavra, a seguir, foi concedida ao autor da dissertação que em cerca de 8:30 minutos procedeu à apresentação de seu trabalho. Terminada a apresentação, cada membro da banca arguiu o examinado, tendo-se adotado o sistema de diálogo sequencial. Terminada a fase de arguição, procedeu-se à avaliação da dissertação. Tendo-se em vista o que consta na Resolução nº1340 de 2015 do Conselho de Ensino, Pesquisa, Extensão e Cultura (CEPEC), que regulamenta o Programa de Pós-Graduação em Ciências Biológicas, a dissertação foi APROVADA, considerando-se integralmente cumprido este requisito para fins de obtenção do título de Mestre em Ciências Biológicas pela Universidade Federal de Goiás. A conclusão do curso dar-se-á quando da entrega da versão definitiva da dissertação na Secretaria do programa, com as devidas correções sugeridas pela banca examinadora, no prazo de trinta dias a contar da data da defesa. Cumpridas as formalidades de pauta, às 11 horas e 10 minutos, encerrou-se a sessão de defesa e, para constar, eu, Renato César Rodrigues, Assistente em Administração da Secretaria de Pós-graduação do Instituto de Ciências Biológicas da

SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS

Universidade Federal de Goiás, lavrei a presente ata que, após lida e aprovada, será assinada pelos membros da banca examinadora em três vias de igual teor.

Profa. Dra. Rosana Pereira Vianello
Presidenta da Banca
Embrapa Arroz e Feijão

Prof. Dr. Claudio Brondani
Embrapa Arroz e Feijão

Profa. Dra. Anna Cristina Lanna
Embrapa Arroz e Feijão

“Tenho a impressão de ter sido uma criança
brincando à beira-mar, divertindo-me em descobrir
uma pedrinha mais lisa ou uma concha mais bonita
que as outras, enquanto o imenso oceano da
verdade continua misterioso diante de meus olhos”
(Sir Isaac Newton)

Agradecimentos

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa de mestrado.

Ao Centro Nacional de Pesquisa em Arroz e Feijão – (Embrapa Arroz e Feijão) pelo financiamento desse estudo.

À Dra. Rosana Pereira Vianello pela exímia orientação e profissionalismo durante a realização desse trabalho, pelos ensinamentos, conselhos e pela confiança. Agradeço ainda por propiciar um ambiente de oportunidades e colaborações que tanto enriqueceram a minha formação.

Ao Dr. Arthur Tavares de Oliveira Melo pela colaboração na execução deste trabalho, por propiciar o primeiro contato e facilitar a minha inserção no fascinante campo da bioinformática.

Ao Prof. Dr. Alexandre Siqueira Guedes Coelho por fornecer a estrutura de análise necessária para a condução do estudo, pelas dicas e valorosos conhecimentos transmitidos.

Às pesquisadoras Dra. Anna Cristina Lanna, Dra. Tereza Cristina de Oliveira Borba, Dra. Priscila Zaczuk Bassinello e ao pesquisador Dr. Claudio Brondani pelo apoio e pelas contribuições durante a execução deste trabalho.

Às analistas, aos assistentes e estagiários do Laboratório de Biotecnologia da Embrapa Arroz e Feijão pelas contribuições e trocas de conhecimento.

A todos os professores, técnicos e demais servidores que fazem parte do Programa de Pós-Graduação em Ciências Biológicas da Universidade Federal de Goiás pelo empenho em nos garantir uma formação adequada.

Por fim, aos meus pais, Fátima e Luiz, aos familiares, amigos e a todos que de alguma forma contribuíram para minha formação acadêmica e pessoal, meu muito obrigado!

SUMÁRIO

LISTA DE ABREVIATURAS E SIGLAS

Água DEPC - Água livre de RNAse após tratamento com Dicarbonato de dietila (Diethylpyrocarbonate)
bp – *Base pairs*
BQSR - *Base Quality Score Recalibration*
cDNA - DNA complementar ao RNA
CIAT - Centro Internacional de Agricultura Tropical
CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico
CONAB - Companhia Nacional de Abastecimento
CPM – *Count per million*
Cq – *Quantitative cycle*
CV – Coeficiente de variação
DNA - Ácido desoxirribonucleico
DNase – Desoxirribonuclease
Eff - Eficiência
Embrapa - Empresa Brasileira de Pesquisa Agropecuária
ESTs - *Expressed Sequence Tags*
FPKM – *Fragments Per Kilobase Million*
GAII - Genome Analyzer II
GATK - Genome Analyze Toolkit
GO – Gene Ontology Consortium
Indels – Inserção ou Deleção
MAPA - Ministério da Agricultura, Pecuária e Abastecimento
Mercosul - Mercado Comum do Sul
NaCl – *sodium chloride*
Nr – *Non redundant database* (NCBI)
NTC – *Non-template control*
ORFs – *Open Reading Frame*
PCR - *Polymerase Chain Reaction*
qPCR - *Quantitative Real-time Polymerase Chain Reaction*
RI - Realinhamento de indels
RIN – *RNA Integrity Number*
Rn - *Normalized report*
RNA - Ácido ribonucléico
SD – *Standard deviation*
SNPs – *Single Nucleotide Polymorphisms*
ssDNA – *Single-strand(ed) DNA*
Tm – *Melting temperature*

RESUMO

O feijoeiro comum (*Phaseolus vulgaris* L.) é um gênero alimentício de primeira necessidade fortemente associado à segurança alimentar mundial. Dos fatores que afetam a sua produção, episódios de estresses abióticos, como a seca, figuram como de grande risco e impacto, podendo restringir as áreas de cultivo e o rendimento das lavouras. A caracterização em larga escala da expressão diferencial de genes envolvidos na resposta da planta aos estresses abióticos possibilita traçar o perfil dos genes, bem como das rotas metabólicas, envolvidos nos mecanismos de tolerância aos estresses abióticos. Nesse estudo, a expressão de genes relacionados à resposta aos estresses abióticos em tecido foliar e radicular de dois genótipos de feijoeiro comum contrastantes quanto à tolerância ao estresse de seca foi avaliada pela análise de RNA-Seq. Um novo transcritoma de feijoeiro comum foi construído apresentando 1668 novos loci e 23 169 novas isoformas. Ao todo foram identificados 54 807 transcritos distribuídos em 28 640 genes. A análise de expressão diferencial em tecido foliar e radicular foi realizada através do pacote edgeR, Bioconductor, possibilitando identificar 1242 genes diferencialmente expressos (GDE). Destes, 458 genes tiveram expressão diferenciada no genótipo tolerante (191 induzidos e 267 reprimidos) quando comparado ao genótipo sensível. A anotação funcional revelou no genótipo tolerante um predomínio de genes nas categorias enriquecidas *oxidoreductase activity*, *oxidation-reduction process*, *regulation of gene expression*, *regulation of macromolecule metabolic process* e *dioxygenase activity*. Por meio da técnica de qPCR, dos 15 GDEs selecionados para caracterização, 74,42% foram validados quanto à expressão diferencial. Adicionalmente, um total de 151 283 variantes foram identificadas, sendo 135 167 SNPs e 16 115 Indels. A anotação e predição de efeitos das variantes foi realizada com SnpEff v.4.2, onde foram identificados 330 378 efeitos putativos para as variantes encontradas. Os dados utilizados nesse estudo serão disponibilizados no *Sequence Read Archive* (NCBI). Os resultados deste estudo contribuem para a ampliação do conhecimento sobre os mecanismos gênicos, bem como das variantes funcionais, relacionados com a tolerância à estresses abióticos em feijoeiro comum.

Palavras-Chave: Estresse abiótico, transcritoma, genômica funcional, RNA-Seq, SNPs.

ABSTRACT

The common bean (*Phaseolus vulgaris* L.) is an important foodstuff strongly associated with an issue if global food security. Among the factors that affect their agricultural production, abiotic stress episodes listed as high risk and impact, can restrict the area under cultivation and crop yields. The characterization of high-throughput differential gene expression involved in plant response to abiotic stresses enables profile the genes, and the metabolic pathways, involved in the mechanisms of tolerance to abiotic stresses. In this study, the expression of genes related to response to drought stress in leaf and root tissue of two contrasting common bean genotypes for stress tolerance was evaluated by RNA-Seq. A new *Phaseolus vulgaris* transcriptome was built featuring 1,668 new loci and 23,169 new isoforms. In all, 54,807 transcripts were identified, found being distributed in 28,640 genes. Through the edgeR package (Bioconductor), they determined by differential expression analysis of leaf and root 1,242 differentially expressed genes. Of these, 458 genes had different expression in tolerant genotype (191 up-regulates and 267 down-regulated) when compared to the sensitive genotype. Functional annotation revealed in tolerant genotype predominance of genes in the categories enriched oxidoreductase activity, oxidation-reduction process, regulation of gene expression, regulation of macromolecule metabolic process and dioxygenase activity. By qPCR, of the 15 DEGs identified via RNA-Seq and selected for characterization, 74.42% were validated for differential expression. Additionally, a total of 151,283 variants were identified, of which 135,167 are SNPs and 16,115 are indels. Through functional annotation, performed in SnpEff v.4.2, were identified 330,378 effects caused by variants. The data used in this study were deposited in the Sequence Read Archive (NCBI) and will be available. The results of this study contribute to the expansion of knowledge about the gene mechanisms, as well as their functional variants, related to abiotic stresses tolerance in beans.

Keywords: Abiotic stress, transcriptome, functional genomics, RNA-Seq, SNPs.

1. Introdução

O feijoeiro comum pertence à família *Fabaceae*, gênero *Phaseolus* L., espécie *Phaseolus vulgaris* L, sendo, portanto, uma leguminosa (1). As leguminosas constituem a terceira maior família de angiospermas e incluem várias espécies cultiváveis importantes, como ervilhas (*Pisum sativum* L.), amendoim (*Arachis hypogaea* L.), soja (*Glycine max* (L.) Merr.) e o feijoeiro comum (*Phaseolus vulgaris* L.) (2).

Assim como outras leguminosas consumidas na forma de grãos secos, o feijoeiro comum constitui uma importante fonte nutricional para um grande número de pessoas, particularmente na população de baixa renda de países em desenvolvimento. Devido ao seu baixo custo, quando comparado às fontes proteicas derivadas de animais, e ao grande fornecimento de nutrientes como proteínas, fibras dietéticas, carboidratos complexos, vitaminas e minerais, o feijoeiro é considerada a leguminosa mais importante para o consumo humano. Por apresentar alto teor de lisina em suas proteínas, as leguminosas são consideradas um bom complemento para uma dieta baseada em grãos cereais básicos, como o arroz e o milho, que possuem baixo teor proteico e pequena quantidade desse aminoácido essencial (3–6).

Além da importância nutricional do feijoeiro, tem sido demonstrada sua capacidade nutracêutica devido ao seu alto teor de compostos bioativos como amido resistente, polifenóis, oligossacarídeos, frações proteicas não digeríveis e peptídeos bioativos (7). Dentre os efeitos observados dos compostos bioativos estão: indução de atividade protetora de mucosas, ação anti-inflamatória e antioxidante, além do auxílio na prevenção e tratamento de doenças como diabetes, hipertensão e até mesmo na inibição da proliferação de células cancerígenas (8–12).

Aproximadamente, 75% da produção mundial de feijão é proveniente de apenas nove países, sendo o Brasil o terceiro maior país produtor, atrás da Índia e de Myanmar. Dentro do Mercosul, o Brasil é o responsável por cerca de 90% da produção e consumo de feijão, correspondendo a aproximadamente 3,1 milhões de

toneladas ao ano no período de 2009 a 2013 (13). Na safra 2014/2015, a área total plantada foi superior a 3 milhões de hectares, com produtividade estimada em 1025 kg/ha e produção total de 3,12 milhões de toneladas. Os principais estados brasileiros produtores foram o Paraná (640,9 mil toneladas), Minas Gerais (512,4 mil toneladas) e Mato Grosso (484,5 mil toneladas) (13).

Assim como acontece com a maioria dos principais países produtores de feijão, a produção brasileira é praticamente toda destinada ao mercado consumidor interno. Essa característica de produção voltada ao mercado interno, associada com o fato de que os grãos de feijoeiro são pouco consumidos nos países desenvolvidos, gera um pequeno aporte de divisas para o país. Contudo, a sua grande importância comercial e nutricional para os países consumidores, assim como, descobertas recentes do potencial de seus componentes bioativos para o tratamento e prevenção de doenças, tem levado a um expressivo número de estudos com essa espécie.

Mesmo com a sua grande importância, somente recentemente os mecanismos que originaram a espécie foram esclarecidos. Em 2014 foi publicado o genoma do feijoeiro comum Andino, com tamanho estimado em 587 Mpb, com aproximadamente 27 mil genes identificados (14). Já o genoma do feijoeiro comum Mesoamericano foi publicado em 2016, com tamanho estimado em 549,6 Mpb e aproximadamente 30 mil genes identificados (2). Atualmente, considera-se que a espécie *Phaseolus vulgaris* L. teve origem na Mesoamerica, dispersando-se posteriormente pelas Américas, originando os *pools* gênicos Andino e Mesoamericano. Posteriormente, a espécie teria passado por ao menos 2 processos independentes de domesticação, gerando os pool gênicos atuais (2,15).

Estudos recentes com *Phaseolus vulgaris* L., englobam pesquisas em genética genômica para avanço do conhecimento (14,16), redução na perda de qualidade do grão em decorrência do armazenamento (17,18), melhoria nas características nutricionais (4), resistência e/ou tolerância aos estresses biótico e abiótico(19–24). Adicionalmente, pesquisas destinadas à aplicação de seus componentes na área da saúde (8,10,12) e na produção de alimentos funcionais (25,26) também vêm sendo conduzidas.

Com a ampla difusão de estudos com essa espécie, cresce a demanda por técnicas que permitam gerar e ampliar os conhecimentos básicos e aplicados em relação ao feijoeiro comum. Nesse sentido, os estudos destinados ao entendimento dos perfis de transcrição destacaram-se, evoluindo de técnicas que examinavam um pequeno número de genes até tecnologias capazes de avaliar todo o perfil transcricional em um dado momento (transcritoma), como as baseadas no sequenciamento de nova geração. Essas tecnologias reduziram tanto o custo por reação, quanto o tempo de execução dispendido para o sequenciamento, tornando a técnica acessível para uma gama de abordagens experimentais (27). Dessa forma, a transcritômica consolidou-se tendo como objetivos principal catalogar todas espécies de transcritos, incluindo os mRNAs, *non-coding* RNAs e *small* RNAs, visando determinar a estrutura dos genes, em relação aos sítios de início da transcrição, regiões UTR 5' e 3', sítios de *splicing,* dentre outros (28).

A estratégia de análise por RNA-Seq depende do organismo a ser estudado, das informações já disponíveis, bem como dos objetivos desejados. Por exemplo, quando não há genoma de referência disponível, é possível a construção do transcritoma através do montagem *de novo.* Uma vez que exista a disponibilidade do genoma de referência para a espécie, sendo este bem anotado, é possível realizar a identificação de novos transcritos através do mapeamento de *reads* no genoma. Paralelamente, os pesquisadores podem estar interessados em realizar apenas a quantificação dos níveis de mRNA ou microRNAs (miRNAs), bem como na identificação de variantes estruturais, como os SNPs e Indels (29).

A despeito da grande difusão dessa tecnologia, os estudos utilizando RNA-Seq ainda estão sujeitos a uma série de vieses que podem ser introduzidos em todas as etapas do processo para obtenção dos dados, desde o planejamento experimental, preparo das bibliotecas até no sequenciamento propriamente dito (29). Igualmente, diversos métodos e algoritmos estão disponíveis e são implementados para análise dos dados de RNA-Seq em suas várias etapas, como o alinhamento dos *reads*, montagem de transcritos e análise de expressão gênica diferencial (incluindo procedimentos de normalização, modelagem estatística, determinação da variância, dentre outros) (30). Sabidamente nenhum método de análise é livre de erros e a validação dos resultados obtidos é necessária. Dessa forma, para os experimentos de quantificação dos perfis de transcrição, a validação de alguns

genes diferencialmente expressos (GDEs) por PCR quantitativa (qPCR) é amplamente relatada como forma de validação dos resultados de RNA-Seq (22,30–32).

A PCR quantitativa (qPCR) é uma das técnicas com maior difusão e aplicação em estudos envolvendo diversas espécies, seja para mensurar a expressão de genes de interesse ou validar experimentos de microarranjo e RNA-Seq. A técnica baseia-se na amplificação *in vitro* de cópias do DNA complementar ao RNA (cDNA), monitorando em tempo real a quantidade de moléculas produzidas em cada ciclo (33). Sua simplicidade conceitual e facilidade de execução, aliadas com a velocidade, sensibilidade e especificidade em um ensaio homogêneo, a transformaram no padrão para a quantificação de ácidos nucleicos (34,35).

Como descrito por Navarro et al. (36), nos últimos 15 anos um grande número de métodos para detecção de DNA em qPCR foram desenvolvidos e comercializados. As principais metodologias implementadas se baseiam em um dos seguintes métodos:

a) Corante de ligação ao DNA (*DNA binding dyes*): são moléculas que se ligam à fita do DNA e, quando excitadas por determinado comprimento de onda, emitem fluorescência que é então captada e utilizada para a quantificação. Existe uma grande variedade dessas moléculas para comercialização, como Brometo de etídio, YO-PRO-1, SYBR® Green I, SYBR® Gold, SYTO, BEBO e EvaGreen, sendo que a mais utilizada é o SYBR® Green I.

b) Oligonucleotídeos marcados com fluoróforo: nesse sistema, os fluoróforos, pequenas moléculas fluorescentes, estão ligados a oligonucleotídeos que funcionam como sondas, ligando-se a uma região específica do fragmento a ser amplificado. Nesse sistema de detecção, utiliza-se a atividade nucleásica 5’ da Taq DNA Polimerase para gerar o sinal fluorescente durante a qPCR, sendo o sistema comercial TaqMan™ o mais conhecido.

Com a possibilidade de aplicação dessas tecnologias em pesquisas de genética e genômica, muitos estudos têm sido realizados visando à caracterização da genômica funcional do feijoeiro comum, agregando conhecimento a respeito dos genes e rotas metabólicos presentes em *Phaseolus vulgaris* L.. Por exemplo, Patel

et al. (37) realizaram a montagem e a notação do transcritoma da variedade SRR1283084 de feijoeiro comum, enquanto O'Rourke et al. (32) identificaram 26,964 genes distribuídos em sete tecidos diferentes, constituindo um atlas da expressão gênica para feijoeiro, onde foram verificados a ocorrência de genes diferencialmente expressos em vários estádios de desenvolvimento e em três condições de suplementação de nitrogênio.

Embora a espécie seja amplamente estudada, muitas questões sobre o feijoeiro permanecem não esclarecidas. Do ponto de vista agronômico, muitos aspectos necessitam ser avaliados, visando o aumento do conhecimento sobre caracteres importantes e o fornecimento de subsídios para os programas de melhoramento de feijoeiro. Dentre os fatores relevantes, a tolerância aos estresses abióticos tem papel fundamental nas pesquisas com feijoeiro comum, principalmente com a expectativa de aumento na frequência desse tipo de estresse, em decorrência das mudanças climáticas.

De acordo com Wani et. al. (38), os principais fatores ambientais que impõem estresse às plantas são a seca, salinidade, calor, frio, congelamento, ozônio e a incidência da radiação ultra violeta (UV). Destes, os mais frequentes e significativos mundialmente são os estresses por seca, salinidade e temperaturas extremas, que afetam o crescimento, produtividade e frequentemente levam a mudanças morfológicas, fisiológicas e bioquímicas nas plantas (39). Adicionalmente, estresses abióticos relacionados com a deficiência de nutrientes, bem como com o excesso de metais como alumínio, cadmio e mercúrio são impactantes sob as espécies cultiváveis em geral.

Dentre os estresses abióticos que acometem as cultivares, o estresse por seca talvez seja o mais severo, estando relacionado com a redução da produção agrícola mundial (38). De fato, a seca é o fator ambiental que desencadeia as maiores perdas na produção de feijoeiro comum. Devido ao seu sistema radicular limitado e superficial, o feijoeiro comum é bastante suscetível à seca, principalmente nos períodos de floração (40). Igualmente, nesse período a exposição dessa espécie às altas temperaturas leva a redução na produção de vagens e, consequentemente, de grãos (41). A ocorrência desses estresses, de forma isolada ou em conjunto com outros estresses abióticos ou bióticos, afeta o feijoeiro em vários estádios de

desenvolvimento, podendo levar a perdas de até 100% na produção dos grãos, dependendo da intensidade e período de ocorrência (42).

Para a agricultura, a seca está relacionada à deficiência de umidade no solo, impossibilitando que sejam atendidas as demandas da planta para o seu crescimento ou produção (43). Estima-se que ondas de seca e calor provocaram prejuízo de aproximadamente 210 bilhões de dólares entre 1980 e 2011, considerando-se apenas os Estados Unidos (43). A seca pode ocorrer em períodos moderados ou muito longos e extremamente danosos, restringindo as colheitas e afetando a produção de alimentos (44). Esse fenômeno tem se tornado mais intenso, elevando a preocupação com a seguridade alimentar à medida que são esperados aumentos na demanda por alimentos nas próximas décadas (45). Espera-se que em 2050 mais de 50% das áreas cultiváveis serão afetadas por esse fenômeno (44).

Durante o processo evolucionário, as mudanças climáticas têm imposto pressão seletiva sob as populações biológicas e a sua diversidade genética, selecionando variações biológicas que desempenham um importante papel na reposta adaptativa às mudanças ambientais (46). Essa pressão seletiva pode conferir aumento na frequência de alelos vantajosos para a resposta adaptativa às mudanças climáticas (47, 48). O impacto da seca sobre a agricultura tem acelerado a busca por variedades cultivadas melhor adaptadas às mudanças climáticas atuais, visando reduzir a vulnerabilidade dos sistemas produtivos (49,50). A capacidade adaptativa às condições de seca é uma característica sofisticada, sendo dependente de adaptações em diferentes níveis, incluindo alterações na morfologia da planta, nas estruturas anatômicas e nas reações bioquímicas e fisiológicas (51).

Plantas submetidas ao déficit hídrico ativam diferentes mecanismos de resistência, visando o equilíbrio entre crescimento e sobrevivência sob estas condições (52). Esses mecanismos visam a tolerância aos períodos de baixa disponibilidade de água e podem ser classificados em mecanismos de prevenção a desidratação, escape, tolerância e recuperação à seca (51). Algumas espécies possuem melhor capacidade para lidar com a seca do que outras, por exemplo, o feijão-caupi (*Vigna unguiculata* L.), considerado bem adaptado às condições de seca

e de alta temperatura, quando comparado a outras espécies, incluindo *Phaseolus vulgaris* L. (53).

Genótipos de feijoeiro mais produtivos sob condições de seca foram desenvolvidos a partir da combinação de variedades Mesoamericanas, possibilitando uma ótima adaptação dos genótipos com grão do tipo Mesoamericano no Brasil e na América Central em detrimento das variedades Andinas. Esforços têm sido realizados para identificar o amplo conjunto de mecanismos associados com tolerância à seca em genótipos de feijoeiro Mesoamericano, como características fisiológicas, fenotípicas e moleculares (54). A despeito dos esforços globais, devido à grande complexidade da resposta das plantas à seca, a correlação precisa entre os mecanismos genéticos e a tolerância à seca ainda não foi totalmente esclarecida. Contudo, tem se demonstrado que a resposta genética ao estresse por déficit hídrico está sob regulação de um grande número de genes, os quais sofrem forte influência dos fatores ambientais (48,55).

A busca por genes responsivos à seca tem sido realizada através de estudos de expressão gênica, como a geração de *expressed sequence tags* (ESTs) e avaliação dos perfis de transcrição via RT-qPCR, possibilitando a identificação de centenas de genes induzidos ou suprimidos em condições de deficiência hídrica. Recchia et al. (56) identificaram 1120 genes diferencialmente expressos em raiz de BAT 477, quando submetido a interrupção da irrigação durante seu desenvolvimento. Usando o mesmo genótipo tolerante (BAT 477), Müller et al. (22) identificaram em folha 802 diferentes EST entre os estádios de floração e enchimento de grãos. Recentemente, Wu et al. (31) construíram um transcritoma *de novo* para *Phaseolus vulgaris* sob condição de seca, onde foram obtidos 49 Mpb de transcritos únicos, dos quais 42% foram anotados. Nesse estudo, foi verificado a ocorrência de 8932 genes diferencialmente expressos entre as cultivares tolerante (Long 22-0579) e suscetível (Naihua).

Os estudos com estresses abióticos específicos são amplamente desenvolvidos e de fundamental importância, ampliando o conhecimento dos mecanismos envolvidos para obtenção de tolerância. No entanto, em condições de cultivo, as plantas estão submetidas à ocorrência de vários estresses abióticos, e bióticos, simultaneamente ou sequencialmente, sendo esta uma ameaça importante

enfrentada pelas plantas (23, 57, 58). A reposta da plantas a combinação de dois diferentes estresses abióticos não pode ser diretamente extrapolada a partir da resposta das plantas a cada um dos estresses isoladamente (59). Nesse sentido, tem se demonstrado ainda que estresses que ocorrem simultaneamente, como seca e calor, seca e salinidade e calor e metais pesados, são mais danosos para a planta do que a ocorrência dos estresses separadamente (58). Dessa forma, a obtenção de cultivares resistentes a mais de um tipo de estresse é extremamente desejável.

Com os avanços em tecnologias de genômica funcional, como o sequenciamento de nova geração e o fornecimento de ferramentas de bioinformática, cada vez mais robustas para melhor explorar o transcritoma *in silico*, espera-se um crescimento exponencial de dados de expressão gênica para feijoeiro, fornecendo uma melhor compreensão dos complexos mecanismos envolvidos na resposta aos estresses abióticos. Neste estudo, foram aplicadas ferramentas de bioinformática para explorar os potenciais genes responsivos à seca em folhas e raízes sob condições de estresse abióticos, com particular interesse no estresse de seca.

Através do sequenciamento de RNA (RNA-Seq) foi realizada a caracterização da expressão gênica de dois genótipos contrastantes quanto a resposta ao estresse de seca, BAT 477 (tolerante à seca) e Pérola (sensível à seca). Adicionalmente, foram verificados os genes diferencialmente expressos avaliando-se os genótipos e tecidos, sob três condições de estresse. Como estudo complementar, foi realizada a caracterização de mais de 150 mil variantes (SNPs e indels) presentes nos dois genótipos. Mais de 130 mil SNPs foram caracterizados quanto aos seus potenciais efeitos com o software SnpEff. Adicionalmente, foram identificados os genes afetados pela ocorrência dos SNPs, os quais foram categorizados de acordo com os termos do *Gene Ontology Consortium* (GO *terms*). Igualmente, todos os genes diferencialmente expressos foram categorizados através dos GO *terms*.

Os resultados obtidos, bem como os dados utilizados para a análise, serão disponibilizados e contribuirão para o enriquecimento do conhecimento dos genes e mecanismos moleculares que estão potencialmente relacionados a resposta aos estresses abióticos.

2. Objetivos

Construir um transcritoma de feijoeiro específico para condição de estresses abióticos, compará-lo ao transcritoma de referência para o feijoeiro (Phytozome v.11.0) e, se possível, identificar novos loci e isoformas.

Identificar genes diferencialmente expressos entre dois genótipos de feijoeiro com fenótipo contrastante para deficit hídrico, bem como relacioná-los com os possíveis mecanismos fisiológicos envolvidos na resposta de tolerância aos estresses abióticos.

Identificar as variantes, SNPs e indels, presentes nos genótipos utilizados, determinar os seus efeitos e, se possível, estabelecer suas relações com a divergência fenotípica entre os genótipos.

Selecionar um conjunto de SNPs com potencial aplicação para construção de um painel de genotipagem para *Phaseolus vulgaris* L..

3. Metodologia

3.1. Material vegetal

Foram utilizados dois genótipos de feijoeiro comum (*Phaseolus vulgaris L.*), BAT 477 e Pérola, bem caracterizados quanto a sua performance sob condições de déficit hídrico, conforme demonstrado por Lanna et al. (60). O genótipo BAT 477 foi desenvolvido pelo Centro Internacional de Agricultura Tropical (CIAT, Cali, Colômbia) e é considerada tolerante à seca (61), enquanto que Pérola é uma cultivar desenvolvida pelo programa de melhoramento genético da Empresa Brasileira de Pesquisa Agropecuária (Embrapa) em 1994, sendo caracterizada como um genótipo suscetível a seca (62). Ambos os genótipos possuem origem mesoamericana e apresentam hábitos de crescimento tipo III (63,64).

3.2. Delineamento experimental

O experimento foi realizado em casa de vegetação na unidade da Embrapa Soja em 2010, usando delineamento em blocos casualizados, com cinco repetições. As plantas dos dois genótipos foram crescidas sob condições ambientais controladas em sistema de hidroponia, utilizando solução nutritiva proposta por Hewitt (65), conforme recomendado por Martins et al., (66). As sementes foram pré-germinadas em papel filtro úmido, em ambiente escuro, com controle da umidade relativa (~ 65%) e da temperatura (~ 25ºC). Os brotos foram então transferidos para os suportes de forma que as raízes fossem completamente submersas na solução nutritiva. Durante 15 dias as mudas foram mantidas na casa de vegetação a aproximadamente 25ºC e 60% de umidade relativa, com período de incidência luminosa de 12h.

Após 15 dias de crescimento, as plantas de ambos os genótipos foram submetidas aos estresses abióticos, através da retirada da solução nutriente, e mantidas em câmara escura. A amostragem foi realizada em três tempos: 0 min após a indução do estresse (condição controle), 75 min após a indução do estresse (T_{75} - condição de estresse intermediário) e 150 min após a indução do estresse (T_{150} - condição de estresse intenso). Conforme estabelecido previamente para soja, esses tempos foram baseados nas taxas de fotossíntese e respiração, em que a taxa máxima de fotossíntese foi encontrada no tempo 0 e o maior valor de respiração foi obtido após 150 min de indução do estresse (66).

3.3. Extração de RNA, preparo das bibliotecas e sequenciamento

Amostras de folhas e raízes jovens foram coletadas após o 15º dia de crescimento para ambos os genótipos, em cada um dos três tempos (T_0, T_{75} e T_{150}), e foram armazenadas em freezer a -80ºC, totalizando doze amostras. O RNA total foi individualmente extraído para o preparo das bibliotecas, utilizando o produto comercial PureLink® RNA Mini Kit (Ambion®, Carlsbad, California), conforme as instruções do fabricante, seguido pela digestão do DNA genômico contaminante através do tratamento com a enzima PureLink® DNase (Ambion®, Carlsbad, California). A qualidade e pureza de cada amostra de RNA foram estimados através de espectrofotometria utilizando o NanoVue™ Plus Spectrophotometer (General Electric Company, GE). Adicionalmente, a integridade das amostras de RNA foi verificada usando tecnologia de microfluidos com o equipamento Agilent 2100 Bioanalyzer instrument (General Electric Company, GE). As bibliotecas de RNA-Seq foram preparadas usando Illumina TruSeq RNA kit, seguindo as instruções do fabricante e sequenciadas como replicas técnicas nas plataformas de sequenciamento Genome Analyzer II e HiSeq2000, ambas fabricadas pela Illumina. Uma *lane* de sequenciamento com 2x100 *cycles* (*paired-end*) foi utilizada para sequenciar as 12 bibliotecas em multiplex em ambas as plataformas.

3.4. Diagnóstico e controle de qualidade dos dados

O diagnóstico de qualidade dos dados gerados pelos sequenciamentos foi realizado através do software FastQC v.0.11.3 (67). Após a interpretação das estimativas de qualidade, foram definidos os parâmetros para o controle de qualidade dos *reads* e realizada a retirada das bases com baixa qualidade associada. Para a remoção das sequências provindas de adaptadores e bases com estimativas de qualidade insatisfatórias, a partir da extremidade 5', as 15 primeiras bases dos *reads* oriundos do GAII, assim como as 17 bases iniciais dos *reads* provindos do sequenciamento no HISeq2000 foram descartadas. O software utilizado para o controle de qualidade dos dados de sequenciamento foi o Trimmomatic v.0.33 (68), onde além do descrito acima, os parâmetros implementados foram: LEADING:39 SLIDINGWINDOW:4:30 TRAILING:30 MINLEN:50.

Ao término do controle de qualidade, os *reads* que apresentavam menos do que 50 pb de tamanho foram eliminados. Dessa forma, para cada biblioteca foram recuperados 2 conjuntos de dados com alta qualidade que foram utilizados nas etapas posteriores, um formado pelos *reads "paired-end"* e outro composto pelos "*single reads*" ou *reads* órfãos (*reads* cujo os pares correspondentes foram descartados).

3.5. Montagem de transcritoma para *P. vulgaris* sob estresse abiótico

Como etapa previa a análise de expressão gênica diferencial, foi montado um transcritoma representativo para *P. vulgaris* em resposta aos estresses abióticos induzidos pela remoção do meio de crescimento. Para a obtenção deste transcritoma, foi utilizado o fluxo de trabalho sugerido para o software Cufflinks v2.2.1 (69). Inicialmente, para cada biblioteca, foi realizado o mapeamento dos

reads no genoma de referência para feijoeiro (14) através do software TopHat v.2.1.0 (70). Foi utilizada a versão do genoma com máscara para as regiões repetitivas (Pvulgaris_218_v1.0.hardmasked.fa, disponível para acesso no site Phytozome v.11.0 (71), visando impossibilitar o mapeamento de *reads* oriundos de regiões de DNA repetitivo. Na etapa seguinte, foi realizada a retirada dos *reads* duplicados, produzidos por viés na etapa de "PCR em ponte" ou no preparo das bibliotecas, através do algoritmo "MarkDuplicates" implementado no software Picard-Tools v. 1.139 (disponível em: https://broadinstitute.github.io/picard/, acesso em 03/2016).

Após a retirada dos *reads* duplicados, Cufflinks v2.2.1 foi utilizado para a montagem de um transcritoma por biblioteca sequenciada, totalizando 24 transcritomas. Em cada montagem, tanto o genoma de referência (*Pvulgaris_218_v1.0.hardmasked.fa*) quanto as informações sobre a anotação de genes fornecidas pelo arquivo GFF3 (*Pvulgaris_218_v1.0.gene.gff3,* disponível para acesso no site Phytozome v.11.0) foram utilizadas como guia durante a montagem. Adicionalmente, foram consideradas como isoformas apenas os transcritos que possuíam ao menos 20% da abundância encontrada para a isoforma principal (-F =0.2) e algoritmo para correção de viés descrito por Roberts, et al., (72) foi implementado através da inclusão do argumento "–b" na linha de comando.

Para a construção de um transcritoma representativo, os transcritomas de todas as bibliotecas foram combinados através do algoritmo Cuffmerge utilizando os parâmetros definidos por padrão. Após a obtenção de um único transcritoma de feijoeiro em condições de seca, o algoritmo Cuffcompare foi utilizado para comparar os *loci* e transcritos montados com os genes e transcritos conhecidos disponibilizados na base de dados Phytozome v.11.0 (Pvulgaris_218_v1.0.gene.gff3). Ambos, Cuffmerge e Cuffcompare são algoritmos inclusos em Cufflinks tools.

3.6. Análise de genes diferencialmente expressos

A análise de expressão diferencial foi aplicada visando determinar genes diferencialmente expressos entre o genótipo tolerante à seca (BAT477) e o genótipo suscetível (Pérola), assim como, para encontrar os genes diferencialmente expressos dentro de cada genótipo em relação ao tempo de submissão ao estresse. Adicionalmente, a expressão gênica diferencial foi avaliada em folha e raiz. Inicialmente, os *reads* que passaram pelo controle de qualidade foram mapeados ao genoma com TopHat v2.1.0. O alinhamento foi direcionado a partir das informações contidas no arquivo merged.gtf, gerado pelo Cuffmerge e que corresponde ao transcritoma específico para condições de estresse abiótico. Os *reads* duplicados, geralmente vieses de sequenciamento, foram retirados logo após a etapa de mapeamento, utilizando o algoritmo “MarkDuplicates” implementado no software Picard-Tools v.1.139.

A abundância dos *reads* alinhados foi quantificada para cada gene através do software HTSeq (73). A análise de expressão diferencial foi realizada com o software edgeR (74), onde foram avaliados somente os genes com contagem por milhão maior do que 1 (CPM > 1) em ao menos duas das amostras avaliadas. A dispersão no conjunto de dados foi estimada através da função “estimateCommonDisp” (edgeR), a partir de um conjunto de genes selecionados por não responderem ao tratamento implementado (~3900 genes). A expressão diferencial foi realizada para as doze bibliotecas através de comparações par-a-par (*pairwise*), entre os diferentes genótipos, tecidos vegetais e tempos de estresse fisiológico (Tabela 1), e os níveis dos genes diferencialmente expressos foram estimados usando como critério o valor de FPKM (*Fragments Per Kilobase Million*). Foram consideradas como diferencialmente expressos os genes com |log do *Fold Change*| ≥ 2 e FDR < 0.01.

3.7. Anotação Funcional

A anotação gênica foi realizada para todos os novos *loci* descobertos durante a montagem do transcritoma de seca, bem como para todos os genes considerados como diferencialmente expressos em ao menos uma das comparações realizadas. Inicialmente, os genes foram comparados com o banco de dados de sequências de

proteínas não redundantes (Nr) através do algoritmo BLASTx v.2.2.30 (75) usando as configurações padrão. Para os novos loci que não foram identificados com BLASTx, a comparação com o Nr foi realizada com o algoritmo PSI-BLAST (76). O conjunto de novos *loci* que não foram identificados nas etapas anteriores, foi submetido à predição de ORFs (*open reading frame*), utilizando o software Transdecoder v.2.0.1. (77). Os bancos de dados "Pfam-A.hmm.gz" (modificado em 10/11/15) e "Uniref90" (modificado em 01/21/16) foram utilizados para identificação das sequências com ORFs.

A ferramenta BLAST2GO toolv3.1 (*Free PRO Trial*) (78) foi utilizada para categorização das sequências pelos termos do *Gene Ontology Consortium* (GO), utilizando a base de dados disponível na versão go_201508-assocdb-data. Para identificar os termos *gene ontology* enriquecidos, foi utilizado o teste exato de Fisher, tendo como critério de corte FDR < 0.05. Durante o teste exato de Fisher, é realizada a busca por termos mais representados em um conjunto de genes (grupo teste) do que em um grupo de referência. Nesse estudo, foram comparados os grupos de GDE em cada comparação (Tabela 1) com o grupo de referência formado por todos os genes descritos para o genoma do feijoeiro (14), somados aos novos *loci* identificados nesse trabalho. Adicionalmente, para sequências com até 2000pb, o banco de dados Rfam (79) (disponível em: http://rfam.xfam.org) foi utilizado na busca de RNAs não codantes.

A anotação dos genes foi realizada também através da busca por homologia com as sequências disponíveis no banco de dados da KEGG (*Kyoto Encyclopedia of Genes and Genomes*)*(80)*, visando identificar genes correspondentes as enzimas envolvidas em rotas metabólicas especificas que pudessem estar relacionadas com a resposta à condição de seca.

3.8. Identificação de variantes estruturais

A análise para descoberta de variantes foi realizada seguindo o fluxo de trabalho denominado "*The GATK best practices for variant calling on RNA-seq*",

proposto para aplicação com o software Genome Analyze Toolkit (GATK) v.3.4-46 (81). Inicialmente, foi realizado o mapeamento dos *reads* que passaram pelo controle de qualidade através do software STAR v.2.4.1 (82). Em seguida, foi realizada a marcação dos *reads* duplicados com o algoritmo "MarkDuplicates", implementado no software Picard-Tools v1.119. A chamada das variantes (*Variant Calling*) foi realizada com o software GATK v.3.4-46 aplicando o algoritmo "HaplotypeCaller". Visando aumentar a qualidade das variantes descobertas, foram promovidas as etapas de realinhamento de indels (RI) e recalibração dos índices de qualidade das bases sequenciadas (*base quality score recalibration* – BQSR).

Para as etapas de recalibração de indels e BQSR, devido à ausência de um conjunto de SNPs conhecidos para feijoeiro, foi implementada uma abordagem descrita pelo "*GATK team*" para espécies que não são consideradas referências no estudo genômico. Inicialmente, a análise de *Variant Calling* foi realizada suprimindo as etapas de RI e BSRQ. O conjunto de variantes obtido nessa etapa foi submetido a um controle de qualidade onde foram aplicados os parâmetros *Fisher Strand* (FS) > 30 e *Quality Depth* (QD) < 2. Igualmente, foram aceitos apenas SNPs identificados com no mínimo 35 pb de distância física entre os 2 SNPs mais próximos, de acordo com as recomendações do GATK (83,84). As variantes que atenderam aos critérios de filtragem foram então utilizadas como grupo de variantes conhecidas nas etapas de BQRS e RI em uma segunda análise de *Variant Calling*.

É recomendado que, ao fim da segunda análise de *Variant Calling* seja implementado novamente o controle de qualidade descrito anteriormente, de forma que o conjunto de variantes obtido seja utilizado como referência em uma nova análise de *Variant Calling*. Esse procedimento deve ser realizado sucessivamente até que seja verificada uma convergência no número de variantes obtido. Após determinada a convergência no número de variantes identificadas, foi realizada a predição dos seus efeitos através do software SnpEff v.4.2 (85).

4. Artigo: A drought transcriptional map reveals new genes for the Mesoamerican common bean germplasm

Wendell Jacinto Pereira[1,2], Arthur Tavares de Oliveira Melo[3], Alexandre Siqueira Guedes Coelho[3], Fabiana Aparecida Rodrigues[4], Sujan Mamidi[5], Sérgio Amorim de Alencar[6], Anna Cristina Lanna[7], Paula Arielle Mendes Ribeiro Valdisser[7], Claudio Brondani[1,3,7], Ivanildo Ramalho do Nascimento-Júnior[8], Tereza Cristina de Oliveira Borba[7], Rosana Pereira Vianello[1,7 *]

[1] Universidade Federal de Goiás, Instituto de Ciências Biológicas, Goiânia, GO, Brazil.

[2] Universidade de Brasília, Departamento de Biologia Celular, Brasília, DF, Brazil.

[3] Universidade Federal de Goiás, Escola de Agronomia, Goiânia, GO, Brazil.

[4] EMBRAPA Soja, Rod. Carlos João Strass, s/n, Londrina, PR, Brazil

[5] Genome Sequencing Center, HudsonAlpha Institute for Biotechnology, Huntsville, AL, United States

[6] Programa de Pós-Graduação em Ciências Genômicas e Biotecnologia, Universidade Católica de Brasília, Brasília, DF, Brazil

[7] EMBRAPA Arroz e Feijão, Rod. GO - 462, Km 12, Santo Antônio de Goiás, GO, Brazil.

[8] Programa de Pós-Graduação em Genética e Melhoramento de Plantas, Universidade Estadual Paulista, Jaboticabal, SP, Brazil.

Wendell Jacinto Pereira – wendelljpereira@gmail.com

Arthur Tavares de Oliveira Melo - arthurmelobio@gmail.com

Alexandre Siqueira Guedes Coelho – alexandre.coelho@me.com

Fabiana Aparecida Rodrigues - fabiana.rodrigues@colaborador.embrapa.br

Sujan Mamidi – sujan.mamidi@gmail.com

Sérgio Amorim de Alencar - sergiodealencar@gmail.com

Anna Cristina Lanna – anna.lanna@embrapa.br

Paula Arielle Mendes Ribeiro Valdisser – paula.valdisser@embrapa.br

Claudio Brondani – claudio.brondani@embrapa.br

Ivanildo Ramalho do Nascimento-Júnior – ivanildoramalho@gmail.com

Tereza Cristina de Oliveira Borba - tereza.borba@embrapa.br

* Rosana Pereira Vianello (Corresponding author) – rosana.vianello@embrapa.br

ABSTRACT

Background: The common bean (*Phaseolus vulgaris* L.) is an important foodstuff, and its cultivation is strongly associated with global food security. Drought episodes are listed as high risk and high impact, affecting the agricultural production of the common bean. In this study, a broad characterization of genes related to the response to drought stress in leaf and root tissue was evaluated by RNA sequencing (RNA-Seq).

Results: In total, 54,750 transcripts were identified, representative of 28,590 genes, of which 1,618 were first described for the common bean and 1,242 were differentially expressed genes (DEGs) in the leaves and roots of genotypes contrasting to drought. Of these, 458 genes showed different expression in the tolerant genotype (191 up- and 267 down-regulated) with a predominance of genes in categories of oxidative stress, responses to stimulus, and kinase activity, among others. Through quantitative real-time PCR (qPCR), 88.64% of the evaluated DEGs were validated for differential expression. A total of 151,283 variants with high quality was identified among the expressed genes, of which 135,167 were SNPs and 16,116 were Indels. From 183,033 SNP variant effects identified, 49.67% were annotated as the modifier type, while the remaining 50.33% were classified as low (29.21%), high (1.91%) and

moderate impact (19.21%) types. Functional effects were estimated for 85,780 variants, most as silent (59.49%). The raw data files were submitted to the NCBI Sequence Read Archive (SRA) database.

Conclusions: In this study, new genes were revealed for drought tolerance, and a *P. vulgaris* transcriptome map of Mesoamerican origin was built featuring 1,618 new loci and 23,112 new isoforms. The increasing availability of genetic information supports a better understanding of the drought tolerance mechanisms in beans, contributing to the research progress.

Keywords: *Phaseolus vulgaris*, water deficit, transcriptome, functional genomics, RNA-Seq, SNPs.

Background

The climatic phenomenon of drought has become more intense and worrisome because the world demand for agricultural products is expected to increase over the coming decade [1]. During the evolutionary process, climate changes have imposed a selective pressure under biological populations, and genetic diversity, considered as the source of allelic variation, plays an important role in the adaptive response to environmental changes [2]. This directional selection pressure would provide an increased allelic frequency that is most advantageous for the adaptive response [3,4]. However, most cultivated species strongly need to rely on a combination of approaches to overcome readily the impact of global warming from ongoing climate changes. This impact under the farming systems has accelerated the search for cultivated varieties better adapted to actual climate changes to reduce vulnerabilities in cropping systems [5,6]. According to Food and Agriculture Organization of the United Nations (FAO), the term “drought” in agriculture means an insufficient amount of water in the soil to meet plant demands at a particular time. Consequently, plants submitted to drought stress activate different mechanisms of resistance, maintaining a balance to optimize the plant yield under drought conditions [7] and seeking to survive.

The common bean (*Phaseolus vulgaris* L.) is a new world crop [8], grown extensively in all continents as a monoculture or in intercropping systems, from lowland tropics to semiarid regions in temperate

environments, with or without irrigation [9]. Among the legumes, the common bean is quoted as the major source of nutrients for millions of people in developing countries, representing an important origin of total protein, energy and micronutrients consumed around the world [10]. Drought has a great impact on the global production of common beans, resulting in yield losses of over 60% under terminal or intermittent drought stress, with more critical effects when the soil water depletion level reaches 60-70% during the grain-filling period [11,12]. In Latin America, the areas of bean growth most affected by drought are northeastern Brazil and the central and northern highlands of Mexico [13]. Lines with higher yield in drought environments have resulted from a combination of races from the Mesoamerican genepool, Durango and Mesoamerica [14], followed by excellent adaptation of the small-seeded beans in Central America and Brazil [11]. Despite the global efforts, due to the highly complex plant response to drought episodes, the precise correlation between genes and drought tolerance remains to be demonstrated.

Efforts have been made to identify the broad set of mechanisms associated with enhanced drought tolerance under water restriction in Mesoamerican dry bean genotypes, such as physiological, phenological and molecular traits [11]. Regarding physiological indicators of tolerance to water deficit, the robustness of the root system, osmotic adjustment, relative water content and stomatal conductance are noted as important mechanisms related to drought tolerance [15,16]. Concerning phenological traits, shortening of the number of days to maturity was observed under drought stress in resistant common bean cultivars [17]. The phenotypic traits and molecular data association resulted in the identification of quantitative trait loci (QTL) related to drought stress in the common bean [18,19]. The genetic responses to drought stress are under the effect of numerous genes whose expression undergoes a strong environmental influence. Progress from QTLs towards the identification of candidate genes has been tried using the linkage disequilibrium (LD), through a methodology that combines comparative *in silico* analysis based on the large volume of genomic information currently available [20]. However,

candidate genes have not delivered as much as anticipated for crop breeding, especially for drought tolerance [21].

An increasing set of common bean genomic data is becoming available for the scientific community. In 2014, the Andean common bean genome was made available, estimated at 587 mega base pairs (Mbp) in size, with ~27 thousand genes, of which 91% were retained in synteny blocks with *Glycine max* [22]. More recently, the Mesoamerican genome was estimated at 549.6 Mbp, with ~30 thousand genes, of which 94% were functionally annotated [23]. These reference genomes represent a valuable and useful resource in plant breeding, allowing more accurate and predictive research based on genomic information. The identification of drought-responsive genes through expression profiling studies is favoured as more genome information is available. Blair et al. [24] realized that, under drought abiotic stresses, low soil phosphorus and high soil aluminium toxicity, 4,219 unigenes derived from cDNA libraries prepared using contrasting genotypes (BAT 477 as drought tolerant (DT) were identified and characterized, contributing to thousands of newly described sequences for common beans. Subsequently, Recchia et al. [25], using a suppression subtractive hybridization (SSH) library, identified 1,120 DEGs in the roots of a DT genotype (BAT 477) when submitted to an interruption of irrigation during its development. Using the same tolerant genotype (BAT 477), Müller et al. [26] identified 802 differentially ESTs (expressed sequence tags) in the leaf in the flowering and grain-filling developmental stages. Recently, cDNA libraries contrasting the tolerant (Long 22-0579) and sensitive (Naihua) genotypes from different gene pools were constructed to identify drought-responsive genes in the leaf [27] and generated 49 Mbp of unique transcriptome sequences, of which 42% were annotated, 18% were assigned into Gene Ontology (GO) terms and only 8,932 were DEGs. Using qPCR, several differential expression profiles of identified drought tolerant genes were investigated for the stress-related response. Many of the identified and validated genes encode, among others, transcription factors, oxidative stress-responsive proteins and genes related to the components of photosystem II [25,27,28].

Here, we report the expression profiling of the two drought-responsive contrasting genotypes, BAT 477 (DT) and Pérola (drought sensitive - DS), using RNA-Seq and multiple bioinformatics resources to explore the potential candidate drought-responsive genes in leaves and roots tissues. New genes and isoforms were identified through this study, providing a drought transcriptome map for common beans, and more than 150,000 variants (SNPs and Indels) were characterized according to the SNP effect. In addition, non-coding RNAs and Long non-coding RNAs (lncRNAs) were identified as differentially expressed in response to drought conditions. From Kyoto Encyclopedia of Genes and Genomes (KEGG), 70 metabolic pathways were identified for new genes, and 97 metabolic pathways for DEGs. The results provided a broad overview under the transcriptional regulation in response to drought, enriching the public database of genes potentially involved in abiotic stress.

METHODS

Plant material

Two well-known genotypes of the common bean, BAT 477 and Pérola, characterized according to their contrasting performance under drought stress, were used in the present study [15]. The lineage BAT 477, characterized as DT, was developed at the International Center for Tropical Agriculture (CIAT, Cali, Colombia), while the Pérola cultivar, characterized as DS, was developed at the EMBRAPA (Brazilian Company for Agricultural Research) breeding programme in 1994 [29]. Both lineages have Mesoamerican origin and present growth habit type III [30,31].

Experimental conditions

Both varieties were maintained in a controlled environment (hydroponics system) with a specific nutrient solution, according to the recommendations of Martins et al. [32]. Briefly, the plants were grown in plastic containers with an aerated pH 6.6-balanced nutrient solution [33]. The seeds were pre-germinated on moist filter paper in the dark room with controlled temperature and relative humidity. The seedlings were placed in tray supports, keeping their roots completely immersed in nutritive solution,

and the hydroponic system was maintained in a greenhouse at 25°C and 60% relative humidity, under natural day light. After 15 days, the seedlings were submitted to the treatments in which they were completely removed from the hydroponic solution for 0 min (T_0; control), 75 min (T_{75}), and 150 min (T_{150}). As previously established, these times were based on the photosynthesis rates for which the minimum and maximum values were obtained at time zero and 150 min after being subjected to water stress, respectively [32]. The experiment was conducted in the form of random complete blocks with five replications, in a double factorial arrangement, and two sampled genotypes and water regime as factors.

Physiological evaluation and statistical analysis

The physiological evaluations consisted of seven treatments: 0 min (T_0; control), 25 min (T_{25}), 50 min (T_{50}), 75 min (T_{75}), 100 min (T_{100}), 125 min (T_{125}) and 150 min (T_{150}). In all treatments, the photosynthetic rate (µmol CO_2 m^{-2} s^{-1}), stomatal conductance (mmol H_2O m^{-2} s^{-1}), leaf temperature (°C) and leaf transpiration rate (mmol H_2O m^{-2} s^{-1}) were evaluated using a LI-6400 Portable Photosynthesis System model (LI-COR Inc., Lincoln, Nebraska, USA). Measurements were taken on the expanded leaf subjected to progressive water deficit treatments, for the periods of 0 min, 25 min, 50 min, 75 min, 100 min, 125 min, and 150 min of dehydration. The osmotic (Ψ_s) and hydric potential (Ψ_w) in Mega Pascal (Mpa), were taken according to the methodology described by Bajji et al. [34]. The data were submitted to ANOVA analysis using the SAS (Statistical Analysis System v.7.0, SAS Institute Inc., Cary, North Carolina, USA) program, and the treatment means were compared using Tukey's test at 95% of statistical significance.

Library preparation and RNA-Seq

Fresh leaves and roots from three treatments (T_0, T_{75} and T_{150}) for both lineages (BAT 477 and Pérola) were sampled from 15-day seedlings and were stored at -80 °C, totalling twelve different RNA-Seq

libraries (Figure 1). Total RNA was individually extracted from those libraries using a commercial PureLink® RNA Mini Kit (Thermo Fisher Scientific Inc., Waltham, Massachusetts, USA), following the manufacturer's instructions. The RNA quantity and purity were estimated through spectrophotometry using a NanoVue™ Plus Spectrophotometer (General Electric Co. , Boston, Massachusetts, USA), and the integrity was verified using microfluidics technology and the Agilent 2100 Bioanalyzer instrument (General Electric Co., Boston, Massachusetts, USA)). The RNA-Seq libraries were prepared using the Illumina's TruSeq RNA kit (Illumina, Inc., San Diego, California, USA) following the manufacturer's instructions and were paired-end sequenced in a technical replicate system using two Illumina platforms (GAII and HiSeq 2000).

Pre-processing data

The quality of sequenced paired-end reads was first visualized using the FastQC v.0.11.3 software [35]. The sequences were trimmed to eliminate adapters and low quality bases, with the cutting of the first 15 bases (GAII) and first 17 bases (HiSeq 2000) at the 5' portion, allowing a maximum of two mismatches between the adapter sequences and reads. In addition, sequences of four contiguous nucleotides with an average score $Q \leq 30$ and reads shorter than 50 base pair (bp) were trimmed. These procedures were performed using Trimmomatic software v.0.35 [36]. Finally, the high-quality paired-end and single-end (orphan) reads from each RNA-Seq sequenced libraries were retained for further analysis.

***P. vulgaris* drought-specific transcriptome assembly**

The high-quality reads were mapped to the reference genome using TopHat v2.1.0. [37] and duplicated reads were marked and removed using Picard tools [38]. The transcriptome was assembled using reads that mapped in each of those twelve RNA-Seq libraries using the Cufflinks v2.2.1 [39,40] software in the RABT mode [41]. An alternative isoform was set when the abundance was, at least, 20% compared with the principal isoform (-F = 0.2). The reference genome and gene annotation (GFF3 file), available at Phytozome v.11.0 for *P. vulgaris* [22], were used to guide the assembly. In addition, reads that represent

repetitive DNA sequence regions were masked, and, to obtain a single representative transcriptome for all libraries, the individual assembled transcripts were merged using the Cuffmerge algorithm following the default parameters. A comparative analysis was also performed using the Cuffcompare algorithm to evaluate the transcripts overlapped between the transcriptomes available in the Phytozome v.11.0 database, and the drought-specific transcriptome was assembled. Finally, a set of particular reads from the drought-specific transcriptome was appropriately sampled.

Differential Gene Expression (DGE) analysis

The DEGs were identified using the edgeR [42] Bioconductor package. The abundance of reads was quantified using HTSeq [43], and only genes with counts-per-million (CPM) > 1 in at least two samples were kept in the analysis. The estimation of common dispersion was conducted by the function "estimateCommonDisp" from a set of all genes that did not respond to the implemented treatment (~3900 genes). This strategy was applied to improve the variation estimate without biological replicate samples, according to the edgeR manual. Levels of DEGs were estimated based on their relative abundance against the reference sequence, using the FPKM (Fragments Per Kilobase Million) parameter. We defined an FDR (False discovery rate) < 0.01 and the |log2 ratio| > 2 as a threshold to judge the significance of differences in gene expression. The DGE analyses were performed across all RNA-Seq libraries, in a pairwise comparison among different genotypes, plant tissue and physiological stress time (Table 1).

Table 1. Description of the comparisons performed between different samples of transcripts derived from the 12 sequenced libraries, including the number of up- and down-regulated genes in each comparison.

Library comparison ID	Genotype(s)	Tissue	Treatment	Number of DEGs	Up-regulated DEGs	Down-regulated DEGs
1	Pérola x BAT477	Root	All	21	1	20
2	Pérola x BAT477	Leaf	All	139	65	74
3.1	BAT447	Root	T_0 x T_{75}	192	177	15

3.2	BAT447	Root	T_{75} x T_{150}	4	0	4
3.3	BAT447	Root	T_0 x T_{150}	687	569	118
4.1	Pérola	Root	T_0 x T_{75}	123	110	13
4.2	Pérola	Root	T_{75} x T_{150}	31	13	18
4.3	Pérola	Root	T_0 x T_{150}	405	338	67
5.1	BAT447	Leaf	T_0 x T_{75}	2	2	0
5.2	BAT447	Leaf	T_{75} x T_{150}	14	13	1
5.3	BAT447	Leaf	T_0 x T_{150}	251	206	45
6.1	Pérola	Leaf	T_0 x T_{75}	0	0	0
6.2	Pérola	Leaf	T_{75} x T_{150}	25	25	0
6.3	Pérola	Leaf	T_0 x T_{150}	208	149	59
7	Pérola x BAT477	Root	T_0	46	20	26
8	Pérola x BAT477	Root	T_{75}	17	7	10
9	Pérola x BAT477	Root	T_{150}	5	1	4
10	Pérola x BAT477	Leaf	T_0	76	41	35
11	Pérola x BAT477	Leaf	T_{75}	97	31	66
12	Pérola x BAT477	Leaf	T_{150}	57	25	32

Gene Annotation

The gene annotation was carried out for the drought-specific transcriptome assembly and DGEs in pairwise comparisons across the RNA-Seq libraries. Initially, the major transcripts for each gene were compared with the NCBI's non-redundant protein database (NR) through the BLASTx v.2.2.30+ algorithm [44] using the standard configurations. For new loci without hits in BLASTx, the comparison was made through PSI-BLAST [45]. The prediction of open reading frames (ORFs) on raw reads was conducted with Transdecoder v2.0.1 [46]. In addition, the new loci with no hit were annotated against the Mesoamerican transcriptome and lncRNAs [23] available in [http://denovo.cnag.cat/genomes/bean/] using BLASTn v.2.2.30+. The BLAST2GO tool v4.1 (BioBam®, Valencia, Spain; Free PRO Trial, [47])

was used for GO categorization (E-value < 10^{-6}, annotation cutoff = 55). The GO terms with significant changes in abundance were identified using Fisher's exact test, with FDR correction ($p \leq 0.05$). The identification of genes involved in specific metabolic pathways was performed using the KEGG database [48]. Additionally, for sequences up to 2,000 bp, the Rfam database [49], was used to search non-coding RNA.

Variant call and characterization procedure

The Genome Analyze Toolkit (GATK) v.3.4-46 [50] performed the calls of SNPs and Indels (using "HaplotypeCaller" algorithm), as well as analysis of base quality score recalibration and Indel realignment. The mapping step was performed using STAR v.2.4.1 [51], and duplicate reads were identified and removed using Picard tools v1.119 [38]. Quality control filters were used such as a Fisher Strand (FS) > 30, a Quality Depth (QD) < 2, and a minimum of 35 bp of physical distance between two consecutive SNPs according to GATK recommendations [52,53]. The annotation and variant prediction of those high-quality SNPs was performed using the SnpEff v.4.2 software [54] and an "in-home" genome database built from the genome Phaseolus_vulgaris_218 v.1 [22] combined with Pvulgaris_218_v1.0.gene_exons.gff3 (Phytozome v.11.0). Putative SNPs effects categorized as high were annotated based on GO terms and were integrated in the KEGG pathway through BLAST2GO tool v4.1.

Differentially expressed gene validation by qPCR

To perform validation using qPCR, 15 target differentially expressed genes were evaluated using the 12 cDNA samples taken to develop the libraries. The amplifications based on the TaqMan® Gene Expression assay (Thermo Fisher Scientific Inc., Waltham, Massachusetts, USA) were performed using the Applied Biosystems® 7500 Real-Time PCR System device (Thermo Fisher Scientific Inc., Waltham, Massachusetts, USA), followed by the determination of the C_q (quantification cycle) values using 7500 Software v.2.3 (Thermo Fisher Scientific Inc., Waltham, Massachusetts, USA). The C_q values were

subjected to the $2^{-\Delta\Delta C}{}_T$ method [55], and the significance was determined by ANOVA and Tukey's test ($p < 0.05$) with R [56] using the packages agricolae v.1.2-4 and multcomp v.1.4-6. A set of three reference genes was used (Elongation factor, assay ID AI20SMX; 18S ribosomal RNA, assay ID AI39QS5 and PvT197, assay ID AIRR961), for which the stability was assessed with the geNorm Plus application (M values) integrated into the qbase^{PLUS}/qPCR data analysis software package (Biogazelle, Zulte, Belgium; [57]).

RESULTS

Evaluation of physiological conditions

The physiological parameters measured on plants grown hydroponically exposed to a short-term water deficit are presented in Additional file 1. In the hydroponic experiments, the photosynthetic rate (mol CO_2 $m^{-2}s^{-1}$), stomatal conductance (µmol H_2O $m^{-2}s^{-1}$) and leaf transpiration rate (mmol H_2O $m^{-2}s^{-1}$) showed decreasing values, while the leaf temperature (°C) was increased, and the internal CO_2 concentration (mol CO_2 mol^{-1}) varied across the dehydration process. These data indicated that the dehydration treatments were effective because it triggered changes in the plant physiological behaviour at both genotypes. In general, the tolerant genotype (BAT 477) presented lower values of the photosynthetic rate (p-value < 0.05 at T_0 and T_{50}) and stomatal conductance (p-value < 0.05 at T_{50}) than the sensitive Pérola genotype. For leaf transpiration, BAT 477 presented a reduced value at T_0 (p-value < 0.05). Similar values of the leaf temperature and CO_2 concentration were obtained for both genotypes.

Providing a drought-specific transcriptome for *P.* vulgaris

All raw sequence data generated a total of ~ 580,000,000 paired-end 100-bp reads, totalling 5.8 Giga base pairs (Gbp) of RNA-Seq data, of which ~329 and ~250 millions of reads were obtained, and 82.76% and 96.13% were retained (Q30) for further assembly in HiSeq and GAII, respectively (Additional file 2). The raw data were summarized in Additional file 2 and deposited in the SRA/NCBI database repository through the accession number SRP077562

(http://www.ncbi.nlm.nih.gov/sra/SRP077562). The coverages of the bean genome per library were ~ 4x, while that of the total RNA-Seq was 94.96x and that of the coding DNA sequence (CDS) was 476.39x. The total mapped reads per library was 87.52% (SD ± 3.72%). Here, we propose a new *P. vulgaris* drought-specific transcriptome that was assembled into 28,590 loci with 54,750 transcripts, of which 1,618 new *loci* (5.66%) and 23,112 new isoforms were described.

Newly discovered genes and functional annotation

From the newly identified 1,668 genes, 839 showed hits with previous sequences evaluated using both BLASTx and PSI-BLAST, while the remaining 829 (49.70%) showed no homology with the NR. The greatest sequence similarities were identified with *P. vulgaris* (42%), *Vigna angularis* (adzuki bean) (21.06%) and *G. max* (10.35%). Fifth loci corresponded to repetitive elements (48 loci as transposons, retrotransposons and rRNAs), and bacterial rRNAs (two) were removed from the DGE analysis. A set of 58 *loci* categorized as non-coding RNA was classified into seven biotypes: a) Gene, snRNA, snoRNA, CD-box (56.8%); b) Gene, miRNA (25.3%); c) Gene, snRNA, snoRNA, HACA-box (11.6%); d) Intron (3.2%); e) Gene (1.1%); f) Gene, antisense (1.1%); g) Gene, tRNA (1.1%).

From the new loci, 1,886 GO terms (GO IDs) were identified, and 466 GO terms were directly annotated (Additional file 3 and 4), of which 1,301 revealed a significantly relative abundance compared with the previously annotated *P. vulgaris* genes and 1,300 terms were over-represented (Additional file 5) in biological process (69.02%), molecular function (18.52%) and cellular component (12.45%). Through KEGG analysis 70 metabolic pathways were identified containing 77 associated enzymes (Additional file 6). For the remaining 769 loci, 261 presented high similarity with transcripts encoding proteins, and 136 had lncRNAs according to the Mesoamerican genome [23] homology (Additional file 7). Altogether, 372 new loci had no hit with previous sequences. From these, 23 potential protein-encoding loci were identified through ORF prediction (12 complete and 11 partial ORFs). A group of 349 loci remained without information.

Differentially expressed genes (DEGs)

For the whole set of genes (28,590), the comparative transcriptome analysis revealed differences in gene expression patterns between BAT 477 (DT) and Pérola (DS), even under control conditions (Additional file 8). For the leaves and roots, the number of expressed genes (CPM >1) was similar, although the set of genes and tissue distribution varied considerably. In total, 21,696 genes were identified for DT, with 17,425 (80.31%) common to both tissues (leaf and root), 2,423 (11.17%) exclusive in the root, and 1,848 (8.52%) exclusive in the leaves. For Pérola, a set of 21,699 was identified under control conditions, of which 17,484 (80.58%) were expressed in both tissues, 2,550 (11.75%) were exclusive in the roots, and 1,665 (7.67%) were exclusive in the leaves. Between the contrasting genotypes, considering the leaf tissues, 816 genes were expressed only in BAT 477 and 692 in Pérola (Additional file 8); for the root, 489 were expressed only in BAT 477 and 675 in Pérola.

Compared with the control condition (T_0), the T_{75} condition showed an increased in the overall number of expressed genes in DS (326) and DT (42). Interestingly, within genotypes, while DS was increased in both tissues (189), DT was increased in the root (332) with reduction in the leaves. Between the genotypes, an increase in the total number of genes expressed specifically in DS (an additional of 232 for leaves and 198 for roots) was observed. By contrast, at T_{150}, there was an increase in the number of expressed genes in DT (301), mainly in the root (63 genes), compared with the susceptible genotype (30 genes), of which the genes were preferentially expressed in the leaves. Overall, between the genotypes, a higher number of genes was expressed in the leaves for DS (900) than for the tolerant genotype (653); however, in the root, an opposite trend was noted: more genes were observed for DT (615) than for DS (536) (Additional file 8). In general, a greater number of KEGG pathways linked to DEGs was observed in the comparison of T_0 *vs* T_{150} within genotypes for the up- and down-regulated genes, in both tissues and genotypes (Additional file 6).

For DGE analysis, in each pairwise comparison (Figure 2 and Additional file 9), the value of the dispersion utilized was 0.191, according to that estimated in edgeR. In the root tissue, at T_{75}, 192 DEGs for DT and 123 for DS were observed, of which 177 (91.71%) and 110 (89.43%) were up-regulated,

respectively (Figure 2). At the T_{150}, in the root tissue, 681 DEGs were reported in DT (569 up- and 112 down-regulated), of which 184 (27.02%) were common to the T_{75} treatment. However, in DS, 403 DEGs were identified (338 up and 65 down-regulated), of which 103 (25.56%) were common to both treatments (Figure 2). For the leaf tissue, at T_{150}, a set of 251 DEGs was identified in DT (82.07% up-regulated) and 208 in DS (71.64% up-regulated). From the DEGs, only one lncRNA was verified in the leaf tissue, from the DS genotype; considering the root tissue, six lncRNAs were identified that were responsive to drought stress with high reliability (fold change ≥ 4 and FDR <0.01), four in DT, of which three were up-regulated, and two in DS (Additional file 10). For the leaf tissue, only a small portion of DEGs identified in T_{150} was differentially expressed in relation to T_{75}, being 5.58% for DT and 11.53% for DS (Figure 2).

Among the DEGs in the roots, 73 in T_{75} were common to both genotypes (69 up-regulated and four down-regulated); however, in T_{150}, this total was 245 (227 up- and 18 down-regulated). For the leaves, from the 353 DGEs in T_{150}, 106 were the same in both genotypes (92 up- and 14 down-regulated) (Additional file 11). Between the genotypes, 17 DEGs were identified in the roots at T_{75}, and only five were identified in T_{150}; however, in the leaves, at T_{75} and T_{150}, 97 and 57 DEGs were identified, respectively (Table 1).

Functional annotation of differentially expressed genes (DEGs)

The GO categorization provided by BLAST2GO for all 1,242 DEGs is described in Additional file 3. The most representative distributions filtered by node score inside biological process were oxidation-reduction process, proteolysis, response to light stimulus and response to oxidative stress; for molecular function, the most representative distributions were protein serine/threonine kinase activity, iron ion binding, heme binding, transcription factor activity – sequence-specific DNA binding and signal transducer activity (Additional file 12). Through KEGG pathway analysis, 97 metabolic processes were identified containing 145 associated enzymes (Additional file 6).

Regarding the functional annotation, within the genotypes and root tissue at T_{75}, 27 enriched terms were associated with DT up-regulated genes, some of them related to the oxidoreductase activity (GO:0016491, GO:0016705 and GO:0016706), oxidation-reduction process (GO:0055114), among others. For the DS, only one term related to the oxidoreductase activity was reported (Additional file 5). No enriched terms were identified for down-regulated DEGs in the roots at $T_{75,}$ as well as in the leaves at T_{75} and T_{150}. By contrast, in the leaf tissue at T_{75}, only two DEGs were reported for DT, a translocator protein homologue (Phvul.001G205900) and low-temperature-induced genes (Phvul.007G069800); none was reported for Pérola.

For DT, over the long dehydration treatment (T_{150}) 51 over-represented GO terms for up-regulated genes were identified in the roots. The DEGs were involved in the same terms described for T_{75}, in addition to other terms such as antiporter activity (GO:0015297), growth factor activity (GO:0008083), response to biotic stimulus (GO:0009607), antiporter activity (GO:0015297), nucleic acid-templated transcription (GO:0097659), receptor binding (GO:0005102) and 41 other terms related to regulatory processes (Additional file 5). Down-regulated genes were also categorized in GO terms, such as oxidoreductase activity (GO:0016705 and GO:0016706). For DS, enriched GO terms similar to those previously related to the T_{75} and T_{150} treatments in the DT genotype were reported, such as oxidoreductase activity (GO:0016705, GO:0016706 and GO:0016491), oxidation-reduction process (GO:0055114), dioxygenase activity (GO:0051213), heme binding (GO:0020037), tetrapyrrole binding (GO:0046906), metabolic process (GO:0008152) and iron ion binding (GO:0005506)(Additional file 5).

Regarding KEGG analysis, within the genotypes, in the DT root tissues, 75 pathways for the up- and 47 for the down-regulated genes were identified between the treatments; however, for DS, it was 49 and 24, respectively. In the leaf tissue, 51 and 18 pathways for the up- and down- regulated genes among treatments were reported for DT; however, for DS, it was 55 and 14, respectively. In both comparisons for the root tissue, pathways exclusively represented in up-regulated DEGs were verified from DT, 17 for T_0 *vs* T_{75} and 28 for T_0 *vs* T_{150}; however, for DS, one and two KEGG pathways were identified for the

same comparisons. Regarding DEGs identified in leaf tissues, a total of four and eight KEGG pathways were exclusively identified for T_0 vs T_{150} for DT and DS, respectively, and another 12 were verified in DEGs from the T_{75} *vs* T_{150} comparison from DS (Additional File 6).

For KEGG analysis based on DEGs between genotypes, in leaf tissues, nine metabolic pathways were identified in DT. Among these, four were exclusive in DT (Tropane, piperidine and pyridine alkaloid biosynthesis; Lysine degradation; Cysteine and methionine metabolism; Sulphur metabolism), being two common to all three treatments for the DT (Cysteine and methionine metabolism and Sulphur metabolism). For DS, 15 pathways were identified, five common to all three treatments and 10 exclusively in DS. For roots, 10 pathways were identified in DS and none in DT (Additional file 6).

Variant call and characterization

Initially, the variant calling was performed suppressing the base quality score recalibration (BSRQ) and Indel realignment (IR) analysis, and the variants that passed on imposed quality filters (FS > 30.0, QD < 2.0, -window 35 and -cluster 3) were used as references for the BSRQ and IR in the second step of variant calling. This second group of variants identified served as a standard in a third variant calling step. The results evidenced a convergence of the numbers of SNPs and Indels called, indicating that the increase in the number of steps of BSRQ and IR would not lead to quality improvement of the variants identified (Table 2).

Table 2. Number of variants identified after each Base Quality Score Recalibration (BQSR) step.

Variants \ Haplotype caller step		**Without BSRQ**	**First BSRQ**	**Second BSRQ**
BAT 477	**SNPs**	119,371	119,573	119,572
	Insertions	7,141	6,596	6,591
	Deletions	7,560	7,260	7,261
Pérola	**SNPs**	96,061	96,198	96,192
	Insertions	6,180	5,643	5,638
	Deletions	6,697	6,411	6,412
All	**SNPs**	134,870	135,175	135,167
	Insertions	8,194	7,656	7,652
	Deletions	8,749	8,462	8,464

	Total	151,813	151,293	151,283

A total of 151,283 variants was identified, 135,167 SNPs and 16,116 Indels, of which 120,726 SNPs and 13,631 were placed inside of one gene. For DT, 119,572 SNPs and 13,852 Indels were identified; however, for DS, 96,192 SNPs and 12,050 Indels were identified (Table 2). The variants were widely distributed in the *P. vulgaris* genome, with one variant for each 3,413 bases (Figure 3). Furthermore, variants affecting 71.80% of the 27,197 loci described for the reference genome of common bean were found [22]. In total, 183,033 putative effects were identified classified as modifier (49.68%), low (29.21%), high (1.91%) and moderate impact (19.21%) types. Classification by the functional class was reported for 85,780 variants and revealed 59.49% as silent, 40.04% as missense and 0.47% as nonsense. The largest proportion of the predicted effects occurred in exons (45.76%), and a significant number of effects (20.15%) was also observed in the 5' and 3' UTR regions, suggesting possible changes in the regulatory regions (Table 3). The Ts\Tv ratio was 1.34.

Table 3. Number of effects by type according to SnpEff classification.

Type of effect	Count	Percent (%)
3_prime_UTR_variant	26,765	14.48%
5_prime_UTR_premature_start_codon_gain_variant	1074	0.58%
5_prime_UTR_variant	9,041	4.89%
disruptive_inframe_deletion	317	0.17%
disruptive_inframe_insertion	223	0.12%
frameshift_variant	2,831	1.53%
inframe_deletion	166	0.09%
inframe_insertion	203	0.11%
initiator_condon_variant	8	0.01%
intergenic_region	17,465	9.46%
intron_variant	38,910	21.04%
missense_variant	34,257	18.53%
non_conding_transcript_variant	10	0.01%
splice_acceptor_variant	10	0.01%
splice_donor_variant	83	0.05%

splice_region_variant	1,815	0.98%
start_lost	66	0.04%
stop_gained	432	0.23%
stop_lost	83	0.05%
stop_retained_variant	82	0.04%
synonymous_variant	50,951	27.56%

Validation of the DEG by qPCR

Validation through qPCR of a set of 15 DEGs constituted a biological replicated experiment. The gene expression profile obtained by qPCR was 88.64%, concordant with the results obtained in RNA-Seq analysis for the DEGs in implemented comparisons (Figure 4; Additional file 14). From 15 DEGs identified by RNA-Seq, most of them (60%; POX, LEA, OPDA, CWI, MYB, GST, HSTF, FHOS2C) correlated well with the qPCR results, presenting high correspondence (100%) and a strong coefficient of *correlation* (≥ 60%), which were important technical validation parameters of the RNA-Seq experiments (Additional file 15). Among the genes with high correspondence between RNA-Seq and qPCR methodologies, those that were shown to be up-regulated for DT with more hydric stress were glutathione S-transferase (GST; Phvul.008G113700) and peroxidase (POX; Phvul.009G140700). GST, a well-characterized detoxification enzyme family member involved in stress tolerance, presented increased expression in the leaf tissue of DT at T_0, T_{75} and T_{150}. The POX enzyme presented an up-regulation in T_0 and T_{75} in DT, followed by weak up-regulation in subsequent T_{150} treatments in the leaves, while DS showed the opposite pattern of expression—increased up-regulation at T_{150}. In the roots, the expression level of peroxidase was gradually increased over the treatments, with significant levels at T_{150} of the DT.

DISCUSSION

For plants to grow and utilize their nutrients efficiently, water must be adequately supplied. Episodes of drought with a temporary decrease in water availability lead to stress and induce changes in several morphological, physiological, biochemical and molecular processes in plants [58]. Hydroponics is a

useful system to study plant responses to abiotic stresses, including drought. However, in the soil condition, during drought stress, a slower water deprivation process occurs, allowing the plant to adapt to the water deficit; in the hydroponic systems, water deprivation occurs abruptly by removing the plant from the nutrient solution [59,60]. As an advantage, the hydroponic systems overcome the effects of several abiotic stresses other than water stress, such as the problems of heterogeneity and drainage [61], allowing the maintenance of constant conditions such as temperature, relative humidity, and the light regimen [32]. A more uniform water stress can facilitate the discovery of more related genes and, consequently, the recognition of mechanisms more directly involved in the response to the environmental changes. Although it may be found that hydroponic and soil-grown plants induce different responses, studies have shown that the responses should be the same but expressed differently in time [61,62]. In a comparative overview of the datasets generate from hydroponically *versus* soil-grown plants similar activation of metabolic processes was demonstrated in response to stress, triggering similar expression of many genes [62]. In addition, the expression pattern of soybean homologues to *Arabidopsis* drought genes in roots under the two water deficit systems revealed that the genes were differentially induced, and there is an urgent need to access both systems [60]. Consequently, the use of hydroponics system is a valid option to understand the mechanisms of the plant response to water stress [62]. Additionally, experiments with root tissue samples are favoured because this organ can be easily accessed and collected for analysis under different stress conditions [63].

As previously reported by Lanna et al. [15], in soil, osmotic stress was shown to be an important component of drought tolerance for the BAT 477 and Pérola genotypes. In the present study, hydroponic stress failed to predict significant differences in the physiological parameter of the osmotic potential, among treatments and genotypes, probably due to the short periods of dehydration condition, which is critical to the plant´s solute concentration and/or amount of water control in the metabolic fluids. Otherwise, it was shown that the process of dehydration reduced the values of gas exchange, such as the photosynthetic rate, transpiration rate, stomatal conductance at values close to zero, which may be

a consequence of the decline in photosynthesis due to darkness and dehydration condition. Chloroplasts are sensitive to a broad set of stressful environments such as drought and light intensity, among others, reducing the photosynthetic rate by stress-induced stomatal or nonstomatal limitations, as reviewed by Ashraf and Harris [64].

Functional annotation of newly discovered transcripts

Approximately 3.87% of the transcripts represented newly discovered genes (5.66% of all genes) within the Andean and Mesoamerican gene pools [22,23]. However, all the new genes described are involved directly or indirectly in the mechanisms of plants affected by water stress; the most interesting was that several reported enriched GO terms were related to abiotic stimulus and the defence response. These are especially important because there is intense interest in exploring the potential of candidate target genes that confers tolerance to drought for the implementation of breeding strategies. Among the discovered genes, a considerable number (34 of 466 new genes with GO terms) was annotated to protein kinase (PK) GO terms (Additional file 4). These enzymes play a fundamental role in various regulatory mechanisms of the cell through signal transduction pathways during biotic and abiotic stress [65]. In addition, different types of PK not previously described for common beans were identified, including various receptor-like kinase genes (20 new for beans) that could be strongly associated with the mechanisms of the response of external stimuli such as water deficit, activating signalling networks to adapt the plant to the changing environment, as described previously [66].

In the present study, three novel genes involved in response to abscisic acid (ABA) (XLOC_002778, XLOC_009245, XLOC_017074) and one in ABA biosynthesis (XLOC_011282) were reported. ABA is a well-characterized phytohormone that triggers several pathways for plant stress responses, mainly by the activation of a transcriptional regulatory network involved in drought [67]. Transcription factors (TF) were also newly reported for common beans, such as MYB (XLOC_009355) and WRKY (XLOC_002066), which have been proposed to regulate several processes, such as hormonal induction

during stress and responses to diverse biotic or abiotic environments [68,69]. Among others, a less-specific TF was also identified, including GTE1 (XLOC_005373, XLOC_011757), which functions as a general transcription factor [70], and ethylene-responsive TF (XLOC_013281), which is differentially regulated by ethylene and various forms of abiotic stress [71]. Genes related to jasmonic (JA), ethylene (ET) and salicylic acid (AS) were also newly described and related to the GO terms of biosynthetic process and mediated signalling pathway (XLOC_013281; XLOC_002778; XLOC_024793; XLOC_026696; XLOC_006232; XLOC_006232; XLOC_028474; XLOC_028574; XLOC_011121), in addition to auxins (XLOC_018037) and cytokinins (XLOC_001761, XLOC_019264). The role of these hormones in abiotic stress tolerance [72–74] and plant defence, as well as plant development and growth through efficient signalling networks and an elaborate crosstalk, has been broadly reviewed [75].

These findings also revealed a new class of chaperones (four new loci) for previous common beans transcriptional studies. Chaperones play an important role in protein homeostasis and, although constitutively expressed, are extremely active in the signalling of stresses such as drought, temperature extremes and salinity [76]. There are several diverse groups of chaperones in the organisms, sharing the property to bind to unstable substrate protein [77,78]. These co-chaperones identified in the present study, known as DnaJ/HSP40 proteins (XLOC_010633, XLOC_028205) and p23 proteins (XLOC_010687), were described as key regulatory factors that modulate the activity of HSP70 and HSP90 chaperones, respectively [79–81]. Finally, several genes related to plant disease resistance in several ways were newly identified for common beans (more than 30 genes), contributing to the enrichment of the biotic stress-related gene database. This would be expected because, during the drought stress, the plant becomes more predisposed to diseases/infections and susceptible to insect and pest infestations [82].

Long non-coding RNAs

LncRNAs have been identified as regulators of many fundamental biological processes such as tissue development or the response to external stimuli [83–88], as well as the response to osmotic, saline and drought stress [84–87]. Efforts have been made and many stress-responsive lncRNA transcripts have been predicted and identified in some species, such as *Medicago truncatula* [86] and *P. vulgaris* [23] and wheat [89]. To date, only a few plant lncRNAs with potential roles in response to drought stress have been characterized, such as *Arabidopsis [90]*, foxtail millet [81] and maize [87]. In this study, lncRNAs were identified as DEGs responsive to drought stress with high reliability (fold change ≥ 4 and FDR <0.01), of which three were specifically detected and up-regulated in the DT root tissues (Additional file 11). The recently demonstrated function of lncRNA in cell differentiation and development [91] has suggested that lncRNA expression could play a critical role in the molecular mechanisms for the ability of root development under drought stress in BAT477 because this line demonstrated a greater ability for this function [92]. Our results suggest that these lncRNAs could be potential targets to explore the functional role of lncRNAs for the bean mechanisms of tolerance. More recently, a genome-scale screening platform based on CRISPR-mediated interference (CRISPRi) has been used for the identification of lncRNA function in human cells [93], opening a new perspective to functionally test other organisms.

DEGs

At T_0, a specific profile of gene expression by genotype was observed, as well as a considerable variation between the root and leaf tissue modulated by the activation of different genes along the dehydration periods (T_{75} - T_{150}). This differential expression pattern over time demonstrated the effectiveness of the stress-induced protocol implemented. In the root tissue, there was an increased number of DEGs up-regulated in the DT genotype (T_{75} = 108 genes; T_{150} = 342) compared with that in the DS genotype (T_{75} = 41; T_{150} = 111) with a predominance of enriched categories related to osmotic and redox processes, signal transduction mechanisms, transcription factors, as well as the development of cellular metabolism. Regarding the physiological aspects, in dry conditions, the DT genotype

develops the root tissue for access to deeper layers of the soil, as an important adaptive advantage in the mechanism of bean tolerance [15], although this potential is dependent on favourable environmental conditions such as soil type [94]. BAT 477 (DT) stands out for presenting differentiated root growth, suggesting a strong effect of genetic factors controlling these traits. Asfaw & Blair [92] identified 15 regions containing QTLs related to root development under dry conditions, from which the most favourable alleles were derived from the parental BAT477. A detailed annotation of these QTLs regions, integrating the information of the genes identified in this study, together with the high linkage disequilibrium in the bean genome [95], may be used for the selection of the traits related to root development and will be effective in the breeding programme for drought tolerance.

Plants can adapt to environments with large variations in climate conditions to survive and grow, which has driven the evolution of a highly flexible metabolism [96]. This environmental condition of stress leads to the formation of ROS (reactive oxygen species) such as superoxide, hydrogen peroxide (H_2O_2), hydroxyl radical, and singlet oxygen [97], which are regulated by a cellular antioxidant defence system consisting of enzymatic and non-enzymatic antioxidants that protect the cell from oxidative damage. In this study, it was clearly observed in the DT genotype that the gene expression involved in redox and related processes showed an increase in leaf and root from treatment T_0, the time immediately subsequent to the plants' removal from the nutrient solution. Approximately, 15% of enriched terms were directly related to these, such as oxidation-reduction, oxidoreductase activity, electron carrier activity and iron ion binding. These lines of evidence suggest that, in the DT genotype, the oxidative stress may have been the primary factor to trigger the signalling pathway mediating several other responses, such as cell wall modification and pH regulation, activation of transmembrane transporters and growth factors, the production of heat shock proteins, LEA and chaperones, and ethylene synthesis, signals that act on anatomical structures and induce the production of factors involved in programmed cell death. Non-enzymatic antioxidants, such as ascorbate, glutathione, proline and betaine, were also identified in the present study as DEGs (Additional file 4). Although the redox process has been strongly

activated in the roots and leaves, the associated DEGs were completely different. However, in the roots, the main DEGs were peroxidases, lipoxygenases (linoleato 13s-lipoxygenase-3-chloroplatic-like), dioxygenases (Gibberellin and leucoantocianidins), oxidases (1-aminociclopropano-1-carboxilato), *NADP-dependent malic* enzymes and *Cytochrome P450* (CYP) *proteins*; in the leaves, the main DEGs were cationic peroxidase, oxidases (poliphenol oxidase e long chain alcohol oxidase), lipoxygenases (seed linoleato 9s-lipoxygenase), lysine-ketoglutarate reductase saccharopine dehydrogenase, zinc-binding alcohol dehydrogenase family protein and protein hothead-like. These DEGs identified between tissues indicate that the DT triggers distinct mechanisms of drought tolerance, inducing expressive antioxidant responses.

The dehydration stress in DT drove changes in expression of genes related to developmental and cellular growth processes associated with the contexts of cell division, differentiation, structuration. Additionally, many terms related to kinases play a central role in the regulation of cell function [Krebs et al. 1985 - [98] and adaptive responses [99]. The overexpression of CDPK enhances crop stress resistance/tolerance to cold, salt and drought, such as that for rice [100] and pepper [101], and has been involved in specific tissue growth and development [102]. In the root tissue of the DT still highlighted in T_{75} terms related to metabolic processes of polysaccharides, aminoglycans, chitin, and glucosamine, which are related to plant adaptation process under stress, such as restructuring of the cell wall at lower water content and secondary wall formation [103,104]. In addition, the GO terms related to the chitin process were linked to the same set of differentially expressed genes (four genes) in the root tissue under dehydration (T_{75} and T_{150}) for DT and DS. The enzymatic hydrolysis of chitin involves an initial cleavage by chitinases into chitin oligosaccharides and further cleavage to N-acetylglucosamine (GlcNAc), and monosaccharides by chitobiases (reviewed by Hamid et al. [105]). In the present study, the enzyme chitinase (Phvul.005G155800) displayed increased expression after drought stress in both genotypes and tissues, highlighting the DT genotype in T_{75} for leaf tissue and T_{150} for the root. In plants, cell wall glycoproteins containing GlcNAc appear to be an endogenous substrate

for plant chitinases [106]. Their important role in plant defence against biotic stress is well known; moreover, it has been demonstrated that they are expressed under normal conditions and act in response to abiotic agents (ethylene, jasmonato) and conditions (cold, drought) [107,108].

We also observed that most genes related to oxidation-reduction terms were up-regulated genes in DT roots in early stress (T_{75}); however, in DS, it was identified in late stress (T_{150}) (Figure 5), suggesting that the susceptibility also seemed to be related to the late response to stress. On the other hand, the tolerance could be attributed to the ability to activate genes promoting an immediate response at the beginning of stress, increasing the chances of adaptation minimizing and/or fighting the effects of the lack of water. These findings are supported by previous results of the physiological characterization of Bat477 and Pérola genotypes evaluated under water deficit [15], in which the DS genotype triggered late perception and signalling pathways induced by water deficit compared with DT. In addition, the terms associated with the antioxidant system early activated in DT can be an important mechanism of tolerance in beans, inhibiting and/or reducing the damage caused by the deleterious effects of ROS to cells and tissues [109]. Among the stress-inducible genes, those involved in direct protection from stress, including the synthesis of regulatory proteins such as transcription factors, protein kinases, and phosphatase, were activated early in DT. In the DS plants (commercial cultivar Pérola), several genes related to the response to biotic stimulus in roots were identified, an expected finding because disease resistance is one of the main pillars of genomic breeding of common bean lines [110].

KEGG pathways of *DEGs*

Our data showed that, among the KEGG pathways enriched by DEGs up-regulated in DT, two were common to all treatments in leaves. These pathways are related to the regulation of sulphur-containing amino acid metabolism (cysteine and methionine) and sulphur metabolism, which are essential elements for plant growth [111]. In the present study, these findings suggested that these pathways are important to improve stress tolerance in the common bean due to the formation of many sulphur-containing defence compounds involved in plant defence signalling [112]. These processes of sulphur

metabolism consist of the uptake of inorganic sulphur (sulphate) by roots, followed by sulphate reduction and assimilation in leaves, leading to the synthesis of sulphur-containing amino-acids (reviewed by Wirtz and Droux [113]). Increased demand for reduced sulphur is seen involved in regulating essential processes in the plant, including protection against stresses, through the induction of S-containing compounds with cysteine as precursors for essential biomolecules [114]. Of interest is the antioxidant glutathione (GSH), which has been a central molecule determinant of cellular redox homeostasis [115]. Moreover, the biosynthesis of methionine from cysteine resulted in the hormone ethylene, polyamines and nicotine amine all involved at variable degrees in the modulation of the plant responses to stresses [116–118].

qPCR validation

Our results validated a high proportion of DEGs in RNA-seq experiments using qPCR (88.64%) with an overall strong correlation (61%) between both methods. As the experiment was carried out as biological replicates with the same genotypes and design used for the RNA-seq study, this similarity was expected. Although qPCR-based methods have been considered the gold standard for gene expression, it is negatively affected by several factors, besides being sensitive to low expression genes, and the specificity of amplification [119]. In addition, the differential isoform expression not considered for primer designing is a challenging situation and may interfere significantly compared with sampling different genes [120]. Of the 15 DEGs identified by RNA-Seq, 60% correlated well with qPCR results (100% of correspondence), an important technical validation of the RNA-Seq experiments. Among those eight validated genes, the antioxidant enzymes GST and POX, which are involved in maintaining cellular redox balance [121,122], have been up-regulated in the DT in several treatments of leaves and root tissues. The enzyme POX, up-regulated in the leaves of DT (T_0, T_{75}) and roots (T_{150}), have been shown to be closely associated with the response to abiotic stresses, such as drought tolerance [123]. Increased expression of POX was reported in the root tissue of soybean promoting drought tolerance (Mohammadi et al. [124]), as well as in other plants (Katam et al. [125]; Sheoran et al. [126]). Our study

also showed an up-regulation of GST in DT at all conditions in the leaf tissue. Xu et al [127] reported an overexpression of GST from Tomato (LeGSTU2) in *Arabidopsis*, resulting in increased activity of enzymes related to antioxidant responses and improving tolerance to salinity and drought stresses. These findings represent evidence that these enzymes have a favourable impact on the drought tolerance mechanism of the common bean and could be used as targets in the search and development of more tolerant genotypes.

For the enzyme involved in cell wall degradation, GH10 Xylanase, the results of RNA-Seq and qPCR were not perfectly correlated, but in the same direction and showing a similar trend, being DEG between DT and DS and up-regulated in the roots of DT at T_0 and T_{75} (confirmed using both methodologies). In addition, by qPCR, this enzyme was up-regulated in DT leaves and roots for all treatments in the present study. This enzyme presents predominantly an endo-beta-1,4-xylanase activity (GO:0005975) and hydrolase activity (GO:0004553), previously described in soybean as LOC100801147, and acts in the breakdown of hemicellulose, suggesting the disassembly of cell wall components during drought stress. The high levels of expression in the roots of DT over all treatments probably imply the cell expansion necessary to maintain the root growth under drought to absorb water in the deeper layers of the soil and, therefore, maintain the water status of the plant. In leaves, although the gene expression in DT was greater than that in DS, down-regulation was observed during the treatments (greater than 50%), and the expression level was three times smaller than that in the root. Loss of leaf area is the most important morphological adaptation to drought in common beans [128]. During the drought stress, while the root is strategically stimulated to grow to explore the water in the soil, the leaves become reduced in cell turgor caused by osmotic stress, and the mechanical power of the cell is also reduced to expand the polysaccharide network [129].

Genomic regions with variants of high impact

Through genome re-sequencing, numerous SNPs have been identified for species [130], providing more subsidies for genome wide association analysis, identification of genomic regions of agronomic interest, genomic selection and studies of genetic diversity. In this study, 3,067 transcripts, belonging to 2,501 genes, were under high impact (~2%). Although these are significantly lower in number, the functional annotation revealed that the affected genes are mostly involved in the cellular response to stimulus, response to stress, oxidation-reduction stress and regulation of gene expression (Figure 6; Additional files 4 and 6). These genes are targets for subsequent functional studies because they can be related to the divergence resulting from the adaptive process to which the genotypes were imposed during their development process. Bat 477 is a well-known source of drought tolerance developed by CIAT [131], whose adaptive advantage is attributed to the root systems [94], while DS (Pérola) was developed for growing under adequate cultivation conditions. An SNP set carrying variants filtered for high impact was identified in the present study, allowing follow up of the genetic variations and their contribution to phenotypic diversity in a large and representative sample of accessions. In a recent study [132], variants of high impact identified in genes involved in horse fertility were assigned to putative deleterious effects and were recommended for selection against these effects, demonstrating how genetic information can be used to benefit the breeding programmes.

CONCLUSION

In this study, we constructed a bean transcriptome under drought conditions. The presence of these terms from functional annotation resulting from the newly loci sampled in this study would suggest the absence of transcripts related to drought stress in the *P. vulgaris* reference transcriptome, possibly due to the absence or low level of expression under normal conditions. On the other hand, the sampling and identification of hundreds of new loci and thousands of new isoforms in this study resulted from the high coverage obtained through RNA-Seq, increasing the power to detect new genes. As more genes become available, the chance will be greater to perform more refined mining to obtain better knowledge of the control of specific routes related to the response to drought stress, which can be used for

breeding drought-resistant plants. In addition, a broad genome SNP coverage increases the perspective of predicting target genic regions and/or genomic candidate variants, allowing more refined studies. The results of this study contribute to the expansion of knowledge about the gene mechanisms, as well as their functional variants, related to two contrasting common drought-tolerant bean genotypes.

ABBREVIATIONS

ABA: Abscisic acid

AS: Salicylic acid.

bp: base pair

BSRQ: Base quality score recalibration

CDS: Coding DNA sequence

CIAT: International Center for Tropical Agriculture

C_q: Quantification cycle

CPM: Counts-per-million

DEG(s): Differentially expressed gene(s)

DGE: Differential Gene Expression

DS: Drought sensitive

DT: Drought tolerant

ET: Ethylene acid

EMBRAPA: Brazilian Company for Agricultural Research

FAO: Food and Agriculture Organization of the United Nations

FDR: False discovery rate

FPKM: Fragments Per Kilobase Million

FS: Fisher Strand

GO: Gene Ontology

GATK: Genome Analyze Toolkit

Gbp: Giga base pairs

Indel: Insertion or deletion

IR: Indel realignment

JA: Jasmonic acid

KEGG: Kyoto Encyclopedia of Genes and Genomes

LD: Linkage disequilibrium

lncRNAs: Long non-coding RNAs

Mbp: Mega base pairs

Mpa: Mega Pascal

NR: NCBI's non-redundant protein database

ORF: Open reading frame

PK: Protein kinase

QD: Quality Depth

qPCR: Quantitative real-time PCR (qPCR)

QTL: Quantitative trait loci

SD: Standard Deviation

SNP: Single nucleotide polymorphism

SRA: Sequence Read Archive

SSH library: Suppression subtractive hybridization library

T_0: Control condition

T_{25}: 25 minutes after being subjected to water stress

T_{50}: 50 minutes after being subjected to water stress

T_{75}: 75 minutes after being subjected to water stress

T_{100}: 100 minutes after being subjected to water stress

T_{125}: 125 minutes after being subjected to water stress

T_{150}: 150 minutes after being subjected to water stress

TF: Transcription factor

Ψw: Hydric potential

Ψs: Osmotic potential

DECLARATIONS

Ethics approval and consent to participate

The *common bean* genotypes used in the present study are derived from the Brazilian common bean core collection available at EMBRAPA Rice and Beans. The two plants used for transcriptome analysis were accessed with the knowledge of the institution for scientific research purposes only, with an official authorization from Brazilian authorities (IBAMA authorization number 02001.008430/2012-37).

Consent for publication

Not applicable.

Availability of data and material

The dataset(s) supporting the conclusions of this article are available in the SRA/NCBI database repository through the accession numbers SRP077562 (http://www.ncbi.nlm.nih.gov/sra/SRP077562). Annotation files, as GTF (General Feature Format) files containing the transcriptome annotation and VCF (Variant Call Format) files containing variant information, as well as other relevant data, are available on FigShare(https://doi.org/10.6084/m9.figshare.5057863).

Competing interests

The authors declare that they have no competing interests.

Funding

The National Council for Scientific and Technological Development (CNPq) for the grants to WJP, CB and RPV; and Coordination for the Improvement of Higher Education Personnel (CAPES) for the grant to ATOM. This work was supported by the Brazilian Agricultural Research Corporation (EMBRAPA - 02.12.12.005.00.00) and it was developed as described and approved in the institutional project. The contents of this publication, such as the results and conclusions, are the sole responsibility of the authors.

Authors' contributions

RPV, SAA, FAR and CB conceived and designed the study. FAR conducted the hydroponic experiment and took the physiological data. IRN-J performed statistical analyses of the physiological data. PAMRV performed qPCR experiments. WJP, ATOM, ASGC, SM and SAA performed the analysis of the RNA-seq data. WJP, ATOM and ASGC interpreted the results and wrote the paper. ACL interpreted the annotation results. WJP, CB, TCOB and RPV wrote the manuscript. All authors read and approved the final version of the manuscript.

Acknowledgements

We thank Barbara Müller for suggestions for the analysis of RNA-seq data and we are also grateful to Georgios Pappas for the access to the bioinformatics facility at the University of Brasilia.

REFERENCES

1. FAO LAND & WATER. Drought. http://www.fao.org/docrep/017/aq191e/aq191e.pdf. Accessed 02 August 2017.

2. Ravenscroft CH, Whitlock R, Fridley JD. Rapid genetic divergence in response to 15 years of simulated climate change. Glob. Change Biol. 2015;21:4165–76.

3. Anderson JT, Panetta AM, Mitchell-Olds T. Evolutionary and Ecological Responses to Anthropogenic Climate Change Update on Anthropogenic Climate Change. Plant Physiol. 2012;160:1728–40.

4. Jump AS, Marchant R, Peñuelas J. Environmental change and the option value of genetic diversity. Trends Plant Sci. 2009;14:51–8.

5. Alemayehu FR, Bendevis MA, Jacobsen S-E. The Potential for Utilizing the Seed Crop Amaranth (*Amaranthus* spp.) in East Africa as an Alternative Crop to Support Food Security and Climate Change Mitigation. J. Agron. Crop Sci. 2015;201:321–9.

6. Fisher M, Abate T, Lunduka RW, Asnake W, Alemayehu Y, Madulu RB. Drought tolerant maize for farmer adaptation to drought in sub-Saharan Africa: Determinants of adoption in eastern and southern Africa. Clim. Change. 2015;133:283–99.

7. Zhang Q. Strategies for developing Green Super Rice. Proc. Natl. Acad. Sci. 2007;104:16402–9.

8. Gentry HS. Origin of the Common Bean, *Phaseolus vulgaris*. Econ. Bot. 1969;23:55–69.

9. Thung M, Rao IM. Integrated Management of Abiotic Stresses. In: Singh SP, editor. Common Bean Improv. Twenty-First Century. Springer Netherlands; 1999. p. 331–70.

10. Petry N, Boy E, Wirth JP, Hurrell RF. Review: The Potential of the Common Bean (*Phaseolus vulgaris*) as a Vehicle for Iron Biofortification. Nutrients. 2015;7:1144–73.

11. Beebe S, Rao I, Blair M, Acosta J. Phenotyping common beans for adaptation to drought. Plant Physiol. 2013;4:35.

12. White JW, Singh SP. Sources and inheritance of earliness in tropically adapted indeterminate common bean. Euphytica. 1991;55:15–9.

13. Wortmann CS, Kirkby RA, Eledu CA, Allen DJ. Atlas of common bean (*Phaseolus vulgaris* L.) production in Africa. Centro Internacional de Agricultura Tropical (CIAT), Cali, CO. 131 p. (CIAT publication no. 297)

14. Singh SP, Terán H, Gutiérrez JP. Registration of SEA 5 and SEA 13 drought tolerant dry bean germplasm. Crop Sci. 2011;41:276–277.

15. Lanna AC, Mitsuzono ST, Terra TGR, Vianello RP, Carvalho MA de F. Physiological characterization of common bean (*Phaseolus vulgaris* L.) genotypes, water-stress induced with contrasting response towards drought. Aust. J. Crop Sci. 2016;10:1–6.

16. Rosales MA, Cuellar-Ortiz SM, de la Paz Arrieta-Montiel M, Acosta-Gallegos J, Covarrubias AA. Physiological traits related to terminal drought resistance in common bean (*Phaseolus vulgaris* L.). J. Sci. Food Agric. 2013;93:324–31.

17. Rosales-Serna R, Kohashi-Shibata J, Acosta-Gallegos JA, Trejo-López C, Ortiz-Cereceres J, Kelly JD. Biomass distribution, maturity acceleration and yield in drought-stressed common bean cultivars. Field Crops Res. 2004;85:203–11.

18. Mukeshimana G, Butare L, Cregan PB, Blair MW, Kelly JD. Quantitative Trait Loci Associated with Drought Tolerance in Common Bean. Crop Sci. 2014;54:923.

19. Blair MW, Soler A, Cortés AJ. Diversification and Population Structure in Common Beans (*Phaseolus vulgaris* L.). PLoS ONE. 2012;7:e49488.

20. Tuberosa R, Salvi S. From QTLs to genes controlling root traits in maize. Frontis. 2007;21:13–22.

21. Mir RR, Zaman-Allah M, Sreenivasulu N, Trethowan R, Varshney RK. Integrated genomics, physiology and breeding approaches for improving drought tolerance in crops. Theor. Appl. Genet. 2012;125:625–45.

22. Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, et al. A reference genome for common bean and genome-wide analysis of dual domestications. Nat. Genet. 2014;46:707–13.

23. Vlasova A, Capella-Gutiérrez S, Rendón-Anaya M, Hernández-Oñate M, Minoche AE, Erb I, et al. Genome and transcriptome analysis of the Mesoamerican common bean and the role of gene duplications in establishing tissue and temporal specialization of genes. Genome Biol. 2016;17:32.

24. Blair MW, Fernandez AC, Ishitani M, Moreta D, Seki M, Ayling S, et al. Construction and EST sequencing of full-length, drought stress cDNA libraries for common beans (*Phaseolus vulgaris* L.). BMC Plant Biol. 2011;11:171.

25. Recchia GH, Caldas DGG, Beraldo ALA, da Silva MJ, Tsai SM. Transcriptional Analysis of Drought-Induced Genes in the Roots of a Tolerant Genotype of the Common Bean (*Phaseolus vulgaris* L.). Int. J. Mol. Sci. 2013;14:7155–79.

26. Müller BS de F, Sakamoto T, Silveira RDD, Zambussi-Carvalho PF, Pereira M, Jr GJP, et al. Differentially Expressed Genes during Flowering and Grain Filling in Common Bean (*Phaseolus vulgaris*) Grown under Drought Stress Conditions. Plant Mol. Biol. Report. 2013;32:438–51.

27. Wu J, Wang L, Li L, Wang S. De Novo Assembly of the Common Bean Transcriptome Using Short Reads for the Discovery of Drought-Responsive Genes. PLoS ONE. 2014;9:e109262.

28. Müller BSF, Sakamoto T, Menezes IPP, Prado GS, Martins WS, Brondani C, et al. Analysis of BAC-end sequences in common bean (*Phaseolus vulgaris* L.) towards the development and characterization of long motifs SSRs. Plant Mol. Biol. 2014;86:455–70.

29. Guimarães CM, Stone LF, Moreira JAA. Soil compaction in a bean crop. II: effect on root and shoot development. Rev. Bras. Eng. Agríc. E Ambient. 2002;6:213–8.

30. Gonçalves RJ de S, Abreu Â de FB, Ramalho MAP, Bruzi AT. Strategies for recommendation of common bean lines tested for value of cultivation and use in different environments. Crop Breed. Appl. Biotechnol. 2009.

31. Terán H, Singh SP. Comparison of Sources and Lines Selected for Drought Resistance in Common Bean. Crop Sci. 2002;42:64–70.

32. Martins PK, Jordão BQ, Yamanaka N, Farias JRB, Beneventi MA, Binneck E, et al. Differential gene expression and mitotic cell analysis of the drought tolerant soybean (*Glycine max* L. Merrill *Fabales*, *Fabaceae*) cultivar MG/BR46 (Conquista) under two water deficit induction systems. Genet. Mol. Biol. 2008;31:512–21.

33. Hewitt EJ. Mineral nutrition of plants in culture media. Plant Physiol. Academic Press; 1963. p. 97–134.

34. Bajji M, Lutts S, Kinet J-M. Water deficit effects on solute contribution to osmotic adjustment as a function of leaf ageing in three durum wheat (*Triticum durum* Desf.) cultivars performing differently in arid conditions. Plant Sci. 2001;160:669–81.

35. Andrews S. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc. Accessed 02 August 2017.

36. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics. 2014;btu170.

37. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25:1105–11.

38. Picard Tools - By Broad Institute. http://broadinstitute.github.io/picard/. Accessed 02 August 2017.

39. Roberts A, Pimentel H, Trapnell C, Pachter L. Identification of novel transcripts in annotated genomes using RNA-Seq. Bioinformatics. 2011;27:2325–9.

40. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 2010;28:511–5.

41. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L. Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biol. 2011;12:R22.

42. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.

43. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31:166–9.

44. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.

45. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–402.

46. TransDecoder (Find Coding Regions Within Transcripts). https://transdecoder.github.io/. Accessed 02 August 2017.

47. Conesa A, Götz S. Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. Int. J. Plant Genomics. 2008. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2375974/. Accessed 02 August 2017.

48. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2015;gkv1070.

49. Nawrocki EP, Burge SW, Bateman A, Daub J, Eberhardt RY, Eddy SR, et al. Rfam 12.0: updates to the RNA families database. Nucleic Acids Res. 2015;43:D130–7.

50. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303.

51. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.

52. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 2011;43:491–8.

53. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinforma. Ed. Board Andreas Baxevanis Al. 2013;11:11.10.1-11.10.33.

54. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly (Austin). 2012;6:80–92.

55. Livak KJ, Schmittgen TD. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. Methods. 2001;25:402–8.

56. The Comprehensive R Archive Network. https://cran.r-project.org/. Accessed 02 August 2017.

57. Hellemans J, Mortier G, Paepe A, Speleman F, Vandesompele J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol. 2007;8.

58. Mehrotra R, Bhalothia P, Bansal P, Basantani MK, Bharti V, Mehrotra S. Abscisic acid and abiotic stress tolerance – Different tiers of regulation. J. Plant Physiol. 2014;171:486–96.

59. Tripathi P, Rabara RC, Reese RN, Miller MA, Rohila JS, Subramanian S, et al. A toolbox of genes, proteins, metabolites and promoters for improving drought tolerance in soybean includes the metabolite coumestrol and stomatal development genes. BMC Genomics. 2016;17:102.

60. Neves-Borges AC, Guimarães-Dias F, Cruz F, Mesquita RO, Nepomuceno AL, Romano E, et al. Expression pattern of drought stress marker genes in soybean roots under two water deficit systems. Genet. Mol. Biol. 2012;35:212–21.

61. Munns R, James RA, Sirault XRR, Furbank RT, Jones HG. New phenotyping methods for screening wheat and barley for beneficial responses to water deficit. J. Exp. Bot. 2010;61:3499–507.

62. Tripathi P, Rabara RC, Shulaev V, Shen QJ, Rushton PJ. Understanding Water-Stress Responses in Soybean Using Hydroponics System—A Systems Biology Perspective. Front. Plant Sci. 2015. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4685135/. Accessed 02 August 2017.

63. Shavrukov Y, Gupta NK, Miyazaki J, Baho MN, Chalmers KJ, Tester M, et al. HvNax3—a locus controlling shoot sodium exclusion derived from wild barley (*Hordeum vulgare* ssp. *spontaneum*). Funct. Integr. Genomics. 2010;10:277–91.

64. Ashraf M, Harris PJC. Photosynthesis under stressful environments: An overview. Photosynthetica. 2013;51:163–90.

65. Ho HL. Functional Roles of Plant Protein Kinases in Signal Transduction Pathways during Abiotic and Biotic Stress. J Biodivers Biopros Dev. 2015;

66. Osakabe Y, Yamaguchi-Shinozaki K, Shinozaki K, Tran L-SP. Sensing the environment: key roles of membrane-localized kinases in plant perception and response to abiotic stress. J. Exp. Bot. 2013;64:445–58.

67. Tuteja N. Abscisic Acid and Abiotic Stress Signaling. Plant Signal. Behav. 2007;2:135–8.

68. Ambawat S, Sharma P, Yadav NR, Yadav RC. MYB transcription factor genes as regulators for plant responses: an overview. Physiol. Mol. Biol. Plants Int. J. Funct. Plant Biol. 2013;19:307–21.

69. Rushton PJ, Somssich IE, Ringler P, Shen QJ. WRKY transcription factors. Trends Plant Sci. 2010;15:247–58.

70. Duque P, Chua N-H. IMB1, a bromodomain protein induced during seed imbibition, regulates ABA- and phyA-mediated responses of germination in Arabidopsis. Plant J. 2003;35:787–99.

71. Fujimoto SY, Ohta M, Usui A, Shinshi H, Ohme-Takagi M. Arabidopsis Ethylene-Responsive Element Binding Factors Act as Transcriptional Activators or Repressors of GCC Box–Mediated Gene Expression. Plant Cell. 2000;12:393–405.

72. Miura K, Tada Y. Regulation of water, salinity, and cold stress responses by salicylic acid. Front. Plant Sci. 2014. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3899523/. Accessed 02 August 2017.

73. Santino A, Taurino M, Domenico SD, Bonsegna S, Poltronieri P, Pastor V, et al. Jasmonate signaling in plant development and defense response to multiple (a)biotic stresses. Plant Cell Rep. 2013;32:1085–98.

74. Kazan K. Diverse roles of jasmonates and ethylene in abiotic stress tolerance. Trends Plant Sci. 2015;20:219–29.

75. Verma V, Ravindran P, Kumar PP. Plant hormone-mediated regulation of stress responses. BMC Plant Biol. 2016;16:86.

76. Park C-J, Seo Y-S. Heat Shock Proteins: A Review of the Molecular Chaperones for Plant Immunity. Plant Pathol. J. 2015;31:323–33.

77. Gupta D, Tuteja N. Chaperones and foldases in endoplasmic reticulum stress signaling in plants. Plant Signal. Behav. 2011;6:232–6.

78. Trivedi A, Mavi PS, Bhatt D, Kumar A. Thiol reductive stress induces cellulose-anchored biofilm formation in *Mycobacterium tuberculosis*. Nat. Commun. 2016;7:11392.

79. Felts SJ, Toft DO. p23, a simple protein with complex activities. Cell Stress Chaperones. 2003;8:108–13.

80. Kampinga HH, Craig EA. The HSP70 chaperone machinery: J proteins as drivers of functional specificity. Nat. Rev. Mol. Cell Biol. 2010;11:579–92.

81. Qi X, Xie S, Liu Y, Yi F, Yu J. Genome-wide annotation of genes and noncoding RNAs of foxtail millet in response to simulated drought stress by deep sequencing. Plant Mol. Biol. 2013;83:459–73.

82. Rosenzweig C, Iglesias A, Yang XB, Epstein PR, Chivian E. Climate Change and Extreme Weather Events; Implications for Food Production, Plant Diseases, and Pests. Glob. Change Hum. Health. 2:90–104.

83. Liu X, Hao L, Li D, Zhu L, Hu S. Long Non-coding RNAs and Their Biological Roles in Plants. Genomics Proteomics Bioinformatics. 2015;13:137–47.

84. Lu X, Chen X, Mu M, Wang J, Wang X, Wang D, et al. Genome-Wide Analysis of Long Noncoding RNAs and Their Responses to Drought Stress in Cotton (*Gossypium hirsutum* L.). PLOS ONE. 2016;11:e0156723.

85. Muthusamy M, Uma S, Backiyarani S, Saraswathi MS. Genome-wide screening for novel, drought stress-responsive long non-coding RNAs in drought-stressed leaf transcriptome of drought-tolerant and -susceptible banana (*Musa* spp) cultivars using Illumina high-throughput sequencing. Plant Biotechnol. Rep. 2015;9:279–86.

86. Wang T-Z, Liu M, Zhao M-G, Chen R, Zhang W-H. Identification and characterization of long non-coding RNAs involved in osmotic and salt stress in *Medicago truncatula* using genome-wide high-throughput sequencing. BMC Plant Biol. 2015;15:131.

87. Zhang W, Han Z, Guo Q, Liu Y, Zheng Y, Wu F, et al. Identification of Maize Long Non-Coding RNAs Responsive to Drought Stress. PLOS ONE. 2014;9:e98958.

88. Quan M, Chen J, Zhang D. Exploring the Secrets of Long Noncoding RNAs. Int. J. Mol. Sci. 2015;16:5467–96.

89. Zhang H, Hu W, Hao J, Lv S, Wang C, Tong W, et al. Genome-wide identification and functional prediction of novel and fungi-responsive lincRNAs in Triticum aestivum. BMC Genomics. 2016. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4791882/. Accessed 02 August 2017.

90. Ben Amor B, Wirth S, Merchan F, Laporte P, d'Aubenton-Carafa Y, Hirsch J, et al. Novel long non-protein coding RNAs involved in Arabidopsis differentiation and stress responses. Genome Res. 2009;19:57–69.

91. Fatica A, Bozzoni I. Long non-coding RNAs: new players in cell differentiation and development. Nat. Rev. Genet. 2014;15:7–21.

92. Asfaw A, Blair MW. Quantitative trait loci for rooting pattern traits of common beans grown under drought stress versus non-stress conditions. Mol. Breed. 2011;30:681–95.

93. Liu SJ, Horlbeck MA, Cho SW, Birk HS, Malatesta M, He D, et al. CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. Science. 2017;355:eaah7111.

94. Sponchiado BN, White JW, Castillo JA, Jones PG. Root Growth of Four Common Bean Cultivars in Relation to Drought Tolerance in Environments with Contrasting Soil Types. Exp. Agric. 1989;25:249–257.

95. Valdisser PAMR, Pappas GJ, Menezes IPP de, Müller BSF, Pereira WJ, Narciso MG, et al. SNP discovery in common bean by restriction-associated DNA (RAD) sequencing for genetic diversity and population structure analysis. Mol. Genet. Genomics. 2016;291:1277–91.

96. Shao H-B, Chu L-Y, Lu Z-H, Kang C-M. Primary antioxidant free radical scavenging and redox signaling pathways in higher plant cells. Int. J. Biol. Sci. 2007;4:8–14.

97. Noctor G, Foyer CH. Intracellular redox compartmentation and ROS-related communication in regulation and signaling. Plant Physiol. 2016;pp.00346.2016.

98. Krebs EG. The phosphorylation of proteins: a major mechanism for biological regulation. Biochem. Soc. Trans. 1985;13:813–20.

99. Lehti-Shiu MD, Shiu S-H. Diversity, classification and function of the plant protein kinase superfamily. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 2012;367:2619–39.

100. Wei S, Hu W, Deng X, Zhang Y, Liu X, Zhao X, et al. A rice calcium-dependent protein kinase OsCPK9 positively regulates drought stress tolerance and spikelet fertility. BMC Plant Biol. 2014;14:133.

101. Cai H, Cheng J, Yan Y, Xiao Z, Li J, Mou S, et al. Genome-wide Identification and Expression Analysis of Calcium-dependent Protein Kinase and Its Closely Related Kinase Genes in Capsicum annuum. Front. Plant Sci. 2015. http://journal.frontiersin.org/article/10.3389/fpls.2015.00737/full. Accessed 02 August 2017.

102. Ivashuta S, Liu J, Liu J, Lohar DP, Haridas S, Bucciarelli B, et al. RNA Interference Identifies a Calcium-Dependent Protein Kinase Involved in *Medicago truncatula* Root Development. Plant Cell. 2005;17:2911–21.

103. Moore JP, Vicré-Gibouin M, Farrant JM, Driouich A. Adaptations of higher plant cell walls to water loss: drought vs desiccation. Physiol. Plant. 2008;134:237–45.

104. Isolation and characterization of new polymorphic microsatellite markers from the cuttlefish *Sepiella maindroni* (*Cephalopoda*; *Sepiidae*). GMR Genet. Mol. Res. 2015. http://www.geneticsmr.com/articles/2326. Accessed 02 August 2017.

105. Hamid R, Khan MA, Ahmad M, Ahmad MM, Abdin MZ, Musarrat J, et al. Chitinases: An update. J. Pharm. Bioallied Sci. 2013;5:21.

106. Dyachok JV, Wiweger M, Kenne L, von Arnold S. Endogenous Nod-Factor-Like Signal Molecules Promote Early Somatic Embryo Development in Norway Spruce. Plant Physiol. 2002;128:523–33.

107. Passarinho PA, Hengel AJV, Fransz PF, Vries SC de. Expression pattern of the Arabidopsis thaliana AtEP3/AtchitIV endochitinase gene. Planta. 2001;212:556–67.

108. Ahmed NU, Park J-I, Jung H-J, Kang K-K, Hur Y, Lim Y-P, et al. Molecular characterization of stress resistance-related chitinase genes of *Brassica rapa*. Plant Physiol. Biochem. 2012;58:106–15.

109. Gill SS, Tuteja N. Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. Plant Physiol. Biochem. PPB Société Fr. Physiol. Végétale. 2010;48:909–30.

110. Souza TLPO, Faleiro FG, Dessaune SN, Paula-Junior TJ de, Moreira MA, Barros EG de. Breeding for common bean (*Phaseolus vulgaris* L.) rust resistance in Brazil. Trop. Plant Pathol. 2013;38:361–74.

111. Hawkesford MJ, De Kok LJ. Managing sulphur metabolism in plants. Plant Cell Environ. 2006;29:382–95.

112. Capaldi FR, Gratão PL, Reis AR, Lima LW, Azevedo RA. Sulfur Metabolism and Stress Defense Responses in Plants. Trop. Plant Biol. 2015;8:60–73.

113. Wirtz M, Droux M. Synthesis of the sulfur amino acids: cysteine and methionine. Photosynth. Res. 2005;86:345–62.

114. Romero LC, Aroca MÁ, Laureano-Marín AM, Moreno I, García I, Gotor C. Cysteine and Cysteine-Related Signaling Pathways in *Arabidopsis thaliana*. Mol. Plant. 2014;7:264–76.

115. Noctor G, Mhamdi A, Chaouch S, Han Y, Neukermans J, Marquez-Garcia B, et al. Glutathione in plants: an integrated overview. Plant Cell Environ. 2012;35:454–84.

116. Liu J-H, Wang W, Wu H, Gong X, Moriguchi T. Polyamines function in stress tolerance: from synthesis to regulation. Front. Plant Sci. 2015. http://journal.frontiersin.org/article/10.3389/fpls.2015.00827/full. Accessed 02 August 2017.

117. Burstenbinder K, Sauter M. Early events in the ethylene biosynthetic pathway–regulation of the pools of methionine and S-adenosylmethionine. Ann Plant Rev Plant Horm. Ethyl. 2012;44:22.

118. Lee D-K, Chung PJ, Jeong JS, Jang G, Bang SW, Jung H, et al. The rice OsNAC6 transcription factor orchestrates multiple molecular mechanisms involving root structural adaptions and nicotianamine biosynthesis for drought tolerance. Plant Biotechnol. J. 2017;15:754–64.

119. Everaert C, Luypaert M, Maag JLV, Cheng QX, Dinger ME, Hellemans J, et al. Benchmarking of RNA-sequencing analysis workflows using whole-transcriptome RT-qPCR expression data. Sci. Rep. 2017;7:1559.

120. Teixeira FK, Menezes-Benavente L, Galvão VC, Margis R, Margis-Pinheiro M. Rice ascorbate peroxidase gene family encodes functionally diverse isoforms localized in different subcellular compartments. Planta. 2006;224:300.

121. Galant A, Preuss ML, Cameron J, Jez JM. Plant Glutathione Biosynthesis: Diversity in Biochemical Regulation and Reaction Products. Front. Plant Sci. 2011. http://journal.frontiersin.org/article/10.3389/fpls.2011.00045/full. Accessed 02 August 2017.

122. Asada K. Ascorbate peroxidase – a hydrogen peroxide-scavenging enzyme in plants. Physiol. Plant. 1992;85:235–41.

123. Rossel JB, Walter PB, Hendrickson L, Chow WS, Poole A, Mullineaux PM, et al. A mutation affecting ASCORBATE PEROXIDASE 2 gene expression reveals a link between responses to high light and drought tolerance. Plant Cell Environ. 2006;29:269–81.

124. Mohammadi PP, Moieni A, Hiraga S, Komatsu S. Organ-specific proteomic analysis of drought-stressed soybean seedlings. J. Proteomics. 2012;75:1906–23.

125. Katam R, Sakata K, Suravajhala P, Pechan T, Kambiranda DM, Naik KS, et al. Comparative leaf proteomics of drought-tolerant and -susceptible peanut in response to water stress. J. Proteomics. 2016;143:209–26.

126. Sheoran S, Thakur V, Narwal S, Turan R, Mamrutha HM, Singh V, et al. Differential Activity and Expression Profile of Antioxidant Enzymes and Physiological Changes in Wheat (*Triticum aestivum* L.) Under Drought. Appl. Biochem. Biotechnol. 2015;177:1282–98.

127. Xu J, Xing X-J, Tian Y-S, Peng R-H, Xue Y, Zhao W, et al. Transgenic Arabidopsis Plants Expressing Tomato Glutathione S-Transferase Showed Enhanced Resistance to Salt and Drought Stress. PLOS ONE. 2015;10:e0136960.

128. Acosta-Gallegos JA. Selection of Common Bean (*Phaseolus vulgaris* L) Genotypes with Enhanced Drought Tolerance and Biological Nitrogen Fixation. Michigan State University; 1988.

129. Tenhaken R. Cell wall remodeling under abiotic stress. Front. Plant Sci. 2015. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4285730/. Accessed 02 August 2017.

130. Song Q, Hyten DL, Jia G, Quigley CV, Fickus EW, Nelson RL, et al. Fingerprinting Soybean Germplasm and Its Utility in Genomic Research. G3 GenesGenomesGenetics. 2015;5:1999–2006.

131. White JW, Castillo JA. Relative Effect of Root and Shoot Genotypes on Yield of Common Bean under Drought Stress. Crop Sci. 1989;29:360–2.

132. Schrimpf R, Gottschalk M, Metzger J, Martinsson G, Sieme H, Distl O. Screening of whole genome sequences identified high-impact variants for stallion fertility. BMC Genomics. 2016;17:288.

FIGURES

Figure 1. A representative scheme detailing the twelve RNA-Seq libraries evaluated in the present study. Two different plant tissues (leaves and roots) from each contrasting genotype (BAT 477 *vs*. Pérola) were submitted to drought stress, and the plant tissues were sampled at three different times.

Figure 2. Differentially expressed genes between treatments for each genotype and tissue. a) DEGs from BAT 477 in the root; b) DEGs from Pérola in the root; c) DEGs from BAT 477 in the leaf; d) DEGs from Pérola in the leaf.

Figure 3. Distribution of variants (SNPs and Indels) in *Phaseolus vulgaris* chromosomes.

Figure 4. Comparisons between the gene expression profiles obtained for DEGs by RNA-Seq and qPCR analyses.

Figure 5. Comparison of genes associated with the terms related to oxidative stress obtained for BAT 477 and Pérola.

Figure 6. Functional annotation showing the most relevant GO terms associated with transcripts under high-impact predicted effects, according to SnpEff. The terms were filtered according to the node score. The numbers represent the total transcripts linked to each term.

Aditional file 1. Physiological parameters measured on plants grown hydroponically exposed to short-term water deficit. Statistical analysis presents the mean, standard deviation (SD) and coefficient of variation (CV) as a percentage of physiological analyses performed on two bean genotypes at seven distinct times of treatment, as described in the methodology. (XLSX 10.8 kB)

Additional file 2. Information of raw sequences derived from each library sequenced on the Illumina Genome Analyzer and Illumina HiSeq2000 platforms, including the total number of reads in base pairs, the paired- and single-end total resulting reads and the quality control (cc) data. The numbers shown represent the pair of reads, each pair with 2x150 bp before QC control.(XLSX 9.3 kB)

Additional file 3. Functional annotation based on Blast2GO with *high* BLAST *e-value* predicted for the new loci constructed by Cufflinks, the differentially expressed genes (DEGs) and the most relevant GO terms associated with transcripts under high-impact predicted effects, according to SnpEff. (XLSX 4.7 MB)

Additional file 4. Classification of new loci on the “Biological Process” (A), “Molecular Function” (B) and “Cellular Component” (C) GO categories. The values indicated on the plot refer to the "Node Score", which is one of the criteria used to define the most relevant terms. (JPG 4.7 MB)

Additional file 5. Description of the G*ene Ontology (GO) enrichment analysis for* the new loci, DEGs and genes under high-impact effects using Fisher's exact test corrected for multiple testing using FDR <0.05. (XLSX 142.7 KB)

Additional file 6. Metabolic pathways identified through KEGG analysis for the new loci, DEGs and SNPs under high-impact effects. (XLSX 80.2 KB)

Additional file 7. Description of the new loci with no hits found using the Blastx or PSI-Blast against the NR database, the Mesoamerican transcriptome sequences and lncRNAs described by Vlasova et al. (2016). (XLSX 347.2 KB)

Additional file 8. Comparison of the number of genes expressed in leaf and root tissues, as well as between genotypes, in the control condition (without induction of physiological stress - A), after 75 min of physiological stress (B) and 150 min of physiological stress (C). Comparisons were performed with treatments between the leaf and root of the BAT 477 genotype (a), between the leaf and root of the Pérola genotype (b), between BAT 477 and Pérola considering the leaf tissue (c) and between BAT 477 and Pérola considering the root tissue (d). (PNG 2.6 MB)

Additional file 9. DEG list including the up- and down-regulated genes based on the pairwise comparison among the libraries, for which the identification was based on the Cufflinks loci ID and corresponding ID to the *P. vulgaris* v.1.0 annotation files (Phytozome). (XLSX 75.4 KB)

Additional file 10. Expression profile from the Long non-conding RNAs (lncRNAs) categorized as DEGs (Fold Change > 4 and FDR < 0.01). (PNG 577.6 kB)

Additional file 11. Comparison of the differentially expressed genes (DEGs) found for BAT 477 and Pérola genotypes in response to the treatments: a) DEGs up-regulated between T_{75} and T_0 for the root tissue; b) DEGs down-regulated between T_{75} and T_0 for the root tissue; c) DEGs up-regulated between T_{150} and T_0 for the leaf tissue; d) DEGs down-regulated between T_{150} and T_0 for the leaf tissue; e) DEGs up-regulated between T_{150} and T_0 for the root tissue; f) DEGs down-regulated between T_{150} and T_0 for the root tissue. (PNG 2.2 MB)

Additional file 12. Classification of DEGs on the “Biological Process” (A); “Molecular Function” (B) and “Cellular Component” GO categories. The values indicated on the plot refer to the "Node Score", which is one of the criteria used to define the most relevant terms. (PNG 8.2 MB)

Additional file 13. Descriptions of TaqMan probes with the identification of their efficiency of amplification and corresponding annotation based on the BioMart tool available in Phytozome v.12. (XLSX 10.8 kB)

Additional file 14. Expression profile for the 15 genes selected for qPCR validation. The data were analysed using the Delta-Delta Ct *method* and the DS genotype under the control condition (T_0) as the reference sample. The statistical tests were performed separately with each sample and tissue using ANOVA and Turkey's test with a 95% confidence interval and correction by FDR < 0.05 (the significances were represented by letters). a) Phvul.006G171000 (aquaporin NIP – AQP), b) Phvul.001G205900 (Peripheral-type benzodiazepine receptor and related proteins – PBR), c) Phvul.006G060700 (DNAJ Homologue subfamily C member – DNAJC), d) Phvul.005G158500 (Beta-fructofuranosidase/Insoluble isoenzyme CWINV1-Related – GH32), e) Phvul.005G155800 (Class IV Chitinase – CHIA), f) Phvul.001G075400 (Protein phosphatase 2C – PP2C), g) Phvul.008G113700 (Glutathione S-transferase – GST), h) Phvul.007G061800 (Heat shock transcription factor – HSF), i) Phvul.007G259400 (Late embryogenesis abundant (LEA) group 1 – LEA), j) Phvul.002G228700 (Linoleate 13S-lipoxygenase – LOX), k) Phvul.002G184600 (Myb proto-oncogene protein, plant – MYB), l) Phvul.005G084500 (No apical meristem (NAM) protein / NAC Domain-containing protein 9-related – NAM_NAC), m) Phvul.003G131500 (12-oxophytodienoate reductase – NADH), n) Phvul.009G140700 (Peroxidase – POX) and o) Phvul.009G120500 (Glycosyl hydrolase family 10 – GH10). (XLSX 335.8 kB)

Additional file 15. Proportion of the correspondence between DEGs identified by RNA-Seq and qPCR results and correlation coefficient measurement using the Pearson's and Spearman's methods. (XLSX 335.8 k)

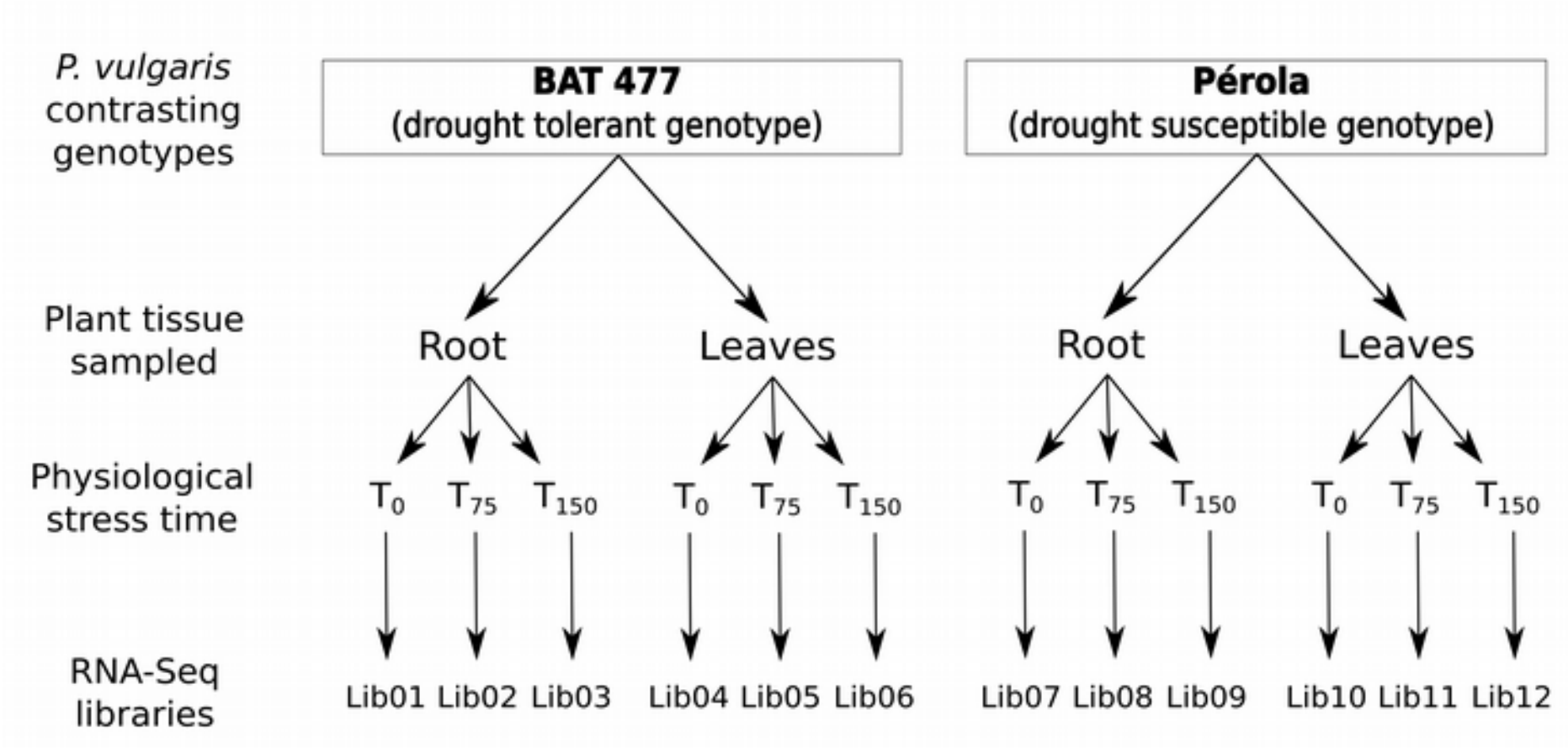
P. vulgaris
contrasting
genotypes
BAT 477
(drought tolerant genotype)
Pérola
(drought susceptible genotype)
Plant tissue
sampled
Root
Leaves
Root
Leaves
Physiological
stress time
T0 T75 T150
T0 T75 T150
T0 T75 T150
T0 T75 T150
RNA-Seq
libraries
Lib01 Lib02 Lib03
Lib04 Lib05 Lib06
Lib07 Lib08 Lib09
Lib10 Lib11 Lib12

Figure 1.

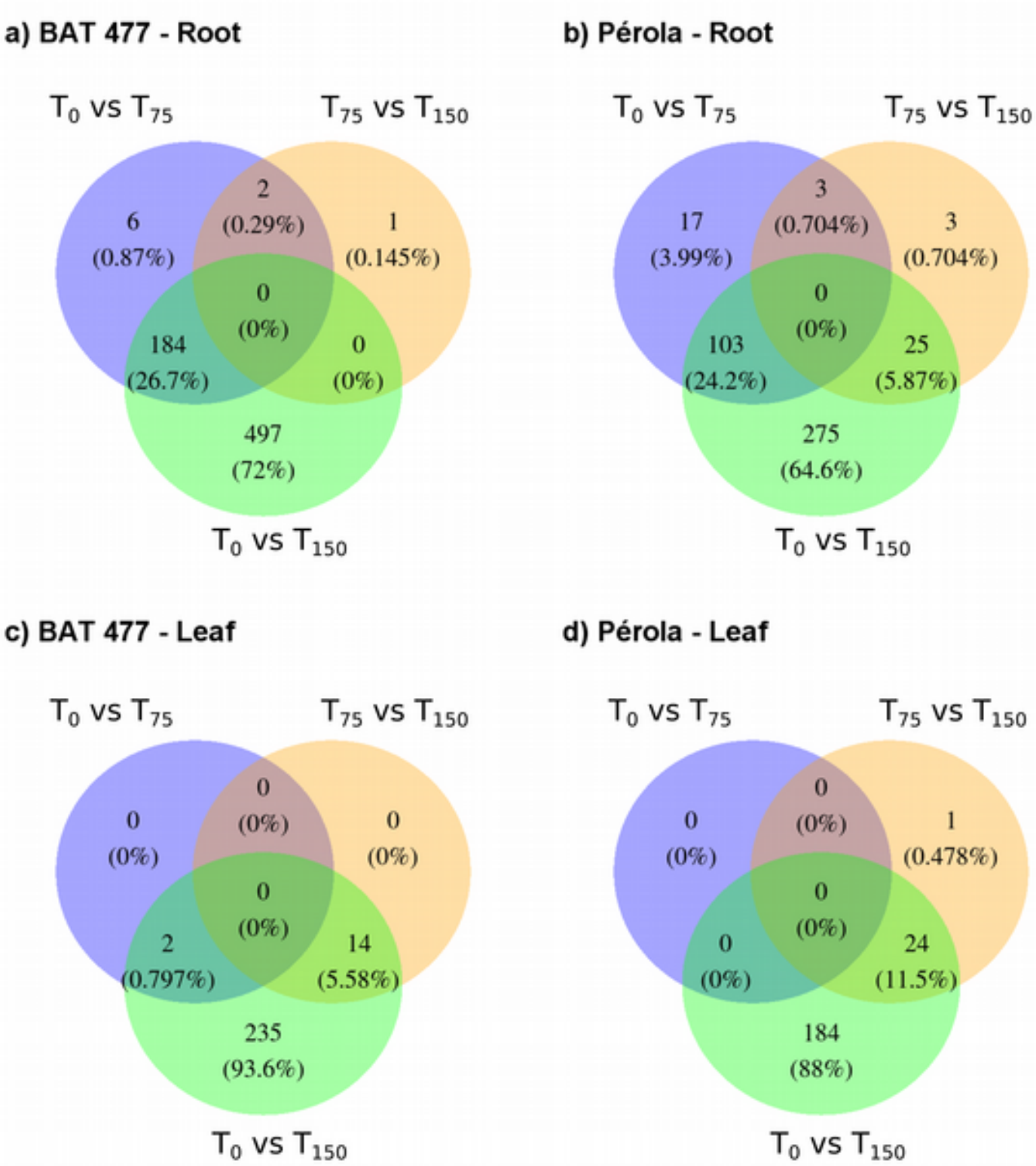
a) BAT 477 - Root
T0 vs T75
T75 vs T150
6 (0.87%)
2 (0.29%)
1 (0.145%)
0 (0%)
184 (26.7%)
0 (0%)
497 (72%)
T0 vs T150
b) Pérola - Root
T0 vs T75
T75 vs T150
17 (3.99%)
3 (0.704%)
3 (0.704%)
0 (0%)
103 (24.2%)
25 (5.87%)
275 (64.6%)
T0 vs T150
c) BAT 477 - Leaf
T0 vs T75
T75 vs T150
0 (0%)
0 (0%)
0 (0%)
0 (0%)
2 (0.797%)
14 (5.58%)
235 (93.6%)
T0 vs T150
d) Pérola - Leaf
T0 vs T75
T75 vs T150
0 (0%)
0 (0%)
1 (0.478%)
0 (0%)
0 (0%)
24 (11.5%)
184 (88%)
T0 vs T150

Figure 2.

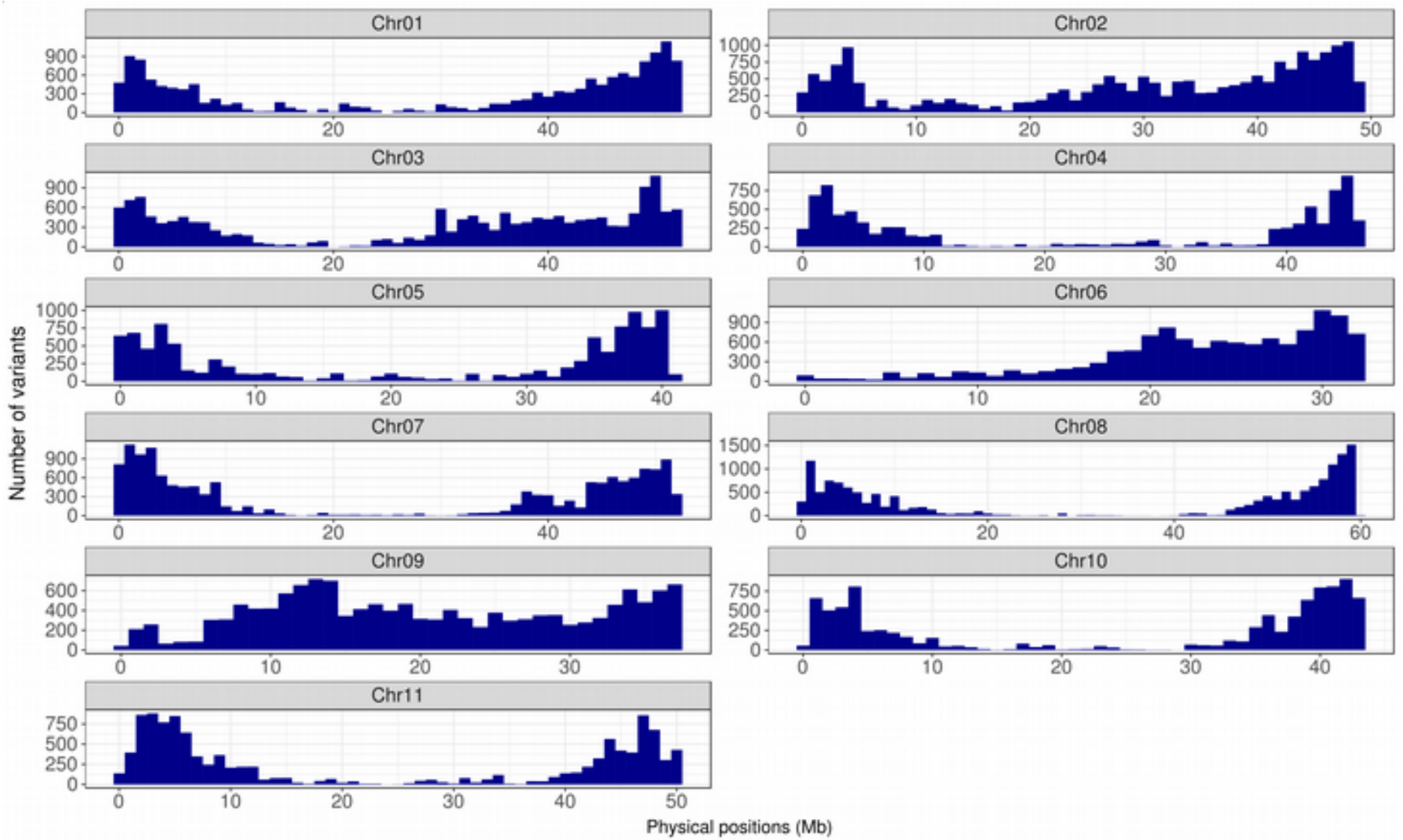
Chr01
Chr02
Chr03
Chr04
Chr05
Chr06
Chr07
Chr08
Chr09
Chr10
Chr11
Number of variants
Physical positions (Mb)

Figure 3.

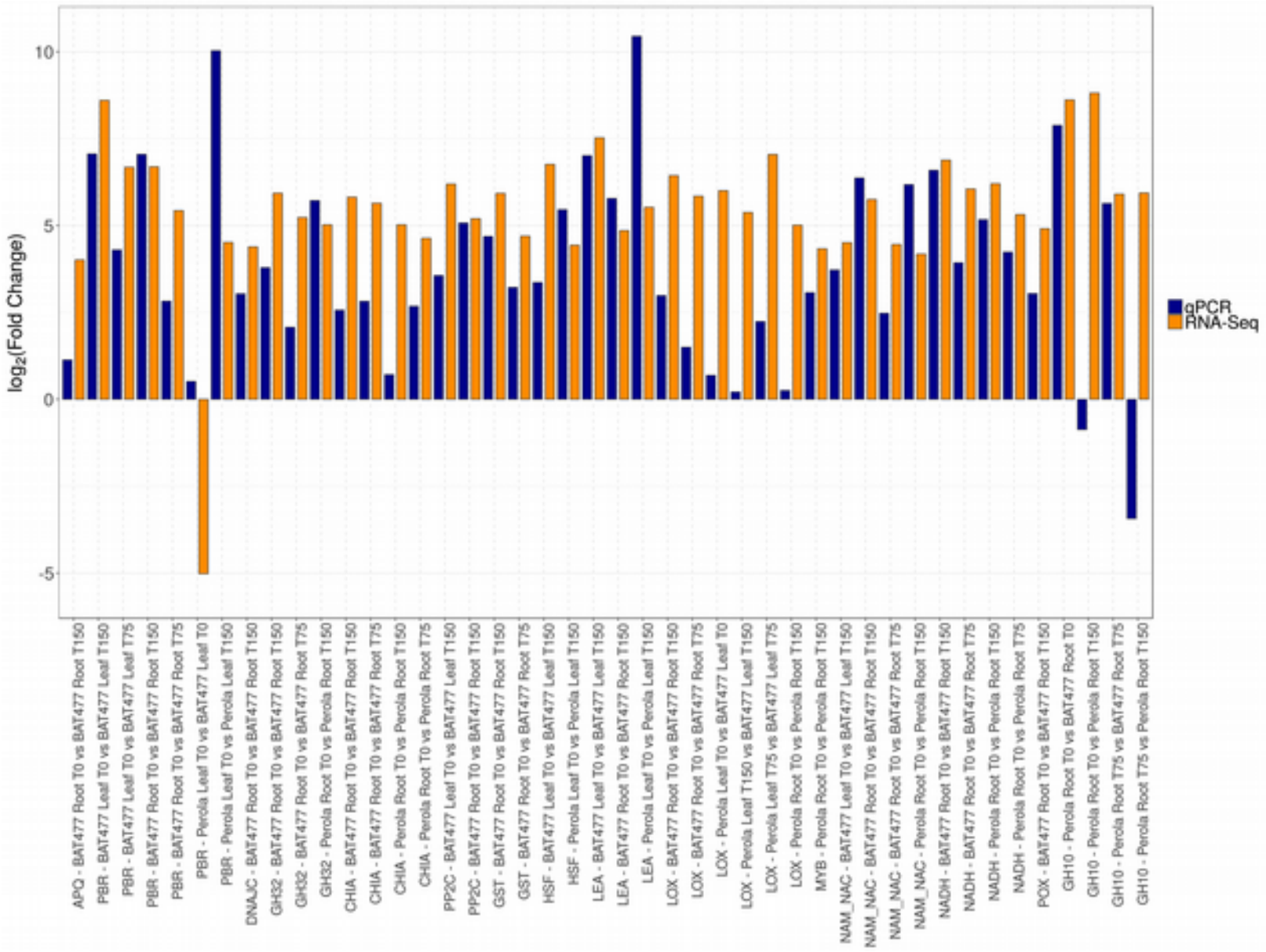
log2(Fold Change)
10
5
0
-5
qPCR
RNA-Seq
APQ - BAT477 Root T0 vs BAT477 Root T150
PBR - BAT477 Leaf T0 vs BAT477 Leaf T150
PBR - BAT477 Leaf T0 vs BAT477 Leaf T75
PBR - BAT477 Root T0 vs BAT477 Root T150
PBR - BAT477 Root T0 vs BAT477 Root T75
PBR - Perola Leaf T0 vs BAT477 Leaf T0
PBR - Perola Leaf T0 vs Perola Leaf T150
DNAJC - BAT477 Root T0 vs BAT477 Root T150
GH32 - BAT477 Root T0 vs BAT477 Root T150
GH32 - BAT477 Root T0 vs BAT477 Root T75
GH32 - Perola Root T0 vs Perola Root T150
CHIA - BAT477 Root T0 vs BAT477 Root T150
CHIA - BAT477 Root T0 vs BAT477 Root T75
CHIA - Perola Root T0 vs Perola Root T150
CHIA - Perola Root T0 vs Perola Root T75
PP2C - BAT477 Leaf T0 vs BAT477 Leaf T150
PP2C - BAT477 Root T0 vs BAT477 Root T150
GST - BAT477 Root T0 vs BAT477 Root T150
GST - BAT477 Root T0 vs BAT477 Root T75
HSF - BAT477 Leaf T0 vs BAT477 Leaf T150
HSF - Perola Leaf T0 vs Perola Leaf T150
LEA - BAT477 Leaf T0 vs BAT477 Leaf T150
LEA - BAT477 Root T0 vs BAT477 Root T150
LEA - Perola Leaf T0 vs Perola Leaf T150
LOX - BAT477 Root T0 vs BAT477 Root T150
LOX - BAT477 Root T0 vs BAT477 Root T75
LOX - Perola Leaf T0 vs BAT477 Leaf T0
LOX - Perola Leaf T150 vs BAT477 Leaf T150
LOX - Perola Leaf T75 vs BAT477 Leaf T75
LOX - Perola Root T0 vs Perola Root T150
MYB - Perola Root T0 vs Perola Root T150
NAM_NAC - BAT477 Leaf T0 vs BAT477 Leaf T150
NAM_NAC - BAT477 Root T0 vs BAT477 Root T150
NAM_NAC - BAT477 Root T0 vs BAT477 Root T75
NAM_NAC - Perola Root T0 vs Perola Root T150
NADH - BAT477 Root T0 vs BAT477 Root T150
NADH - BAT477 Root T0 vs BAT477 Root T75
NADH - Perola Root T0 vs Perola Root T150
NADH - Perola Root T0 vs Perola Root T75
POX - BAT477 Root T0 vs BAT477 Root T150
GH10 - Perola Root T0 vs BAT477 Root T0
GH10 - Perola Root T0 vs Perola Root T150
GH10 - Perola Root T75 vs BAT477 Root T75
GH10 - Perola Root T75 vs Perola Root T150

Figure 4.

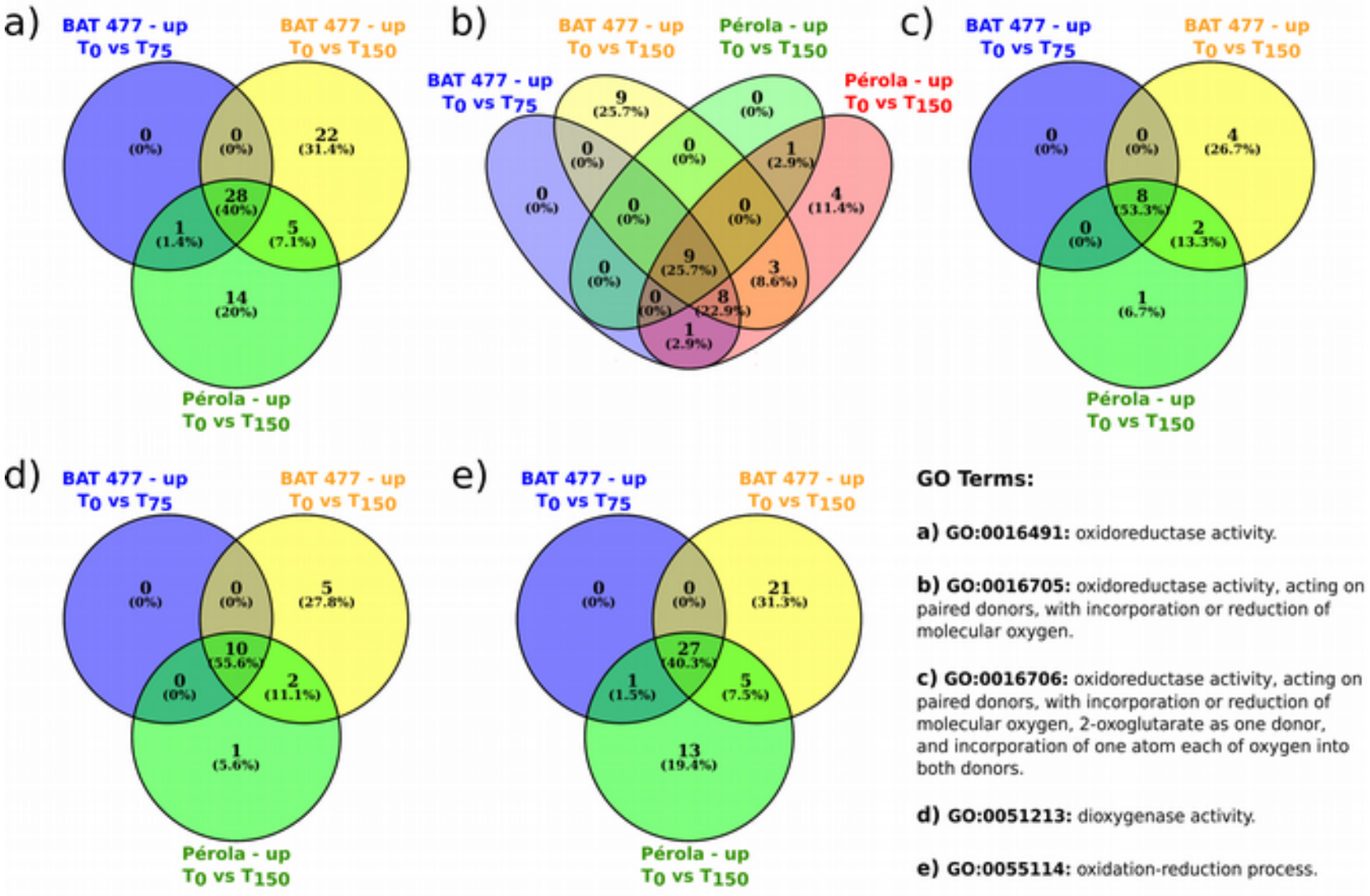
a)
BAT 477 - up T0 vs T75
BAT 477 - up T0 vs T150
Pérola - up T0 vs T150
0 (0%)
0 (0%)
22 (31.4%)
28 (40%)
1 (1.4%)
5 (7.1%)
14 (20%)
b)
BAT 477 - up T0 vs T150
Pérola - up T0 vs T150
BAT 477 - up T0 vs T75
Pérola - up T0 vs T150
9 (25.7%)
0 (0%)
0 (0%)
0 (0%)
1 (2.9%)
0 (0%)
0 (0%)
0 (0%)
4 (11.4%)
9 (25.7%)
0 (0%)
3 (8.6%)
0 (0%)
8 (22.9%)
1 (2.9%)
c)
BAT 477 - up T0 vs T75
BAT 477 - up T0 vs T150
Pérola - up T0 vs T150
0 (0%)
0 (0%)
4 (26.7%)
8 (53.3%)
0 (0%)
2 (13.3%)
1 (6.7%)
d)
BAT 477 - up T0 vs T75
BAT 477 - up T0 vs T150
Pérola - up T0 vs T150
0 (0%)
0 (0%)
5 (27.8%)
10 (55.6%)
0 (0%)
2 (11.1%)
1 (5.6%)
e)
BAT 477 - up T0 vs T75
BAT 477 - up T0 vs T150
Pérola - up T0 vs T150
0 (0%)
0 (0%)
21 (31.3%)
27 (40.3%)
1 (1.5%)
5 (7.5%)
13 (19.4%)
GO Terms:
a) GO:0016491: oxidoreductase activity.
b) GO:0016705: oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen.
c) GO:0016706: oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors.
d) GO:0051213: dioxygenase activity.
e) GO:0055114: oxidation-reduction process.

Figure 5.

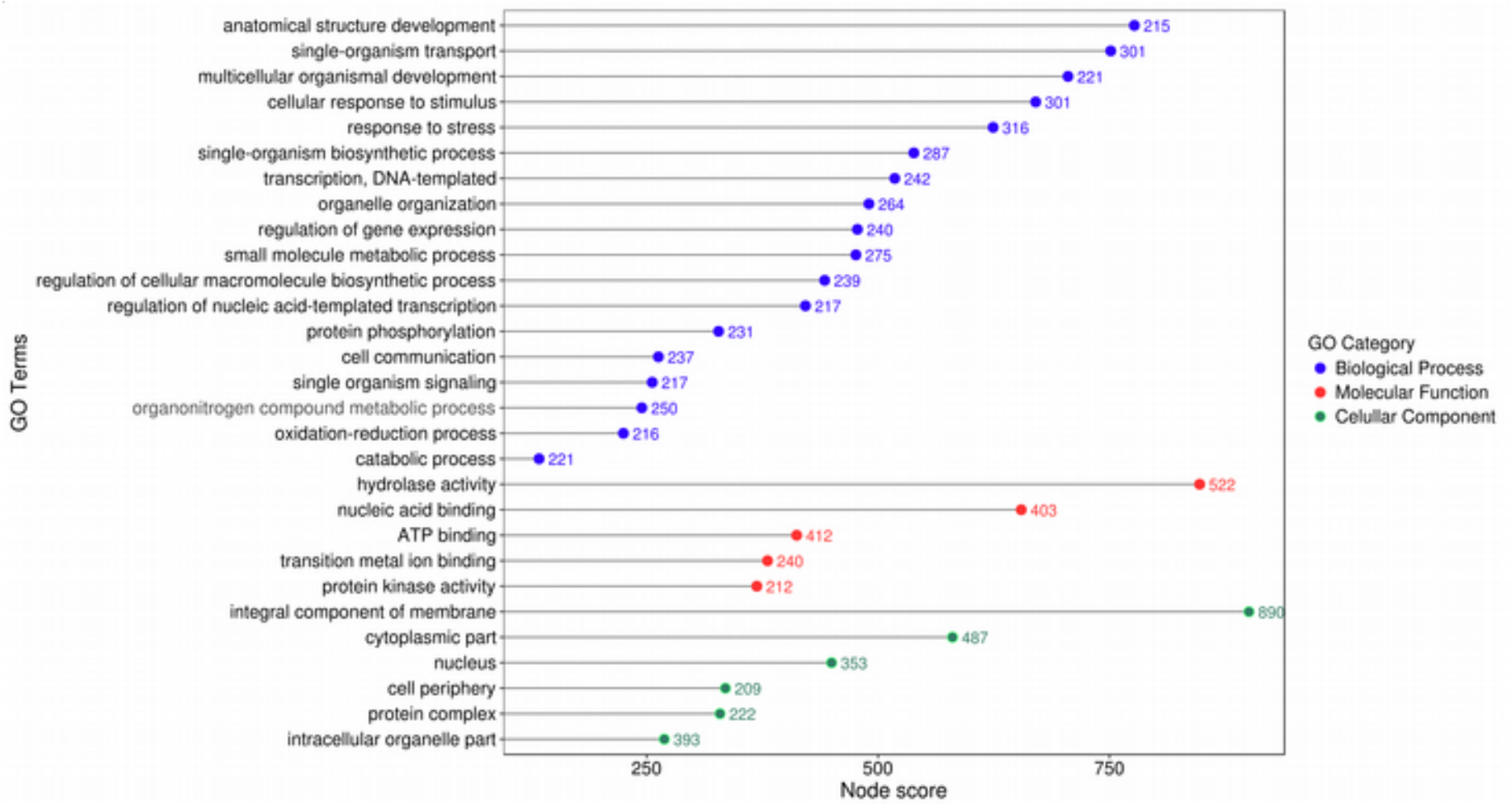
GO Terms
anatomical structure development
single-organism transport
multicellular organismal development
cellular response to stimulus
response to stress
single-organism biosynthetic process
transcription, DNA-templated
organelle organization
regulation of gene expression
small molecule metabolic process
regulation of cellular macromolecule biosynthetic process
regulation of nucleic acid-templated transcription
protein phosphorylation
cell communication
single organism signaling
organonitrogen compound metabolic process
oxidation-reduction process
catabolic process
hydrolase activity
nucleic acid binding
ATP binding
transition metal ion binding
protein kinase activity
integral component of membrane
cytoplasmic part
nucleus
cell periphery
protein complex
intracellular organelle part
215
301
221
301
316
287
242
264
240
275
239
217
231
237
217
250
216
221
522
403
412
240
212
890
487
353
209
222
393
250
500
750
Node score
GO Category
Biological Process
Molecular Function
Celullar Component

Figure 6.

5. Conclusões

Nesse estudo, foi construído um transcritoma específico de feijoeiro sob condições de estresse abiótico, contendo 1668 novos loci e 23 169 novas isoformas, totalizando um conjunto de 54 807 transcritos. Dos 1668 novos loci, 791 foram identificados e anotados, sendo genes que ainda não haviam sido relatados para o genoma de referência. Adicionalmente, foram verificados 1242 genes diferencialmente expressos (GDEs), estando distribuídos entre as comparações implementadas. Dos 1242 GDEs, quinze genes foram analisados através da técnica de qPCR, onde 74,4% dos resultados obtidos em RNA-Seq foram validados.

Foram identificadas 151 283 variantes, sendo 135 167 SNPs e 16 115 indels. Todos os SNPs verificados tiveram anotação funcional realizada com SnpEff v.4.2, demonstrando a ocorrência de 330 378 efeitos causais em decorrência da presença das variantes.

Os dados utilizados nesse estudo foram depositados no *Sequence Read Archive* (NCBI) e serão disponibilizados publicamente. Os resultados deste estudo contribuirão para a ampliação do conhecimento sobre os mecanismos gênicos, bem como de suas variantes funcionais, relacionados com a tolerância a seca em feijoeiro.

6. Referências

1. Classification | USDA PLANTS [Internet]. [citado em Março de 2016]. Disponível em:http://plants.usda.gov/java/ClassificationServlet?source=display&classid=PHVU
2. Vlasova A, Capella-Gutiérrez S, Rendón-Anaya M, Hernández-Oñate M, Minoche AE, Erb I, et al. Genome and transcriptome analysis of the Mesoamerican common bean and the role of gene duplications in establishing tissue and temporal specialization of genes. Genome Biol. 2016;17:32.
3. Chen PX, Tang Y, Marcone MF, Pauls PK, Zhang B, Liu R, et al. Characterization of free, conjugated and bound phenolics and lipophilic antioxidants in regular- and non-darkening cranberry beans (*Phaseolus vulgaris* L.). Food Chem. 2015 Oct 15;185:298–308.
4. Njoroge DM, Kinyanjui PK, Makokha AO, Christiaens S, Shpigelman A, Sila DN, et al. Extraction and characterization of pectic polysaccharides from easy- and hard-to-cook common beans (*Phaseolus vulgaris*). Food Res Int. 2014;64:314–22.
5. Pedrosa MM, Cuadrado C, Burbano C, Muzquiz M, Cabellos B, Olmedilla-Alonso B, et al. Effects of industrial canning on the proximate composition, bioactive compounds contents and nutritional profile of two Spanish common dry beans (*Phaseolus vulgaris* L.). Food Chem. 2015 Jan 1;166:68–75.
6. Petry N, Boy E, Wirth JP, Hurrell RF. Review: The Potential of the Common Bean (*Phaseolus vulgaris*) as a Vehicle for Iron Biofortification. Nutrients. 2015 Feb 11;7(2):1144–73.
7. Luna-Vital DA, Mojica L, González de Mejía E, Mendoza S, Loarca-Piña G. Biological potential of protein hydrolysates and peptides from common bean (*Phaseolus vulgaris* L.): A review. Food Res Int. 2015 Outubro;76, Part 1:39–50.
8. García-Lafuente A, Moro C, Manchón N, Gonzalo-Ruiz A, Villares A, Guillamón E, et al. In vitro anti-inflammatory activity of phenolic rich extracts from white and red common beans. Food Chem. 2014 Outubro;161:216–23.

9. Luna Vital DA, González de Mejía E, Dia VP, Loarca-Piña G. Peptides in common bean fractions inhibit human colorectal cancer cells. Food Chem. 2014 Agosto;157:347–55.

10. Monk JM, Zhang CP, Wu W, Zarepoor L, Lu JT, Liu R, et al. White and dark kidney beans reduce colonic mucosal damage and inflammation in response to dextran sodium sulfate. J Nutr Biochem. 2015 Jul 1;26(7):752–60.

11. Moreno-Jiménez MR, Cervantes-Cardoza V, Gallegos-Infante JA, González-Laredo RF, Estrella I, García-Gasca T de J, et al. Phenolic composition changes of processed common beans: their antioxidant and anti-inflammatory effects in intestinal cancer cells. Food Res Int. 2015 Outubro;76, Part 1:79–85.

12. Oseguera-Toledo ME, Gonzalez de Mejia E, Amaya-Llano SL. Hard-to-cook bean (*Phaseolus vulgaris* L.) proteins hydrolyzed by alcalase and bromelain produced bioactive peptide fractions that inhibit targets of type-2 diabetes and oxidative stress. Food Res Int. 2015 Outubro;76, Part 3:839–51.

13. Safra 2015/16 - Quinto levantamento. Acomp Safra Bras Grãos. 2016 Feb;3(5):1–182.

14. Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, et al. A reference genome for common bean and genome-wide analysis of dual domestications. Nat Genet. 2014;46:707–13.

15. Valdisser PAMR, Jr GJP, Menezes IPP de, Müller BSF, Pereira WJ, Narciso MG, et al. SNP discovery in common bean by restriction-associated DNA (RAD) sequencing for genetic diversity and population structure analysis. Mol Genet Genomics. 2016 Mar 1;1–15.

16. Müller BSF, Jr GJP, Valdisser PAMR, Coelho GRC, Menezes IPP de, Abreu AG, et al. An Operational SNP Panel Integrated to SSR Marker for the Assessment of Genetic Diversity and Population Structure of the Common Bean. Plant Mol Biol Report. 2015 Mar 19;33(6):1697–711.

17. Lima RAZ, Tomé LM, Abreu CMP de, Lima RAZ, Tomé LM, Abreu CMP de. Vacuum package: its effect on the hardening and darkening of the beans during storage. Ciênc Rural. 2014 Sep;44(9):1664–70.

18. Siqueira BS, Pereira WJ, Batista KA, Oomah BD, Fernandes KF, Bassinello PZ. Influence of Storage on Darkening and Hardening of Slow- and Regular-Darkening Carioca Bean (*Phaseolus vulgaris* L.) Genotypes. J Agric Stud. 2014 Jul 26;2(2):87–104.

19. Ambachew D, Mekbib F, Asfaw A, Beebe SE, Blair MW. Trait associations in common bean genotypes grown under drought stress and field infestation by BSM bean fly. Crop J. 2015 Agosto;3(4):305–16.

20. Borges A, Melotto M, Tsai SM, Caldas DGG. Changes in spatial and temporal gene expression during incompatible interaction between common bean and anthracnose pathogen. J Plant Physiol. 2012;169:1216–20.

21. Budi M, Saboti J, Megli V, Kos J, Kidri M. Characterization of two novel subtilases from common bean (*Phaseolus vulgaris* L.) and their responses to drought. Plant Physiol Biochem. 2013;62:79–87.

22. Müller BSF, Sakamoto T, Silveira RDD, Zambussi-Carvalho PF, Pereira M, Papas-Júnior GJ, et al. Diferentially expressed genes during flowering and grain filling in common bean (*Phaseolus vulgaris*) growth under drought stress conditions. Plant Mol Biol Rep. 2014;32:438–51.

23. Ramegowda V, Senthil-Kumar M. The interactive effects of simultaneous biotic and abiotic stresses on plants: Mechanistic understanding from drought and pathogen combination. J Plant Physiol. 2015 Mar 15;176:47–54.

24. Ruiz-Nieto JE, Aguirre-Mancilla CL, Acosta-Gallegos JA, Raya-Perez JC, Piedra-Ibarra E, Vazquez-Medrano J, et al. Photosynthesis and chloroplast genes are involved in water-use efficiency in common bean. Plant Physiol Biochem. 2015;86:166–73.

25. Batista KA, Prudêncio SH, Fernandes KF. Wheat Bread Enrichment with Hard-to-Cook Bean Extruded Flours: Nutritional and Acceptance Evaluation. J Food Sci. 2011 Jan 1;76(1):S108–13.

26. Câmara CRS, Urrea CA, Schlegel V. Pinto Beans (*Phaseolus vulgaris* L.) as a Functional Food: Implications on Human Health. Agriculture. 2013 Feb 22;3(1):90–111.

27. Sagar S. Patel, Dipti B. Shah, Hetalkumar J. Panchal. De Novo RNA Seq Assembly and Annotation of *Phaseolus vulgaris* L. (SRR1283084). Genomics Appl Biol. 2014 Oct 22;5(5):1–6.

28. O'Rourke JA, Iniguez LP, Fu F, Bucciarelli B, Miller SS, Jackson SA, et al. An RNA-Seq based gene expression atlas of the common bean. BMC Genomics. 2014;15:866.

29. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009 Jan;10(1):57–63.

30. Heid CA, Stevens J, Livak K, Williams PM. Real time quantitative PCR. Genome Methods. 1996;6:986–94.

31. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, et al. The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. Clin Chem. 2009;55:611–22.

32. Hu Y, Chen H, Luo C, Dong L, Zhang S, He X, et al. Selection of reference genes for real-time quantitative PCR studies of kumquat in various tissues and under abiotic stress. Sci Hortic. 2014 Jul 22;174:207–16.

33. Navarro E, Serrano-Heras G, Castaño MJ, Solera J. Real-time PCR detection chemistry. Clin Chim Acta. 2015 Jan 15;439:231–50.

34. VanGuilder HD, Vrana KE, Freeman WM. Twenty-five years of quantitative PCR for gene expression analysis. BioTechniques. 2008 Apr;44(5):619–26.

35. Dheda K, Huggett JF, Bustin SA, Johnson MA, Rook G, Zumla A. Validation of housekeeping genes for normalizing RNA expression in real-time PCR. BioTechniques. 2004;37:112–9.

36. Dheda K, Huggett JF, Chang JS, Bustin SA, Johnson MA, Rook GAW, et al. The implications of using an inappropriate reference gene for rela-time reverse transcription PCR data normalization. Anal Biochem. 2005;344:141–3.

37. Yin G, Xu H, Liu J, Gao C, Sun J, Yan Y, et al. Screening and identification of soybean seed-specific genes by using integrated bioinformatics of digital differential display, microarray, and RNA-seq data. Gene. 2014;546:177–86.

38. Artico S, Nardeli SM, Brilhante O, Grossi-de-Sa MF, Alves-Ferreira M. Identification and evaluation of new reference genes in *Gossypium hirsutum* for accurate normalization of real-time quantitative RT-PCR data. BMC Plant Biol. 2010;10:1–12.

39. Hu R, Fan C, Li H, Zhang Q, Fu YF. Evaluation of putative reference genes for gene expression normalization in soybean by quantitative real-time RT-PCR. BMC Mol Biol. 2009;10.

40. Løvdal T, Lillo C. Reference gene selection for quantitative real-time PCR normalization in tomato subjected to nitrogen, cold, and light stress. Anal Biochem. 2009;287:238–42.

41. Manoli A, Sturaro A, Trevisan S, Quaggiotti S, Nonis A. Evaluation of candidate reference genes for qPCR in maize. J Plant Physiol. 2012;169(807-815).

42. Schmidt GW, Delaney SK. Stable internal reference genes for normalization of real-time RT-PCR in tobacco (*Nicotiana tabacum*) during development and abiotic stress. Mol Genet Genomics. 2010 Jan 23;283(3):233–41.

43. Hao Z, Hong Y, Xia Y, Singh VP, Hao F, Cheng H. Probabilistic drought characterization in the categorical form using ordinal regression. J Hydrol. 2016 Abril;535:331–9.

44. Vurukonda SSKP, Vardharajula S, Shrivastava M, SkZ A. Enhancement of drought stress tolerance in crops by plant growth promoting rhizobacteria. Microbiol Res. 2016 Mar;184:13–24.

45. FAO LAND & WATER. Drought [Internet]. 2013. [citado em Março de 2016], Disponível em: http://www.fao.org/docrep/017/aq191e/aq191e.pdf

46. Ravenscroft CH, Whitlock R, Fridley JD. Rapid genetic divergence in response to 15 years of simulated climate change. Glob Change Biol. 2015 Nov 1;21(11):4165–76.

47. Anderson JT, Panetta AM, Mitchell-Olds T. Evolutionary and Ecological Responses to Anthropogenic Climate Change Update on Anthropogenic Climate Change. Plant Physiol. 2012 Dec 1;160(4):1728–40.

48. Jump AS, Marchant R, Peñuelas J. Environmental change and the option value of genetic diversity. Trends Plant Sci. 2009 Jan 1;14(1):51–8.

49. Alemayehu FR, Bendevis MA, Jacobsen S-E. The Potential for Utilizing the Seed Crop Amaranth (*Amaranthus* spp.) in East Africa as an Alternative Crop to Support Food Security and Climate Change Mitigation. J Agron Crop Sci. 2015 Outubro;201(5):321–9.

50. Fisher M, Abate T, Lunduka RW, Asnake W, Alemayehu Y, Madulu RB. Drought tolerant maize for farmer adaptation to drought in sub-Saharan Africa: Determinants of adoption in eastern and southern Africa. Clim Change. 2015 Jul 4;133(2):283–99.

51. Fang Y, Xiong L. General mechanisms of drought response and their application in drought resistance improvement in plants. Cell Mol Life Sci. 2014 Oct 22;72(4):673–89.

52. Zhang Q. Strategies for developing Green Super Rice. Proc Natl Acad Sci. 2007 Oct 16;104(42):16402–9.

53. Ehlers JD, Hall AE. Cowpea (*Vigna unguiculata* L. Walp.). Field Crops Res. 1997 Jul;53(1–3):187–204.

54. Beebe S, Rao I, Blair M, Acosta J. Phenotyping common beans for adaptation to drought. Plant Physiol. 2013;4:35.

55. Lanna AC, Mitsuzono ST, Terra TGR, Vianello RP, Carvalho MA de F. Physiological characterization of common bean (*Phaseolus vulgaris* L.) genotypes, water-stress induced with contrasting response towards drought. Aust J Crop Sci. 2016 Jan;10:1–6.

56. Rosales MA, Cuellar-Ortiz SM, de la Paz Arrieta-Montiel M, Acosta-Gallegos J, Covarrubias AA. Physiological traits related to terminal drought resistance in common bean (*Phaseolus vulgaris* L.). J Sci Food Agric. 2013 Jan 30;93(2):324–31.

57. Rosales-Serna R, Kohashi-Shibata J, Acosta-Gallegos JA, Trejo-López C, Ortiz-Cereceres J, Kelly JD. Biomass distribution, maturity acceleration and yield in drought-stressed common bean cultivars. Field Crops Res. 2004 Fevereiro;85(2–3):203–11.

58. Franks SJ. Plasticity and evolution in drought avoidance and escape in the annual plant *Brassica rapa*. New Phytol. 2011 Apr 1;190(1):249–57.

59. Blair MW, Fernandez AC, Ishitani M, Moreta D, Seki M, Ayling S, et al. Construction and EST sequencing of full-length, drought stress cDNA libraries for common beans (*Phaseolus vulgaris* L.). BMC Plant Biol. 2011;11:171.

60. Recchia GH, Caldas DGG, Beraldo ALA, da Silva MJ, Tsai SM. Transcriptional Analysis of Drought-Induced Genes in the Roots of a Tolerant Genotype of the Common Bean (*Phaseolus vulgaris* L.). Int J Mol Sci. 2013 Mar 28;14(4):7155–79.

61. Wu J, Wang L, Li L, Wang S. De Novo Assembly of the Common Bean Transcriptome Using Short Reads for the Discovery of Drought-Responsive Genes. PLoS ONE. 2014 Outubro;9(10):e109262.

62. Miklas PN, Kelly JD, Beebe SE, Blair MW. Common bean breeding for resistance against biotic and abiotic stresses: From classical to MAS breeding. Euphytica. 2006 Jan;147(1-2):105–31.

63. Guimarães CM, Stone LF, Moreira JAA. Soil compaction in a bean crop. II: effect on root and shoot development. Rev Bras Eng Agríc E Ambient. 2002;6(2):213–8.

64. Terán H, Singh SP. Comparison of Sources and Lines Selected for Drought Resistance in Common Bean. Crop Sci. 2002 Jan;42(1):64–70.

65. Gonçalves RJ de S, Abreu Â de FB, Ramalho MAP, Bruzi AT. Strategies for recommendation of common bean lines tested for value of cultivation and use in different environments. Crop Breed Appl Biotechnol [Internet]. 2009 [citado em Março de 2016]; Disponível em: http://repositorio.ufla.br/jspui/handle/1/1293

66. Hewitt EJ. Mineral nutrition of plants in culture media. In: Plant Physiology. Academic Press; 1963. p. 97–134.

67. Martins PK, Jordão BQ, Yamanaka N, Farias JRB, Beneventi MA, Binneck E, et al. Differential gene expression and mitotic cell analysis of the drought tolerant soybean (*Glycine max* L. Merrill *Fabales*, *Fabaceae*) cultivar MG/BR46 (Conquista) under two water deficit induction systems. Genet Mol Biol. 2008;31(2):512–21.

68. Andrews S. FastQC: a quality control tool for high throughput sequence data [Internet]. 2010 [citado em Março de 2016]. Disponível em: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

69. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics. 2014 Apr 1;btu170.

70. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010 Maio;28(5):511–5.

71. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14(4):R36.

72. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 2012 Jan;40(Database issue):D1178–86.

73. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L. Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biol. 2011;12(3):R22.

74. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015 Jan 15;31(2):166–9.

75. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010 Jan 1;26(1):139–40.

76. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009 Dec 15;10:421.

77. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997 Sep 1;25(17):3389–402.

78. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-Seq: reference

generation and analysis with Trinity. Nat Protoc [Internet]. 2013 Aug [citado em Março de 2016]; 8(8). Disponível em: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3875132/

79. Conesa A, Götz S. Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. Int J Plant Genomics [Internet]. 2008 [citado em Março de 2016];2008. Disponível em: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2375974/

80. Nawrocki EP, Burge SW, Bateman A, Daub J, Eberhardt RY, Eddy SR, et al. Rfam 12.0: updates to the RNA families database. Nucleic Acids Res. 2015 Jan 28;43(D1):D130–7.

81. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2015 Oct 17;gkv1070.

82. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010 Sep 1;20(9):1297–303.

83. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013 Jan 1;29(1):15–21.

84. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011 May;43(5):491–8.

85. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinforma Ed Board Andreas Baxevanis Al. 2013 Oct 15;11(1110):11.10.1–11.10.33.

86. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly (Austin). 2012 Apr 1;6(2):80–92.