UNIVERSIDADE DE SÃO PAULO
CENTRO DE ENERGIA NUCLEAR NA AGRICULTURA

ENÉAS RICARDO KONZEN

Expression profiling and sequence diversity of novel *DREB* genes from common bean (*Phaseolus vulgaris* L.) and their association with drought-related traits

Piracicaba
2015

ENÉAS RICARDO KONZEN

Expression profiling and sequence diversity of novel *DREB* genes from common bean (*Phaseolus vulgaris* L.) and their association with drought-related traits

Tese apresentada ao Centro de Energia Nuclear na Agricultura da Universidade de São Paulo para a obtenção do título de Doutor em Ciências

Área de Concentração: Biologia na Agricultura e no Ambiente

Orientadora: Profa. Dra. Siu Mui Tsai

Co-orientador: Prof. Dr. Paul Gepts

Piracicaba

2015

Dados Internacionais de Catalogação na Publicação (CIP)

Seção Técnica de Biblioteca - CENA/USP

Konzen, Enéas Ricardo

Expression profiling and sequence diversity of novel DREB genes from common bean (*Phaseolus vulgaris* L.) and their association with drought-related traits / Enéas Ricardo Konzen; orientadora Siu Mui Tsai, co-orientador Paul Gepts. - - Piracicaba, 2015.

222 p.: il.

Tese (Doutorado – Programa de Pós-Graduação em Ciências. Área de Concentração: Biologia na Agricultura e no Ambiente) – Centro de Energia Nuclear na Agricultura da Universidade de São Paulo.

1. Estresse oxidativo 2. Genética de populações 3. Fenótipos 4. Mapeamento genético 5. Marcador molecular 6. Melhoramento genético vegetal 7. Reação em cadeia por polimerase 8. Seca I. Título

CDU (631.52 + 575.17) : 635.652

There is no greater inspiration I can find than with my mom and dad. They have taught me about the greatness of a simple life. If there is a reason for me being who I am… It's because of them.

To JOÃO and ELAINE KONZEN

"The important thing is to not stop questioning. Curiosity has its own reason for existence. One cannot help but be in awe when he contemplates the mysteries of eternity, of life, of the marvelous structure of reality. It is enough if one tries merely to comprehend a little of this mystery each day."

Albert Einstein

AKNOWLEDGMENTS

A special acknowledgment to Prof. Tsai Siu Mui, whose enthusiasm, advisory, innovative character, brilliancy and strength has encouraged me to go through all the steps of this project. A friend, a person willing to talk to you anytime… Much more to say, but let's keep it humble as she is! Thank you.

My gratitude to Prof. Paul Gepts for his valorous advice and encouragement during all this period, especially for the wonderful opportunity that I had while working at UC-Davis, in California.

To Prof. Beatriz Madalena Januzzi Mendes for all support during the first year.

My special thank you to Gustavo Henrique Recchia and Fernanda Cassieri. They were always outstanding, helping in every single experiment from the very beginning. *Nós sempre estávamos meio 'perdidos' com tantas coisas a serem feitas!!!!* And not only speaking of science, they are great friends that gave me motivation to go ahead on all the steps to be taken during these years. I really couldn't miss this space to describe how much I am thankful to you guys.

A very special acknowledgment to Jorge Carlos Berny and Viviana Medina, my California science and happy hour mates. I can't tell how much I've learned from you and how much fun we've been having together. From the LICOR afternoons in the hot Davis fields to the labeling lab nights…guys, thank you so much for being part of this thesis and my life. *Amaaaaaaaaaaaaaazing! Takatakatakataka!* I will always be visiting around, you know that for sure.

The financial support of FAPESP scholarship (process 2011-12.484-7) and from the Science Without Borders program from CNPq (process 249992/2013-2) were essential and without them this work would not have been accomplished. I extend this acknowledgment to CAPES, USDA, BeanCAP and California Bean Advisory Board

Thanks to all BIOMOL/CENA lab mates: Acacio, Aline, Andressa, Bia, Caio, Carol, Clóvis, Danielle, Dennis, Fabi, Felipe, Fernanda Nakamura, Jonathas, Juliane, Leandro, Lucas Mendes, Lucas Palma, Ludmila, Maju, Marília, Marina, Nayara, Rosineide.

Thanks to all California lab mates: Amanda, Andrea, Antonia, Arthur, Erin, Kay, Jamily, Karla, Laura, Manisha, Ninh, Paige, Regina, Sarah Dohle, Saarah Kuzay, Tamara, Travis, Vanessa. I had such a nice time and learned a lot from you guys.

To Mariana Bellotti and Prof. Victor Vitorello, for helping with photosynthesis measurements.

I could not forget my precious friends which have accompanied me through all this period: my housemates and great friends Wellington, Ezio, Nunes, Arnaldo, Juan and Henrique. My estimated friend Ana Beraldo. Two friends I have spent unique and funny moments: Franz and Naissa. To Jonathas, my colleague and great friend. To my estimated friends Prof. Maria Zucchi, Prof. José Baldin Pinheiro and Carolina Grando.

Thanks to all the friends from California. Baris and Dawit, thanks for all the laughing and happy moments. Raul, Stefano, Ziv, Angel, Jaime, Andrés, Syed… I will never forget the precious dinners and trips we had together has housemates. A special thank you to Stephanie Trasher for her hospitality, integrity and friendship. I appreciate every moment spent in California with all of you.

I could not forget the assistance and friendship of the technical staff from BIOMOL/CENA: Fabio, Wagner, Elias. A special acknowledgment to Chiquinho (*in memoriam*), from whom I learned a lot about beans and life. My acknowledgement extends to Steve, from UC-Davis, for all the help in setting up the greenhouse experiments, and to Najwa for all visa paperwork and general assistance.

I must also highlight all support given by CPG/CENA: Fábio, Neuda, Daiane, Sônia. Thank you to Marília for all support given to final formatting and editing of the thesis. Thank you to all people from CENA and CPG.

My family, my origin, my roots. A very special thank you to my mother, father and grandmother.

This last acknowledgment is addressed to anyone else that directly or indirectly contributed to my work and personal growth.

ABSTRACT

KONZEN, E. R. **Expression profiling and sequence diversity of novel DREB genes from common bean (*Phaseolus vulgaris* L.) and their association with drought-related traits**. 2015. 222 p. Tese (Doutorado) – Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, Piracicaba, 2015.

Common bean is a major dietary component in several countries, but its productivity is negatively affected by abiotic stresses. Dissecting candidate genes involved in abiotic stress tolerance is a paramount step toward the improvement of common bean performance under such constraints. Thereby, this thesis presents a systematic analysis of the *DEHYDRATION RESPONSIVE ELEMENT-BINDING* (*DREB*) gene subfamily, which encompasses genes that regulate several processes during stress responses, but with limited information for common bean. First, a series of *in silico* analyses with sequences retrieved from the *P. vulgaris* genome on Phytozome supported the categorization of 54 putative *PvDREB* genes distributed within six phylogenetic subgroups (A-1 to A-6), along the 11 chromosomes. Second, we cloned four novel *PvDREB* genes and determined their inducibility-factors, including the dehydration-, salinity- and cold-inducible genes *PvDREB1F* and *PvDREB5A*, and the dehydration- and cold-inducible genes *PvDREB2A* and *PvDREB6B*. Afterwards, nucleotide polymorphisms were searched through Sanger sequencing along those genes, revealing a high number of single nucleotide polymorphisms within *PvDREB6B* by the comparison of Mesoamerican and Andean genotypes. The nomenclature of *PvDREB6B* is discussed in details. Furthermore, we used the BARCBean6K_3 SNP platform to identify and genotype the closest SNP to each one of the 54 *PvDREB* genes. We selected *PvDREB6B* for a broader study encompassing a collection of wild common bean accessions of Mesoamerican origin. The population structure of the wild beans was accessed using sequence polymorphisms of *PvDREB6B*. The genetic clusters were partially associated with variation in latitude, altitude, precipitation and temperature throughout the areas such beans are distributed. With an emphasis on drought stress, an adapted tube-screening method in greenhouse conditions enabled the phenotyping of several drought-related traits in the wild collection. Interestingly, our data revealed a correlation between root depth, plant height and biomass and the environmental data of the location of the accessions. Correlation was also observed between the population structure determined through *PvDREB6B* and the environmental data. An association study combining data from the SNP array and *DREB* polymorphisms enabled the detection of SNP associated with drought-related traits through a compressed mixed linear model (CMLM) analysis. This thesis highlighted important features of *DREB* genes in common bean, revealing candidates for further strategies aimed at improvement of abiotic stress tolerance, with emphasis on drought tolerance.

Keywords: *DREB1*. *DREB2*. Expression profiling. SNP. Association mapping. Phenotyping. Drought tolerance.

RESUMO

KONZEN, E. R. **Expressão gênica e diversidade nucleotídica de novos genes *DREB* em feijoeiro (*Phaseolus vulgaris* L.) e sua associação com parâmetros de déficit hídrico.** 2015. 222 p. Tese (Doutorado) – Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, Piracicaba, 2015.

O feijoeiro é um componente essencial na dieta em diversos países, no entanto, sua produção é afetada negativamente por estresses abióticos. O estudo de genes candidatos envolvidos na adaptação aos estresses é uma etapa fundamental para o melhoramento da performance do feijoeiro sob tais estresses. Desse modo, esta tese apresenta uma análise sistemática da subfamília de genes *DEHYDRATION RESPONSIVE ELEMENT-BINDING* (*DREB*), que reúne genes envolvidos em diversos processos em resposta a estresses, mas pouco estudados no feijoeiro. Primeiramente, uma série de análises *in silico* com sequências de feijoeiro obtidas da plataforma Phytozome possibilitaram a categorização de 54 genes *PvDREB* putativos, distribuídos em seis subgrupos (A-1 até A-6) nos 11 cromossomos da espécie. Posteriormente, quatro novos genes *PvDREB* foram clonados e seus padrões de inducibilidade foram determinados. *PvDREB1F* e *PvDREB5A* foram induzidos por desidratação, baixa temperatura e salinidade, enquanto *PvDREB2A* e *PvDREB6B* foram predominantemente induzidos por desidratação e baixa temperatura. Polimorfismos de nucleotídeos foram buscados através de sequenciamento por método derivado de Sanger, revelando elevado número de SNP no gene *PvDREB6B*. A nomenclatura desse gene foi discutida detalhadamente ao longo da tese. A plataforma de marcadores SNP BARCBean6K_3 foi acessada para identificar o SNP mais próximo de cada um dos 54 *PvDREB*. O gene *PvDREB6B* foi selecionado para um estudo mais amplo, envolvendo uma coleção de acessos selvagens de origem Mesoamericana. A estrutura populacional destes genótipos foi analisada a partir de polimorfismos na sequência de *PvDREB6B*. Os grupos genéticos apresentaram associação parcial com variação da latitude, altitude, precipitação e temperatura das áreas em que os acessos naturalmente ocorrem. Com ênfase no estudo do déficit hídrico, uma plataforma de fenotipagem destes acessos em casa de vegetação, utilizando um sistema de tubos plásticos, foi elaborada para a análise de diversos parâmetros relacionados ao estresse por déficit hídrico. Os dados revelaram correlação entre profundidade de raízes, altura das plantas e a biomassa e as variáveis ambientais de cada local. A correlação também foi detectada entre a estrutura populacional estudada por *PvDREB6B* e os dados ambientais. Finalmente, um estudo de associação genética foi realizado entre os SNP da plataforma e ligados a *DREB* e os parâmetros fenotípicos, permitindo a identificação de marcadores SNP associados a caracteres específicos, usando um modelo linear misto (CMLM). Esta tese apresentou importantes aspectos sobre os genes *DREB* em feijoeiro, revelando candidatos para seu uso em estratégias de melhoramento para tolerância a estresses abióticos, com ênfase em déficit hídrico.

Palavras-chave: *DREB1*. *DREB2*. Expressão gênica. SNP. Mapeamento associativo. Fenotipagem. Tolerância ao déficit hídrico.

SUMMARY

1. INTRODUCTION

Common bean (*Phaseolus vulgaris* L.) is the grain legume with the highest volume of direct consumption in the world (BEEBE et al., 2014). It constitutes a source of protein (~22%), vitamins and mineral for human diets, which is important for developing countries (BEEBE, 2012). Brazil is the highest consumer worldwide and fits among the largest producers as well. Bean production in the 2014/2015 harvest season in the country was estimated at approximately 3.27 million tons in an area of 3.1 million hectares. The average productivity in the period was of 1,059 kg ha^{-1} (CONAB, 2015), which is considered low, given the potential yielding of this crop.

Among the factors affecting dry beans productivity are abiotic stresses (drought, salinity, low temperature, phosphorous and nitrogen deficiencies). Common bean requires approximately 200 to 400 mm of rainfall during growth and development. An estimate of 73% of the total production of beans in Latin America and 40% of the microclimates that occur in Africa have moderate to severe drought episodes during the growth season (BROUGHTON et al., 2003). Furthermore, arid or semiarid regions in Northeast Brazil are frequently affected by high soil salinity. In these areas, leaching and transport of soluble salts are restricted due to low rainfall and high evaporation rates, a process accelerated by poorly sized irrigation and drainage (GHEIY, 2000). South and part of Southeast Brazil and Argentina, USA, Uruguay among other areas have periods with very low temperatures, many times below 0ºC, which restricts its cultivation.

Nutritional deficiencies such as of phosphorus and nitrogen have a big impact on the cultivation of common bean (MIKLAS et al., 2006). About 60% of production in Latin America and Africa are affected by phosphorus deficiency. Estimates account for losses from 300 to 400 tons of beans over constraints such lacking of phosphorous and nitrogen along with water deficit (WORTMANN et al., 1998).

In a scenario of climate change, issues related to drought and other abiotic stresses have been increasingly debated. It draws the attention for the importance of studying the effects of such constraints to common bean by unraveling the biochemical and molecular signatures that influence plants from germination to maturity. Then, proper strategies might be designed toward the improvement of stress tolerance mechanisms.

Abiotic stresses affect over and belowground growth reducing cell expansion, stomatal conductance, photosynthesis and dry matter accumulation (HASANUZZAMAN et al., 2013). Regardless, several mechanisms enable plants to develop tolerance against stresses such as

drought, salinity and low temperature. Adaptation to stress has a quantitative genetic background, depending on various physiological and morphological characteristics (DUQUE; ALMEIDA; SILVA, 2013). Thus, genotypes differing in stress-tolerance are likely to have qualitative and quantitative variability for the expression of various genes. In general, the majority of studies concerned with abiotic stress responses have given focus on the shoot organs of plants. Nevertheless, more emphasis has recently been driven to the root system as well (RECCHIA et al., 2013; THAO et al., 2013). Classic breeding and genetic engineering have been moving forward to improve root traits of interest (JEONG et al., 2010; COMAS et al., 2013). In this thesis, we highlighted both root and shoot, and how gene expression and DNA polymorphisms are associated with abiotic stresses responses and adaptation.

Specifically at the gene level, *cis* and *trans* regulatory elements have important roles in the expression of stress-responsive genes. Yamaguchi-Shinozaki and Shinozaki (1994) identified a *cis*-regulatory element, DRE (Dehydration Responsive Element), in the promoter region of COR78/RD29A. It was found involved in the response to drought, high salinity and low temperature, further named as DREB transcription factor (DRE-Binding). These proteins are capable of binding DRE activating the expression of genes of the stress-signaling pathway. DREB transcription factors are unique to plant species, and so far, several genes have been described in *Arabidopsis* and other plants. Liu et al. (1998) isolated the two first genes of the group, the cold-inducible *DREB1A* and the drought-inducible *DREB2A*.

Each one of the DREB transcription factors contain an EREBP/AP2 conserved functional domain, constituted of approximately 60 amino acids. EREBP/AP2 was initially identified in APETALA2 protein, encoded by *APETALA2* (*AP2*) (OKAMURO et al., 1997). *AP2* plays an important role in controlling flowering and seed development in *Arabidopsis* (JOFUKU et al., 1994). The EREBP/AP2 domain was also found in proteins encoded by other regulatory genes in *Arabidopsis*, such as TINY (WILSON et al., 1996); DREBP (LIU et al., 1998) and the CBF family (MEDINA et al., 1999), and other genes involved in stress acclimation and retrograde signaling (DIETZ; VOGEL; VIEHHAUSER, 2010).

In *Glycine max*, the genes *GmDREBa*, *GmDREBb* and *GmDREBc* were the first isolated. *GmDREBa* and *GmDREBb* were induced by salinity, drought and low temperature in leaves of soybean seedlings. *GmDREBc* was slightly induced in roots by drought, salinity and abscisic acid (LI et al., 2005). Another gene, *GmDREB2*, was induced by drought, salinity, low temperature and abscisic acid (CHEN et al., 2007). *GmDREB3* was activated only by treatment with low temperature (CHEN et al., 2009).

In a broad panel, numberless efforts have been done to understand the characteristics and mechanisms of *DREB* genes in plants. However, few reports are available for *P. vulgaris*. The first full paper emphasizing common bean *DREB* loci was of Cortés et al. (2012), evaluating two genes and their nucleotide diversity patterns across wild and domesticated accessions. Although, a prior publication first cited these two genes along with several orthologs from other species (NAYAK et al., 2009). Regardless, no information about expression profiles and no whole-genome categorization of the *DREB* gene family are available for common bean.

This thesis takes part in a broader research line concerned with the identification of candidate genes for breeding programs aimed at drought tolerance improvement in common bean. Some *DREB* genes were isolated and characterized at the Laboratory of Cellular and Molecular Biology (CENA/USP). Genes were named *PvDREB2C, PvDREB5A, PvDREB5AB* and *PvDREB6A* and their expression profiles have been examined under drought, high salinity and low temperature (data not published). Further study has unraveled the molecular and functional basis of *PvDREB6A*, proving it a strong candidate for improving tolerance to abiotic stresses (ZAKIR-PEREIRA et al., submitted).

In this work, the first genome-wide categorization of the *DREB* gene family in common bean is proposed along with a series of *in silico* analyzes that give prospects for further studies for their dissection. Gene expression profiles were determined under dehydration, high salinity, low temperature and abscisic acid treatments in a temporal and spatial analysis. Moreover, this study was extended to five genotypes with contrasting levels of abiotic stresses tolerance. Cooperation has been established with Dr. Paul Gepts and his team from University of California Davis in order to investigate DNA variants along *DREB* genes, their diversity patterns across a set of common bean accessions and its association with phenotypic traits. An innovative root phenotyping experiment is described involving a collection of wild beans obtained from the Plant Introduction Station (USDA) and the Centro Internacional de Agricultura Tropical (CIAT). Association mapping analyses between *DREB*-associated SNP markers and the root traits are presented.

1.1 Hypotheses and objectives

This work was based on the following hypotheses:

1. *DREB* genes are differentially expressed under abiotic stresses in common bean, and their expression is time and tissue-dependent. *DREB1* genes are likely to be induced mainly by low temperatures while *DREB2* genes may have increased the accumulation of transcripts due water deficit and high salinity. Groups A-5 and A-6 might be induced under multiple abiotic stresses, including drought, salinity and low temperature;

2. The expression of *DREB* genes may be differentially modulated in distinct genotypes of common bean, which could be related to their level of tolerance to abiotic stresses;

3. The nucleotide variation along the promoter, intron, UTR or Open Reading Frame regions of *PvDREB* genes might have association with the geographic distribution of natural populations of *P. vulgaris* and adaptation mechanisms to environmental adversities;

4. Nucleotide variants within or nearby *PvDREB* genes could be associated with variation in drought tolerance or with specific traits. Variation for root deepening under drought might be a trait influenced by such genes.

The general objective of this study was to categorize the *DREB* gene family in common bean, investigate their expression profiles under selective treatments and the SNP variability of some representative genes as well as possible association between DNA variants and phenotypic traits under water deficit.

The specific objectives we wanted to accomplish are described as follows:

1. Identify and categorize all putative members of the *DREB* gene family in common bean through a genome-wide *in silico* search and provide a map with their chromosomal location;

2. Analyze the expression profile of four specific *PvDREB* transcripts under dehydration, high salinity, low temperature and abscisic acid treatment with temporal and spatial resolution in common bean genotypes contrasting for abiotic stress tolerance;

3. Identify DNA polymorphisms along the promoter, intron, Open Reading Frame and UTR regions of these four *PvDREB* genes and the closest SNP to each member of *DREB* based on a 6k SNP array;

4. Investigate the nucleotide diversity and structure of *PvDREB6B* and its association with geographic distribution of wild accessions of common bean;

5. Perform an association analysis between *DREB*-associated SNP markers and phenotypic variation of wild beans under drought-induced stress, testing general linear models (GLM) and mixed linear models (MLM).

1.2 Introdução

O feijoeiro-comum (*Phaseolus vulgaris* L.) é um dos grãos mais consumidos mundialmente (BEEBE et al., 2014). É importante fonte de proteínas (~22%), vitaminas e minerais para a dieta humana, aspecto de grande importância para países em desenvolvimento (BEEBE, 2012). O Brasil é o maior consumidor e figura entre os maiores produtores de feijão. A produção total do país na safra de 2014/2015 foi estimada em aproximadamente 3,27 milhões de toneladas em uma área de 3,1 milhões de hectares. A produtividade média estimada no período foi de 1.059 kg ha^{-1} (CONAB, 2015), considerada reduzida e aquém do potencial produtivo da cultura.

Dentre os fatores que afetam a produção de feijoeiro estão os estresses abióticos (déficit hídrico, estresse salino, baixa temperatura, deficiências de fósforo e nitrogênio). O feijoeiro necessita entre 200 e 400 mm de chuva durante seu crescimento e desenvolvimento. Estima-se que 73% da produção total de feijão da América Latina e 40% do total da África ocorram em microclimas que apresentam de moderado a severo déficit hídrico durante a estação de cultivo (BROUGHTON et al., 2003). Regiões de clima árido ou semi-árido, como em parte do Nordeste brasileiro, além do déficit hídrico, apresentam elevada salinidade. Nesses locais, a lixiviação e o transporte de sais solúveis são restritos devido às baixas precipitações e às altas taxas de evaporação, processo acelerado por irrigações mal dimensionadas e drenagem deficiente (GHEIY, 2000). O Sul e o Sudeste do Brasil, Argentina, Estados Unidos, Uruguai e outras regiões apresentam períodos do ano com temperaturas muito reduzidas, por vezes inferiores à 0ºC, prejudicando o desenvolvimento da cultura do feijoeiro.

Além disso, deficiências nutricionais como de fósforo ou nitrogênio são problemas frequentes para o cultivo do feijoeiro (MIKLAS et al., 2006). Cerca de 60% da produção na América Latina e África é afetada pela deficiência de fósforo. Na África, estima-se que a deficiência de fósforo, de nitrogênio e o déficit hídrico sejam responsáveis por perdas de 300 – 400 toneladas de feijão (WORTMANN et al., 1998).

Em um cenário de mudanças climáticas, problemas relacionados a secas e outros estresses abióticos têm sido intensamente debatidos. Desse modo, é extremamente importante o conhecimento dos efeitos dos estresses ambientais no feijoeiro, elucidando-os sob os enfoques bioquímico e molecular; e tendo em vista o delineamento de estratégias de melhoramento genético para tolerância a estes fatores.

Os estresses abióticos afetam o crescimento da parte aérea e da raiz das plantas, causando redução na expansão celular, na condutância estomática, na fotossíntese e, consequentemente, redução no acúmulo de matéria seca (HASANUZZAMAN et al., 2013). Apesar disto, existem vários mecanismos que permitem às plantas tolerar a falta de água, salinidade e baixa temperatura, sendo a adaptação a estas condições uma função multigênica, dependente de várias características fisiológicas e morfológicas (DUQUE; ALMEIDA; SILVA, 2013). Desse modo, genótipos que diferem na tolerância a esses estresses provavelmente apresentam variabilidade qualitativa e/ou quantitativa da expressão de diversos genes. No geral, têm sido enfatizados estudos com a parte aérea da planta. No entanto, recentemente, mais atenção tem sido dada ao sistema radicular (RECCHIA et al., 2013; THAO et al., 2013) e o melhoramento clássico e a engenharia genética têm sido voltados ao melhoramento de caracteres de interesse no sistema radicular (JEONG et al., 2010; COMAS et al., 2013). Nesta tese, os trabalhos foram dirigidos a ambas partes aérea e radicular, e como perfis de expressão gênica e polimorfismos de DNA podem estar associados a respostas a estresses abióticos.

Elementos regulatórios de ação *cis* e *trans* apresentam importantes papeis na expressão de genes responsivos a estresses abióticos. Yamaguchi-Shinozaki e Shinozaki (1994) identificaram um elemento regulatório *cis*, DRE (*Dehydration Responsive Element*), presente no promotor do gene COR78/RD29A e envolvido na resposta à seca, alta salinidade e baixa temperatura. Fatores de transcrição DREB (*Dehydration Responsive Element-Binding*) são capazes de se ligar a DRE para ativar a expressão de genes da via de sinalização a estresses. Vários fatores de transcrição DREB foram descritos em *Arabidopsis* (LIU et al., 1998). Liu et al. (1998) isolaram os dois primeiros genes do grupo, *DREB1A* e *DREB2A*, que codificam fatores de transcrição DREB. A expressão de *DREB1A* foi induzida por baixa temperatura, e a expressão de *DREB2A* foi induzida pelo déficit hídrico.

Cada um dos fatores de transcrição DREB contêm o domínio funcional conservado EREBP/AP2, composto de aproximadamente 60 aminoácidos, identificado inicialmente na proteína APETALA2, codificada pelo gene APETALA2 (AP2) (OKAMURO et al., 1997). O gene AP2 desempenha importante papel no controle do florescimento e no desenvolvimento da semente em *Arabidopsis* (JOFUKU et al., 1994). O domínio EREBP/AP2 também foi encontrado em proteínas codificadas por outros genes regulatórios em *Arabidopsis*, tais como TINY (WILSON et al., 1996); DREBPs (LIU et al., 1998) e a família CBF (MEDINA et al., 1999), envolvidos em aclimatação ao estresse e sinalização retrógrada (DIETZ; VOGEL; VIEHHAUSER, 2010).

Na família Fabaceae, genes *DREB* começaram a ser isolados em soja. Três homólogos de genes *DREB* foram isolados na espécie, *GmDREBa*, *GmDREBb* e *GmDREBc*. Os fatores de transcrição de *GmDREBa* e *GmDREBb* foram induzidos por salinidade, seca e baixa temperatura em folhas de plântulas de soja. Contrariamente, *GmDREBc* não foi afetado significativamente nas folhas, no entanto, foi aparentemente induzido nas raízes por déficit hídrico, salinidade e tratamento com ácido abscísico (LI et al., 2005). Outro gene, *GmDREB2*, também em soja, foi induzido por déficit hídrico, salinidade, baixa temperatura e ácido abscísico (CHEN et al., 2007). *GmDREB3* foi induzido apenas por tratamento com baixa temperatura (CHEN et al., 2009).

Em geral, diversos esforços têm sido realizados para o melhor entendimento das características e mecanismos dos genes *DREB* em plantas. No entanto, poucos estudos estão disponíveis para o feijoeiro. O primeiro artigo científico acerca de genes *DREB* em feijoeiro foi de Cortés et al. (2012), que considerou somente dois genes e seus padrões de diversidade nucleotídica em acessos selvagens e domesticados de feijoeiro. Uma publicação anterior primeiramente citou estes dois genes (NAYAK et al., 2009), mas sem detalhamentos sobre suas características. Além disso, não há informações disponíveis acerca de perfis de expressão e categorizações genômicas da subfamília *DREB*.

Esta tese integra os objetivos de uma linha de pesquisa concentrada na identificação de genes candidatos para programas de melhoramento objetivando o melhoramento do feijoeiro para tolerância à seca. Alguns genes *DREB* (*PvDREB2C*, *PvDREB5A* e *PvDREB5AB*, *PvDREB6A*) têm sido isolados e caracterizados no Laboratório de Biologia Celular e Molecular do CENA/USP, concentrando os estudos nos efeitos dos estresses por déficit hídrico, alta salinidade e baixa temperatura (CALDAS et al., dados não publicados). Outro estudo investigou a base molecular e funcional do gene *PvDREB6A*, indicando-o como um forte candidato para o aumento da tolerância a estresses abióticos (ZAKIR-PEREIRA et al., submetido).

Neste trabalho, a primeira categorização genômica da subfamília *DREB* foi realizada para o feijoeiro, mostrando-se diversas análises *in silico*. Perfis de expressão dos genes sob tratamentos de desidratação, alta salinidade, baixa temperatura e ácido abscísico foram determinados em escala temporal e espacial. Além disso, este estudo foi estendido a cinco genótipos contrastantes para tolerância à seca. Cooperação foi realizada com o Dr. Paul Gepts e sua equipe da Universidade da Califórnia Davis com o objetivo de investigar variantes de DNA ao longo de genes *DREB*, seus padrões de diversidade em acessos de feijoeiro e sua associação com caracteres fenotípicos. Um experimento inovador visando a fenotipagem de

sistema radicular sob déficit hídrico foi realizado com uma coleção de acessos selvagens obtidos do Plant Introduction Station (USDA) e do Centro Internacional de Agricultura Tropical (CIAT). Análises de mapeamento associativo entre marcadores SNP para os *DREB* e caracteres do sistema radicular são apresentadas.

1.2.1 Hipóteses e objetivos

Este trabalho foi baseado nas seguintes hipóteses:

1. Os genes *DREB* são diferencialmente expressos sob diferentes estresses abióticos em feijoeiro e sua expressão é provavelmente tecido e tempo-dependente. Os genes *DREB1* normalmente são induzidos sob baixas temperaturas enquanto os genes *DREB2* apresentam acumulação de transcritos devido a déficit hídrico e alta salinidade. Os integrantes dos grupos A-5 e A-6 podem ser induzidos sob múltiplos estresses;

2. A expressão dos genes *DREB* pode ser diferencialmente modulada entre genótipos de feijoeiro, o que pode estar relacionado ao nível de tolerância a estresses abióticos;

3. A variação nucleotídica ao longo do promotor, intron, UTR e região codante dos genes *DREB* pode ter associação com a distribuição geográfica e adaptação a adversidades ambientais de populações naturais de *P. vulgaris*;

4. Polimorfismos dentro ou próximos dos genes *PvDREB* podem estar associados com variação para tolerância à seca ou com caracteres específicos. A variação nas taxas de crescimento da raíz sob seca é uma característica que pode ser influenciada por estes genes;

O objetivo geral deste estudo foi categorizar a subfamília de genes *DREB* no feijoeiro, investigando perfis de expressão sob tratamentos seletivos e a variabilidade de marcadores SNP em alguns genes representantes, assim como possíveis associações entre os variantes de DNA e caracteres fenotípicos sob déficit hídrico.

Os objetivos específicos são descritos como segue:

1. Identificar e categorizar todos os membros potenciais da subfamília *DREB* em feijoeiro através de uma busca genômica *in silico*, apresentando um mapa de sua localização cromossômica;

2. Analisar o perfil de expressão de quatro transcritos *PvDREB* sob desidratação alta salinidade, baixa temperatura e ácido abscísico com resolução temporal e espacial em genótipos de feijoeiro contrastantes para tolerância a estresses abióticos;

3. Identificar polimorfismos de DNA dentro da região promotora, intron, *Open Reading Frame* e UTR dos quatro *PvDREB* e o SNP mais próximo de cada *DREB* baseado em uma plataforma de SNP de 6 kb;

4. Investigar a diversidade nucleotídica e estrutura do gene *PvDREB6B* e sua associação com distribuição geográfica em acessos selvagens de feijoeiro;

5. Realizar análise de associação entre marcadores SNP ligados a *DREB* e a variação fenotípica de acessos selvagens de feijoeiro sob déficit hídrico, testando modelos lineares gerais e mistos (GLM and MLM).

2. LITERATURE REVIEW

2.1 Genetic diversity of common bean: a broad panel

Beans belong to the Fabaceae family (Leguminosae, Papilionoidae) and genus *Phaseolus*. About 55 species of *Phaseolus* are described, but essentially five are cultivated: *P. vulgaris, P. acutifolius*, *P. lunatus*, *P. polyanthus* and *P. coccineus* (VIEIRA et al., 2005). *P. vulgaris is* naturally distributed in a wide area from northern Mexico to northeastern Argentina (GENTRY, 1969).

High morphological and allozyme diversity was found among wild populations of *P. vulgaris from* one to the other extreme of the geographical distribution of the species (KOENIG; GEPTS, 1989; SINGH; GEPTS; DEBOUCK, 1991). This variability is observed in different leaf shapes, growth habits, flower colors, but especially for seeds in terms of colors, shapes and sizes (SINGH; GEPTS; DEBOUCK, 1991). Further studies with DNA markers confirmed the high genetic diversity of common bean at the molecular level: RAPD (CARVALHO et al., 2008; KUMAR et al., 2014), microsatellites (BLAIR et al., 2006; BENCHIMOL et al., 2007; BURLE et al., 2010; PERSEGUINI et al., 2011; GIL-LANGARICA et al., 2011), AFLP (MACIEL et al., 2003; GIL-LANGARICA et al., 2011; PERSEGUINI et al., 2011), ISSR (GONZÁLEZ et al., 2005) and SNP (CORTÉS; CHAVARRO; BLAIR, 2011; GALEANO et al., 2012; BLAIR et al., 2013; SCHMUTZ et al., 2014). Furthermore, an investigation of the variability at the transcriptome level showed how domestication has shaped nucleotide and gene expression diversity (BELLUCCI et al., 2014).

Several of these studies recognized two major ecogeographical gene pools of wild beans: Mesoamerican and Andean. However, the geographic structure of the wilds reveals more complexity, with an additional third pool between Peru and Ecuador, characterized by a particular storage seed protein, phaseolin type I (DEBOUCK et al., 1993; KAMI et al., 1995). Further examinations showed wild populations from Colombia to be intermediates. A marked geographic structure has been found in populations from the Mesoamerican pool (GEPTS; BLISS, 1986; PAPA; GEPTS, 2003). Originally, the population of northern Peru and Ecuador was considered an ancestral population from which *P. vulgaris* originated. From this core location, beans probably were spread north and south, resulting in the Mesoamerican and Andean pools, respectively (KAMI et al., 1995; FREYRE et al., 1996).

Conversely, there is an alternative hypothesis, which considers that ancestral beans were distributed through Mesoamerica (DELGADO-SALINAS; BIBLER; LAVIN, 2006). The

high genetic diversity encountered within these gene pools has been used to support this hypothesis. Furthermore, the Mesoamerican origin of the common bean has been suggested based on an analysis of sequence data of five small gene fragments (BITOCCHI et al., 2012). A whole-genome comparison of pooled 30 individuals from each Mesoamerican and Andean wild populations showed high genetic differentiation among gene pools. A demographic inference for the Andean gene pools suggested it was derived from a Mesoamerican population with only a few thousands of individuals (SCHMUTZ et al., 2014). However, the debate on the origin of the species remains and more studies are on their way to solve the issue of deciphering the origin of the species.

The domestication process of *P. vulgaris* has been another matter of debate and extensive molecular studies. Initially, morphological and enzyme analyses showed the existence of two major centers of bean domestication: Mesoamerica and Andean, encompassing six races (Jalisco, Durango and Mesoamerica in the Mesoamerican gene pool; and Chile, Peru and Nueva Granada in the Andean gene pool) (SINGH; GEPTS; DEBOUCK, 1991). There are indications that nearly 8,000 years ago common bean was independently domesticated in Mexico and South America (MAMIDI et al., 2011; BITOCCHI et al., 2012; BITOCCHI et al., 2013). Domestication was followed by local adaptation resulting in landraces with different characteristics (SCHMUTZ et al., 2014). However, many questions remain on this matter. The recent application of genomic, transcriptomic, proteomic and metabolomic research is providing data to address several remaining questions about domestication (GEPTS, 2014).

In Brazil, common bean was probably introduced from northern South America, with little influence from the Andes, as some archaeological and genetic data indicate (FREITAS, 2006; BURLE et al., 2010). An intermediate level of genetic diversity was found within common landraces from Brazil when compared to the overall diversity of the species (BURLE et al., 2010). A significant morphological variability has been verified as well, although not all morphological variability observed for *P. vulgaris* (BURLE et al., 2011). The study of geo-referenced landraces from Andean and Mesoamerican background revealed the latter was four times more frequent than the former (BURLE et al., 2011).

2.2 Nucleotide diversity analyses of single genes: placing *DREB* in the context

Besides molecular fingerprinting and SNP analyzes at the genome level, several studies have addressed evolutionary and domestication processes of common bean based on polymorphisms within individual genes. One of the first studies on gene sequencing was of Kami and Gepts (1994), which evaluated the sequence diversity of the T-type phaseolin. It led to the further identification of the ancestral DNA sequences coding for this protein (KAMI et al., 1995). McClean, Lee and Miklas (2004) studied the nucleotide diversity patterns within intron 1 of the gene encoding for Dihydroflavonol reductase (DFR), important for the synthesis of anthocyanins. As introns are under less selection pressure than coding sequences and thus accumulate more mutations, high nucleotide diversity was detected, generating two major clusters (Mesoamerican and Andean gene pools), subdivided into races.

Nucleotide diversity analysis within *SHATTERPROOF* (*PvSHP1*) revealed high levels of polymorphism, useful toward studying domestication of common bean. The authors confirmed the gene pool structure traditionally observed for common bean and showed a dual domestication process, derived from each gene pool (NANNI et al., 2011).

In another study, polymorphisms were investigated within *PvTFL1y*, a gene responsible for the natural variation for determinacy (growth habit) in common bean (REPINSKI; KWAK; GEPTS, 2012). The analysis of *PvTFL1y* nucleotide variation showed that indeterminate types (wild and domesticated) presented only synonymous nucleotide substitutions. On the other hand, determinate types (only among domesticated accessions) showed both synonymous and son-synonymous mutations, as well as INDEL, putative intron-splicing failures, retrotransposon insertions or even deletion of the entire locus (KWAK et al., 2012). Thereby, polymorphisms within this gene were able to be linked to the variation of a trait of interest in common bean.

Another study revealed the nucleotide patterns of *PvIND*, homolog to the *Arabidopsis INDEHISCENT*. Andean and Mesoamerican genotypes were separated by structure analysis, but some admixture was encountered; and domesticated and wilds were mixed within each subgroup (GIOIA et al., 2013). The authors tried to associate the polymorphisms of this gene to pod shattering. However, the lack of correlation between polymorphisms and phenotypes suggested it may not be directly involved in shattering (GIOIA et al., 2013).

Considering the potential application of the discovery of gene-specific variants and their association with phenotypic variation, the present thesis gives prospects for the analysis of variation of the dehydration-responsive genes such as *DREB* and their effect on traits affected

by major abiotic stresses inducers. Limited literature is available about *DREB* genes in common bean. Nayak et al. (2009) analyzed partial sequences of two genes, *DREB2A* and *DREB2B* in three species of cereals and two legumes, among them *P. vulgaris*. The authors evaluated eight bean cultivars and detected eight SNPs for *DREB2A*, belonging to two haplotypes, however, no SNP was found within *DREB2B*. In *DREB2A*, one of the haplotypes was found in the Mesoamerican pool and the other in the Andean. The nomenclature of these genes is systematically discussed on Section 3 of this thesis.

Nayak et al. (2009) mentioned that Andean genotypes in general exhibit higher susceptibility to drought stress than the Mesoamericans. Beans from the commercial classes Dark, Red Kidney and White (Andean) showed to be the most intolerant in a study conducted in the western United States (SINGH et al., 2001). In contrast, cultivars belonging to race Durango (Mesoamerican) presented higher tolerance (TERAN; SINGH, 2002). Then, the dissection of candidate genes associated with abiotic stresses such as those caused by water deficit seems important to understand how they shape tolerance mechanisms.

The first full paper involving SNP markers associated with *DREB* genes in common bean was developed by Cortés et al. (2012), which isolated two *DREB* and analyzed patterns of nucleotide diversity in a collection of wild and domesticated beans from Andean and Mesoamerican origin. Structure analysis revealed clear differences among Andean and Mesoamerican genotypes, but no evident association was found with phenotypic traits evaluated under drought stress. It was the first report aimed at an association study between *DREB* polymorphisms and phenotypic traits in common bean. However, a partial sequence of each gene was analyzed since the genome of the species was not yet available at the time.

2.3 Effects of abiotic stresses, role of abscisic acid and stress-responsive genes categories

Plants have the natural ability to adjust to circadian rhythms and various environmental factors. However, environmental changes cause various types of stress during their growth and development. Water restrictions lead plants to water deficit. Conversely, an excess of water by rain or irrigation leads to waterlogging or flooding. Moreover, plants growing in saline soils may show symptoms of stress. In a similar manner, low temperature also affects different cultures (GROVER et al., 2001).

Most of the important agricultural crops are in the process of selection for tolerance to abiotic stresses. Nevertheless, adaptability of such plants to environmental adversities is still limited (GROVER et al., 2001). In case of common bean, every season abiotic stresses cause

major production losses in Brazil (see CONAB reports from 2012; 2013; 2014; 2015 for checking oscillation in productivity in Brazil) and other countries. The recent discussions over global warming have raised increasing concerns about genetic improvement for abiotic stresses tolerance. Several areas throughout the world are likely to suffer dramatic changes in average temperatures and rainfall, so then the development of tolerant bean cultivars is essential to withstand such new dynamics.

Abiotic stresses negatively affect common bean survival, production and accumulation of biomass and grain yield. At the cellular level, stresses decrease the availability of water, important to vital functions and maintenance of turgor pressure (SHINOZAKI; YAMAGUCHI-SHINOZAKI, 1997). Drought stress leads to stomatal closure and consequently a reduced ability to assimilate carbon from the air (CHAVES et al., 2002). Stress limits photosynthetic capacity and plant growth. Furthermore, abiotic stresses lead to the production of reactive oxygen species (ROS) in chloroplasts, mitochondria and peroxisomes, causing irreversible histological and cell damage (BRAY et al., 2000).

Nevertheless, plants have developed various mechanisms of adaptation and detoxification in stress episodes, which involve complex gene networks and crosstalk among molecular pathways (BASU, 2012). Significant advances were made toward the understanding of activation mechanisms of signal transduction pathways due abiotic stresses in plants as well as molecular and physiological responses (CRAMER et al., 2011). One of the effects is the increased synthesis of abscisic acid, which is important in mechanisms of stress tolerance (SHINOZAKI; YAMAGUCHI-SHINOZAKI, 1997; HUBBARD et al., 2010; LATA; PRASAD, 2011). The increase in ABA levels in the plant depends on various environmental factors and is known to be a physiological signal that activates stress responses (HUBBARD et al., 2010; LATA; PRASAD, 2011). Therefore, ABA is involved in the adaptation to several stress-inducible factors such as genes for osmotic adjustment, ion compartmentalization, hydraulic conductivity in roots, regulation of shoot and root growth and control of transpiration rate (VERSLUES; ZHU, 2005).

Genes induced by abiotic stress are usually categorized into two groups: those that directly protect from environmental stresses and others that regulate gene expression and signal transduction during stress response (SHINOZAKI; YAMAGUCHI-SHINOZAKI, 1997; LATA; PRASAD, 2011). The first group comprises proteins that function in protecting cells against dehydration, such as enzymes required for the production of various osmolytes, LEA proteins (late embryogenesis abundant), antifreeze proteins, chaperones and detoxification enzymes (SHINOZAKI; YAMAGUCHI-SHINOZAKI, 1997; INGRAM;

BARTELS, 1996; BRAY, 1997). In the second group are included transcription factors, protein kinases and enzymes involved in phosphoinositide metabolism (SHINOZAKI; YAMAGUCHI-SHINOZAKI, 1997).

2.4 *Cis* regulatory elements, ABA influence on genes and DREB transcription factors

Analyses of expression profiles of stress-responsive genes suggest that transcriptional modulation is one of the most important steps toward the adaptation of plants to environmental conditions (ZHU, 2002; NAKASHIMA et al., 2006). Several stress-inducible genes have been characterized in *Arabidopsis* and are indicators of the transmission and perception of signals in plant cells. Analysis of promoters showed that several types of *cis*-regulatory elements have a key role in controlling gene expression under stress conditions (SHINOZAKI; YAMAGUCHI-SHINOZAKI, 2007).

Using *Arabidopsis* as reference, studies have shown that at least four independent signaling pathways operate in the induction of genes responsive to abiotic stress, in case of dehydration (SHINOZAKI; YAMAGUCHI-SHINOZAKI, 2000). Among the pathways, some are dependent on abscisic acid and the others are independent. As previously stated, ABA plays a role in the induction of several stress-related genes (VERSLUES; ZHU, 2005).

On the other hand, some genes are not ABA-responsive. DRE/CRT is a *cis*-element that has been described independent of ABA signaling (NAKASHIMA; ITO; YAMAGUCHI-SHINOZAKI, 2009). DRE was originally discovered by Yamaguchi-Shinozaki and Shinozaki (1994) while studying two genes involved in abiotic stress responses, *rd29A* and *rd29B*. Within *rd29A* two *cis*-elements were detected, one mediated by ABA and the other induced by changes in osmotic potential due water deficit or high salinity. Promoter analysis led to the identification of the DRE (dehydration responsive element), containing 9 bp (TACCGACAT), involved in the first response to the stress-inducing factor. Proteins bound specifically to DRE were isolated and referred as DREB (DRE-Binding).

Afterward, two clones coding proteins bound to DRE were isolated, *AtDREB1A* and *AtDREB2A*. *DREB1A* was responsive to low temperature while *DREB2A* was predominantly induced by drought stress. So far, both genes showed to be ABA-independent (LIU et al., 1998).

Nevertheless, new homologs have been isolated from *Arabidopsis* and other plants such as rice, soybean and *Medicago*. For *Arabidopsis*, Sakuma et al. (2002) showed *DREB2A, DREB2C* and *DREB2E* were also induced by ABA, leading to changes in the models

proposed. Hereafter, new studies showed how ABA could mediate the expression of some *DREB* genes. Even though, some references continued to point *DREB* activity as independent from ABA (NAKASHIMA; ITO; YAMAGUCHI-SHINOZAKI, 2009), since this feature was observed in *Arabidopsis*. As other genes have been isolated in a many other species, though, ABA-responsive models have been shown. Regulatory elements DRE/CRT and ABRE, the latter involved in ABA response, were proved to operate concomitantly in some genes (NARUSAKA et al., 2003; ZHAO et al., 2010), therefore indicating a cross-talk between ABA-dependent and ABA-independent paths (ROYCHOUDHURY; PAUL; BASU, 2013). Furthermore, a compilation described by Lata and Prasad (2011) shows that the signal perception might be mediated by ABA or be ABA-independent for *DREB* genes.

2.5 What features define *DREB* genes and make them different from others?

The primary condition to refer a gene as a *DREB* is the presence of a highly conserved protein domain, EREBP/AP2. It was discovered within *APETALA2*, which plays an important role in flowering and seed development in *Arabidopsis* (JOFUKU et al. 1994). Several proteins have been discovered to contain this domain along their amino acid chain, consisting of a repeated motif of approximately 60 amino acids (JOFUKU et al., 1994; OKAMURO et al., 1997; SAKUMA et al., 2002). The EREBP/AP2 domain is mostly encountered in a single position along the proteins. However, some proteins have more than one copy (NAKANO et al., 2006).

EREBP/AP2 genes comprise a large superfamily, which has been divided into three families referred as AP2, ERF, and RAV, considering their sequence similarity and the number of EREBP/AP2 domains (NAKANO et al., 2006). AP2 proteins contain two EREBP/AP2 domains and genes of this family participate in the regulation of developmental processes (ELLIOTT et al., 1996, CHUCK; MEELEY; HAKE, 1998; BOUTILIER et al., 2002). RAV proteins contain an EREBP/AP2 and a B3 domain and are involved in the response to ethylene (ALONSO et al., 2003), brassinosteroids (HU et al., 2004), biotic and abiotic stresses (SOHN et al., 2006).

The ERF protein family contains only one EREBP/AP2 domain and is subdivided into two subfamilies, CBF/DREB and ERF (SAKUMA et al., 2002). The amino acids 14 and 19 of the EREBP/AP2 domain distinguish DREB (valine and glutamic acid, respectively) from ERF (alanine and aspartic acid, respectively) (SAKUMA et al., 2002). Besides, ERF genes are primarily involved in responses to biotic stresses such as pathogenesis by

recognizing the *cis*-regulatory element AGCCGCC, known as GCC box (HAO et al., 1998). In contrast, *DREB* genes have crucial role in the response of plants to abiotic stress by recognizing the Dehydration Responsive Element (DRE), which has the motif A/GCCGAC (YAMAGUCHI-SHINOZAKI; SHINOZAKI, 1994). DRE was found essential for the regulation of gene expression due to dehydration (YAMAGUCHI-SHINOZAKI; SHINOZAKI, 1994) and is encountered in the promoter region of other genes induced by several abiotic stresses (LATA; PRASAD, 2011; AKHTAR et al., 2012).

2.6 *DREB* genes are divided into six subgroups

DREB genes have been divided into six subgroups (A-1 - A-6). This categorization has been based on phylogenetic analyzes as well as particular features related to their induction. Genes *DREB1/CBF* belong to subgroup A-1 and have been characterized as induced by low-temperature (LIU et al., 1998). A-1 genes are frequently referred as *CBF* (C-repeat binding factors) along with *DREB*. In *Arabidopsis*, at least six genes are characterized within this subgroup. Among them, *AtDREB1A/CBF3*, *AtDREB1B/CBF1* and *AtDREB1C/CBF2* are rapidly induced by low temperature. Plants overexpressing these genes showed increased tolerance to cold, drought and high salinity (LIU et al., 1998; GILMOUR et al., 1998; JAGLO-OTTOSEN et al., 1998). Comparisons of gene expression levels among different ecotypes of *Arabidopsis* revealed a significant correlation between expression levels and cold tolerance (HANNAH et al., 2006). However, other three genes in A-1, *DREB1D/CBF4*, *DREB1E/DWARF AND DELAYED FLOWERING* (*DDF2)* and *DREB1F/DDF1*, have different inducibilities. *DREB1D* is dehydration and ABA-inducible, and *DREB1F* is salinity-inducible (SAKUMA et al., 2002; HAAKE et al., 2002; MAGOME et al., 2004). Therefore, genes within A-1 play important roles in other stress responses than just freezing.

DREB2 genes (subgroup A-2) are primarily involved in responses to osmotic stress (dehydration and salinity) (SAKUMA et al. 2002; NAKASHIMA; YAMAGUCHI-SHINOZAKI, 2006). *AtDREB2A* and *AtDREB2B* are induced by water deficit, heating and high salinity, while *AtDREB2C* expresses more under heating, but later than the formers (LIM et al., 2007).

Although genes from subgroups A-1 and A-2 are the primary representatives of *DREB* genes and are involved in crucial responses due to abiotic stresses, new studies have been performed with genes from groups A-3 to A-6. These transcripts are involved in stress signaling and have potential cross-talk with alternative pathways, thus, sharing functions in

other cellular processes beyond stress responses (MIZOI; SHINOZAKI; YAMAGUCHI-SHINOZAKI, 2012).

The A-3 subgroup comprises *ABI4* and orthologues. Although similar to *DREB2* genes, according to phylogenetic analyses (NAKANO et al., 2006), they have particular functions in ABA, sugar and retrograde signaling (ARENAS-HUERTERO et al., 2000; NIU; HELENTJARIS; BATE, 2002).

Phylogenetic analyzes show the A-4 subgroup is related to A-1 (SAKUMA et al., 2002; LATA; PRASAD, 2011). They share a common protein motif and could be included in only one group. However, their function is still unclear. In general, they do not have considerable stress-inducibility (KARABA et al., 2007).

Subgroup A-5 consists of stress-inducible genes. These genes might have a role as negative regulators of *DREB1* and *DREB2* genes. When overexpressed, *DREB1* and *DREB2* were repressed under low temperature and water deficit (TSUTSUI et al., 2009).

Finally, A-6 subgroup comprises stress-inducible genes as well. *RAP2.4* and *RAP2.4B* are induced in response to dehydration, high salinity and heating (LIN; PARK; WHANG, 2008; RAE; LAO; KAVANAGH, 2011). Microarray analysis showed these genes regulate several aquaporins, suggesting a functional role in plant hydric homeostasis (RAE; LAO; KAVANAGH, 2011). Thereby, A-6 genes are involved in stress responses, but regulate a different set of genes than those controlled by *DREB1* or *DREB2* (MIZOI; SHINOZAKI; YAMAGUCHI-SHINOZAKI, 2012).

2.7 Isolation of *DREB* genes in other species than *Arabidopsis*

DREB loci have been isolated and characterized in *Arabidopsis*, rice, soybeans and many other plants (LIU et al., 1998; DUBOUZET et al., 2003; CHEN et al., 2007). Soybean has been targeted for many of such studies. Chen et al. (2007) isolated *GmDREB*2, an A-5 member inducible under drought, high salinity, low temperature and abscisic acid treatment. Transgenic *Arabidopsis* lines overexpressing this gene showed increased drought and salinity tolerance and no growth retardation was observed in plants. *GmDREB3* was only induced by low temperature and its overexpression in transgenic plants increased not only cold but also drought and high salinity tolerance (CHEN et al. 2009a). Other species are described in Table 2.1.

Table 2.1 – Examples of *DREB* genes isolated in the past years in plants, including model species. Common bean is included, however, no expression data are available since only nucleotide diversity analysis was performed

Gene	Species	Factor of inducibility	Reference
MtDREB2A	*Medicago truncatula*	High salinity and water deficit in roots	Chen et al. (2009b)
GhDREB	*Gossypium hirsutum*	High salinity, water deficit and low temperature	Gao et al. (2009)
SbDREB2A	*Salicornia brachiata*	High salinity, water deficit and high temperature	Gupta et al. (2010)
PgDREB2A	*Penissetum glaucum*	Osmotic stress	Agarwal et al. (2010)
OsDREB2B	*Oryza sativa*	High and low-temperature	Matsukura et al. (2010)
SbDREB2	*Sorgum bicolor*	Water deficit	Bihani et al. (2011)
SiDREB2	*Setaria itálica*	High salinity e water deficit	Lata et al. (2011)
MtCBF4	*Medicago truncatula*	High salinity, water deficit, low-temperature and ABA	Li et al. (2011)
PvDREB2A	*Phaseolus vulgaris*	No expression analysis	Cortés et al. (2012)
PvDREB2B	*Phaseolus vulgaris*	No expression analysis	Cortés et al. (2012)
PeDREB2a	*Populus euphratica*	Water deficit, low-temperature, high salinity, gibberellic acid, 1-naftalene acetic acid and 6-benzil amino purine	Zhou et al. (2012)
MsDREB2C	*Malus sieversii*	Water deficit, high salinity, low and high temperature and ABA	Zhao et al. (2013)
BdDREB2	*Buchloe dactyloides*	High salinity e water deficit	Zhang et al. (2014)
PnDREB68	*Populus nigra*	Low-temperature and water deficit in leaves; low temperature in stem and leaves	Chu et al. (2014)
PnDREB69	*Populus nigra*	Low-temperature in roots, stem and leaves	Chu et al. (2014)

In another Fabaceae member, *Medicago truncatula*, *MtDREB1C* gene was isolated and an east two-hybrid assay showed it binds specifically to the DRE motif and activates the expression of reporter genes *HIS3* and *LacZ*. Overexpressing the gene in transgenic *Medicago* and *Rosa chinensis*, it was found to increase tolerance to low-temperature stress (CHEN et al., 2010).

2.8 Sources of drought tolerance and their selection in common bean

Although other abiotic stresses are studied in this thesis, all chapters emphasize drought stress and its effects and association with *DREB* genes. Drought is one of the main issues for common bean cultivation since an estimated 60% of its production is at risk of water deficit. Many breeding programs have been enrolled in selection for drought tolerance of *P. vulgaris*, especially in Mexico, Honduras, Brazil, and at Centro Internacional de Agricultura Tropical (CIAT) in Colombia (BEEBE et al., 2013).

Recently, several drought nurseries have been performed in Latin America and Africa, focused on germplasm evaluations. Two sets of germplasm have been tested for seed yielding in drought areas of Mexico under insufficient and erratic rainfall conditions. One set constituted of more than 7,000 accessions from INIFAP (Instituto Nacional de Investigaciones Forestales, Agricolas y Pecuarias) grown over three years at two locations. Durango and Mesoamerica races exhibited the highest drought tolerance, whereas those from Jalisco race were drought-sensitive. The other set included 800 genotypes from CIAT, also revealing Durango race as the most drought-tolerant (BEEBE et al., 2013).

Some very specific cultivars have been released and tested as drought-tolerant. A number of studies have showed BAT 477 is considerably adapted to drought conditions (SPONCHIADO et al., 1989; WHITE et al., 1994) and it was the drought-tolerant parent for the mapping population derived from the cross BAT 477 x DOR 364, the latter susceptible to drought (BLAIR et al., 2012). Alternatively, the work of Singh (1995) showed a specific line, SEA-5, to be highly drought-tolerant. Moreover, cultivars Pinto Villa and Pinto Saltillo were released as drought-tolerant varieties for the northern highlands of Mexico (ACOSTA-GALLEGOS et al., 1995; SÁNCHEZ-VALDEZ et al., 2004).

In Brazil, a drought-tolerant Carioca bean has been released by the Instituto Agronômico de Campinas (IAC) with an average yield of almost 2,300 kg ha^{-1} but with potential up to 4,600 kg ha^{-1}, considering 17 environments evaluated (CHIORATO et al., 2012). Furthermore, a study by Molina et al. (2001) showed the cultivar IPR88 Uirapuru and

two Carioca lines, LP 97-13 and LP 97-4 have good performances under drought as well. The country also holds a huge variety of landraces that have been traditionally cultivated by small farmers over years (BURLE et al., 2010). There are indications that landraces from the group Mulatinho, in northern Minas Gerais state, have considerable drought tolerance compared with other groups but more studies need to be performed toward its understanding (BURLE et al., 2011).

CIAT, in Colombia, has been developing new breeding lines with higher drought tolerance in small red, small black, cream and cream-striped classes. Some new cultivars presented up to 36% more yield and some presented better yield in a phosphorous deficient environment (BEEBE et al., 2008).

Some traits have been used to indicate whether bean lines are more drought-tolerant or drought-sensitive. Among a collection of breeding lines evaluated in Africa, those presenting higher pod harvest index (PHI) were more drought-tolerant than the ones with reduced values, suggesting the importance of remobilization of photosynthates from pod wall to seed (ASSEFA et al., 2013).

The last efforts in breeding for drought tolerance have also focused on the identification of sources of drought tolerance in wild beans spanning the natural area of distribution of *P. vulgaris* (CORTÉS et al., 2013). However, reliable estimations of drought tolerance in wild beans are not easy to establish, and attempts toward the development of new methods have been done. Potential evapotranspiration models coupled with precipitation regimes were used to define a drought index for a series of wild bean accessions. Considering this factor along with the population structure might be a useful tool to analyze the levels of drought tolerance and use these materials for introgression of alleles of interest (CORTÉS et al., 2013).

A suitable alternative to improve drought tolerance in common bean relies on interspecific crosses with other species of *Phaseolus*, such as tepary beans (*P. acutifolius*). They natural span from the desert highlands of northwest Mexico to the southwest of the USA and thus they are convenient sources of drought, heat and cold tolerance (MARTINEZ-ROJO et al., 2007). An interesting feature of tepary beans is on their root system. It reveals extremely fine roots with rapid penetration in the soil with profuse branching, which enables quick access to limited soil water (BUTARE et al., 2011).

2.9 Genetic mapping of common bean and QTL: from basic traits to drought tolerance

Several genetic maps have been constructed for common bean for traits of economic importance. The first genetic maps in the early 1990's were combined in a consensus genetic map of common bean (FREYRE et al., 1998), compiled from various RAPD, RFLP, isozyme and morphological markers. The consensus map was constructed from the population of RILs from the cross BAT-93 x Jalo EEP558 (FREYRE et al., 1998). Subsequently, the map has received more marker positions such with microsatellites, AFLP, SCAR, STS, ISSR (YU et al., 2000; McCLEAN et al., 2002; BLAIR et al., 2003; PEREZ-VEGA et al., 2010). Furthermore, microsatellites derived from EST (Expressed Sequence Tags), RGA (Resistant Gene Analogs) and AFLP were included in the consensus map of common bean (HANAI et al., 2010).

Other maps have been developed for specific traits or for evaluating how domestication has shaped genes. Koinange et al. (1996) genotyped a F8 population derived from the crossing Midas x G12873, to map QTLs associated with the 'domestication syndrome' of *P. vulgaris*. Midas is an Andean cultivar and G12873 is a wild genotype. Therefore, the two genotypes contrast in several characteristics, especially those that have been selected in the process of domestication. Other genetic maps were constructed to find markers linked to pathogen resistance such as anthracnose, bacterial blight and rust (ADAM-BLONDON et al., 1994; JUNG et al., 1996; TAR'AN et al., 2002) and nodulation (TSAI et al., 1998).

In Brazil, a genetic map was developed using a F_{10} population originated from the cross IAC-UNA x CAL 143. This genetic map includes microsatellite markers (gene and genomic) and three morphological markers (CAMPOS et al., 2010).

Despite the availability of several genetic maps for common bean, few studies have been done so far to map loci for tolerance to abiotic stresses. An initial publication on the matter was of Schneider et al. (1997), which studied QTL associated with drought tolerance in common bean. The authors genotyped two populations of RILs (Sierra x AC1028, Sierra x Lef-2RB, with genotypes differing in tolerance and adaptation to water deficit), using RAPD markers. Five markers were found associated with QTL in the populations. Assisted selection predictions were performed, showing that there was an 11% increase in performance under stress and 8% under no stress conditions. In contrast, conventional selection based on field performance was not satisfactory, resulting in no genetic gain. In this same study, the RAPD

loci were associated with QTL located on linkage group 9. In the population Sierra x Lef-2RB five markers were found on linkage group 1.

Afterwards, a QTL map for traits linked to drought stress was constructed based on AFLP, RAPD, microsatellite and morphological markers for a population derived from the cross BAT477 (drought-tolerant) x DOR364 (susceptible). The QTL on this map are distributed in all the bean linkage groups except Group 1 (B1) (BLAIR et al., 2012). The same population was further used to map QTL associated with rooting pattern traits (ASFAW; BLAIR, 2012). QTL were found for total root length, fine roots, thick roots, root volume and root biomass, and those were co-localized, explaining significant amount of genetic variance. The authors assumed that QTL affecting root traits in *P. vulgaris might* be based on constitutive expression of genes and drought avoidance (ASFAW; BLAIR, 2012). A third study was focused on QTL for photosynthates acquisition, remobilization and accumulation with the same population. QTL found could explain up to 37% of the phenotypic variation, and SPAD (chlorophyll measurement) and pod partitioning index were the most consistent. However, QTL x environment interactions were significant (ASFAW; BLAIR; STRUIK; 2012).

QTL have also been found for phenology and seed weight (SW) traits under drought stress for an intergene pool population originated from SEA5 x CAL96. One specific QTL was linked to SW, which could be useful for selection for seed yield and size in intergene pool crosses of *P. vulgaris* (MUKESHIMANA et al., 2014).

2.10 Association mapping: trends for candidate genes and examples with genes associated with abiotic stress responses

Association or linkage disequilibrium mapping has arisen to elucidate the complex variation of traits at the DNA level. It can be achieved through the study of recombination events at the population level and under evolutionary concepts (NORDBORG; TAVARE, 2002). In other words, genetic diversity is associated with phenotypic variation in a natural population based on the linkage disequilibrium principle.

As an alternative to the traditional linkage maps, association mapping offers three main advantages, according to Yu and Buckler (2006). First, it enables high-resolution mapping results, since it is based on partial or complete genome sequences. Very accurate linkage disequilibrium calculations can be obtained, including sizes of disequilibrium blocks.

Along with this, it provides a high number of alleles. Furthermore, less time is required for research, since there is no need for developing segregant populations.

The first association study reported for plants showed significant association among polymorphisms of *Dwarf8*, in maize, with flowering time (THORNSBERRY et al., 2001). Even though this work was performed with inbred lines, it used statistical methods specifically applied for association mapping.

Linkage disequilibrium refers to the non-random association between two markers or genes. Association mapping establishes if there is an association between a determined marker or a set of markers with phenotypic traits. In this way, it is an application of the concept of linkage disequilibrium. Two markers in linkage disequilibrium represent an association between alleles but not necessarily implies they are related to a particular phenotype (SOTO-CERCA; CLOUTIER, 2012).

Furthermore, it is important to establish the difference between gene linkage and linkage disequilibrium. Gene linkage refers to two or more genes/alleles that are located on the same chromosome. On the other hand, linkage disequilibrium shows the correlation among alleles in a population (FLINT-GARCIA; THORNSBERRY; BUCKLER, 2003). Even if two linked loci present high LD, the same can happen with distant loci on the same chromosome or different chromosomes (SOTO-CERCA; CLOUTIER, 2012).

Association studies can be broadly divided into two main categories. The first concerns genome-wide association mapping (GWAS) and the other is the candidate-gene association mapping. In this thesis, we dedicated the analysis to candidate-association mapping. Some examples from literature involving stress-responsive genes can be pointed out. A collection of 192 accessions of *Lolium perenne* was evaluated to determine associations among 346 SNP markers and parameters of drought tolerance (YU; JIANG, 2013). These SNP were detected within genes involved in antioxidant activity, dehydration and water transport across membranes. A significant association was found between leaf water content and *LpFeSOD*, which codes for an isoform of Superoxide dismutase. Another gene, LpCyt Cu-ZnSOD, was associated with the fluorescence of chlorophyll (YU; JIANG, 2013). Furthermore, similar analyses were performed in wheat, including a *DREB1* gene as the candidate. In total, 126 lines were phenotyped and genotyped, and a significant association was detected between biomass and *DREB1* (EDAE et al., 2013).

In common bean, Cortés et al. (2013) explored the diversity of *DREB* genes and gave directions for association mapping studies by electing wilds and domesticated bean accessions. Additionally, a collection of accessions suitable for association mapping in common bean has been recently proposed encompassing 180 varieties (MORINI et al., 2015). In this work, we performed an association analysis among SNP markers within candidate genes with drought-related traits from a wild bean collection.

REFERENCES

ACOSTA-GALLEGOS, J.; WHITE, J. W. Phenological plasticity as an adaptation by common bean to rainfed environments. **Crop Science**, Madison, v. 35, n. 1, p. 199-204, 1995.

ADAM-BLONDON, A.; SE´VIGNAC, M.; DRON, M. A genetic map of common bean to localize specific resistance genes against anthracnose. **Genome**, Ottawa, v. 37, n. 6, p. 915–924, 1994.

AGARWAL, P.; AGARWAL, P. K.; JOSHI A. J.; SOPORY, S. K.; REDDY, M. K. Overexpression of *PgDREB2A* transcription factor enhances abiotic stress tolerance and activates downstream stress-responsive genes. **Molecular Biology Reports**, Dordrecht, v. 37, n. 2, p. 1125–1135, 2010.

AKHTAR, M.; JAISWAL, A.; TAJ, G.; JAISWAL, J. P.; QURESHI, M. I.; SINGH, N. K. DREB1/CBF transcription factors: their structure, function and role in abiotic stress tolerance in plants. **Journal of Genetics**, London, v. 91, n. 3, p. 385-395, 2012.

ALONSO, J. M.; STEPANOVA, A. N.; LEISSE, T. J.; KIM, C. J.; CHEN, H.; SHINN, P.; STEVENSON, D. K.; ZIMMERMAN, J.; BARAJAS, P.; CHEUK, R.; GADRINAB, C.; HELLER, C.; JESKE, A.; KOESEMA, E.; MEYERS, C. C.; PARKER, H.; PREDNIS, L.; ANSARI, Y.; CHOY, N.; DEEN, H.; GERALT, M.; HAZARI, N.; HOM, E.; KARNES, M.; MULHOLLAND, C.; NDUBAKU, R.; SCHMIDT, I.; GUZMAN, P.; AGUILAR-HENONIN, L.; SCHMID, M.; WEIGEL, D.; CARTER, D. E.; MARCHAND, T.; RISSEEUW, E.; BROGDEN, D.; ZEKO, A.; CROSBY, W. L.; BERRY, C. C.; ECKER, J. R. Genome-wide insertional mutagenesis of *Arabidopsis thaliana***. Science,** Washington, DC, v. 301, n. 5633, p. 653–657, 2003.

ARENAS-HUERTERO, F.; ARROYO, A.; ZHOU, L.; SHEEN, J.; LEÓN, P. Analysis of *Arabidopsis* glucose insensitive mutants, *gin5* and *gin6*, reveals a central role of the plant hormone ABA in the regulation of plant vegetative development by sugar. **Genes and Development**, Woodbury, v. 14, n. 16, p. 2085–9096, 2000.

ASFAW, A.; BLAIR, M. Quantitative trait loci for rooting pattern traits of common beans grown under drought stress versus non-stress conditions. **Molecular Breeding**, Dordrecht, v. 30, n. 2, p. 681-695, 2012.

ASFAW, A.; BLAIR, M. W.; STRUIK, P. C. Multienvironment quantitative trait loci analysis for photosynthate acquisition, accumulation, and remobilization traits in common bean under drought stress. **Genes, Genomes & Genetics**, Bethesda, v. 2, n. 5, p. 579-595, 2012.

ASSEFA, T.; BEEBE, S. E.; RAO, I. M.; CUASQUER, J. B.; DUQUE, M. C.; RIVERA, M.; BATTISTI, A.; LUCCHIN, M. Pod harvest index as a selection criterion to improve drought resistance in white pea bean. **Field Crops Research**, Amsterdam, v. 148, p. 24–33, 2013.

BASU, U. Identification of molecular processes underlying abiotic stress plants adaptation using "Omics" Technologies. In: BENKEBLIA N. (Ed.). **Sustainable agriculture and new technologies**. Boca Raton: CRC Press, 2012. p. 149-172.

BEEBE, S. Common bean breeding in the tropics. **Plant Breeding Reviews,** Hoboken, v. 36, p. 357-426, 2012.

BEEBE, S. E.; RAO, I. M.; BLAIR, M. W.; ACOSTA-GALLEGOS, J. A. Phenotyping common beans for adaptation to drought. **Frontiers in Plant Physiology**, Lausanne, v. 4, n. 35, p. 1-20, 2013

BEEBE, S. E.; RAO, I. M.; CAJIAO, C.; GRAJALES, M. Selection for drought resistance in common bean also improves yield in phosphorus limited and favorable environments. **Crop Science**, Madison, v. 48, n. 2, p. 582-592, 2008.

BEEBE, S. E.; RAO, I. M.; DEVI, M. J.; POLANIA, J. Common beans, biodiversity, and multiple stresses: challenges of drought resistance in tropical soils. **Crop and Pasture Science**, Clayton, v. 65, n. 7, p. 667-675, 2014.

BELLUCCI, E.; BITOCCHI, E.; FERRARINI, A.; BENAZZO, A.; BIAGETTI, E.; KLIE, S.; MINIO, A.; RAU, D.; RODRIGUEZ, M.; PANZIERA, A.; VENTURINI, L.; ATTENE, G.; ALBERTINI, E.; JACKSON, S. A.; NANNI, L.; FERNIE, A. R.; NIKOLOSKI, Z.; BERTORELLE, G.; DELLEDONNE, M.; PAPA, R. Decreased nucleotide and expression diversity and modified coexpression patterns characterize domestication in the common bean. **The Plant Cell**, Rockville, v. 26, p. 1901–1912, 2014.

BENCHIMOL, L. L.; CAMPOS, T.; CARBONELL, S. A. M.; COLOMBO, C. A.; CHIORATTO, A. F.; FORMIGHIERI, E. F.; GOUVÊA, L. R. L.; SOUZA, A. P. Structure of genetic diversity among common bean (*Phaseolus vulgaris* L.) varieties of Mesoamerican and Andean origins using new developed microsatellite markers. **Genetic Resources and Crop Evolution**, Dordrecht, v. 54, p. 1747-1762, 2007.

BITOCCHI, E.; NANNI, L.; BELLUCCI, E.; ROSSI, M.; GIARDINI, A., ZEULI, P. S.; PAPA, R. Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. **Proceedings of the National Academy of Sciences of the USA**, Washington, DC, v. 109, n. 14, p. E788-E796, 2012. doi: 10.1073/pnas.1108973109.

BITOCCHI, E.; BELLUCCI, E.; GIARDINI, A.; RAU, D.; RODRIGUEZ, M.; BIAGETTI, E.; SANTILOCCHI, R.; SPAGNOLETTI ZEULI, P.; GIOIA, T.; LOGOZZO, G.; ATTENE, G.; NANNI, L.; PAPA, R. Molecular analysis of the parallel domestication of the common bean (*Phaseolus vulgaris*) in Mesoamerica and the Andes. **New Phytologist**, Oxford, v. 197, n. 1, p. 300–13, 2013.

BIHANI, P.; CHAR, B.; BHARGAVA, S. Transgenic expression of sorghum *DREB2* in rice improves tolerance and yield under water limitation. **The Journal of Agricultural Science**, Cambridge, v. 149, n. 1, p. 95-101, 2011.

BLAIR, M. W.; GALEANO, C. H.; TOVAR, E.; TORRES, M. C. M.; CASTRILLÓN, A. V.; BEEBE, S. E.; RAO, I. M. Development of a Mesoamerican intra-gene pool genetic map for quantitative trait loci detection in a drought tolerant x susceptible common bean (*Phaseolus vulgaris* L.) cross. **Molecular Breeding**, Dordrecht, v. 29, n. 1, p. 1-18, 2012.

BLAIR, M. W.; CORTÉS, A. J.; PENMETSA, R. V.; FARMER, A.; CARRASQUILHA-GARCIA, N.; COOK, D. R. A high-throughput SNP marker system for parental polymorphism screening, and diversity analysis in common bean (*Phaseolus vulgaris* L.). **Theoretical and Applied Genetics**, Heidelberg, v. 126, n. 2, p. 535-548, 2013.

BLAIR, M. W.; GIRALDO, M. C.; BUENDIA, H. F.; TOVAR, E.; DUQUE, M. C.; BEEBE, S. E. Microsatellite marker diversity in common bean (*Phaseolus vulgaris* L.). **Theoretical and Applied Genetics**, Heidelberg, v. 113, p. 100–109, 2006.

BLAIR, M. W.; PEDRAZA, F.; BUENDIA, H. F.; GAITA´N-SOLI´S, E.; BEEBE, S. E.; GEPTS, P.; TOHME, J. Development of a genome-wide anchored microsatellite map for common bean (*Phaseolus vulgaris* L.). **Theoretical and Applied Genetics**, Heidelberg, v. 107, p. 1362–1374, 2003.

BOUTILIER, K.; OFFRINGA, R.; SHARMA, V. K.; KIEFT, H.; OUELLET, T.; ZHANG, L.; HATTORI, J.; LIU, C. M.; VAN LAMMEREN, A. A.; MIKI, B. L.; CUSTERS, J. B.; VAN LOOKEREN CAMPAGNE, M. M.Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. **The Plant Cell**, Rockville, v. 14, n. 8, p. 1737–1749, 2002.

BRAY, E. A. Plant responses to water deficit. **Trends in Plant Science**, Oxford, v. 2, n. 2, p. 48–54, 1997.

BRAY, E. A.; BAILEY-SERRES, J.; WERETILNYK, E. Responses to abiotic stresses. In: BUCHANAN, B.; GRUISSEM, W.; JONES, R. (Ed.). **Biochemistry & molecular biology of plants**. Hoboken: Wiley, 2000. p. 1158-1203.

BROUGHTON, W. J.; HERNÁNDEZ, G.; BLAIR, M.; BEEBE, S.; GEPTS, P.; VANDERLEYDEN, J. Beans (*Phaseolus* spp.) - model food legumes. **Plant and Soil**, Dordrecht, v. 252, p. 55–128, 2003.

BURLE, M. L.; FONSECA, J. R.; KAMI, J. A.; GEPTS, P. Microsatellite diversity and genetic structure among common bean (*Phaseolus vulgaris* L.) landraces in Brazil, a secondary center of diversity. **Theoretical and Applied Genetics**, Heidelberg, v. 121, n. 5, p. 801-813, 2010.

BURLE, M. L.; FONSECA, J. R.; JOSÉ DEL PELOSO, M.; MELO, L. C.; TEMPLE, S. R.; GEPTS, P. Integrating phenotypic evaluations with a molecular diversity assessment of a Brazilian collection of common bean landraces. **Crop Science**, Madison, v. 51, n. 6, p. 2668-2680, 2011.

BUTARE, L.; RAO, I.; LEPOIVRE, P.; POLANIA, J.; CAJIAO, C.; CUASQUER, J.; BEEBE, S. New genetic sources of resistance in the genus *Phaseolus* to individual and combined aluminium toxicity and progressive soil drying stresses. **Euphytica**, Dordrecht, v. 181, n. 3, p. 385-404, 2011.

CAMPOS, T.; OBLESSUC, P. R.; SFORÇA, D. A.; CARDOSO, J. M. K.; BARONI, R. M.; SOUSA, A. C. B.; CARBONELL, S. A. M.; CHIORATTO, A. F.; GARCIA, A. A. F.; RUBIANO, L. B.; SOUZA, A. P. Inheritance of growth habit detected by genetic linkage analysis using microsatellites in the common bean (*Phaseolus vulgaris* L.). **Molecular Breeding**, Dordrecht, v. 27, n. 4, p. 549–560, 2011.

CARVALHO, M. F.; CRESTANI, M.; FARIAS, F. L.; COIMBRA, J. L. M.; BOGO, A.; GUIDOLIN, A. F. Caracterização da diversidade genética entre acessos crioulos de feijão (*Phaseolus vulgaris* L.) coletados em Santa Catarina por marcadores RAPD. **Ciência Rural**, Santa Maria, v. 38, n. 6, p. 1522-1528, 2008.

CHAVES, M. M.; PEREIRA, J. S.; MAROCO, J.; RODRIGUES, M. L.; RICARDO, C. P. P.; OSÓRIO, M. L.; CARVALHO, I.; FARIA, T.; PINHEIRO, C. How plants cope with water stress in the field. Photosynthesis and growth. **Annals of Botany**, Oxford, v. 89, n. 7, p. 907-916, 2002.

CHEN, M.; WANG, Q. Y.; CHENG, X. G.; XU, Z. S.; LI, L. C.; YE, X. G.; XIA, L. Q.; MA, Y. Z. *GmDREB2*, a soybean DRE-binding transcription factor, conferred drought and high-salt tolerance in transgenic plants. **Biochemical and Biophysical Research Communications**, New York, v. 353, n. 2, p. 299–305, 2007.

CHEN, M.; XU, Z.; XIA, L.; LI, L.; CHENG, X.; DONG, J.; WANG, Q.; MA, Y. Cold-induced modulation and functional analyses of the DRE-binding transcription factor gene, *GmDREB3*, in soybean (*Glycine max* L.). **Journal of Experimental Botany**, Oxford, v. 60, n. 1, p. 121-135, 2009.

CHEN, J. R.; LÜ, J. J.; WANG, T. X.; CHEN, S. Y.; WANG, H. F. Activation of a DRE-binding transcription factor from *Medicago truncatula* by deleting a Ser/Thr-rich region. **In Vitro Cellular and Developmental Biology - Plant**, Heidelberg, v. 45, n. 1, p. 1–11, 2009.

CHEN, J.; LU, J.; LIU, R.; XIONG, X.; WANG, T.; CHEN, S.; GUO, L.; WANG, H. *DREB1C* from *Medicago truncatula* enhances freezing tolerance in transgenic *M. truncatula* and China Rose (*Rosa chinensis* Jacq). **Plant Growth Regulation**, Dordrecht, v. 60, n. 3, p. 199-211, 2010.

CHIORATO, A. F.; CARBONELL, S. A. M.; CARVALHO, C. R. L.; NISHIJIMA, V. L.; BARROS, P. de; BORGES, W. L. B.; TICELLI, M.; GALLO, P. B.; FINOTO, E. L.; SANTOS, C. B. `IMPERADOR´: early maturity " carioca " bean cultivar. **Crop Breeding and Applied Biotechnology**, Londrina, v. 13, p. 297–300, 2012.

CHU, Y.; HUANG, Q.; ZHANG, B.; DING, C.; SU, X. Expression and molecular evolution of two *DREB1* genes in black poplar (*Populus nigra*). **Plos One**, San Francisco, v. 9, n. 6, e98334, 2014.

CHUCK, G.; MEELEY, R. B.; HAKE, S. The control of maize spikelet meristem fate by the *APETALA2*-like gene indeterminate *spikelet1*. **Genes and Development**, Woodbury, v. 12, n. 8, p. 1145–54, 1998.

COMAS, L. H.; BECKER, S. R.; VON MARK, V. C.; BYRNE, P. F.; DIERIG, D. A. Root traits contributing to plant productivity under drought. **Frontiers in Plant Science**, Lausanne, v. 4, p. 1-16, 2013.

COMPANHIA NACIONAL DE ABASTECIMENTO - CONAB. **Produção de feijão 2012-2015**. Brasília, DF, 2015.

CORTÉS, A. J.; CHAVARRO, M. C.; BLAIR, M. W. SNP marker diversity in common bean (*Phaseolus vulgaris* L.). **Theoretical and Applied Genetics**, Heidelberg, v. 123, n. 5, p. 827–845, 2011.

CORTÉS, A. J.; MONSERRATE, F. a.; RAMÍREZ-VILLEGAS, J.; MADRIÑÁN, S.; BLAIR, M. W. Drought tolerance in wild plant populations: the case of common beans (*Phaseolus vulgaris* L.). **Plos One**, San Francisco, v. 8, n. 5, e62898, 2013. doi:10.1371/journal.pone.0062898.

CORTÉS, A. J.; THIS, D.; CHAVARRO, M. C.; MADRIÑAN, S.; BLAIR, M. W. Nucleotide diversity patterns at the drought related DREB encoding genes in wild and cultivated common bean (*Phaseolus vulgaris* L.). **Theoretical and Applied Genetics**, Heidelberg, v. 125, n. 5, p. 1069–1085, 2012.

CRAMER, G. R.; URANO, K.; DELROT, S.; PEZZOTTI, M.; SHINOZAKI, K. Effects of abiotic stress on plants: a systems biology perspective. **BMC Plant Biology**, London, v. 11, n. 1, p. 163, 2011.

DEBOUCK, D. G.; TORO, O.; PAREDES, O. M.; JOHNSON, W. C.; GEPTS, P. Genetic diversity and ecological distribution of *Phaseolus vulgaris* (Fabaceae) in northwestern South America. **Economic Botany**, New York, v. 47, n. 4, p. 408-423, 1993.

DELGADO-SALINAS, A.; BIBLER, R.; LAVIN, M. Phylogeny of the genus *Phaseolus* (Leguminosae): a recent diversification in an ancient landscape. **Systematic Botany**, Laramie, v. 31, n. 4, p. 779-791, 2006.

DIETZ, K. J.; VOGEL, M. O.; VIEHHAUSER, A. AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signalling. **Protoplasma**, Wien, v. 245, n. 1-4, p. 3-14, 2010.

DUBOUZET, J. G.; SAKUMA, Y.; ITO, Y.; KASUGA, M.; DUBOUZET, E. G.; MIURA, S., SEKI, M.; SHINOZAKI, K.; YAMAGUCHI-SHINOZAKI, K. *OsDREB* genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. **Plant Journal**, Chichester, v. 33, p. 751–763, 2003.

DUQUE, A. S.; ALMEIDA, A. M.; SILVA, A. B.; SILVA, J. M., FARINHA, A. P.; SANTOS; D.; FEVEREIRO, P.; ARAÚJO, S. S. Abiotic stress responses in plants: unraveling the complexity of genes and networks to survive. In: VAHDATI, K.; LESLIE, C. (Ed.). **Abiotic stresses** – Plant responses and applications in agriculture. Rijeka, Croatia: InTech, 2013. p. 49-101.

EDAE, E. A.; BYRNE, P. F.; MANMATHAN, H.; HALEY, S. D.; MORAGUES, M.; LOPES, M. S.; REYNOLDS, M. P. Association mapping and nucleotide sequence variation in five drought tolerance candidate genes in spring wheat. **Plant Genome**, Madison, v. 6, p. 1-13, 2013.

ELLIOTT, R. C.; BETZNER, A. S.; HUTTNER, E.; OAKES, M. P.; TUCKER, W. Q.; GERENTES, D.; PEREZ, P.; SMYTH, D. R. *AINTEGUMENTA*, an *APETALA2*-like gene of *Arabidopsis* with pleiotropic roles in ovule development and floral organ growth. **The Plant Cell**, Rockville, v. 8, p. 155–168, 1996.

FLINT-GARCIA, S.; THORNSBERRY, J.; BUKLER, E. Structure of linkage disequilibrium in plants. **Annual Review of Plant Biology**, Palo Alto, v. 54, p. 357-374, 2003.

FREITAS, F. O. Evidências genético-arqueológicas sobre a origem do feijão comum no Brasil. **Pesquisa Agropecuária Brasileira**, Brasília, DF, v. 41, n. 7, p. 1199–1203, 2006.

FREYRE, R.; RÍOS, R.; GUZMÁN, L.; DEBOUCK, D. G.; GEPTS, P. Ecogeographic distribution of *Phaseolus* spp. (Fabaceae) in Bolivia. **Economic Botany**, New York, v. 50, n. 2, p. 195-215, 1996.

FREYRE, R.; SKROCH, P. W.; GEFFROY, V.; ADAM-BLONDON, A. F.; SHIRMOHAMADALI, A.; JOHNSON, W. C.; LLACA, V.; NODARI, R. O.; PEREIRA, P. A.; TSAI, S. M.; TOHME, J.; DRON, M.; NIENHUIS, J.; VALLEJOS, C. E.; GEPTS, P. Towards an integrated linkage map of common bean. 4. Development of a core linkage map and alignment of RFLP maps. **Theoretical and Applied Genetics**, Heidelberg, v. 97, p. 847–856, 1998.

GALEANO, C. H.; CORTÉS, A. J.; FERNÁNDEZ, A. C.; SOLER, Á.; FRANCO-HERRERA, N.; MAKUNDE, G.; VANDERLEYDEN, J.; BLAIR, M. W. Gene-based single nucleotide polymorphism markers for genetic and association mapping in common bean. **BMC Genetics**, London, v. 13, n. 1, p. 48, 2012.

GAO, S. Q.; CHEN, M.; XIA, L. Q.; XIU, H. J.; XU, Z. S.; LI, L. C.; ZHAO, C. P.; CHENG, X. G.; MA, Y. Z. A cotton (*Gossypium hirsutum*) DRE-binding transcription factor gene, *GhDREB*, confers enhanced tolerance to drought, high salt, and freezing stresses in transgenic wheat. **Plant Cell Reports**, Heidelberg, v. 28, n. 2, p. 301–311, 2009.

GENTRY, H. S. Origin of the common bean. **Economic Botany**, New York, v. 23, n. 1, p. 55-69, 1969.

GEPTS, P. The contribution of genetic and genomic approaches to plant domestication studies. **Current Opinion in Plant Biology**, Oxford, v. 18, n. 1, p. 51–59, 2014.

GEPTS P.; BLISS, F. A. Phaseolin variability among wild and cultivated common beans (*Phaseolus vulgaris*) from Colombia. **Economic Botany**, New York, v. 40, n. 4, p. 469–478, 1986.

GHEYI, H. R. Problemas de salinidade na agricultura irrigada. In: OLIVEIRA, T. S.; ASSIS JUNIOR, R. N.; OMERO, R. E.; SILVA, J. R. C. (Ed.). **Agricultura, sustentabilidade e o semi-árido**. Fortaleza: Sociedade Brasileira de Ciências do Solo, 2000. p. 329-346.

GIL-LANGARICA, H. R. Genetic diversity analysis of common beans based on molecular markers. **Genetics and Molecular Biology**, Ribeirão Preto, v. 34, n. 4, p. 595-605, 2011.

GILMOUR, S. J.; ZARKA, D. G.; STOCKINGER, E. J.; SALAZAR, M. P.; HOUGHTON, J. M.; THOMASHOW, M. F. Low temperature regulation of the *Arabidopsis* CBF family of AP2 transcriptional activators as an early step in cold-induced COR gene expression. **The Plant Journal**, Chichester, v. 16, p. 433–442, 1998.

GIOIA, T.; LOGOZZO, G.; KAMI, J.; ZEULI, P. S.; GEPTS., P. Identification and characterization of a homologue to the *Arabidopsis* INDEHISCENT gene in common bean. **Journal of Heredity**, Washington, DC, v. 104, p. 273-286, 2013.

GONZÁLEZ, A.; WONG, A.; DELGADO-SALINAS, A.; PAPA, R.; GEPTS, P. Assessment of inter simple sequence repeat markers to differentiate sympatric wild and domesticated populations of common bean. **Crop Science**, Madison, v. 45, n. 2, p. 606-615, 2005.

GROVER, A.; KAPPOR, A.; LAKSHMI, O. S.; AGARWAL, S.; SAHI, C.; KATIYAR-AGARWALL, S.; AGARWAL; DUBEY, H. Understanding molecular alphabets of the plant abiotic responses. **Current Science**, Rajasthan, v. 80, n. 2, p. 206-216, 2001.

GUPTA, K.; AGARWAL, P. K.; REDDY, M. K.; JHA, B. *SbDREB2A*, an A-2 type DREB transcription factor from extreme halophyte *Salicornia brachiata* confers abiotic stress tolerance in *Escherichia coli*. **Plant Cell Reports**, Heidelberg, v. 29, n. 10, p. 1131–1137, 2010.

GUPTA, P. K.; ROY, J. K.; PRASAD, M. Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. **Current Science**, Rajasthan, v. 80, n. 4, p. 524–535, 2001.

HAAKE, V.; COOK, D.; RIECHMANN, J.; PINEDA, O.; THOMASHOW, M. F.; ZHANG, J. Z. Transcription factor CBF4 is a regulator of drought adaptation in *Arabidopsis*. **Plant Physiology**, Rockville, v. 130, n. 2, p. 639-648, 2002.

HANNAH, M. A.; WIESE, D.; FREUND, S.; FIEHN, O.; HEYER, A. G.; HINCHA, D. K. Natural genetic variation of freezing tolerance in *Arabidopsis*. **Plant Physiology**, Rockville, v. 142, n. 1, p. 98-112, 2006.

HANAI, L. R.; SANTINI, L.; CAMARGO, L. E. A.; FUNGARO, M. H. P.; GEPTS, P.; TSAI, S. M.; VIEIRA, M. L. C. Extension of the core map of common bean with EST-SSR, RGA, AFLP, and putative functional markers. **Molecular Breeding**, Dordrecht, v. 25, p. 25-45, 2010.

HAO, D.; OHME-TAKAGI, M.; SARAI, A. Unique mode of GCC box recognition by the DNA-binding domain of ethylene-responsive element-binding factor (ERF domain) in plant. **Journal of Biological Chemistry**, Bethesda, v. 273, n. 41, p. 26857–26861, 1998.

HASANUZZAMAN, M.; NAHAR, K.; GILL, S. S.; GILL, R.; FUJITA, M. Drought stress responses in plants, oxidative stress, and antioxidant defense. In: TUTEJA, N.; GILL, S. S. (Ed.). Climate change and plant abiotic stress tolerance. Weinheim, Germany: Wiley-VCH Verlag, 2013. p. 209-237.

HU, Y. X.; WANG, Y. X.; LIU, X. F.; LI, J. Y. *Arabidopsis* RAVI is down-regulated by brassinosteroid and may act as a negative regulator during plant development. **Cell Research**, London, v. 14, p. 8–15, 2004.

HUBBARD, K. E.; NISHIMURA, N.; HITOMI, K.; GETZOFF, E. D.; SCHROEDER, J. I. Early abscisic acid signal transduction mechanisms: newly discovered components and newly emerging questions. **Genes and Development**, Woodbury, v. 24, p. 1695–1708, 2010.

INGRAM, J.; BARTELS, D. The molecular basis of dehydration tolerance in plants. **Annual Review in Plant Physiology and Plant Molecular Biology**, Palo Alto, v. 47, p. 377–403, 1996.

JAGLO-OTTOSEN, K. R.; GILMOUR, S. J.; ZARKA, D. G.; SCHABENBERGER, O.; THOMASHOW, M. F. *Arabidopsis CBF1* overexpression induces *COR* genes and enhances freezing tolerance. **Science**, Washington, DC, v. 280, n. 5360, p. 104–106, 1998.

JEONG, J. S.; KIM, Y. S.; BAEK, K. H.; JUNG, H.; HA, S.-H.; DO CHOI, Y.; KIM, M.; REUZEAU, C.; KIM, J.-K. Root-specific expression of *OsNAC10* improves drought tolerance and grain yield in rice under field drought conditions. **Plant Physiology**, Rockville, v. 153, n. 1, p. 185–197, 2010.

JOFUKU, K. D.; DEN BOER, B. G.; VAN MONTAGU, M.; OKAMURO, J. K. Control of *Arabidopsis* flower and seed development by the homeotic gene APETALA2. **The Plant Cell**, Rockville, v. 6, n. 9, p. 1211–1225, 1994.

JUNG, G.; COYNE, D. P.; SKROCH, P. W.; NIENHUIS, J.; ARNAUD-SANTANA, E.; BOKOS, J.; ARIYARATHNE, H. M.; STEADMAN, J. R.; BEAVER, J. S.; KAEPPLER, S. M. Molecular markers associated with plant architecture and resistance to common blight, web blight and rust in common bean (*Phaseolus vulgaris* L.). **Journal of the American Society for Horticultural Science**, Alexandria, v. 121, p. 794–803, 1996.

KAMI, J. A.; GEPTS, P. Phaseolin nucleotide sequence diversity in Phaseolus. I. Intraspecific diversity in *Phaseolus vulgaris*. **Genome**, Ottawa, v. 37, n. 5, p. 751-757, 1994.

KAMI, J.; VELÁSQUEZ, V. B.; DEBOUCK, D. G.; GEPTS, P. Identification of presumed ancestral DNA sequences of phaseolin in *Phaseolus vulgaris*. **Proceedings of the National Academy of Sciences of the USA**, Washington, DC, Philadelphia, v. 92, n. 4, p. 1101-1104, 1995.

KARABA, A.; DIXIT, S.; GRECO, R.; AHARONI, A.; TRIJATMIKO, K. R.; MARSCH-MARTINEZ, N.; KRISHNAN, A.; NATARAJA, K. N.; UDAYAKUMAR, M.; PEREIRA, A. Improvement of water use efficiency in rice by expression of *HARDY*, an *Arabidopsis* drought and salt tolerance gene. **Proceedings of the National Academy of Sciences of the USA**, Washington, Dc, v. 104, n. 39, p. 15270–15275, 2007.

KOENIG, R.; GEPTS, P. Allozyme diversity in wild *Phaseolus vulgaris*: further evidence for two major centers of diversity. **Theoretical and Applied Genetics**, Heidelberg, v. 78, p. 809-817, 1989.

KOINANGE, E. M. K.; SINGH, S. P.; GEPTS, P. Genetic control of the domestication syndrome in common-bean. **Crop Science**, Madison, v. 36, p. 1037–1045, 1996.

KUMAR, A.; SINGH, P. K.; RAI, N.; BHASKAR, G. P.; DATTA, D. Genetic diversity of French bean (*Phaseolus vulgaris* L .) genotypes on the basis of morphological traits and molecular markers. **Indian Journal of Biotechnology**, New Delhi, v. 13, p. 207–213, 2014.

KWAK, M.; TORO, O.; DEBOUCK, D. G.; GEPTS, P. Multiple origins of the determinate growth habit in domesticated common bean (*Phaseolus vulgaris*). **Annals of botany**, Oxford, v. 110, n. 8, p. 1573-1580, 2012.

LATA, C.; BHUTTY, S.; BAHADUR, R. P.; MAJEE, M.; PRASAD, M. Association of an SNP in a novel *DREB2*-like gene *SiDREB2* with stress tolerance in foxtail millet [*Setaria italica* (L.)]. **Journal of Experimental Botany**, Oxford, v. 62, n. 10, p. 3387–3401, 2011.

LATA, C.; PRASAD, M. Role of DREBs in regulation of abiotic stress responses in plants. **Journal of Experimental Botany**, Oxford, v. 62, n. 14, p. 4731–4748, 2011.

LI, X. P.; TIAN, A. G.; LUO, G. Z.; GONG, Z. Z.; ZHANG, J. S.; CHEN, S. Y. Soybean DRE-binding transcription factors that are responsive to abiotic stresses. **Theoretical and Applied Genetics**, Heidelberg, v. 110, n. 8, p. 1355–1362, 2005.

LI, D.; ZHANG, Y.; HU, X.; SHEN, X.; MA, L.; SU, Z.; WANG, T.; DONG, J. Transcriptional profiling of *Medicago truncatula* under salt stress identified a novel CBF transcription factor *MtCBF4* that plays an important role in abiotic stress responses. **BMC Plant Biology**, London, v. 11, n. 1, p. 109, 2011.

LIM, C. J.; HWANG, J. E.; CHEN, H.; HONG, J. K.; YANG, K. A.; CHOI, M. S.; LEE, K. O.; CHUNG, W. S.; LEE, S. Y.; LIM, C. O. Over-expression of the *Arabidopsis* DRE/CRT-binding transcription factor *DREB2C* enhances thermotolerance. **Biochemical and Biophysical Research Communications**, New York, v. 362, n. 2, p. 431–436, 2007.

LIN, R. C.; PARK, H. J.; WANG, H. Y. Role of *Arabidopsis* RAP2.4 in regulating lightand ethylene-mediated developmental processes and drought stress tolerance. **Molecular Plant**, Cambridge, v. 1, n. 1, p. 42–57, 2008.

LIU, Q.; KASUGA, M.; SAKUMA, Y.; ABE, H.; MIURA, S.; YAMAGUCHI-SHINOZAKI, K.; SHINOZAKI, K. Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in *Arabidopsis*. **The Plant Cell**, Rockville, v. 10, n. 8, p. 1391–1406, 1998.

MACIEL, F.; ECHEVERRIGARAY, S.; GERALD, L.; GRAZZIOTIN, F. Genetic relationship and diversity among Brazilian cultivar and of common beans (*Phaseolus vulgaris* L.) revealed by AFLP markers. **Genetic Resources and Crop Evolution**, Dordrecht, v. 50, p. 887–893, 2003.

MAGOME, H.; YAMAGUCHI, S.; HANADA, A.; KAMIYA, Y.; ODA, K. Dwarf and delayed-flowering 1, a novel *Arabidopsis* mutant deficient in gibberellin biosynthesis because of overexpression of a putative AP2 transcription factor. **Plant Journal**, Chichester, v. 37, n. 5, p. 720–729, 2004.

MAMIDI, S.; ROSSI, M.; ANNAM, D.; MOGHADDAM, S.; LEE, R.; PAPA, R.; MCCLEAN, P. Investigation of the domestication of common bean (*Phaseolus vulgaris*) using multilocus sequence data. **Functional Plant Biology**, Clayton, v. 38, n. 12, p. 953-967, 2011.

MARTINEZ-ROJO, J.; GURUSAMY, V.; VANDENBERG, A.; BETT, K. E. Tolerance to sub-zero temperatures in *Phaseolus acutifolius* and development of interspecies hybrids with *P. vulgaris*. **Annual Report of the Bean Improvement Cooperative**, Prosser, WA, v. 50, p. 9-10, 2007.

MATSUKURA, S.; MIZOI, J.; YOSHIDA, T.; TODAKA, D.; ITO, Y.; MARUYAMA, K. Comprehensive analysis of rice *DREB2*-type genes that encode transcription factors involved in the expression of abiotic stress-responsive genes. **Molecular Genetics and Genomics**, Heidelberg, v. 283, 185-196, 2010.

McCLEAN, P.; LEE, R.; OTTO, C.; GEPTS, P.; BASSETT, M. Molecular and phenotypic mapping of genes controlling seed coat pattern and color in common bean (*Phaseolus vulgaris* L.). **Journal of Heredity**, Washington, DC, v. 93, p. 148–152, 2002.

McCLEAN, P. E.; LEE, R. K.; MIKLAS, P. N. Sequence diversity analysis of dihidroflavonol-4-reductase intron 1 in common bean. **Genome**, Ottawa, v. 47, p. 266-280, 2004.

MEDINA, J.; BARGUES, M.; TEROL, J.; PÉREZ-ALONSO, M.; SALINAS, J. The *Arabidopsis* CBF gene family is composed of three genes encoding AP2 domain-containing proteins whose expression is regulated by low temperature but not by abscisic acid or dehydration. **Plant Physiology**, Rockville, v. 119, n. 2, p. 463–470, 1999.

MIKLAS, P. N.; KELLY, J. D.; BEEBE, S. E.; BLAIR, M. W. Common bean breeding for resistance against biotic and abiotic stresses: from classical to MAS breeding. **Euphytica**, Dordrecht, v. 147, p. 105-131, 2006.

MIZOI, J.; SHINOZAKI, K.; YAMAGUCHI-SHINOZAKI, K. AP2/ERF family transcription factors in plant abiotic stress responses. **Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms**, Amsterdam, v. 1819, n. 2, p. 86–96, 2012.

MOLINA, J. C.; MODA-CIRINO, V.; JÚNIOR, N. D. S. F.; FARIA, R. T.; DESTRO, D. Response of common bean cultivars and lines to water stress. **Crop Breeding and Applied Biotechnology**, Londrina, v. 1, n. 4, p. 363-372, 2001.

MORINI, J.; PERSEGUINI, K. C.; MICAELE, G.; SILVA, B.; BACHEGA, J. R.; ROSA, F.; GAZAFFI, R.; MARÇAL, J. F.; MORAIS CARBONELL, S. A.; CHIORATO, A. F.; ZUCCHI, M. I.; FRANCO GARCIA, A. A.; LASRY BENCHIMOL-REIS, L. Developing a common bean core collection suitable for association mapping studies. **Genetics and Molecular Biology**, Ribeirão Preto, v. 78, p. 67–78, 2015.

MUKESHIMANA, G.; BUTARE, L.; CREGAN, P. B.; BLAIR, M. W.; KELLY, J. D. Quantitative trait loci associated with drought tolerance in common bean. **Crop Science**, Madison, v. 54, n. 3, p. 923-938, 2014.

NAKANO, T.; SUZUKI, K.; FUJIMURA, T.; SHINSHI, H. Genome-wide analysis of the ERF gene family in *Arabidopsis* and rice. **Plant Physiology**, Rockville, v. 140, p. 411–432, 2006.

NAKASHIMA, K.; YAMAGUCHI-SHINOZAKI, K. Regulons involved in osmotic stress-responsive and cold stress-responsive gene expression in plants. **Plant Physiology**, Rockville, v. 126, p. 62–71, 2006.

NAKASHIMA, K.; ITO, Y.; YAMAGUCHI-SHINOZAKI, K. Transcriptional regulatory networks in response to abiotic stresses in *Arabidopsis* and grasses. **Plant physiology**, Rockville, v. 149, n. 1, p. 88–95, 2009.

NANNI, L.; BITOCCHI, E.; BELLUCCI, E.; ROSSI, M.; RAU, D.; ATTENE, G.; GEPTS, P.; PAPA, R. Nucleotide diversity of a genomic sequence similar to SHATTERPROOF (PvSHP1) in domesticated and wild common bean (*Phaseolus vulgaris* L.). **Theoretical and Applied Genetics**, Heidelberg, v. 123, n. 8, p. 1341–1357, 2011.

NARUSAKA, Y.; NAKASHIMA, K.; SHINWARI, Z. K.; SAKUMA, Y.; FURIHATA, T.; ABE, H.; NARUSAKA, M.; SHINOZAKI, K.; YAMAGUCHI-SHINOZAKI, K. Interaction between two cis-acting elements, ABRE and DRE, in ABA-dependent expression of *Arabidopsis rd29A* gene in response to dehydration and high-salinity stresses. **Plant Journal**, Chichester, v. 34, n. 2, p. 137–148, 2003.

NAYAK, S. N.; BALAJI, J.; UPADHYAYA, H. D.; HASH, C. T.; KISHOR, P. B. K.; CHATTOPADHYAY, D.; RODRIQUEZ, L. M.; BLAIR, M. W.; BAUM, M.; MCNALLY, K.; THIS, D.; HOISINGTON, D. a.; VARSHNEY, R. K. Isolation and sequence analysis of *DREB2A* homologues in three cereal and two legume species. **Plant Science**, Amsterdam, v. 177, n. 5, p. 460–467, 2009.

NIU, X.; HELENTJARIS, T.; BATE, N. J. Maize ABI4 binds coupling element1 in abscisic acid and sugar response genes. **The Plant Cell**, Rockville, v. 14, n. 10, p. 2565–2575, 2002.

NORDBORG, M.; TAVARE, S. Linkage disequilibrium: What history has to tell us. **Trends in Genetics**, Cambridge, v. 18, p. 83–90, 2002.

OKAMURO, J. K.; CASTER, B.; VILLARROEL, R.; VAN MONTAGU, M.; JOFUKU, K. D. The AP2 domain of APETALA2 defines a large new family of DNA binding proteins in *Arabidopsis*. **Proceedings of the National Academy of Sciences of the USA**, Washington, DC, v. 94, n. 13, p. 7076-7081, 1997.

PAPA, R.; GEPTS, P. Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris* L.) from Mesoamerica. **Theoretical and Applied Genetics**, Heidelberg, v. 106, n. 2, p. 239-250, 2003.

PÉREZ-VEGA, E.; PAÑEDA, A.; RODRÍGUEZ-SUÁREZ, C.; CAMPA, A.; GIRALDEZ, R.; FERREIRA, J. J. Mapping of QTLs for morpho-agronomic and seed quality traits in a RIL population of common bean (*Phaseolus vulgaris* L.). **Theoretical and Applied Genetics**, Heidelberg, v. 120, n. 7, p. 1367-1380, 2010.

PERSEGUINI, J. M. K. C.; CHIORATTO, A. F.; ZUCCHI, M. I.; COLOMBO, C. A.; CARBONELL, S. A. M.; MONDEGO, J. M. C.; GAZAFFI, R.; GARCIA, A. A. F.; DE CAMPOS, T.; DE SOUZA, A. P.; RUBIANO, L. B. Genetic diversity in cultivated carioca common beans based on molecular marker analysis. **Genetics and Molecular Biology**, Ribeirão Preto, v. 34, n. 1, p. 88–102, 2011.

RAE, L.; LAO, N. T.; KAVANAGH, T. A. Regulation of multiple aquaporin genes in *Arabidopsis* by a pair of recently duplicated DREB transcription factors. **Planta**, Berlin, v. 234, n. 3, p. 429–444, 2011.

RECCHIA, G. H.; CALDAS, D. G. G.; BERALDO, A. L. A.; DA SILVA, M. J.; TSAI, S. M. Transcriptional analysis of drought-induced genes in the roots of a tolerant genotype of the common bean (*Phaseolus vulgaris* L.). **International Journal of Molecular Sciences**, Basel, v. 14, n. 4, p. 7155-7179, 2013.

REPINSKI, S. L.; KWAK, M.; GEPTS, P. The common bean growth habit gene *PvTFL1y* is a functional homolog of *Arabidopsis TFL1*. **Theoretical and Applied Genetics**, Heidelberg, v. 124, n. 8, p. 1539-1547, 2012.

ROYCHOUDHURY, A.; PAUL, S.; BASU, S. Cross-talk between abscisic acid-dependent and abscisic acid-independent pathways during abiotic stress. **Plant Cell Reports**, Heidelberg, v. 32, n. 7, p. 985–1006, 2013.

SAKUMA, Y.; LIU, Q.; DUBOUZET, J. G.; ABE, H.; SHINOZAKI, K.; YAMAGUCHI-SHINOZAKI, K. DNA-binding specificity of the ERF/AP2 domain of *Arabidopsis* DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. **Biochemical and Biophysical Research Communications**, New York, v. 290, p. 998–1009, 2002.

SANCHEZ-VALDEZ, I.; ACOSTA-GALLEGOS, J. A.; IBARRA-PEREZ, F. J.; ROSALES-SERNA, R.; SINGH, S. P. Registration of "Pinto Saltillo" common bean. **Crop Science**, Madison, v. 44, n. 5, p. 1865–1866, 2004.

SCHMUTZ, J.; MCCLEAN, P. E.; MAMIDI, S.; WU, G. A.; CANNON, S. B.; GRIMWOOD, J.; JENKINS, J.; SHU, S.; SONG, Q.; CHAVARRO, C.; TORRES-TORRES, M.; GEFFROY, V.; MOGHADDAM, S. M.; GAO, D.; ABERNATHY, B.; BARRY, K.; BLAIR, M.; BRICK, M. A.; CHOVATIA, M.; GEPTS, P.; GOODSTEIN, D. M.; GONZALES, M.; HELLSTEN, U.; HYTEN, D. L.; JIA, G.; KELLY, J. D.; KUDRNA, D.; LEE, R.; RICHARD, M. M. S.; MIKLAS, P. N.; OSORNO, J. M.; RODRIGUES, J.; THAREAU, V.; URREA, C. A.; WANG, M.; YU, Y.; ZHANG, M.; WING, R. A.; CREGAN, P. B.; ROKHSAR, D. S.; JACKSON, S. A. A reference genome for common bean and genome-wide analysis of dual domestications. **Nature**, London, v. 46, n. 7, p. 707–713, 2014.

SCHNEIDER, K. A.; BROTHERS, M. E.; KELLY, J. D. Marker assisted selection to improve drought resistance in common bean. **Crop Science**, Madison, v. 37, p. 51–60, 1997.

SHINOZAKI, K.; YAMAGUCHI-SHINOZAKI, K. Gene expression and signal transduction in water-stress response. **Plant Physiology**, Rockville, v. 115, p. 327–334, 1997.

SHINOZAKI, K.; YAMAGUCHI-SHINOZAKI, K. Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways. **Current Opinion in Plant Biology**, Oxford, v. 3, n. 3, p. 217–223, 2000.

SHINOZAKI, K.; YAMAGUCHI-SHINOZAKI, K. Gene networks involved in drought stress response and tolerance. **Journal of Experimental Botany**, Oxford, v. 58, n. 2, p. 221-227, 2007.

SINGH, S. P.; GEPTS, P.; DEBOUCK, D. Races of common bean (*Phaseolus vulgaris*, Fabaceae). **Economic Botany**, New York, v. 45, p. 379–396, 1991.

SINGH, S. P. Selection for water-stress tolerance in interracial populations of common bean. **Crop Science**, Madison, v. 35, n. 1, p. 118-124, 1995.

SINGH, S. P.; HAYES, R.; ROBISON, C.; DENNIS, M.; POWERS Response of bean (*Phaseolus vulgaris* L.) cultivars to drought stress. **Annual Report of the Bean Improvement Cooperative**, Prosser, WA, v. 44, p. 45–46, 2001.

SOHN, K. H.; LEE, S. C.; JUNG, H. W.; HONG, J. K.; HWANG, B. K. Expression and functional roles of the pepper pathogen-induce transcription factor RAV1 in bacterial disease resistance, and drought and salt stress tolerance. **Plant Molecular Biology**, Dordrecht, v. 61, p. 897–915, 2006.

SOTO-CERCA, B. J. S.; CLOUTIER, S. Association mapping in plant genomes. In: CALISKAN, M. (Ed.). **Genetic diversity in plants**. Rijeka, Croatia: InTech, 2012. p. 29-54.

SPONCHIADO, B. N.; WHITE, J. W.; CASTILLO, J. A.; JONES, P. G. Root growth of four common bean cultivars in relation to drought tolerance in environments with contrasting soil types. **Experimental Agriculture**, Cambridge, v. 25, p. 249-257, 1989.

TAR'AN, B.; MICHAELS, T. E.; PAULS, K. P. Genetic mapping of agronomic traits in common bean. **Crop Science**, Madison, v. 42, p. 544-556, 2002.

TERÁN, H.; SINGH, S. P. Comparison of sources and lines selected for drought resistance in common bean. **Crop Science**, Madison, v. 42, p. 64-70, 2002.

THAO, N. P.; THU, N. B. A.; HOANG, X. L. T.; HA, C. V.; TRAN, L. S. P. Differential expression analysis of a subset of drought-responsive GmNAC genes in two soybean cultivars differing in drought tolerance. **International Journal of Molecular sciences**, Basel, v. 14, n. 12, p. 23828-23841, 2013.

THORNSBERRY, J. M.; GOODMAN, J. D.; KRESOVICH, S.; NIELSEN, D.; BUCKLER, E. S. *Dwarf8* polymorphisms associate with variation in flowering time. **Nature Genetics**, London, v. 28, p. 286–289, 2001.

TSAI, S. M.; NODARI, R. O.; MOON, D. H.; CAMARGO, L. E. A.; VENCOVSKY, R.; GEPTS, P. QTL mapping for nodule number and common bacterial blight in *Phaseolus vulgaris* L. **Plant and Soil**, Dordrecht, v. 204, p. 135-145, 1998.

TSUTSUI, T.; KATO, W.; ASADA, Y.; SAKO, K.; SATO, T.; SONODA, Y.; KIDOKORO, S.; YAMAGUCHI-SHINOZAKI, K.; TAMAOKI, M.; ARAKAWA, K.; ICHIKAWA, T.; NAKAZAWA, M.; SEKI, M.; SHINOZAKI, K.; MATSUI, M.; IKEDA, A.; YAMAGUCHI, J. DEAR1, a transcriptional repressor of DREB protein that mediates plant defense and freezing stress responses in *Arabidopsis*. **Journal of Plant Research**, Tokyo, v. 122, n. 6, p. 633–643, 2009.

VERSLUES P. E.; ZHU, J. K. Before and beyond ABA, upstream sensing and internal signals that determine ABA accumulation and response under abiotic stress. **Biochemical Society Transactions**, London, v. 33, p. 375–379, 2005.

VIEIRA, C.; BORÉM A.; RAMALHO, M. A. P.; CARNEIRO, J. E. S. Melhoramento do feijão. In: BORÉM, A. **Melhoramento de espécies cultivadas**. Viçosa: UFV, 2005. p. 273-349.

WHITE, J. W.; OCHOA, R. M.; IBARRA, F. P.; SINGH, S. P. Inheritance of seed yield, maturity and seed weight of common bean (*Phaseolus vulgaris*) under semi-arid rainfed conditions. **Journal of Agricultural Science**, Cambridge, v. 122, p. 265-273, 1994.

WILSON, K.; LONG, D.; SWINBURNE, J.; COUPLAND, G. A. Dissociation insertion causes a semidominant mutation that increases expression of *TINY*, an *Arabidopsis* gene related to *APETALA2*. **The Plant Cell**, Rockville, v. 8, p. 659–671, 1996.

WORTMAN, C. S.; KIRKBY, R. A.; ALEDU, C.A.; ALLEN, D.J. Atlas of common bean (*Phaseolus vulgaris* L) production in Africa. Cali, Colombia: CIAT, 1998. 131 p.

YAMAGUCHI-SHINOZAKI, K.; SHINOZAKI, K. A novel cis-acting element in an *Arabidopsis* gene is involved in responsiveness to drought, low-temperature, or high-salt stress. **The Plant Cell**, Rockville, v. 6, p. 251–264, 1994.

YU, K.; PARK, S.; POYSA, V.; GEPTS, P. Integration of simple sequence repeat (SSR) markers into a molecular linkage map of common bean (*Phaseolus vulgaris* L.). **Journal of Heredity**, Washington, DC, v. 91, p. 429–434, 2000.

YU, J.; BUCKLER, E. S. Genetic association mapping and genome organization of maize. **Current Opinion in Biotechnology**, Oxford, v. 17, p. 155–160, 2006.

YU, X.; JIANG, Y. **Association of candidate genes with drought tolerance traits in diverse perennial ryegrass accessions**. Purdue: Purdue University Turfgrass Science Program 2012 Annual Report, 2013. p. 1-6.

ZHANG, P.; YANG, P.; ZHANG, Z.; HAN, B.; WANG, W.; WANG, Y.; CAO, Y.; HU, T. Isolation and characterization of a buffalograss (*Buchloe dactyloides*) dehydration responsive element binding transcription factor, *BdDREB2*. **Gene**, Amsterdam, v. 536, p. 123-128, 2014.

ZHAO, L.; HU, Y.; CHONG, K.; WANG, T. ARAG1, an ABA-responsive DREB gene, plays a role in seed germination and drought tolerance of rice. **Annals of Botany**, Oxford, v. 105, p. 401-409, 2010.

ZHAO, K.; SHEN, X.; YUAN, H.; LIU, Y.; LIAO, X.; WANG, Q.; LIU, L.; LI, F.; LI, T. Isolation and characterization of Dehydration-Responsive Element-Binding Factor 2C (*MsDREB2C*) from *Malus sieversii* Roem. **Plant and Cell Physiology**, Oxford, v. 54, n. 9, p. 1415-1430, 2013.

ZHOU, M. L.; MA, J. T.; ZHAO, Y. M.; WEI, Y. H.; TANG, Y. X.; WU, Y. M. Improvement of drought and salt tolerance in *Arabidopsis* and *Lotus corniculatus* by overexpression of a novel *DREB* transcription factor from *Populus euphratica*. **Gene**, Amsterdam, v. 506, p. 10-17, 2012.

ZHU, J. K. Salt and drought stress signal transduction in plants. **Annual Review of Plant Biology**, Palo Alto, v. 53, p. 247–273, 2002.

3 CATALOGUE OF THE *DREB* GENE SUBFAMILY IN COMMON BEAN: GENE CATEGORIZATION, CHROMOSOMAL LOCATION AND EXPRESSION PROFILING OF FOUR NOVEL GENES

Abstract

With increasing plant genomes available for crop plants, the genome-wide categorization of gene families is an important step to indicate potential candidates for dissecting their biological role and in favor of breeding traits of interest. *DREB* (*DEHYDRATION RESPONSIVE ELEMENT-BINDING*) gene subfamily encompasses several genes primarily involved in responses to abiotic stresses such as drought, high salinity and low temperature. However, almost no information of *DREB* genes is available for common bean (*Phaseolus vulgaris* L.), crop of high economic and nutritional value throughout the world, but constantly affected by abiotic stresses in cultivation areas. Hereby, we present the first genome-wide categorization of the putative members of the *DREB* gene subfamily in common bean. The common bean genome deposited on Phytozome was searched for all sequences containing an AP2 protein domain, the primary condition for categorizing a DREB. Following a systematic categorization, we found 54 putative common bean *DREB* genes distributed within six subgroups (A-1 to A-6) along the 11 chromosomes of common bean. A Blast2Go annotation predicted that all sequences have DNA-binding activity and involvement in abiotic stress responses. Of all sequences, we introduce four novel *DREB* genes hereafter named *PvDREB1F*, *PvDREB2A*, *PvDREB5A* and *PvDREB6B*. A temporal and spatial analysis under stress-treatments (dehydration, salinity, low temperature and abscisic acid) showed different patterns of expression for each gene in the stress-adapted genotype BAT 477. *PvDREB1F* and *PvDREB5AF* showed high inducibility under all treatments in roots, stem and leaves. Conversely, *PvDREB2A* was predominantly induced by dehydration and cold in the stem, while *PvDREB6B* expression was essentially modulated by cold and dehydration in leaves. A spatial comparison analysis among five genotypes with different backgrounds showed similar patterns of expression, although with some particular responses. With this work, we outline potential candidate genes and directions for functional analyses as well as association mapping with traits of interest in common bean.

Keywords: Transcription factors. RT-qPCR. Motifs prediction. Phylogeny. Abiotic stresses.

Resumo

Com o aumento do número de genomas disponíveis para plantas de importância agronômica, a categorização de famílias gênicas em escala genômica representa uma etapa importante para indicar potenciais candidatos para o entendimento de suas funções biológicas e em favor de programas de melhoramento para características de interesse. A família *DREB* engloba genes envolvidos principalmente em respostas a estresses abióticos como seca, salinidade e baixa temperatura. No entanto, não há informações suficientes sobre estes genes no feijoeiro (*Phaseolus vulgaris* L.), cultura de ampla importância econômica, mas constantemente afetada por estresses abióticos em áreas de cultivo. Desse modo, neste trabalho foi realizada a primeira categorização em nível genômico dos potenciais membros da subfamília *DREB* em feijoeiro. Todas as sequências contendo um domínio AP2, primeira condição para se considerar um gene *DREB*, foram procuradas no genoma do feijoeiro depositado no Phytozome. De maneira sistemática, foram categorizados 54 prováveis genes *DREB* distribuídos entre seis subgrupos (A-1 a A-6) e pelos 11 cromossomos do feijoeiro. A anotação destas sequências por Blast2Go revelou que todas estão envolvidas em atividade de ligação a DNA e em vias de estresses abióticos. De todos os genes, quatro novos foram isolados e nomeados *PvDREB1F*, *PvDREB2A*, *PvDREB5A* e *PvDREB6B*. Análise temporal e espacial sobre estresses abióticos (desidratação, salinidade, baixa temperatura e ácido abscísico) mostraram diferentes padrões de expressão para cada gene no genótipo BAT 477, adaptado a estresses. *PvDREB1F* e *PvDREB5A* mostraram elevada inducibilidade sob todos os tratamentos em raízes, caule e folhas. Contudo, *PvDREB2A* foi induzido principalmente no caule por desidratação e baixa temperatura, enquanto *PvDREB6B* foi predominantemente induzido por baixa temperatura e desidratação em folhas. Uma análise espacial comparativa entre cinco genótipos de origens distintas mostrou perfis similares de expressão, mas peculiaridades foram observadas. Com este trabalho foram indicados potenciais genes candidatos e direções para análises funcionais assim como mapeamento de associação com características de interesse no feijoeiro.

Palavras-chave: Fatores de transcrição. RT-qPCR. Predição de motivos. Filogenia. Estresses abióticos.

3.1 Introduction

Climate change has been intensively discussed as it may cause long-term changes in the way living organisms interact with their environment. Pulse crops productivity might be negatively affected, bringing new challenges for breeding traits to improve their performance. Extreme weather conditions will probably be much more frequent and intense, such as longer periods of drought or flooding episodes, and heat or cold waves. Moreover, changes in soil salinity and fertility, whether caused by environmental changes or human interference, represent significant challenges for agriculture.

Abiotic stresses have a negative effect on plants, limiting their growth and survival. Abiotic stresses intensify the synthesis of abscisic acid, leading to stomatal closure and thus reducing the activity of the photosynthetic apparatus (HASANUZZAMAN et al., 2013; DASZKOWSKA-GOLEC; SZAREJKO, 2013). As a defense mechanism, several genes are induced for protecting plants against adversities. One category of genes primarily works for mechanic and osmotic adjustment, while another set in involved in a series of regulation processes for developing tolerance to the stresses. The latter group comprises several stress-inducible genes such as *NAC*, *bZIP*, *leucine-rich repeats* (*LRR*) and *EREBP/AP2* (BRAY; BAILEY-SERRES; WERETILNY, 2000).

EREBP/AP2 proteins constitute a large superfamily, which has been divided into three families (AP2, RAV and ERF) based on sequence similarity and the number of EREBP/AP2 domains (NAKANO et al., 2006). AP2 proteins contain two EREBP/AP2 domains and are involved in the regulation of developmental processes (ELLIOTT et al., 1996; CHUCK; MEELEY; HAKE, 1998; BOUTILIER et al., 2002). RAV family contains an EREBP/AP2 domain and a B3 domain. *RAV* genes expression is influenced by ethylene (ALONSO et al., 2003), brassinosteroids (HU et al., 2004) and biotic and abiotic stresses (SOHN et al., 2006).

The ERF family contains only one EREBP/AP2 domain and two subfamilies named CBF/DREB and ERF. The amino acids at positions 14th and 19th from the beginning of EREBP/AP2 domain sequence distinguish DREB (in general valine and glutamic acid, respectively) from ERF (normally alanine and aspartic acid, respectively) proteins (SAKUMA et al., 2002). ERF proteins are primarily involved in responses to biotic stresses, such as pathogenesis, by recognizing the AGCCGCC *cis*-regulatory element, known as GCC box (HAO; OHME-TAKAGI; SARAI, 1998). On the other hand, DREB proteins have crucial role in the response of plants to abiotic stresses by recognizing the dehydration responsive element (DRE), which consists on the conserved motif A/GCCGAC (YAMAGUCHI-SHINOZAKI; SHINOZAKI, 1994). This element has been found essential

for gene regulation due to dehydration (YAMAGUCHI-SHINOZAKI; SHINOZAKI, 1994) but since then, it has also been found in the promoter region of other drought-, salinity- and cold-inducible genes (LIU et al., 1998; SAKUMA et al., 2002; SAKUMA et al., 2006).

DREB genes are usually divided into six subgroups (A-1 - A-6). The general trend observed in *Arabidopsis* is that *DREB1/CBF* (A-1) are induced by low temperature, while *DREB2* genes (A-2) are involved in responses to osmotic stress (dehydration and salinity) (SAKUMA et al., 2002; NAKASHIMA; ITO; YAMAGUCHI-SHINOZAKI, 2009). Studies have shown the members of A-1 and A-2 subgroups are not mediated by ABA-signaling in *Arabidopsis*. On the contrary, *ABI4*, the only member of the A-3 subgroup is involved in ABA and sugar signaling, lipid mobilization in embryo and germinating seeds, chloroplast functioning and retrograde signaling (SHKOLNIK-INBAR; BAR-ZVI, 2011). The most studied members of the A-4 subgroup are *TINY*, which has been shown to be slightly cold-responsive (SAKUMA et al., 2002) and *HARDY*, with low stress inducibility, but with potential for augmenting water use efficiency when overexpressed in rice (NAKANO et al., 2006; KARABA et al., 2007). Moreover, studies suggested that A-4 genes are involved in the cross-talk between abiotic and biotic stress responses by connecting DRE- and ERE-(ethylene-responsive element) mediated signaling pathways (YU et al., 2006; SUN et al., 2008). Genes of A-5 group from *Arabidopsis*, such as *RAP2.1*, exhibit induction by drought and cold stress (DONG; LIU, 2010). In group A-6, *RAP2.4* (salt- and drought-responsive) (LIN; PARK; WANG, 2008) and *RAP2.4B* (heat-responsive) (RAE; LAO; KAVANAGH, 2011) are the most studied.

Besides *Arabidopsis*, *DREB* genes have been isolated and characterized in several other plants such as *Glycine max*. As studies evolved, some *DREB* genes either from A-1 and A-2 subgroups were found to be affected by ABA in other plants. In soybean, *GmDREBa* (A-2) and *GmDREBb* (A-6) were found induced by salinity, drought, and low temperature. *GmDREBc* (A-2) was not affected significantly in leaves, however, was induced in roots by drought, salinity and abscisic acid treatment (LI et al., 2005). Chen et al. (2007) isolated *GmDREB2* (A-5), which was induced by drought, high salinity, low temperature and abscisic acid treatment. Overexpression in *Arabidopsis* lines showed increased tolerance to drought and salinity. Chen et al. (2009) isolated *GmDREB3* (A-5) and found that it was only induced by low temperature and that its overexpression increased tolerance to cold, drought and high salinity.

With the recent release of several plant genomes, genome-wide analyses have been performed to identify *ERF* and *DREB* genes of many species such as *A. thaliana* (NAKANO

et al., 2006), *G. max* (ZHANG et al., 2008), *Malus domestica* (ZHAO; LIANG; WANG, 2012), *Zea mays* (LIU et al., 2013), *Brassica rapa* (SONG; LI; HOU, 2013), *Vitis vinifera* (ZHAO et al., 2014), *Brassica oleracea* (THAMILARASAN et al., 2014), *Setaria italica* (LATA et al., 2014), *Eucalyptus grandis* (CAO et al., 2015), *Salix arbustifolia* (RAO et al., 2015) and *Phyllostachys edulis* (WU et al., 2015).

Nevertheless, no such genomic information is available for the legume common bean (*Phaseolus vulgaris* L.), the most important grain legume for consumption, grown mainly in Latin America, Africa and Asia. Bean grains are widely consumed in these areas since they are a source of proteins, vitamins, and minerals with crucial role in nutrition (BROUGHTON et al., 2003; BEEBE et al., 2013). Several small-farmers consider beans as a complimentary if not the basic food source, especially in Latin America and Africa. However, its production has been severely impaired by a lot of abiotic factors such as constant drought episodes (BEEBE et al., 2013), soil salinity (GAMA et al., 2009), low or high temperatures (RAINEY; GRIFFITHS, 2005) and nutritional deficiencies (MIKLAS et al., 2006). This problem becomes even harder to deal with since most producers have low technology or no systems for proper irrigation and soil management (BROUGHTON et al., 2003). It urges the development of efficient strategies toward breeding for abiotic stress tolerance improvement in common bean varieties throughout the world.

The release of the common bean genome (SCHMUTZ et al., 2014) has opened multiple possibilities for studying the molecular mechanisms involved in the responses to abiotic stresses in the species. In this work, we were concerned with the identification of *DREB* genes along the genome sequence of common bean. Thereby, we performed an *in silico* search of proteins containing the AP2 domain on Phytozome database. Out of all sequences, we classified and separated only putative DREB members. Alignments, phylogenies and motif predictions were generated to designate such proteins/genes, following several criteria based on *Arabidopsis* original categorization as well as *G. max* and other model plants. Furthermore, we isolated four novel *PvDREB* genes and analyzed their expression profiles under dehydration, salinity, low temperature and abscisic acid treatment. Expression analysis was performed in a temporal (five time periods) and spatial scale (leaves, stem and roots) in one common bean genotype. Further comparison of spatial expression profiles were performed among five genotypes with contrasting tolerance levels to abiotic stresses. This work is the first report of a complete categorization for *DREB* genes in common bean.

3.2 Material and Methods

We performed a genome-wide categorization of the *DREB* gene family in common bean by following five basic criteria. First, we checked for the presence of one AP2 conserved domain along the protein structure. Second, ortholog relationships were verified among common bean, *A. thaliana* and *G. max* AP2-containing sequences and *DREB* genes with defined nomenclature. Next, amino acid conservation was verified along the AP2 domain of predicted protein sequences. The fourth verification consisted of a phylogenetic analysis and subgroups division establishment. Ultimately, conserved motifs were searched along the peptide sequences of all putative DREB.

3.2.1 *In silico* search for AP2/ERF proteins and annotation

The first step was to identify all ERF proteins, which presumed the presence of at least one AP2 conserved domain. Phytozome (http://www.phytozome.net) (GOODSTEIN et al., 2012) and GenBank (http//www.ncbi.nlm.nih. gov/genbank) databases were accessed to search all *P. vulgaris* peptide sequences containing the AP2 domain. The database showed 181 unigenes matching this search (Supplementary Table 3.1). All peptides were doubled checked on PFAM (http://pfam.sanger.ac.uk/) and SMART (http://smart.embl-heidelberg.de/) for the presence of AP2 or other domains. We only considered for further analysis those peptides presenting a single AP2 domain, which features one basic aspect of the DREB family. Moreover, a local BLASTp was performed against *G. max* and *A. thaliana* with an *E-value* cutoff of 1×10^{-5} to identify domains and possible orthologs of *DREB* genes (Supplementary Table 3.1).

3.2.2 Alignment and phylogenetic analyses

Alignments and phylogenetic analyzes were performed with full-length protein sequences from common bean. CLC Sequence Viewer version 6 (//http: http://www.clcbio.com/) software was used for sequences alignment, considering default parameters. Global alignment was performed with ClustalW tool, and phylogenetic trees were generated using MEGA 6.0 (TAMURA et al., 2013) by neighbor-joining algorithm with bootstrap analysis with 1000 permutations.

Moreover, all AP2 domain-containing sequences from *P. vulgaris*, *A. thaliana* and *G. max* were downloaded from Phytozome. In total, Nakano et al. (2006) categorized 57 ERF sequences as DREB genes in *Arabidopsis* genome and those were retrieved from TAIR (http://www.*Arabidopsis*.org/). Zhang et al. (2008) classified 98 *ERF* genes in soybean, from

which 36 featured *DREB* genes. However, the analyzes were performed earlier than the release of the genomic sequence of *G. max* (SCHMUTZ et al., 2010). In the current version of the soybean genome (available on Phytozome) 362 sequences are encountered matching the AP2 domain search. Furthermore, the core sequences for DREB proteins already isolated and characterized for *Arabidopsis* and soybean were accessed on Genbank. All these sequences were aligned and grouped with the neighbor-joining algorithm in order to verify which common bean sequences were more similar to the DREB proteins already known for the other species (data not shown in thesis, but presented on derived scientific article).

3.2.3 Alignment of the AP2 domain and amino acid conservation

All putative DREB proteins were aligned to verify if they had conserved the positions 14^{th} and 19^{th} from the beginning of the AP2 domain (positions were determined based on the original *Arabidopsis* sequences), respectively with valine (V) and glutamic acid (E), which have been shown to be essential for DREB proteins binding-specificity, especially valine (SAKUMA et al., 2002). We considered valine (14^{th}) conservation as the prime condition, and sequences not following this criterion were discarded from the analysis. A new alignment and phylogenetic analysis were performed to confirm if the sequences matched orthologs from *G. max* and *Arabidopsis*.

3.2.4 Categorization of putative DREB members into subgroups

The phylogenetic tree with all AP2-containing sequences from common bean, soybean and *Arabidopsis* was analyzed to categorize all common bean DREB proteins into six subgroups (A-1 to A-6), based on Sakuma et al. (2002), Nakano et al. (2006) and Zhang et al. (2008). Subgroups were validated through an analysis of conserved motifs shared among sequences within the same group.

3.2.5 Proteins motifs search

Conserved motif search was performed with MEME tool (BAILEY et al., 2009). Motif search criteria were based on previous studies (NAKANO et al., 2006; ZHANG et al., 2008; ZHAO; LIANG; WANG, 2012), but we also determined the threshold for motifs detection as the maximum number of motifs that could be detected without having significant similarity to each other.

3.2.6 *In silico* mapping

The genomic positions of the *PvDREB* genes were placed on a map that has been created using MapDraw (LIU; MENG, 2003). Gene positions were checked on Phytozome and those that were not separated by more than five gene loci over a 100 kb were considered tandem duplicates (ZHAO et al., 2012).

3.2.7 Gene Ontology (GO) annotation and Phyto Mine expression profiling

The GO annotation of the putative *DREB* genes was investigated through Blast2GO (CONESA; GÖTZ, 2008). Expression profiles for each gene were obtained based on FPKM (fragments per kilobase of transcript per million mapped reads) values retrieved from Phyto Mine on Phytozome, searching the *P. vulgaris* genome. The Phyto Mine database shows a series of RNA-seq da obtained for several plant organs in common bean (flower buds, pods, roots, leaves, stem, flowers, nodules, young trifoliates). All negative FPKM were considered as zero expression values or nearly undetectable transcripts. A heat map was drawn in R, using the package gplots 2.17.0 and the function heatmap.2 (WARNES et al., 2015).

3.2.8 The novel genes *PvDREB1F, PvDREB2A*, *PvDREB5A* and *PvDREB6B*

We used *DREB* genes already characterized in *A. thaliana* (AT2G40220 - *ABI4;* AT5g11590 - *TINY2;* AT1G46788 - *RAP2.1;* AT2G36900 - *RAP2.10;* AT2G40340 - *DREB2C;* AT3G11020 - *DREB2B;* AT5GO5410 - *DREB2A;* AT4G25480 - *DREB1A/CBF3;* AT5G51990 - *DREB1D/CBF4;* AT4G25490 - *DREB1B/CBF1;* AT4G25470 - *DREB1C/CBF2;* AT1G22190 - *RAP2.4;* AT1G78080 - *RAP2.4/WIND1*) to generate a gene expression profile compilation based on microarray data provided by Genevestigator analytical tool (http://www.genevestigator.ethz.ch/). Data were showed as heat maps in red/green coding, which were represented by log ratios (red representing up-regulation and green down-regulation - probe sets in a 22k Affymetrix GeneChip) (Supplementary Figure 3.1). Ortholog genes in *P. vulgaris* were searched. In this step, genes representing the *DREB* subfamily were chosen for expression profiling.

Of all sequences, four novel genes were cloned and referred as *PvDREB1F*, *PvDREB2A*, *PvDREB5A* and *PvDREB6B*, based on the phylogeny. Gene-specific primers (Supplementary Table 3.2) were designed to clone the entire coding region of the four *DREB*. Genes were cloned with a template DNA from the common bean genotype BAT 477. DNA was extracted using a modified CTAB extraction protocol from Doyle and Doyle (1990). PCR was prepared to 25 µl containing 1 X PCR reaction buffer (20 mM Tris-HCl pH 8.4,

50 mM KCl), 1.6 mM $MgCl_2$, 0.12 mM dNTP, 0.2 µM of each primer, 1 U of Taq DNA polymerase (Invitrogen™). Amplification conditions were set as follows: initial denaturation at 94ºC for 2 min; 35 cycles at 94ºC for 30 s (denaturation), 59ºC for 30 s (primer annealing) and 72ºC (extension of fragments); and final extension at 72ºC for 7 min. Fragments were gel-purified using GFX purification kit (GE Healthcare) and cloned into p-GEM vector (Invitrogen) with thermo-competent JM109 *E. coli* cells. Transformed colonies were analyzed by blue/white plaque assays, cultured in circle-growth medium and purified. Sequencing was performed in ABI PRISM® 3130xl Sequencer equipment. Sequences were annotated using BLASTn, BLASTx and BLASTp tools (NCBI/BLAST). Sequences were aligned to those available on Phytozome as a final check for their identity.

3.2.9 Plant materials and stress treatments

Two sets of experiments were carried out for gene expression analyses (*PvDREB1F*, *PvDREB2A*, *PvDREB5A* and *PvDREB6B*) using RT-qPCR in common bean genotypes of different backgrounds: Mesoamerican (derived from crosses among genotypes originated from areas spanning Central America and Colombia) and Andean (derived from genetic materials from the Andes). The first experiment consisted of a temporal (five time-periods) and spatial (roots, stem and leaves) analysis of the *PvDREB* transcripts, using the common bean genotype BAT 477 submitted to abiotic stresses. In the second experiment, besides BAT 477, four other genotypes (BAT 93, Jalo EEP558, IAC-Carioca 80SH and RAB 96) were included. A spatial analysis (roots, stem and leaves) of relative gene expression was performed under the same stress-treatments, but with only one time-period of stress induction (three hours).

The genotype BAT 477 has been used in several studies aimed at screening drought performance. BAT 477 has a Mesoamerican background and was developed at Centro Internacional de Agricultura Tropical (CIAT, Colombia), being used as the drought-tolerant parental line of the mapping population BAT477 x DOR364, which showed QTLs associated with drought in common bean (BLAIR et al., 2012). Previous reports have also shown BAT 477 is considerably adapted to drought conditions (SPONCHIADO et al., 1989; WHITE et al., 1994).

For the first set of experiments, seeds of BAT 477 were surface sterilized in 10% sodium hypochlorite for 3 min and rinsed 3-4 times (1 min each time) in distilled water. Plants were grown in pots with sand/vermiculite (1:1) in a growth chamber at 26ºC ±3 (14 h photoperiod, ~ 60% air moisture, light intensity of 120 µmol.m^{-2}.s) and were normally

watered until first trifoliate leaf was completely expanded (after about 21 days, referred as Vegetative 3 – V3 stage). After that, whole plants were removed from pots and subjected to four stress-treatments: polyethylene glycol (PEG 10%) solution for dehydration stress; NaCl solution (250 mM) for salt stress; 100 µM abscisic acid (ABA) solution (Supplementary Figure 3.2). ABA treatment was included since many reports have shown some *DREB* might be ABA-independent or dependent (LIU et al., 1998; SAKUMA et al., 2002; LATA and PRASAD, 2011), but our study only aimed at showing the responsiveness to such treatment. Furthermore, three plants were incubated in a cold chamber (4ºC). Control treatment consisted in plants placed in distilled water for comparison with the effects of PEG, NaCl and ABA while they were kept in pots only irrigated at room temperature for comparison with the cold-treated plants. Treatments were applied for different periods of time (0 to 12 h), and three plants were collected at each time point for each treatment. All treatments were considered as independent experiments. After each treatment, roots, stem and leaves were separately placed in tubes properly identified and kept in liquid nitrogen until being transferred to ultra freezer (-80ºC).

In the second set of experiments, seeds from the Mesoamerican genotypes BAT 477, BAT 93, IAC-Carioca 80SH and RAB 96, and the Andean Jalo EEP558 were treated in a similar manner than in the first assay. BAT 93 and Jalo EEP558 are the parental lines from the core mapping population of common bean (FREYRE et al., 1998; HANAI et al., 2010). IAC-Carioca 80SH in a drought-sensitive cultivar (RECCHIA et al., 2013) as well as the breeding line RAB 96 (GUIMARÃES; STONE; BRUNINI, 2006). The same four treatments (PEG 10%, NaCl 250 mM, low temperature and ABA 100 µM) were applied to three plants of each genotype. However, all stresses were induced for a three-hour period, an intermediate point selected based on the first experiment. Samples were all collected separately and frozen.

Before sampling, however, in order to give indications that plants were effectively suffering from the imposed stress conditions, we determined the relative water content for all samples. Fully expanded leaves were excised and fresh weight (FW) was recorded; then leaves were soaked in deionized water for 4 hours and turgid weight (TW) was recorded. All samples were placed in an air oven at 60ºC and total dry weight (DW) was recorded after 24 h. RWC was calculated according to Barrs and Weatherley (1962): RWC (%)=[(FW–DW)/(TW–DW)]x100. Furthermore, as a biochemical parameter, Catalase (CAT) activity was determined for each sample. Leaf samples of each treatment were frozen and ground for determinations. A 100 mg leaf tissue sample was used for protein extraction in phosphate solution pH 7.0 with antioxidants (PVPP). Protein quantification was performed using

Bradford reagent (BIORAD) and following the procedures of Bradford (1976). For CAT assay, 100 µL of each protein sample were placed in a cuvette with 3 mL of phosphate buffer and 60 µL of H_2O_2 30% solution were added. Absorbance decrease was monitored with Nanodrop™ 2000c (Thermo Scientific) spectrophotometer for 2 min, with measures at each 10 s. Results were expressed in $\mu.mol.min^{-1}mg^{-1}$ of protein.

3.2.10 RNA extraction and gene expression profiling with RT-qPCR

RNA extraction was performed with 100 mg tissue samples using TRIzol® Reagent (Invitrogen™) and following manufacturer's instructions. Quantification and quality were checked with NanoDrop™ 2000c (Thermo Scientific) spectrophotometer. Gel electrophoresis (agarose 1% in TAE buffer 1 X) was also performed for quality assay. Primers flanking 150-250 bp length across the coding region of the four genes were designed for qRT-PCR (Supplementary Table 3.2). Two reference genes were used for reactions, being chosen according to tissue analyzed and on gene stability, previously studied (Supplementary Table 3.2, BORGES; TSAI; CALDAS, 2012). A 100 ng RNA sample of each treatment was used to synthesize the first cDNA strand using Maxima First Strand cDNA Synthesis kit (Fermentas). Quantitative PCR reactions were prepared using 1 µL of newly synthesized cDNA, 0.25 µM of each primer and 1X SYBR® Green PCR Master Mix (Thermo Scientific). Amplifications were performed on StepOnePlus™ Real Time PCR System (Applied Biosystems) equipment with the following steps: 10 min at 95ºC, 40 cycles of cDNA amplification at 95ºC for 15 s, 59ºC for 30 s and 72ºC for 20 s with fluorescence signal recording at this stage. A final step at 95ºC for 15 s and 60ºC for 1 min, with fluorescence measurements at each 0.7ºC variation (from 60 to 95ºC), was included to obtain the melting curve. All reactions were performed in triplicates.

Raw data with fluorescence levels were submitted to LinRegPCR software (RAMAKERS et al., 2003). Fluorescence was baseline corrected and linear regression analysis was performed for all amplifications. The optimal set of data points (Window-of-Linearity) was defined to allow the calculation of the threshold and quantification cycle (Cq). Samples efficiencies were calculated based on the slope of the line, ranging from 1.8 to 2.0 and with correlation of at least 0.995. Relative expression data were obtained by REST software (PFAFFL; HORGAN; DEMPFLE, 2002) using average values of efficiency and Cq of target and reference genes. This software calculates the concentration of expression (C) by comparing control and treated Cq values and determines the relative expression (RE)

ratio: RE = $C_{target\ gene}$/geometric average $C_{reference\ gene}$. After p-values are obtained by a pairwise reallocation randomization test (bootstrap = 2,000 permutations).

3.3 Results

3.3.1 DREB proteins categorization in common bean

Searching the common bean sequence available on Phytozome, we found 181 unigenes containing at least one AP2 domain. After annotation on NCBI/BLAST and Pfam, three sequences contained one AP2 and one B3 domain (Phvul.003G111800.1, Phvul.007G102800.1 and Phvul.007G002900.1), which fits the basic feature of a RAV protein (ALONSO et al., 2003). Other 20 sequences contained between two and four AP2 domains, similar to AP2 proteins (BOUTILIER et al., 2002). The sequence Phvul.001G131300.1 revealed an AP2 superfamily domain and was annotated as an AP2-like ERF in soybean (e-value = 0). All the remaining 157 sequences had only one AP2 domain, as resulted from the annotation. Thereby, 157 putative *ERF* genes were found distributed along the 11 chromosomes of common bean. Their annotation with an e-value cutoff of 1×10^{-5} provided an initial idea of the putative DREB sequences, but further phylogenetic analysis strengthened the categorization.

Phylogenetic analysis encompassed 181 common bean protein sequences aid by 147 from *A. thaliana* and 359 from *G. max*, and sequences from NCBI, Genbank and *AtDREB* genes deposited on TAIR. Neighbor-joining algorithm was used to group all sequences and generate phylogenetic trees. In general, the analysis showed high homology between two soybean sequences for each one of common bean, which could be explained by the duplication event of the allotetraploid *G. max* after divergence from *P. vulgaris* (SCHMUTZ et al., 2010). In general, one or more sequences from *A. thaliana* were in the same clade as orthologs from soybean and common bean (data not shown). For instance *AtDREB1A*, *AtDREB1B* and *AtDREB1C* were in the same clade as Phvul.004G122000 and Phvul.007G066500. Two soybean sequences were in the same clade for each common bean protein (Glyma09g27180 and Glyma16g32330 with the former; and Glyma10g38440 and Glyma20g29410 with the latter).

In general, sequences from *A. thaliana* fitted their original categorization from A-1 to A-6 subgroups of *AtDREB* (NAKANO et al., 2006), enabling subgroup division for the common bean sequences (data not shown). *AtDREB1A*, *AtDREB1B* and *AtDREB1C* were in the same clade in group A-1. *AtDREB2A, AtDREB2B* and *AtDREB2C* were in the same group as the previously isolated genes *GmDREBa* and *GmDREBc* (LI et al., 2005). A sequence from

P. acutifolius predicted as a *DREB2C*-like was also included in the same group. The only member of the A-3 group in *Arabidopsis* (*AtABI3*) was grouped with two genes from *G. max* and only one from common bean (Phvul.008G222400). Members of the group A-3 were closely related to A-2 (NAKANO et al., 2006). *AtTINY* (A-4) was in the same group as *GmTINY*. *AtRAP2.9, AtRAP2.10* and *AtRAP2.1* were in the same group as *GmDREB2* (CHEN et al., 2007), all from A-5 subgroup. Genes *AtRAP2.4* and *AtRAP2.4B* were also in the same clade (A-6) and the soybean gene *GmDREBb* (LI et al., 2005) fitted in the same phylogenetic group. All common bean sequences within each of the phylogenetic groups (determined based on *Arabidopsis* and soybean) were considered as putative candidates for the *DREB* gene subfamily, giving in total 57 sequences (data not shown).

Another criterion to define the putative DREB proteins was the amino acid conservation along the AP2 domain. Sakuma (2002) demonstrated that *DREB* genes have the positions 14 and 19 from the beginning of the domain conserved, respectively with valine and glutamic acid. However, the 19th amino acid might have some variability among proteins. We extracted the AP2 domain sequence from all sequences on NCBI Domain Finder and performed Clustalw global alignment with the 57 sequences categorized from the phylogenetic tree (Figure 3.1). Out of the total, 54 sequences presented the amino acid valine at the 14th position of the AP2 domain, while the 19th site was represented by glutamic acid in all A-2, A-3 and A-4 members and 14 proteins from A-1 subgroup. Two peptide sequences from A-1 subgroup presented valine at this site (Phvul.003G212800 and Phvul002G153900) while one sequence had glutamine (Phvul.007G222500). Valine has been pointed out as the most important amino acid for binding-affinity at the 14th site (SAKUMA et al., 2002). The other three sequences (Phvul. 006G179800, Phvul.003G292400 and Phvul.008G131500) had different amino acids at the 14th site (alanine and glutamine) and uncommon amino acids for DREB at the 19th (aspartic acid and valine) and, therefore, they were excluded from the list of DREB proteins (Figure 3.1). In fact, alanine and aspartic acid are typical from ERF sequences (SAKUMA et al., 2002).

From our analyses, 54 putative DREB proteins were categorized, fitting subgroups A-1 to A-6, according to previous analyses with *A. thaliana* (NAKANO et al., 2006) and *G. max* (ZHANG et al., 2008). Figure 3.2 shows the final neighbor-joining phylogenetic tree with the putative *PvDREB* genes. Subgroups A-1 and A-4 were phylogenetically more similar as earlier reports have shown for other species (NAKANO et al., 2006, ZHANG et al., 2008; ZHAO et al., 2012). Each group presented different number of proteins (A-1 = 17; A-2 = 8; A-3 = 1; A-4 = 10; A-5 = 10; A-6 = 8).

Figure 3.1 - Alignment of the AP2 domain of 54 putative DREB proteins from common bean. Position 14th and 19th, described as important for protein binding, are separated by spaces along the sequences. Position 14th presents 100% conservation of the amino acid Valine (V), while the 19th varies, although Glutamic acid (E) and Leucine (L) are the most frequent. Other sequences than the 54 DREB were compared in the alignment. Those sequences fitted as putative DREB in the phylogenetic analysis, but not for amino acid conservation.

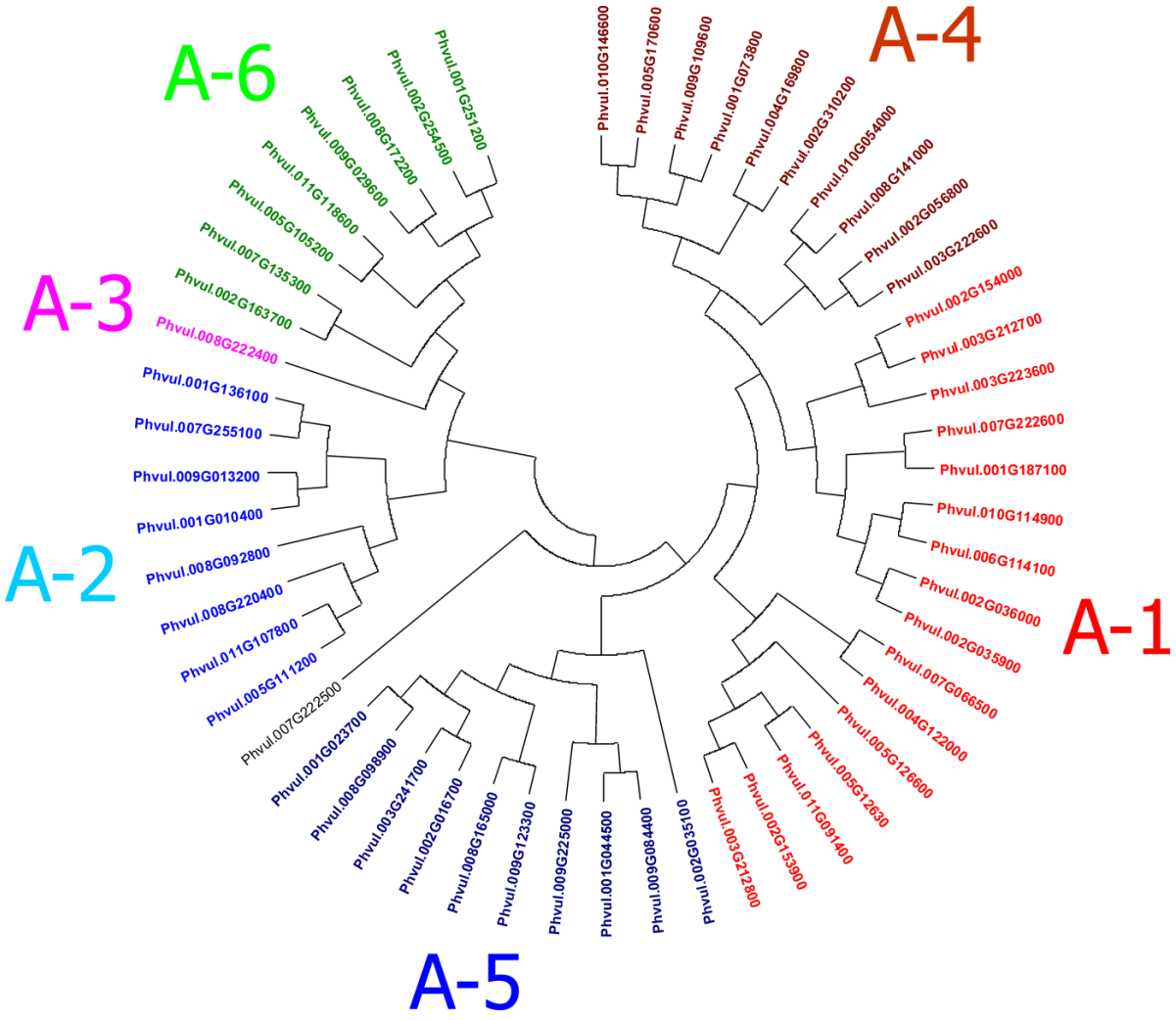


Figure 3.2 – Neighbor-joining tree of 54 putative PvDREB protein sequences. Sequences were retrieved from the common bean (*Phaseolus vulgaris*) genome database on Phytozome. Subgroups of DREB proteins are shown in different colors. Phvul.007G222500 was categorized as an ERF protein and was used as an outlier.

3.3.2 Protein motifs indicate differences among the subgroups of the DREB subfamily

After categorization and subgroup division, a protein motif prediction search was performed with MEME Suite for the 54 putative DREB proteins. In this tool, users define the number of motifs to be searched against databases (BAILEY et al., 2009). Our criterion to define the maximum number of motifs was based on determining the number before there were no similarities or redundancies among motifs in the list. We determined 14 motifs (Supplementary Table 3.3), represented in Figure 3.3.

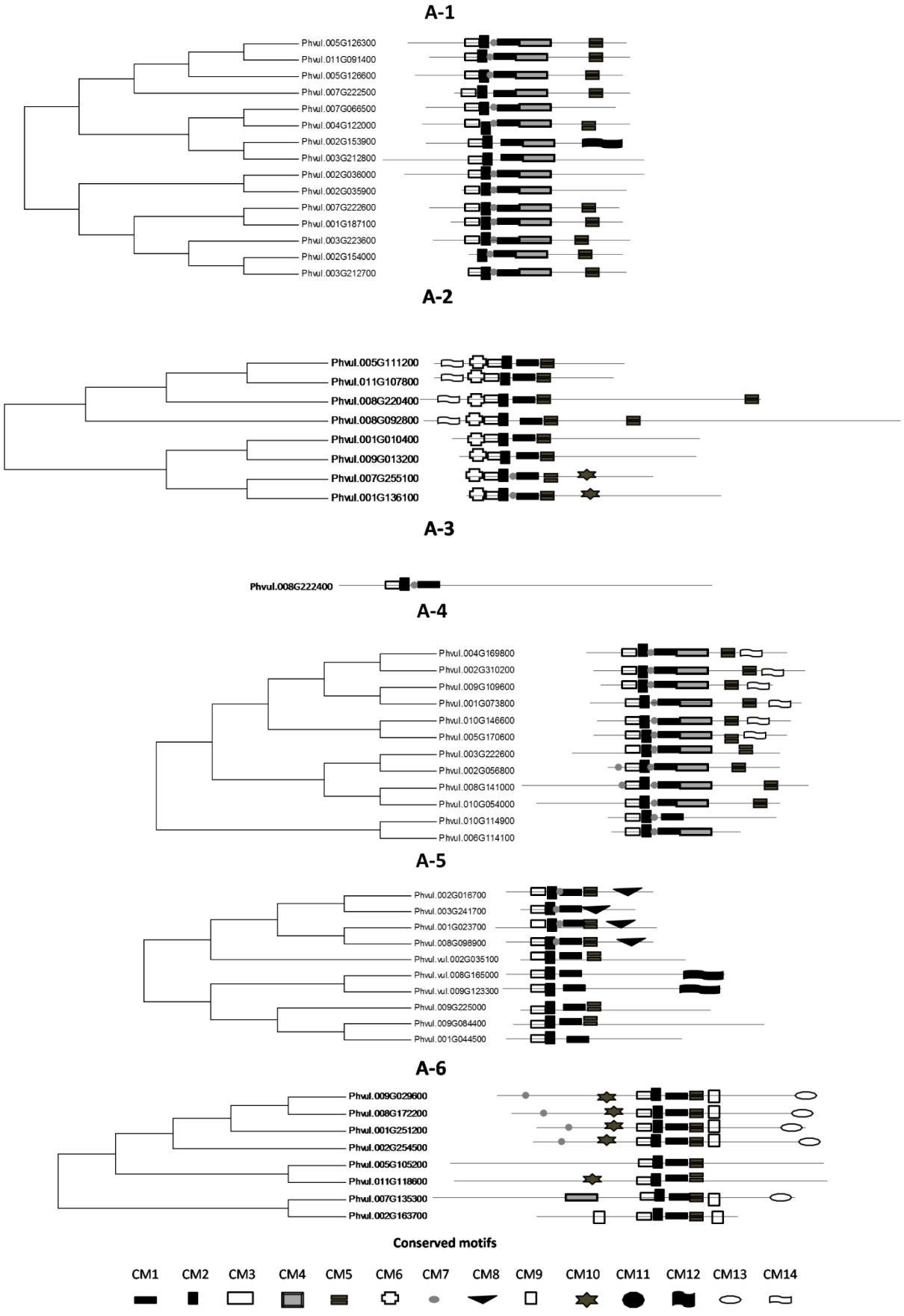


Figure 3.3 - Conserved motifs along the amino acid chain of 54 putative PvDREB sequences divided by subgroups.

Motifs CM1 (RIWLGTFPTPEMAARAYDVAAYCLKG), CM2 (WGKWVCEIR), CM3 (GGPENRHCVYRGVRQR) and CM7 (EPRKK) represented conserved sequences within the AP2 domain (Figure 3.3). All sequences had CM1, CM2 and CM3. On the other hand, CM7 was detected in all DREB subgroups, but not all sequences. In A-6 subgroup, CM7 was observed in four sequences (Phvvul.009G029600, Phvul.008G172200, Phvul.001G251200, Phvul.002G254500) closer to the N-terminus position.

Some motifs were exclusive to specific DREB subgroups (Figure 3.3). Motif CM6 (KKVPAKGWKKGCMRGK) was unique to all sequences from A-2 subgroup. Motif CM8 (DMSADSIRKKATQVGARVDALQTALHHH) was only encountered in four sequences (Phvul.002G016700, Phvul.003G241700, Phvul.001g023700 and Phvul.008G098900) of A-5 subgroup. Another example is the motif CM13 (YWEDDSDHFNLQKYPSYEIDW), only detected in five DREB proteins (Phvul.009G029600, Phvul.008G172200, Phvul.001G251200, Phvul.002G254500 and Phvul.007G135300) from the A-6 subgroup. Motif CM10 (LNHLTPPQVHQIQAQIQIQKQ) was only detected within A-6 sequences as well. Motif 14 (HSKGDGSKSVADTLAKWKEYNAQL) was found in A-2 and A-4 subgroups, but in different positions along the peptide sequence (near N-terminus in A-2 and near C-terminus in A-4).

Exclusive motifs indicated to be good phylogenetic markers for defining DREB subgroups. They might have specific functional roles for each one of the genes. Therefore, motifs identification and categorization of this work are important for further steps aimed at the molecular and functional characterization of *DREB* genes from common bean.

3.3.3 Chromosomal distribution of the *PvDREB* genes

The chromosomal location of all the putative 54 *PvDREB* genes is represented on the map in Figure 3.4. Gene distribution along chromosomes is not separated by *DREB* subgroup (A-1 to A-6), although some groups of genes within the same subgroup were observed such as members of A-1 on chromosomes 2, 3 and 5. Six pairs of genes (Phvul.002G035900 and Phvul.002G036000; Phvul.002G153900 and Phvul.002G154000; Phvul.003G212700 and Phvul.003G212800; Phvul.003G222600 and Phvul.003G223600, Phvul.005G126300 and Phvul.005G126000; Phvul.007G222500 and Phvul.G222600) indicate to be tandemly duplicated (Figure 3.4) since they are located within a distance inferior to 100 kb and are not separated from more than five genes (ZHAO et al., 2012). In general, duplications happened with genes from A-1 subgroup, with one exception between one A-1 and one A-4 gene (Figure 3.4).

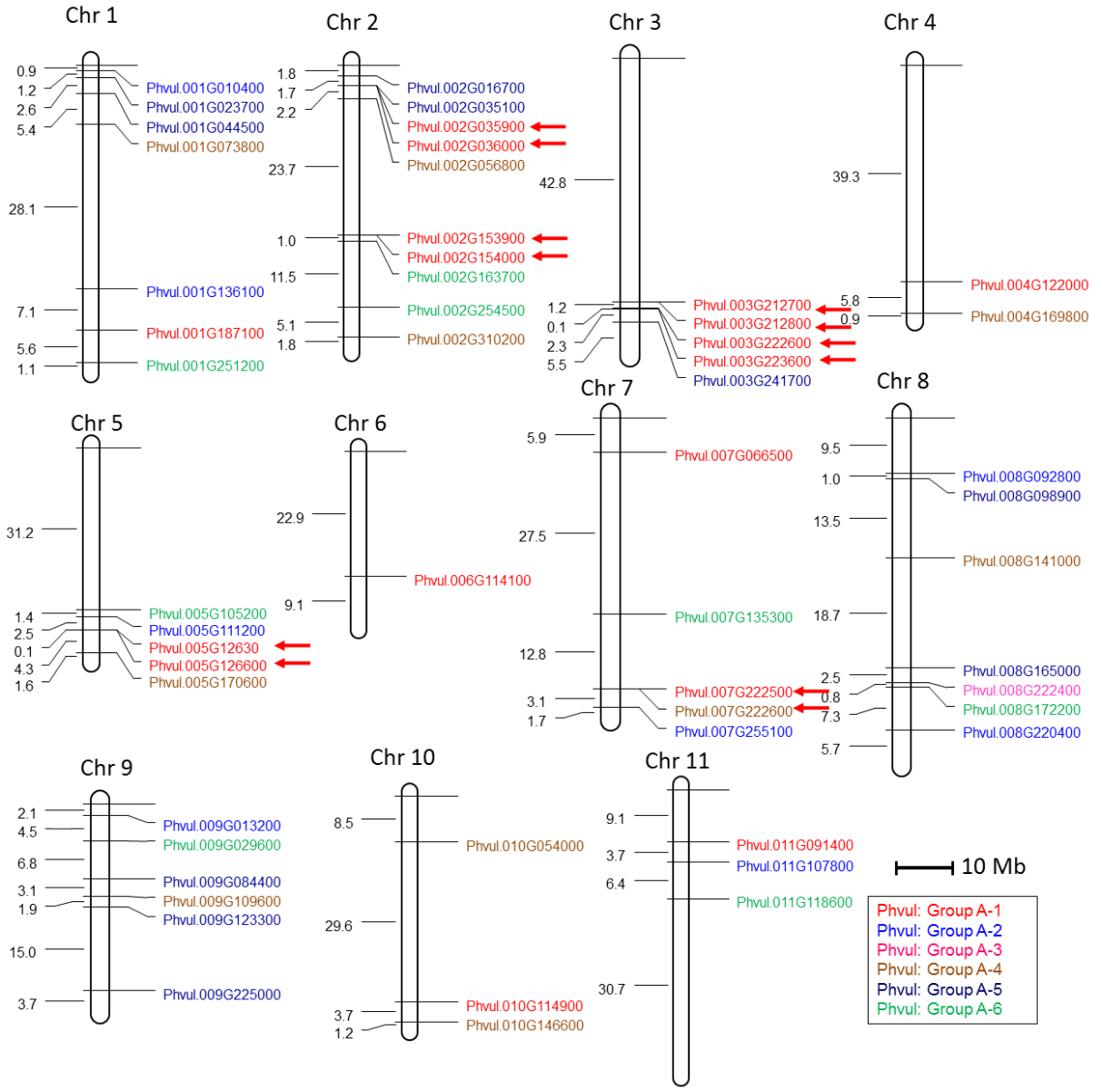


Figure 3.4 – Chromosomal location of 54 putative *PvDREB* genes. Subgroups are represented by different colors. Red arrows indicate tandem duplications.

3.3.4 Gene ontology analysis

The gene ontology analysis on Blast2Go suggested all sequences are involved in sequence-specific DNA binding, the basic characteristic of transcription factors. Furthermore, predictions showed all proteins are localized to the nucleus (Supplementary Table 3.4). Figure 3.5 shows all predictions obtained for basic processes, molecular functions and the GO terms of the putative PvDREB sequences.

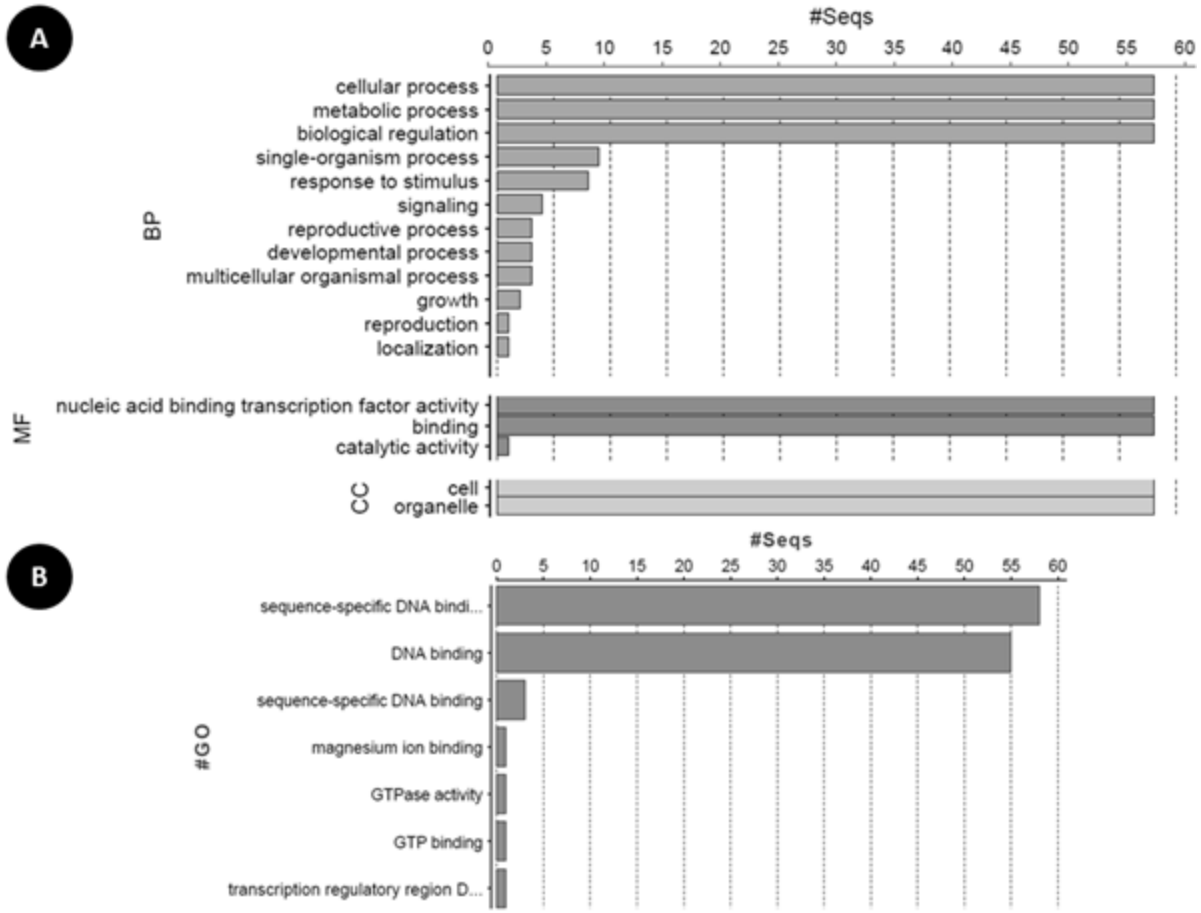


Figure 3.5 – Blast2Go annotation of the putative PvDREB proteins. A – Go distribution by level – processes. B – Direct GO count.

3.3.5 The novel genes *PvDREB1F*, *PvDREB2A*, *PvDREB5A* and *PvDREB6B*

Our group has been working on expression profiling and molecular characterization of several *DREB* genes and here are presented four novel *DREB* transcripts isolated from common bean. *PvDREB1F* (Phvul.003G212800.1) has high homology with the *A. thaliana* genes AT1G12610.1 (*AtDREB1F/DDF1*) and AT1G63030.1 (*AtDREB1E/DDF2*), from subgroup A-1. *PvDREB2A* (Phvul.011G107800.1) is homologous to *GmDREBa* (A-2 subgroup) from soybean and to the *A. thaliana DREB2* genes. *PvDREB5A* (Phvul.008G098900.1) is homologous to *GmDREB2* (A-5 subgroup) and to the *A. thaliana RAP2.1*, from A-5 subgroup. *PvDREB6B* (Phvul.002G254500.1) is homologous to *GmDREBb* (A-6) and the *A. thaliana* genes AT2G22200.1, AT4G39780.1 and AT5G65130.1 (all from A-6). Figure 3.6 shows the alignment and conservation of the AP2 domain of the four novel DREB from common bean with homolog proteins from *A. thaliana* and *G. max*.

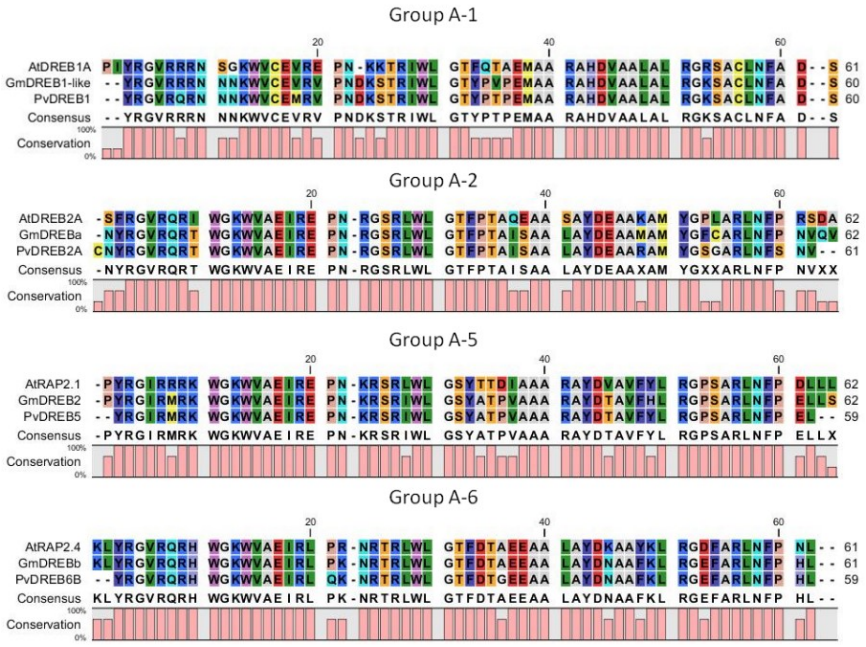


Figure 3.6 – AP2 domain alignment and conservation among the novel proteins PvDREB1, PvDREB2A, PvDREB5 and PvDREB6B and homologs from *Arabidopsis thaliana* and *Glycine max*.

PvDREB6B has been found to be equivalent to *PvDREB2A* in the studies of Nayak et al. (2009) and Cortés et al. (2012). However, at the time, the genomic sequence of common bean was not available and only a few sequences were deposited in GenBank (NCBI), which resulted in limited annotation precision. The current version of the genome shows the complete genomic sequence of the gene and its phylogenetic analysis clearly suggests its homology with A-6 genes. Besides, the annotation suggests it is similar to a *RAP2.4* gene from *A. thaliana*, one of the most studied members of the A-6 subgroup. In Supplementary Figure 3.3, we show a phylogenetic tree with the sequence used by Nayak et al. (2009) and Cortés et al. (2012) and from the current study. Hereby, we proposed the replacement of the name for *PvDREB6B*. *PvDREB6A* is other gene from the A-6 group in analysis (data not published).

3.3.6 Phyto Mine expression profile

The FPKM values searched of Phytozome revealed different basal levels of transcripts in several tissues from common bean (Figure 3.7). In general, most A-1 and A-2 genes had very low levels (FPKM value <= 1). *PvDREB1F* showed negative FPKM values, which were converted to zero in all tissues. *PvDRE2A* had considerable transcript amounts in all tissues

(mean FPKM = 2.79). The only member of the A-3 subgroup (Phvul.008G222400) had negative FPKM in all tissues. Higher amounts of transcripts were detected in most A-4, A-5 and A-6 genes. *PvDREB5A* and *PvDREB6B* had high positive values of FPKM (means FPKM of 3.21 and 3.45, respectively) (Figure 3.7).

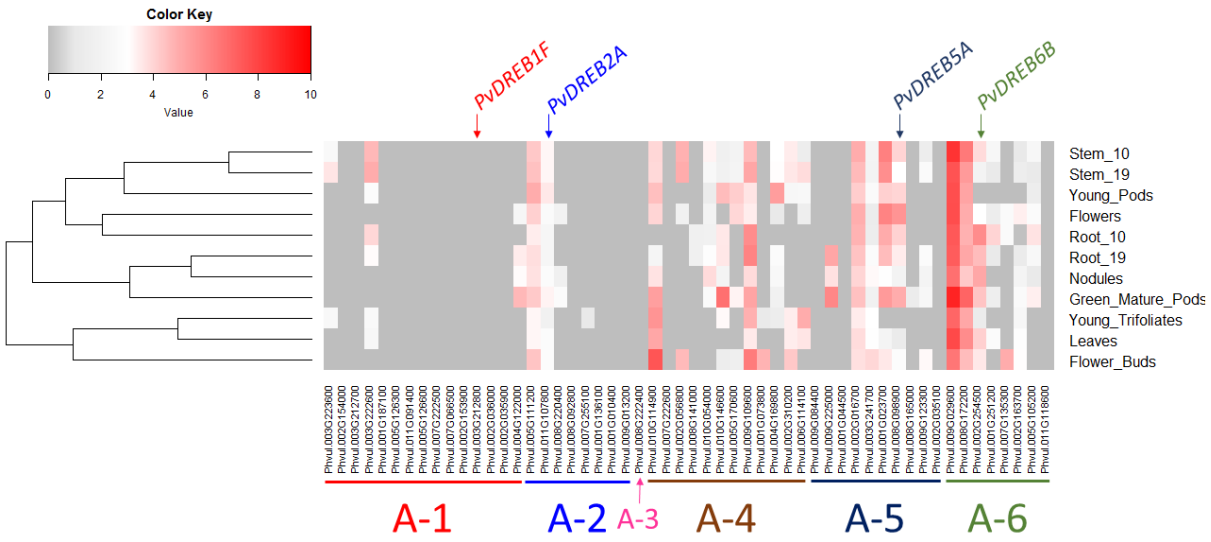


Figure 3.7 – Heat map of converted FPKM values retrieved from Phytozome database (RNA-seq data) for 54 putative *PvDREB* genes. The novel genes *PvDREB1F*, *PvDREB2A*, *PvDREB5A* and *PvDREB6B* are indicated.

3.3.7 Temporal and spatial expression profiling of the four novel *PvDREB* transcripts

Genes *PvDREB1F*, *PvDREB2A*, *PvDREB5A* and *PvDREB6B* exhibited different patterns of expression under the four treatments imposed (PEG 10%, NaCl 250 mM, 4°C and ABA 100 µM) (Figure 3.8). Variation was temporal- and tissue-dependent. Stress treatments reflected in alterations on the relative water content of BAT 477 leaves, as well as in the activity of the ROS-scavenging enzyme catalase (Supplementary Figure 3.4). These results showed that stress-treatments altered physiological and biochemical parameters of BAT 477 plants, reflecting in changes in gene expression.

Transcripts of *PvDREB1F* rapidly accumulated under dehydration (up to 12 fold-change in $\log_2$ units), high salinity (up to 12 fold-change) and ABA treatment (up to 12 fold-change) in all plant organs (roots, stem and leaves), in comparison to the untreated plants (Figure 3.8). In general, relative expression values were lower with the freezing treatment (up to 6.5 fold-change) than with the others. An increased expression has been observed in roots

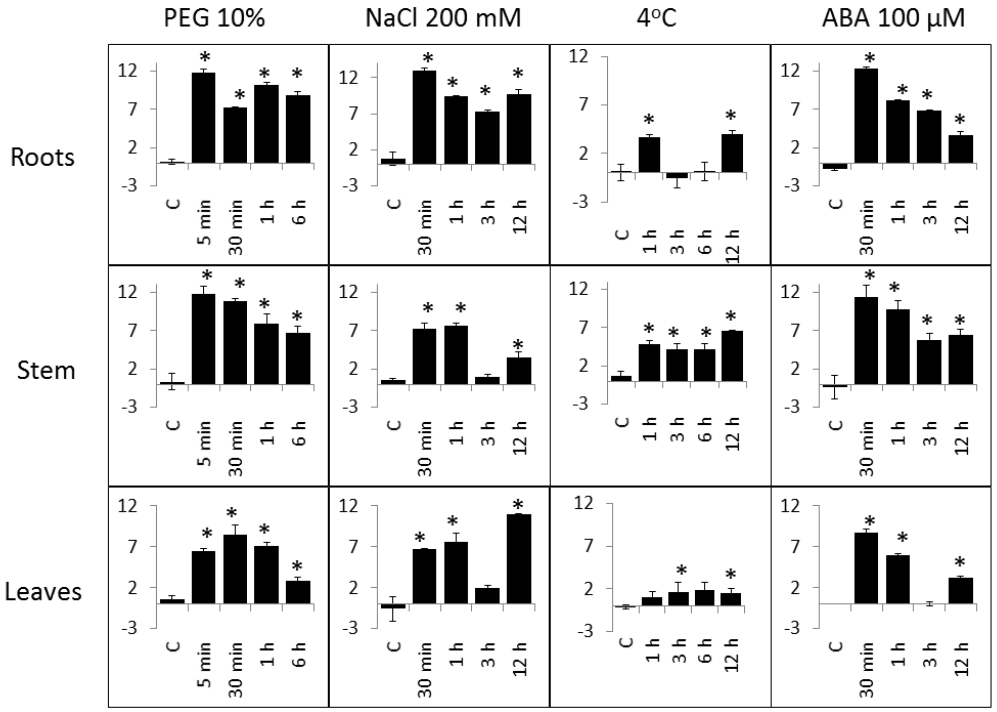
after one hour, but lower levels were observed after three and six hours with a final increase after 12 hours of exposition to cold.

PvDREB2A had low inducibility under the abiotic stresses of the study, with an exception for the dehydration treatment on the stem of BAT 477 (relative expression 3 folds higher than control) (Figure 3.8). Some slight increase in the relative number of transcripts was also observed with the cold treatment (up to 1.25 fold-change). ABA treatment occasioned some oscillation, first with some decrease (until -1 fold-change) followed by an increase (to 0.3 fold-change) in the relative number of transcripts.

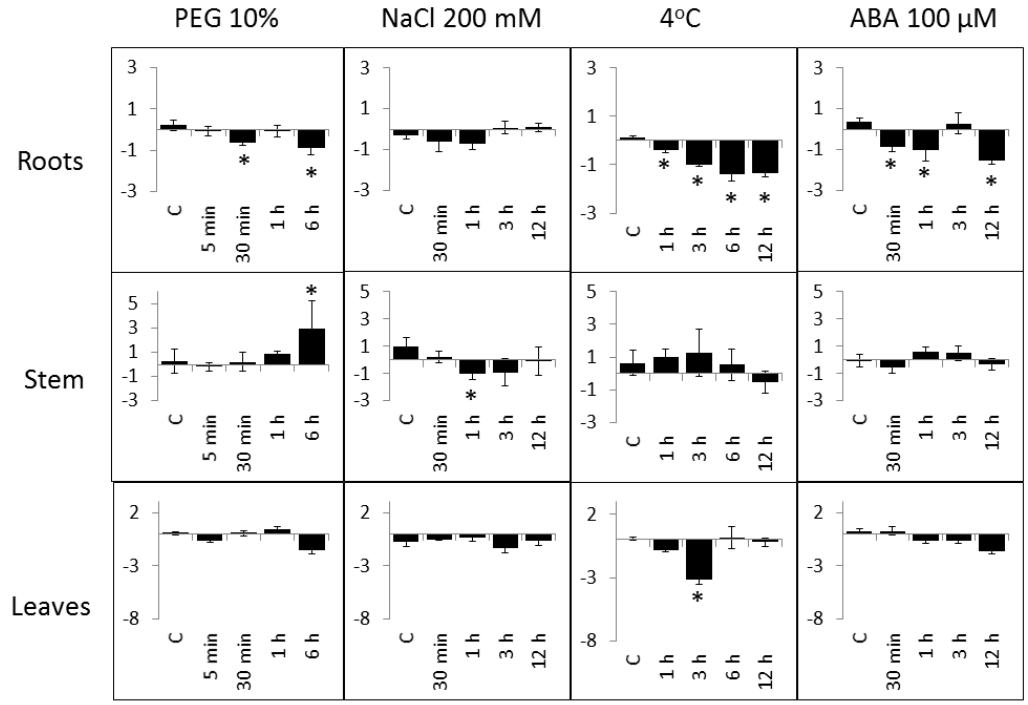
PvDREB5A also revealed to be stress-inducible under all treatments. In general, dehydration led to increased inducibility over time in roots (3.6 fold-change, 6 h), stem (3.7 fold-change, 6 h) and leaves (3 fold-change, 1 h) (Figure 3.8). The same was observed with the treatment with high salinity, with the highest relative expression values after 12 hours of treatment (5.5, 7.3 and 3.3, respectively for roots, stem and leaves). Treatment with cold also increased transcript accumulation in roots (2.9 fold-change) and stem (3.3 fold-change) when compared to the control plants at room temperature. An increase of expression was observed in leaves as well (up to 2.1, 1 h), but it was followed by a high decrease at the periods of six (-1.2 fold-change) and 12 hours (-2.4 fold-change) of stress. ABA mostly led to an increase in the levels of transcripts in roots (3.7 fold-change, 12 h) and stem (3.4 fold-change, 3 h), but a slight and progressive decrease was observed in leaves (up to -1.1 fold-change).

The most significant aspects of the *PvDREB6B* expression profile were an increase of its levels after treatment with dehydration in roots (up to 0.8 fold-change) and leaves (maximum of 1.7 fold-change) (Figure 3.8). Col treatment led to pronounced expression of *PvDREB6B* in leaves, with a progressive increase (up to 1.8 fold-change, 1 h) followed by a decrease (-0.8 fold-change, 12 h). Salinity diminished levels of transcripts in all organs. ABA produced a similar effect, although no significant differences were observed in the stem and some increase in the transcripts relative level was detected after 12 hours of exposition.

PvDREB1F



PvDREB2A



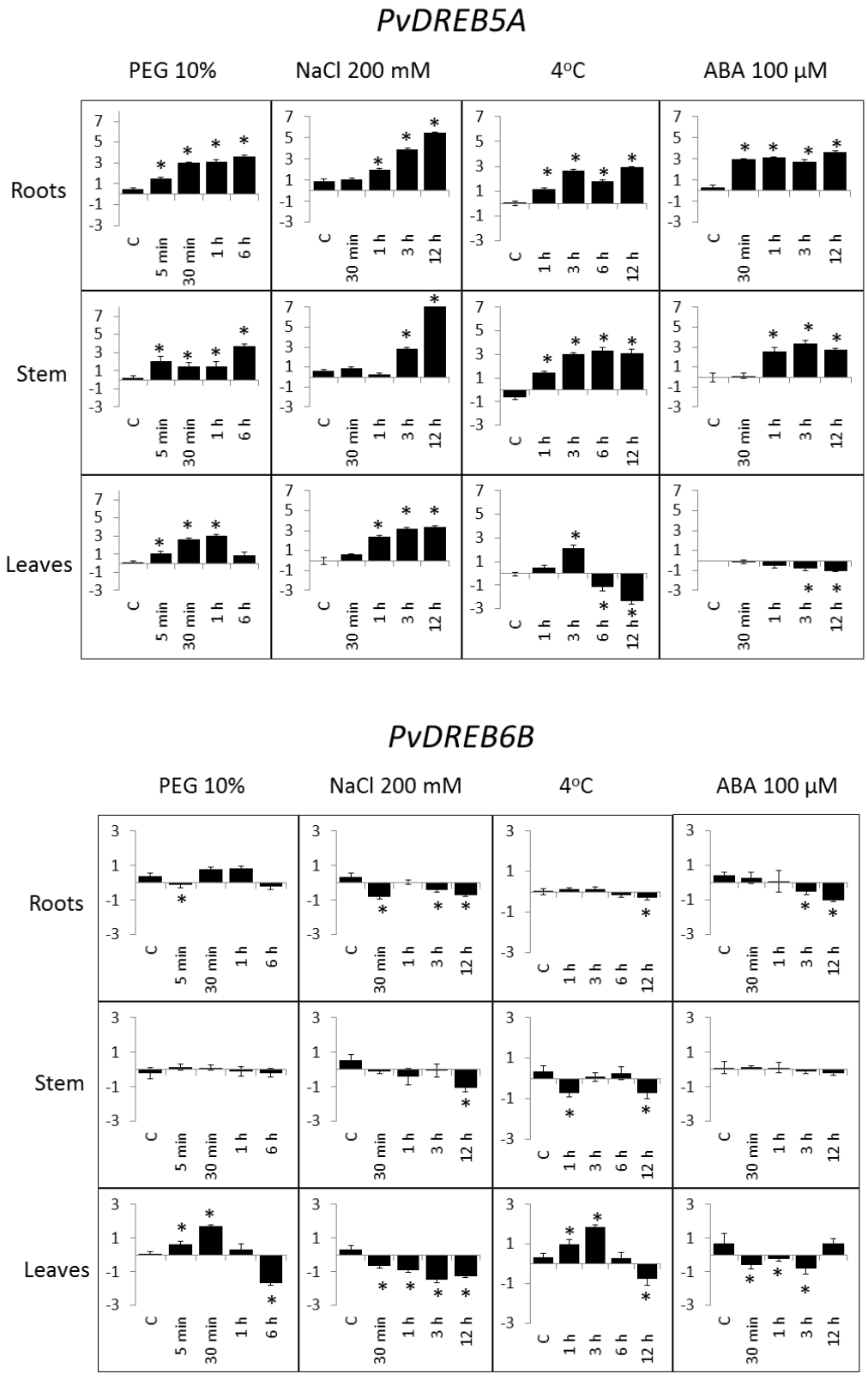
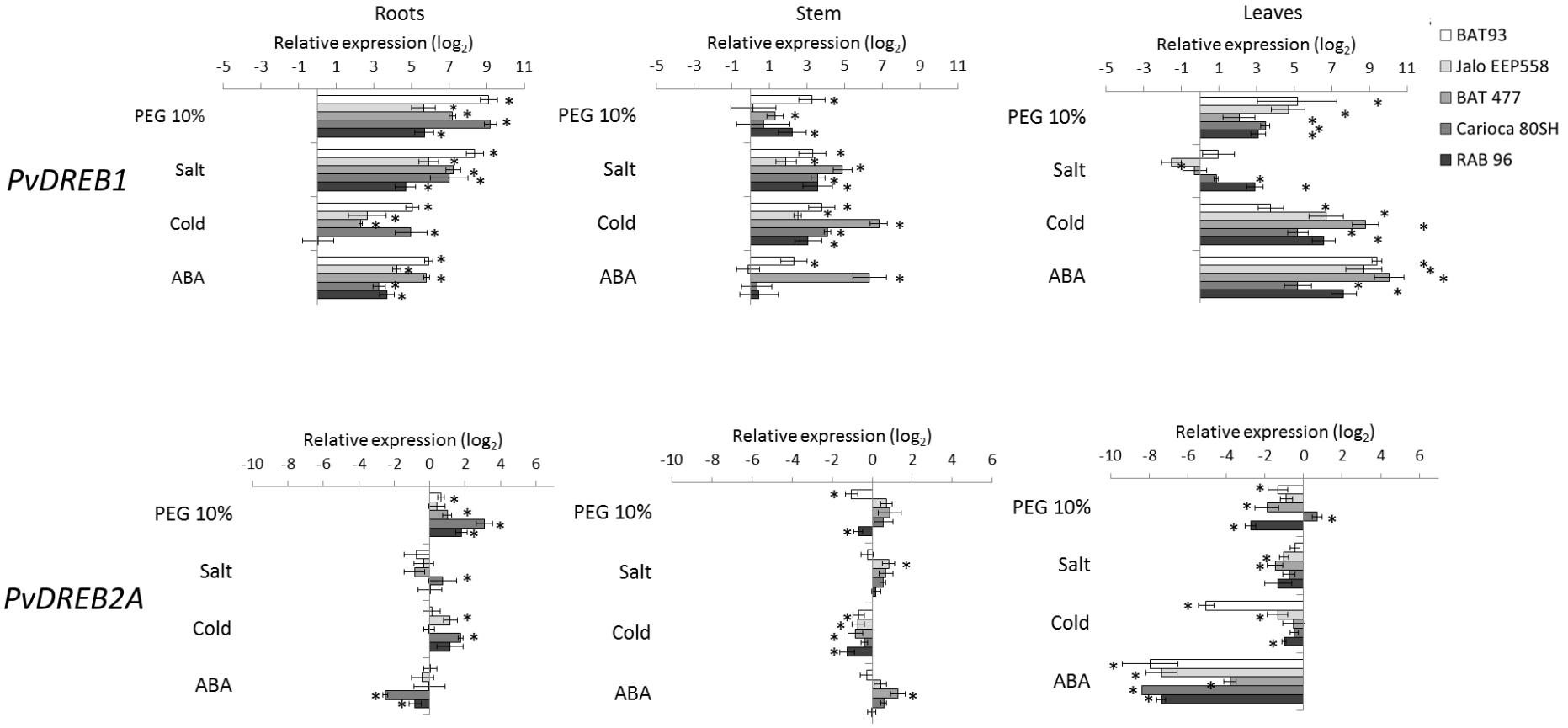


Figure 3.8 - Temporal and spatial scale qRT-PCR gene expression profile of four common bean *DREB* genes (*PvDREB1F*, *PvDREB2A*, *PvDREB5A* and *PvDREB6B*) in BAT 477 (drought tolerant genotype) plants subjected to different abiotic stresses induction: dehydration by using polyethylene glycol (PEG 10%); high salinity by a solution of NaCl 200 mM; cold by incubation at 4ºC; and abscisic acid induction factor (ABA 100 µM solution). Values are expressed in relative terms: expression value of stressed samples is relative to control samples. * indicates significant up or down-regulation of the genes in comparison to their control samples.

3.3.8 Spatial expression profiling in different common bean genotypes

In the second experiment, treatments implicated in different physiological and biochemical responses of each genotype. The relative water content (RWC) was significantly altered after three hours of dehydration, high salinity, low temperature and ABA treatment (Supplementary Figure 3.5 – A). At the enzyme level, treatments affected the ROS-scavenging enzyme catalase. In general, there was a particular increase in the enzyme activity after three hours of cold treatment in all genotypes (Supplementary Figure 3.5 – B).

As in the first experiment, stresses implicated in similar responses of each one of the four novel *PvDREB* genes (*PvDREB1F*, *PvDREB2A*, *PvDREB5A* and *PvDREB6B*), but with some particular differences concerning each genotype. *PvDREB1F* was strongly induced after the three-hour period of stress in all treatments and genotypes, except under salinity in leaves (Figure 3.9). *PvDREB2A* expressed under dehydration in roots and stem of most genotypes (Figure 3.9). The highest relative expression value was observed for the genotype IAC-Carioca 80SH in roots (3.1 fold-change). It was also the only genotype with an increase in the relative transcript levels in leaves (0.7 fold-change). Salinity increased the number of transcripts in roots, as well as ABA in the stem. Exposition to cold augmented expression in Jalo EEP558 (1.2 fold-change), IAC-Carioca 80SH (1.7 fold-change) and RAB 96 (1.1 fold-change). High decrease in transcript levels was observed after ABA treatment in all genotypes.


PvDREB1
PvDREB2A
Roots
Stem
Leaves
Relative expression (log2)
PEG 10%
Salt
Cold
ABA
BAT93
Jalo EEP558
BAT 477
Carioca 80SH
RAB 96

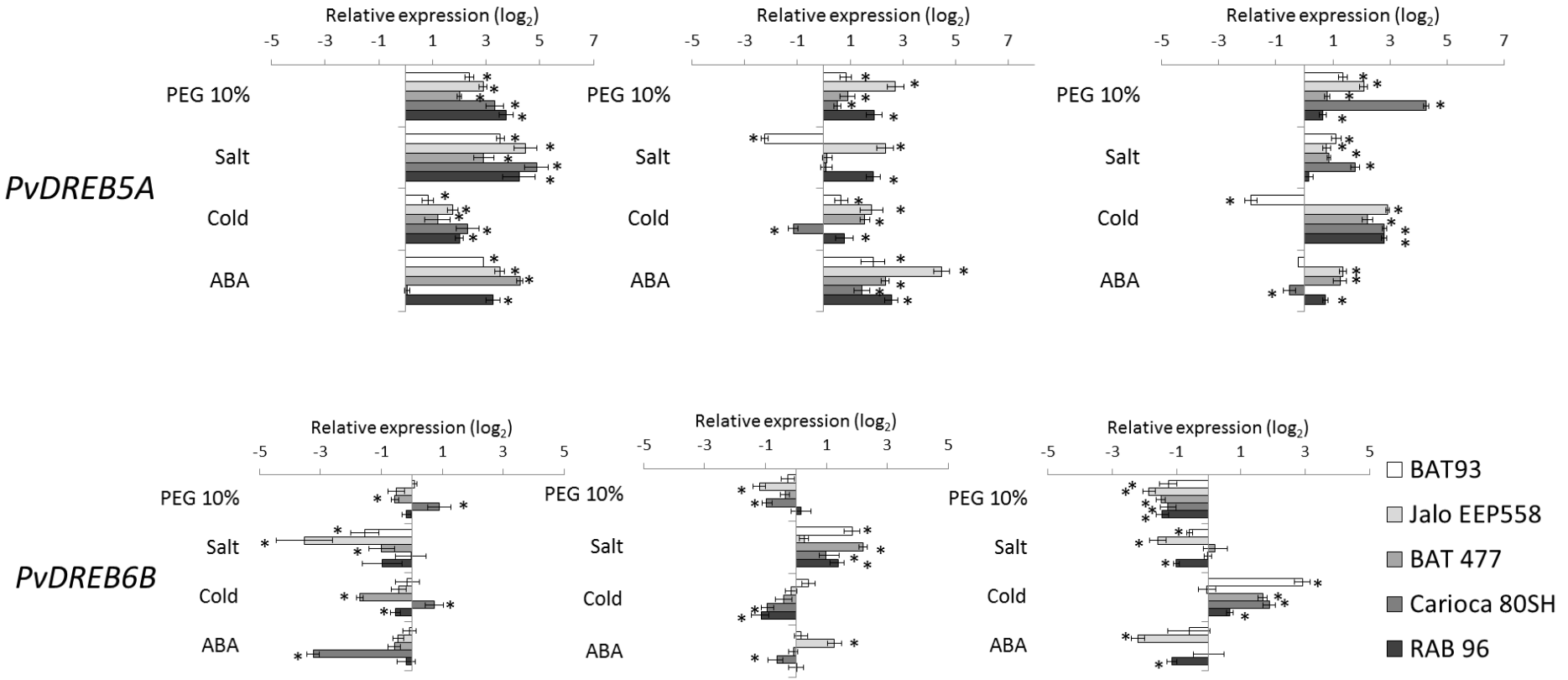


Figure 3.9 – Spatial scale qRT-PCR gene expression profile of four common bean *DREB* genes (*PvDREB1F*, *PvDREB2A*, *PvDREB5A* and *PvDREB6B*) in BAT 93, Jalo EEP558, BAT 477, IAC-Carioca 80SH and RAB 96 plants subjected to different abiotic stresses induction: dehydration by using polyethylene glycol (PEG 10%); high salinity by a solution of NaCl 200 mM; cold by incubation at 4ºC; and abscisic acid induction factor (ABA 100 µM solution). * indicates significant up or down-regulation of the genes in comparison to the control samples.

PvDREB5A was induced by all treatments and genotypes (Figure 3.9), as it was in the temporal analysis with BAT 477 (Figure 3.8). Decreased relative expression level was observed in BAT 93 after exposure to high salinity on the stem (-1.8 fold-change). In the same organ, no difference was observed between control and salt treatment in BAT 477 and IAC-Carioca 80SH, while Jalo EEP 558 and RAB 96 had high inducibility. Additionally, BAT 93 was the only to present decrease in transcript level after cold treatment in leaves.

PvDREB6B transcripts accumulated with salinity treatment after three hours on the stem in all genotypes. As observed in the temporal experiment, inducibility was also detected under low-temperature exposure in leaves (maximum of 2.9 fold-change in BAT 93), with the exception of Jalo EEP558 (-0.05 fold-change). In roots, dehydration and low temperature increased relative transcript levels in IAC-Carioca 80SH (0.7 and 0.9 fold-change, respectively), the opposite of what was observed for the other genotypes (negative values up to -1.7) (Figure 3.9).

3.4 Discussion

3.4.1 Phylogenetic analysis, motifs predictions and expression profiles

Our work provided a detailed genome-wide categorization of the *DREB* gene subfamily in common bean. In total, 54 putative *DREB* genes were catalogued and divided into six subgroups, according to the previous reports for *A. thaliana* (SAKUMA et al., 2002; NAKANO et al., 2006). All proteins feature the common aspects of *DREB* genes, especially the conservation of the 14th and 19th amino acids within the AP2 domain (SAKUMA et al., 2002). The number of *DREB* genes categorized for common bean was similar to *A. thaliana*, which possesses 57 *AtDREB* separated into four main subgroups in the *AP2/ERF* superfamily (NAKANO et al., 2006), or 56 within six subgroups, from A-1 to A-6 (SAKUMA et al., 2002). In soybean, that number has been described to be much lower, with only 36 putative *GmDREB* (ZHANG et al., 2008). However, this study was published before the whole genome sequence of *G. max* (SCHMUTZ et al., 2010). Currently, searching the term “AP2” on Phytozome generates 362 sequences. Moreover, phylogenetic analyses in the current work show that there are two copies in soybean for each *DREB* of common bean, suggesting a higher number of *DREB* loci along the soybean genome (data not shown).

After categorizing *PvDREB* genes, we showed six putative pairs tandemly duplicated in their respective chromosomes (Figure 3.4). Interestingly, all duplication events involved genes from the A-1 subgroup, with one only exception involving one A-1 and one A-4 genes

(Figure 3.4). A previous report has found an overrepresentation of *DREB1/CBF* genes for *Eucalyptus grandis*, which could have been an adaptation response to climates where the species was adapting over time (CAO et al., 2015). It is also well documented that tandem duplications are adaptively relevant to the evolution and function of abiotic and biotic stress responsive genes. Some experimental evidence showed tandem arrays often share regulatory elements and might be coexpressed (SCHMID et al., 2005; FLAGEL; WENDEL, 2009), exhibiting similar functions (BAUMGARTEN et al., 2003). An increased representation *DREB1* genes in common bean might have an adaptive role in a similar manner.

Investigation of protein motifs in all DREB sequences revealed several short conserved regions within DREB subgroups, indicating their potential as phylogenetic markers for each subdivision. The exclusivity of some motifs within subgroups might be related to specific functions in which the protein members are involved. In this work, we have not performed a direct and whole comparison of all motifs with other plant genomes, since our aim was to use such sequences as phylogenetic indicators. However, pre-analyses have shown motifs shared among all the common bean DREB and some isolated genes from *A. thaliana* and *G. max* (Supplementary Figure 3.6). For example, the alanine-rich motif CM4 was found within members of subgroups A-1 and A-4 from common bean, as well as in *AiDREB1A*, *AtDREB1B* and *AtDREB1C* (A-1) and *GmTINY* and *AtTINY* (A-4). This indicates many motifs are conserved among species, which could also have a similar function. Moreover, changes in amino acid structure could have shaped their functions across species, details that need further investigation.

Having defined the putative *PvDREB* genes, annotation and gene ontologies suggested all sequences have DNA-binding ability. However, for common bean, only one *PvDREB* gene has been experimentally tested on this matter (data not published). The present study, then, provides insights for further molecular characterization of *DREB* loci from common bean.

Several transcription factors usually have low basal levels in cells, having their concentration increased when activated by determined stimuli such as abiotic stresses. The FPKM values retrieved from Phytozome database for all the putative *PvDREB* loci showed very low levels of most A-1 and A-2 *PvDREB* members. These are usually the main regulators towards responding to stresses such as drought, salinity and cold (LATA; PRASAD, 2011; AKHTAR et al., 2012). The A-3 member also showed very low levels of transcripts. Most members from A-4, A-5 and A-6 subgroups exhibited higher levels in all tissues analyzed. Genes from A-4 subgroup generally show no consistent stress-inducibility (KARABA et al., 2007), but they may play role in configuring stress responses, although the

mechanisms are not clear so far (MIZOI; SHINOZAKI; YAMAGUCHI-SHINOZAKI, 2012). The stress-inducible A-5 members are known by the presence of an ERF-associated amphiphilic repression (EAR) motif (OHTA et al., 2001). These genes were reported to be upregulated when A-1 and A-2 members were overexpressed (MARUYAMA et al., 2004; SAKUMA et al., 2006) and further evidence show they act as transcriptional repressors downstream of *DREB1* and *DREB2* genes (TSUTSUI et al., 2009; DONG; LIU, 2010). A-6 members are also usually stress-responsive, and microarray analyses have shown the main *Arabidopsis* gene, *RAP2.4*, to be involved in regulation of aquaporins (RAE et al., 2011). The ortholog *PvDREB6A* is also drought-, salt- and cold-responsive (data not published). Therefore, they function in stress regulation, but seem to have different targets from *DREB1* and *DREB2* genes (MIZOI et al., 2012).

3.4.2 Temporal and expression profiles of the novel *PvDREB* genes

The expression profiles of the novel genes *PvDREB1F*, *PvDREB2A*, *PvDREB5A* and *PvDREB6B* were analyzed under dehydration, salinity, low temperature and ABA treatments, considering different time-periods of stress with the genotype BAT 477. *PvDREB1F* showed the highest relative expression values under all treatments (Figure 3.8). However, not very high values were observed with the cold treatment, the usually expected for some *DREB1* genes, as reported in *Arabidopsis* (LATA; PRASAD, 2011; AKHTAR et al., 2012). This might be explained by the fact *PvDREB1F* is phylogenetically closer to the genes *DWARF AND DELAYED FLOWERING 1* (*DREB1F/DDF1*) and *DREB1E/DDF2*, which are mainly induced by salinity in *A. thaliana* (MAGOME et al., 2004; MAGOME et al., 2008). Moreover, there is crosstalk between *DREB1* and *DREB2* genes, which might lead to *DREB1* responsiveness to osmotic stresses as well as low temperature (AKHTAR et al., 2012). In other species, such as *G. max*, a *DREB1*-like (Glyma10g07770.1) gene was also induced by water deficit (MARCOLINO-GOMES et al., 2013). These findings suggest stress-responsiveness of such genes has been shaped in different manners among plant species.

Although high relative expression values were found for *PvDREB1F*, a direct comparison with the patterns of the other transcripts (*PvDREB2A*, *PvDREB5A* and *PvDREB6B*) is not appropriate. *PvDREB1F* showed low basal levels of expression in control conditions in all plant organs. The other genes possessed much higher amounts of transcripts in control plants (Supplementary Figure 3.7). Thereby, the high values of expression of *PvDREB1F* indicate a rapid and greater accumulation of transcripts after stress, which still are lower than the detected for *PvDREB2A*, *PvDREB5A* and *PvDREB6B*. As a result, *PvDREB1F*

remains as the gene with the highest inducibility in this study, but with lower amount of transcript levels than the other genes.

PvDREB2A was predominantly induced in the stem under dehydration, although some increase in relative amounts of transcripts were observed for cold treatment as well. *DREB2* genes have been mostly characterized by their response to osmotic stresses, especially to dehydration and salinity (LIU et al., 1998; SAKUMA et al., 2006). Their engineering in other species has been proved to increase drought tolerance (REIS et al., 2014). The other genes, *PvDREB5A* and *PvDREB6B*, were stress-inducible as reported in literature for member of subgroups A-5 and A-6 (MIZOI et al., 2012). *PvDREB5A* was induced by all treatments was as another A-5 member from soybean, *GmDREB2* (CHEN et al., 2007). Somewhat similar inducibilities were detected for *PvDREB6B* in related genes, *GmDREBb* (LI et al., 2005) and *AtRAP2.4* and *AtRAP2.4B* (RAE et al., 2011).

Gene expression was dependent on the time-period of stress as well as the location in plants. In general, fast responses were observed for all genes after stress induction. It is a typical behavior of *DREB* genes, as observed with *AtDREB1A* and *AtDREB1B*, whose transcripts rapidly augmented after only 15 minutes of exposition to low temperature. Some other genes present slower responses such as *AtDREB1C*, with significant accumulation of transcripts only after 2.5 hours under cold treatment (NOVILLO et al., 2004). Transcript accumulation has also been shown to vary between roots and leaves, such as *AtDREB1*, more frequent in roots under salinity. *AtDREB2A* was abundant in both organs, but *AtDREB2D* was detected primarily under salt in leaves (SAKUMA et al., 2002).

Another critical factor for the analysis of expression of *PvDREB* genes is the developmental stage of plants. Stress-treatments were applied at V3 stage, in which plant metabolism in concentrated in plant growth and investments in leaf area for photosynthesis. It is one of the critical stages, when plants are highly sensitive to abiotic stresses. In comparison, FPKM values retrieved from Phytozome for the four transcripts (*PvDREB1F*, *PvDREB2A*, *PvDREB5A* and *PvDREB6B*) showed different levels of expression in stem, root and leaves. Young trifoliates exhibited low expression in comparison to most of the other tissues analyzed by RNA-seq.

Furthermore, our results showed some differences in expression values among genotypes under the same stress-treatments. Such comparisons might be useful to encounter genes possibly associated with stress tolerance in common bean. In the same manner, genotypes with contrasting expression profiles might be used for further characterization of the regulation patterns of *DREB* genes in different genetic backgrounds. In our study,

however, we only attempted to show that gene expression was temporal-, tissue- and genotype-dependent. One that might be interest in associating gene expression profiles with adaptation to abiotic stress tolerance would probably consider the analysis of a wider set of genotypes, preferably from a wild background, in order to establish accurate correlations.

This study opens the possibility of working with *PvDREB* loci under multiple approaches. We showed several *PvDREB* genes with different structure, coding for proteins with distinctive motifs to be explored for the understanding of their function. Their annotation suggested all sequences are transcription factors involved in stress responses, but experimental analyses need to be performed for proving their function. With their chromosomal location, molecular marker studies such as with SNP might be able to identify molecular signatures associated with traits of interest in common bean. As *DREB* genes are inherently involved with abiotic stress regulation, depositing investments on their research should bring enormous contributions to improving common bean varieties. Adverse conditions of the diverse environments in which beans are grown might be severely intensified and all genomic resources available come to help in the design of proper breeding and engineering strategies.

3.5 Conclusions

In this work, we elaborated a catalogue of the *DREB* gene subfamily in common bean. In total, 54 *DREB* genes were defined according to multiple particularities. All genes fitted six main subgroups (A-1 to A-6) according to previous reports for other model species. Four novel genes were defined and their expression profiles were addressed under the effect of abiotic stress sources (dehydration, salinity and low temperature). The major inducibility factors of *PvDREB1F* (dehydration, salinity and low temperature), *PvDREB2A* (dehydration and cold), *PvDREB5A* (dehydration, salinity and low temperature) and *PvDREB6B* (dehydration and cold) were determined. However, relative expression levels of each transcript was time-, tissue- and genotype-modulated. Our categorization along with the isolation and gene expression profile of novel *PvDREB* genes provides insights for further studies aimed at improvement abiotic stress tolerance in common bean.

References

AKHTAR, M.; JAISWAL, A.; TAJ, G.; JAISWAL, J. P.; QURESHI, M. I.; SINGH, N. K. DREB1 / CBF transcription factors : their structure , function and role in abiotic stress tolerance in plants. **Journal of Genetics**, London, v. 91, n. 3, p. 385–395, 2012.

ALONSO, J. M.; STEPANOVA, A. N.; LEISSE, T. J.; KIM, C. J.; CHEN, H.; SHINN, P.; STEVENSON, D. K.; ZIMMERMAN, J.; BARAJAS, P.; CHEUK, R.; GADRINAB, C.; HELLER, C.; JESKE, A.; KOESEMA, E.; MEYERS, C. C.; PARKER, H.; PREDNIS, L.; ANSARI, Y.; CHOY, N.; DEEN, H.; GERALT, M.; HAZARI, N.; HOM, E.; KARNES, M.; MULHOLLAND, C.; NDUBAKU, R.; SCHMIDT, I.; GUZMAN, P.; AGUILAR-HENONIN, L.; SCHMID, M.; WEIGEL, D.; CARTER, D. E.; MARCHAND, T.; RISSEEUW, E.; BROGDEN, D.; ZEKO, A.; CROSBY, W. L.; BERRY, C. C.; ECKER, J. R. Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. **Science**, Washington, DC, v. 301, n. 5633, p. 653–657, 2003.

BAILEY, T. L.; BODEN, M.; BUSKE, F. a.; FRITH, M.; GRANT, C. E.; CLEMENTI, L.; REN, J.; LI, W. W.; NOBLE, W. S. MEME Suite: Tools for motif discovery and searching. **Nucleic Acids Research**, Oxford, v. 37, n. 2, p. 202–208, 2009.

BARRS, H. D.; WEATHERLEY, P. E. A re-examination of the relative turgidity technique for estimating water deficits in leaves. **Australian Journal of Biological Sciences**, Clayton, v. 15, p. 413-428, 1962.

BAUMGARTEN, A.; CANNON, S.; SPANGLER, R.; MAY, G. Genome-level evolution of resistance genes in *Arabidopsis thaliana*. **Genetics**, Bethesda, v. 165, n. 1, p. 309–319, 2003.

BEEBE, S. E.; RAO, I. M.; BLAIR, M. W.; ACOSTA-GALLEGOS, J. A. Phenotyping common beans for adaptation to drought. **Frontiers in Physiology**, Lausanne, v. 4, n. 3, p. 1–20, 2013.

BLAIR, M. W.; GALEANO, C. H.; TOVAR, E.; TORRES, M. C. M.; CASTRILLÓN, A. V.; BEEBE, S. E.; RAO, I. M. Development of a Mesoamerican intra-gene pool genetic map for quantitative trait loci detection in a drought tolerant × susceptible common bean (*Phaseolus vulgaris* L.) cross. **Molecular Breeding**, Dordrecht, v. 29, n. 1, p. 71–88, 2012.

BORGES, A.; TSAI, S. M.; CALDAS, D. G. G. Validation of reference genes for RT-qPCR normalization in common bean during biotic and abiotic stresses. **Plant Cell Reports**, Heidelberg, v. 31, n. 5, p. 827–838, 2012.

BOUTILIER, K.; OFFRINGA, R.; SHARMA, V. K.; KIEFT, H.; OUELLET, T.; ZHANG, L.; HATTORI, J.; LIU, C.-M.; VAN LAMMEREN, A. A. M.; MIKI, B. L. A.; CUSTERS, J. B. M.; VAN LOOKEREN CAMPAGNE, M. M. Ectopic expression of *BABY BOOM* triggers a conversion from vegetative to embryonic growth. **The Plant Cell**, Rockville, v. 14, n. 8, p. 1737–1749, 2002.

BRADFORD, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. **Analytical Biochemistry**, New York, v. 72, p. 248–254, 1976.

BRAY, E. A.; BAILEY-SERRES, J.; WERETILNYK, E. Responses to abiotic stresses. In: BUCHANAN, B.; GRUISSEM, W.; JONES, R. (Ed.). **Biochemistry & molecular biology of plants**. Hoboken: Wiley, 2000. p. 1158-1203.

BROUGHTON, W. J.; HERNÁNDEZ, G.; BLAIR, M.; BEEBE, S.; GEPTS, P.; VANDERLEYDEN, J. Beans (*Phaseolus* spp.) - Model food legumes. **Plant and Soil**, Dordrecht, v. 252, n. 1, p. 55–128, 2003.

CAO, P. B.; AZAR, S.; SANCLEMENTE, H.; MOUNET, F.; DUNAND, C.; MARQUE, G.; MARQUE, C.; TEULIÈRES, C. Genome-Wide Analysis of the AP2/ERF Family in *Eucalyptus grandis*: An Intriguing Over-Representation of Stress-Responsive *DREB1/CBF* Genes. **Plos One**, San Francisco, v. 10, n. 4, p. e0121041, 2015.

CHEN, M.; WANG, Q. Y.; CHENG, X. G.; XU, Z. S.; LI, L. C.; YE, X. G.; XIA, L. Q.; MA, Y. Z. *GmDREB2*, a soybean DRE-binding transcription factor, conferred drought and high-salt tolerance in transgenic plants. **Biochemical and Biophysical Research Communications**, New York, v. 353, n. 2, p. 299–305, 2007.

CHEN, M.; XU, Z.; XIA, L.; LI, L.; CHENG, X.; DONG, J.; WANG, Q.; MA, Y. Cold-induced modulation and functional analyses of the DRE-binding transcription factor gene, *GmDREB3*, in soybean (*Glycine max* L.). **Journal of Experimental Botany**, Oxford, v. 60, n. 1, p. 121–135, 2009.

CHUCK, G.; MEELEY, R. B.; HAKE, S. The control of maize spikelet meristem fate by the APETALA2-like gene indeterminate *spikelet1*. **Genes and Development**, Woodbury, v. 12, n. 8, p. 1145–1154, 1998.

CONESA, A.; GÖTZ, S. Blast2GO: A comprehensive suite for functional analysis in plant genomics. **International Journal of Plant Genomics**, New York, v. 2008, 2008. doi: 10.1155/2008/619832.

CORTÉS, A. J.; THIS, D.; CHAVARRO, C.; MADRIÑÁN, S.; BLAIR, M. W. Nucleotide diversity patterns at the drought-related DREB2 encoding genes in wild and cultivated common bean (*Phaseolus vulgaris* L.). **Theoretical and Applied Genetics**, Heidelberg, v. 125, n. 5, p. 1069–1085, 2012.

DASZKOWSKA-GOLEC, A.; SZAREJKO, I. Open or close the gate-stomata action under the control of phytohormones in drought stress conditions. **Frontiers in Plant Science**, Lausanne, v. 4, n. 5, p. 138, 2013.

DONG, C.-J.; LIU, J.-Y. The *Arabidopsis* EAR-motif-containing protein *RAP2.1* functions as an active transcriptional repressor to keep stress responses under tight control. **BMC Plant Biology**, London, v. 10, p. 47, 2010.

ELLIOTT, R. C.; BETZNER, A. S.; HUTTNER, E.; OAKES, M. P.; TUCKER, W. Q.; GERENTES, D.; PEREZ, P.; SMYTH, D. R. *AINTEGUMENTA*, an *APETALA2-like* gene of *Arabidopsis* with pleiotropic roles in ovule development and floral organ growth. **The Plant Cell**, Rockville, v. 8, n. 2, p. 155–168, 1996.

FLAGEL, L. E.; WENDEL, J. E. Gene duplication and evolutionaary novelty in plants. **New Phytologist**, Oxford, v. 183, p. 557–564, 2009.

FREYRE, R.; SKROCH, P. W.; GEFFROY, V.; ADAM-BLONDON, A. F.; SHIRMOHAMADALI, A.; JOHNSON, W. C.; LLACA, V.; NODARI, R. O.; PEREIRA, P. A.; TSAI, S. M.; TOHME, J.; DRON, M.; NIENHUIS, J.; VALLEJOS, C. E.; GEPTS, P. Towards an integrated linkage map of common bean. 4. Development of a core linkage map and alignment of RFLP maps. **Theoretical and Applied Genetics**, Heidelberg, v. 97, n. 5-6, p. 847–856, 1998.

GAMA, P. B. S.; TANAKA, K.; ENEJI, A. E.; ELTAYEB, A. E.; SIDDIG, K. El. Salt-induced stress effects on biomass, photosynthetic rate, and reactive oxygen species-scavenging enzyme accumulation in common bean. **Journal of Plant Nutrition**, New York, v. 32, n. 5, p. 837–854, 2009.

GOODSTEIN, D. M.; SHU, S.; HOWSON, R.; NEUPANE, R.; HAYES, R. D.; FAZO, J.; MITROS, T.; DIRKS, W.; HELLSTEN, U.; PUTNAM, N.; ROKHSAR, D. S. Phytozome: A comparative platform for green plant genomics. **Nucleic Acids Research**, Oxford, v. 40, n.D1, p. 1178–1186, 2012.

GUIMARÃES, C. M.; STONE, L. F.; BRUNINI, O. Common bean (*Phaseolus vulgaris* L.) adaptation to drought. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 10, p. 70–75, 2006.

HANAI, L. R.; SANTINI, L.; CAMARGO, L. E. A.; FUNGARO, M. H. P.; GEPTS, P.; TSAI, S. M.; VIEIRA, M. L. C. Extension of the core map of common bean with EST-SSR, RGA, AFLP, and putative functional markers. **Molecular Breeding**, Dordrecht, v. 25, n. 1, p. 25–45, 2010.

HAO, D.; OHME-TAKAGI, M.; SARAI, A. Unique mode of GCC box recognition by the DNA-binding domain of ethylene-responsive element-binding factor (ERF domain) in plant. **Journal of Biological Chemistry**, Bethesda, v. 273, n. 41, p. 26857–26861, 1998.

HASANUZZAMAN, M.; NAHAR, K.; GILL, S. S.; GILL, R.; FUJITA, M. Drought stress responses in plants, oxidative stress, and antioxidant defense. In: TUTEJA, N.; GILL, S. S. (Ed.). Climate change and plant abiotic stress tolerance. Weinheim, Germany: Wiley-VCH Verlag, 2013. p. 209-237.

HU, Y. X.; WANG, Y. X.; LIU, X. F.; LI, J. Y. *Arabidopsis* RAV1 is down-regulated by brassinosteroid and may act as a negative regulator during plant development. **Cell research**, London, v. 14, n. 1, p. 8–15, 2004.

KARABA, A.; DIXIT, S.; GRECO, R.; AHARONI, A.; TRIJATMIKO, K. R.; MARSCH-MARTINEZ, N.; KRISHNAN, A.; NATARAJA, K. N.; UDAYAKUMAR, M.; PEREIRA, A. Improvement of water use efficiency in rice by expression of HARDY, an *Arabidopsis* drought and salt tolerance gene. **Proceedings of the National Academy of Sciences of the USA**, Washington, DC, v. 104, n. 39, p. 15270–15275, 2007.

LATA, C.; MISHRA, A. K.; MUTHAMILARASAN, M.; BONTHALA, V. S.; KHAN, Y.; PRASAD, M. Genome-Wide Investigation and Expression Profiling of AP2/ERF Transcription Factor Superfamily in Foxtail Millet (*Setaria italica* L.). **Plos One**, San Francisco, v. 9, n. 11, p. e113092, 2014.

LATA, C.; PRASAD, M. Role of DREBs in regulation of abiotic stress responses in plants. **Journal of Experimental Botany**, Oxford, v. 62, n. 14, p. 4731–4748, 2011.

LI, X. P.; TIAN, A. G.; LUO, G. Z.; GONG, Z. Z.; ZHANG, J. S.; CHEN, S. Y. Soybean DRE-binding transcription factors that are responsive to abiotic stresses. **Theoretical and Applied Genetics**, Heidelberg, v. 110, n. 8, p. 1355–1362, 2005.

LIN, R. C.; PARK, H. J.; WANG, H. Y. Role of *Arabidopsis RAP2.4* in regulating light and ethylene-mediated developmental processes and drought stress tolerance. **Molecular Plant**, Cambridge, v. 1, n. 1, p. 42–57, 2008.

LIU, Q.; KASUGA, M.; SAKUMA, Y.; ABE, H.; MIURA, S.; YAMAGUCHI-SHINOZAKI, K.; SHINOZAKI, K. Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in *Arabidopsis*. **The Plant Cell**, Rockville, v. 10, n. 8, p. 1391–1406, 1998.

LIU, R. H.; MENG, J. L. MapDraw: a microsoft excel macro for drawing genetic linkage maps based on given genetic linkage data. **Hereditas**, Malden, v. 25, n. 3, p. 317-321, 2003.

LIU, S.; WANG, X.; WANG, H.; XIN, H.; YANG, X.; YAN, J.; LI, J.; TRAN, L. S. P.; SHINOZAKI, K.; YAMAGUCHI-SHINOZAKI, K.; QIN, F. Genome-Wide Analysis of *ZmDREB* Genes and Their Association with Natural Variation in Drought Tolerance at Seedling Stage of *Zea mays* L. **Plos Genetics**, San Francisco, v. 9, n. 9, e1003790, 2013.

MAGOME, H.; YAMAGUCHI, S.; HANADA, A.; KAMIYA, Y.; ODA, K. *Dwarf and delayed-flowering 1*, a novel *Arabidopsis* mutant deficient in gibberellin biosynthesis because of overexpression of a putative AP2 transcription factor. **The Plant Journal**, Chichester, v. 37, n. 5, p. 720–729, 2004.

MAGOME, H.; YAMAGUCHI, S.; HANADA, A.; KAMIYA, Y.; ODA, K. The *DDF1* transcriptional activator upregulates expression of a gibberellin-deactivating gene, *GA2ox7*, under high-salinity stress in *Arabidopsis*. **The Plant Journal**, Chichester, v. 56, n. 4, p. 613–626, 2008.

MARCOLINO-GOMES, J.; RODRIGUES, F. A.; OLIVEIRA, M. C. N.; FARIAS, J. R. B.; NEUMAIER, N.; ABDELNOOR, R. V.; MARCELINO-GUIMARÃES, F. C.; NEPOMUCENO, A. L. Expression Patterns of *GmAP2/EREB*-like Transcription Factors Involved in Soybean Responses to Water Deficit. **Plos One**, San Francisco, v. 8, n. 5, e62294, 2013.

MARUYAMA, K.; SAKUMA, Y.; KASUGA, M.; ITO, Y.; SEKI, M.; GODA, H.; SHIMADA, Y.; YOSHIDA, S.; SHINOZAKI, K.; YAMAGUCHI-SHINOZAKI, K. Identification of cold-inducible downstream genes of the *Arabidopsis* DREB1A/CBF3 transcriptional factor using two microarray systems. **The Plant Journal**, Chichester, v. 38, n. 6, p. 982–993, 2004.

MIKLAS, P. N.; KELLY, J. D.; BEEBE, S. E.; BLAIR, M. W. Common bean breeding for resistance against biotic and abiotic stresses: From classical to MAS breeding. **Euphytica**, Dordrecht, v. 147, n. 1-2, p. 105–131, 2006.

MIZOI, J.; SHINOZAKI, K.; YAMAGUCHI-SHINOZAKI, K. AP2/ERF family transcription factors in plant abiotic stress responses. **Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms**, Amsterdam, v. 1819, n. 2, p. 86–96, 2012.

NAKANO, T.; SUZUKI, K.; FUJIMURA, T.; SHINSHI, H. Genome-Wide Analysis of the ERF Gene Family. **Plant Physiology**, Rockville, v. 140, n. 2, p. 411–432, 2006.

NAKASHIMA, K.; ITO, Y.; YAMAGUCHI-SHINOZAKI, K. Transcriptional regulatory networks in response to abiotic stresses in *Arabidopsis* and grasses. **Plant physiology**, Rockville, v. 149, n. 1, p. 88–95, 2009.

NAYAK, S. N.; BALAJI, J.; UPADHYAYA, H. D.; HASH, C. T.; KISHOR, P. B. K.; CHATTOPADHYAY, D.; RODRIQUEZ, L. M.; BLAIR, M. W.; BAUM, M.; MCNALLY, K.; THIS, D.; HOISINGTON, D. A.; VARSHNEY, R. K. Isolation and sequence analysis of *DREB2A* homologues in three cereal and two legume species. **Plant Science**, Amsterdam, v. 177, n. 5, p. 460–467, 2009.

NOVILLO, F.; ALONSO, J. M.; ECKER, J. R.; SALINAS, J. CBF2/DREB1C is a negative regulator of CBF1/DREB1B and CBF3/DREB1A expression and plays a central role in stress tolerance in *Arabidopsis*. **Proceedings of the National Academy of Sciences of the USA**, Washington, DC, v. 101, n. 11, p. 3985–3990, 2004.

OHTA, M.; MATSUI, K.; HIRATSU, K.; SHINSHI, H.; OHME-TAKAGI, M. Repression domains of class II ERF transcriptional repressors share an essential motif for active repression. **The Plant Cell**, Rockville, v. 13, n. 8, p. 1959–1968, 2001.

PFAFFL, M. W.; HORGAN, G. W.; DEMPFLE, L. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. **Nucleic Acids Research**, Oxford, v. 30, n. 9, p. e36, 2002.

RAE, L.; LAO, N. T.; KAVANAGH, T. A. Regulation of multiple aquaporin genes in *Arabidopsis* by a pair of recently duplicated DREB transcription factors. **Planta**, Berlin, v. 234, n. 3, p. 429–444, 2011.

RAINEY, K. M.; GRIFFITHS, P. D. Differential response of common bean genotypes to high temperature. **Journal of the American Society for Horticultural Science**, Alexandria, v. 130, n. 1, p. 18–23, 2005.

RAMAKERS, C.; RUIJTER, J. M.; LEKANNE DEPREZ, R. H.; MOORMAN, A. F. M. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. **Neuroscience Letters**, Shannon, v. 339, n. 1, p. 62–66, 2003.

RAO, G.; SUI, J.; ZENG, Y.; HE, C.; ZHANG, J. Genome-wide analysis of the AP2/ERF gene family in *Salix arbutifolia*. **FEBS Open Bio**, Amsterdam, v. 5, p. 132–137, 2015.

RECCHIA, G. H.; CALDAS, D. G. G.; BERALDO, A. L. A.; DA SILVA, M. J.; TSAI, S. M. Transcriptional analysis of drought-induced genes in the roots of a tolerant genotype of the common bean (*Phaseolus vulgaris* L.). **International Journal of Molecular Sciences**, Basel, v. 14, n. 4, p. 7155–7179, 2013.

REIS, R. R.; ANDRADE DIAS BRITO DA CUNHA, B.; MARTINS, P. K.; MARTINS, M. T. B.; ALEKCEVETCH, J. C.; CHALFUN-JÚNIOR, A.; ANDRADE, A. C.; RIBEIRO, A. P.; QIN, F.; MIZOI, J.; YAMAGUCHI-SHINOZAKI, K.; NAKASHIMA, K.; CARVALHO, J. D. F. C.; DE SOUSA, C. A. F.; NEPOMUCENO, A. L.; KOBAYASHI, A. K.; MOLINARI, H. B. C. Induced over-expression of *AtDREB2A* improves drought tolerance in sugarcane. **Plant Science**, Amsterdam, v. 221-222, p. 59–68, 2014.

SAKUMA, Y.; LIU, Q.; DUBOUZET, J. G.; ABE, H.; SHINOZAKI, K.; YAMAGUCHI-SHINOZAKI, K. DNA-binding specificity of the ERF/AP2 domain of *Arabidopsis* DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. **Biochemical and Biophysical Research Communications**, New York, v. 290, n. 3, p. 998–1009, 2002.

SCHMID, M.; DAVISON, T. S.; HENZ, S. R.; PAPE, U. J.; DEMAR, M.; VINGRON, M.; SCHÖLKOPF, B.; WEIGEL, D.; LOHMANN, J. U. A gene expression map of *Arabidopsis thaliana* development. **Nature Genetics**, New York, v. 37, n. 5, p. 501–506, 2005.

SCHMUTZ, J.; CANNON, S. B.; SCHLUETER, J.; MA, J.; MITROS, T.; NELSON, W.; HYTEN, D. L.; SONG, Q.; THELEN, J. J.; CHENG, J.; XU, D.; HELLSTEN, U.; MAY, G. D.; YU, Y.; SAKURAI, T.; UMEZAWA, T.; BHATTACHARYYA, M. K.; SANDHU, D.; VALLIYODAN, B.; LINDQUIST, E.; PETO, M.; GRANT, D.; SHU, S.; GOODSTEIN, D.; BARRY, K.; FUTRELL-GRIGGS, M.; ABERNATHY, B.; DU, J.; TIAN, Z.; ZHU, L.; GILL, N.; JOSHI, T.; LIBAULT, M.; SETHURAMAN, A.; ZHANG, X.-C.; SHINOZAKI, K.; NGUYEN, H. T.; WING, R. A.; CREGAN, P.; SPECHT, J.; GRIMWOOD, J.; ROKHSAR, D.; STACEY, G.; SHOEMAKER, R. C.; JACKSON, S. A. Genome sequence of the palaeopolyploid soybean. **Nature**, London, v. 463, n. 7278, p. 178–183, 2010.

SCHMUTZ, J.; MCCLEAN, P. E.; MAMIDI, S.; WU, G. A.; CANNON, S. B.; GRIMWOOD, J.; JENKINS, J.; SHU, S.; SONG, Q.; CHAVARRO, C.; TORRES-TORRES, M.; GEFFROY, V.; MOGHADDAM, S. M.; GAO, D.; ABERNATHY, B.; BARRY, K.; BLAIR, M.; BRICK, M. A.; CHOVATIA, M.; GEPTS, P.; GOODSTEIN, D. M.; GONZALES, M.; HELLSTEN, U.; HYTEN, D. L.; JIA, G.; KELLY, J. D.; KUDRNA, D.; LEE, R.; RICHARD, M. M. S.; MIKLAS, P. N.; OSORNO, J. M.; RODRIGUES, J.; THAREAU, V.; URREA, C. A.; WANG, M.; YU, Y.; ZHANG, M.; WING, R. A.; CREGAN, P. B.; ROKHSAR, D. S.; JACKSON, S. A. A reference genome for common bean and genome-wide analysis of dual domestications. **Nature Genetics**, London, v. 46, n. 7, p. 707–713, 2014.

SHKOLNIK-INBAR, D.; BAR-ZVI, D. Expression of *ABSCISIC ACID INSENSITIVE 4* (ABI4) in developing *Arabidopsis* seedlings. **Plant Signaling & Behavior**, Austin, v. 6, n. 5, p. 694–696, 2011.

SOHN, K. H.; LEE, S. C.; JUNG, H. W.; HONG, J. K.; HWANG, B. K. Expression and functional roles of the pepper pathogen-induced transcription factor *RAV1* in bacterial disease resistance, and drought and salt stress tolerance. **Plant Molecular Biology**, Dordrecht, v. 61, n. 6, p. 897–915, 2006.

SONG, X.; LI, Y.; HOU, X. Genome-wide analysis of the AP2 / ERF transcription factor superfamily in Chinese cabbage (*Brassica rapa* ssp. *pekinensis*). **BMC Genomics**, London, v. 14, n. 1, p. 1, 2013.

SPONCHIADO, B. N.; WHITE, J. W.; CASTILLO, J. A.; JONES, P. G. Root growth of four common bean cultivars in relation to drought tolerance in environments with contrasting soil types. **Experimental Agriculture**, Cambridge, v. 25, p. 249-257, 1989.

SUN, S.; YU, J. P.; CHEN, F.; ZHAO, T. J.; FANG, X. H.; LI, Y. Q.; SUI, S. F. TINY, a dehydration-responsive element (DRE)-binding protein-like transcription factor connecting the DRE- and ethylene-responsive element-mediated signaling pathways in *Arabidopsis*. **Journal of Biological Chemistry**, Bethesda, v. 283, n. 10, p. 6261–6271, 2008.

TAMURA, K.; STECHER, G.; PETERSON, D.; FILIPSKI, A.; KUMAR, S. MEGA6: Molecular evolutionary genetics analysis version 6.0. **Molecular Biology and Evolution**, Chicago, v. 30, n. 12, p. 2725–2729, 2013.

THAMILARASAN, S. K.; PARK, J.-I.; JUNG, H.-J.; NOU, I.-S. Genome-wide analysis of the distribution of AP2/ERF transcription factors reveals duplication and CBFs genes elucidate their potential function in *Brassica oleracea*. **BMC Genomics**, London, v. 15, p. 422, 2014.

TSUTSUI, T.; KATO, W.; ASADA, Y.; SAKO, K.; SATO, T.; SONODA, Y.; KIDOKORO, S.; YAMAGUCHI-SHINOZAKI, K.; TAMAOKI, M.; ARAKAWA, K.; ICHIKAWA, T.; NAKAZAWA, M.; SEKI, M.; SHINOZAKI, K.; MATSUI, M.; IKEDA, A.; YAMAGUCHI, J. DEAR1, a transcriptional repressor of DREB protein that mediates plant defense and freezing stress responses in *Arabidopsis*. **Journal of Plant Research**, Tokyo, v. 122, n. 6, p. 633–643, 2009.

WARNES, A. G. R.; BOLKER, B.; BONEBAKKER, L.; HUBER, W.; LIAW, A.; LUMLEY, T.; MAGNUSSON, A.; MOELLER, S.; SCHWARTZ, M. **Package "gplots"**. Murray Hill: CRAN, 2015.

WHITE, J. W.; OCHOA, R. M.; IBARRA, F. P.; SINGH, S. P. Inheritance of seed yield, maturity and seed weight of common bean (*Phaseolus vulgaris*) under semi-arid rainfed conditions. **Journal of Agricultural Science**, Cambridge, v. 122, p. 265-273, 1994.

WU, H.; LV, H.; LI, L.; LIU, J.; MU, S.; LI, X.; GAO, J. Genome-Wide Analysis of the AP2 / ERF Transcription Factors Family and the Expression Patterns of *DREB* Genes in Moso Bamboo (*Phyllostachys edulis*). **Plos One**, San Francisco, v. 10, n. 5, e0126657, 2015.

YAMAGUCHI-SHINOZAKI, K.; SHINOZAKI, K. A novel cis-acting element in an *Arabidopsis* gene is involved in responsiveness to drought, low-temperature, or high-salt stress. **The Plant Cell**, Rockville, v. 6, n. 2, p. 251–264, 1994.

YU, J.; SUN, S.; JIANG, Y.; MA, X.; CHEN, F.; ZHANG, G.; FANG, X. Single molecule study of binding force between transcription factor *TINY* and its DNA responsive element. **Polymer**, Oxford, v. 47, p. 2533–2538, 2006.

ZHANG, G.; CHEN, M.; CHEN, X.; XU, Z.; GUAN, S.; LI, L. C.; LI, A.; GUO, J.; MAO, L.; MA, Y. Phylogeny, gene structures, and expression patterns of the ERF gene family in soybean (*Glycine max* L.). **Journal of Experimental Botany**, Oxford, v. 59, n. 15, p. 4095–4107, 2008.

ZHAO, T.; LIANG, D.; WANG, P. Genome-wide analysis and expression profiling of the DREB transcription factor gene family in *Malus* under abiotic stress. **Molecular Genetics and Genomics**, Heidelberg, v. 287, n. 5, p. 423–436, 2012.

ZHAO, T.; XIA, H.; LIU, J.; GENE, Á.; VITIS, Á. The gene family of dehydration responsive element-binding transcription factors in grape (*Vitis vinifera*): genome-wide identification and analysis, expression profiles, and involvement in abiotic stress resistance. **Molecular Biology Reports**, Dordrecht, v. 41, n.3, p. 1577-1590, 2014.

4 DISCOVERY OF SNP MARKERS ASSOCIATED WITH *DREB* GENES IN COMMON BEAN THROUGH SANGER SINGLE-GENE SEQUENCING AND THE BARCBean6k_3 BEAD CHIP

Abstract

The identification of genes underlying specific traits through QTL or association mapping have been intensively studied in crop plants. However, limited work has been done with genes associated with abiotic stress tolerance in common bean (*Phaseolus vulgaris* L.). *DREB* genes are majorly involved in the regulation of abiotic stresses, but scarce information is available about their homologs in common bean, which accounts for 54 genes. In this study, we investigated the nucleotide polymorphisms within introns, exons and UTR sequences of four *PvDREB* genes (*PvDREB1F*, *PvDREB2A, PvDREB5A* and *PvDREB6B*) in 17 genotypes representing the two major centers of domestication of the species (Andean and Mesoamerican). Concomitantly, we searched for the nearest SNP markers from each of the 54 *DREB*, based on the SNP-array (BARCBean6K_3 Bead Chip) previously developed for common bean. Our results showed nucleotide variants (SNP, frame shifts, INDELS) in all the four genes, especially *PvDREB6B*, which exhibited the highest number of polymorphisms within the common bean panel (18 SNP sites). We found 51 non-redundant SNP close to the transcription initiation site of all *DREB*. Several SNP sites from the Bead Chip were found within a 10 kb distance from the transcription initiation site of each *DREB*. Principal coordinate analysis with 43 high-quality SNP markers nearby *DREB* loci enabled clear distinction between Mesoamerican and Andean gene pools. Thereby, the SNP chip proved to be a useful platform for identifying variation nearby *DREB* genes in common bean. Exploring the variation of *DREB* genes and identifying markers closely located are helpful resources toward next studies aimed at mapping traits affected by such genes.

Keywords: Nucleotide variability. INDEL. BARCBean6k_3 Bead Chip. Marker-assisted selection.

Resumo

A identificação de genes associados com caracteres específicos através do mapeamento de QTL e do mapeamento associativo tem sido intensamente estudados em plantas de interesse agronômico. No entanto, poucos estudos estão disponíveis para genes associados a respostas a estresses abióticos em feijoeiro (*Phaseolus vulgaris* L.). Genes *DREB* (*Dehydration Responsive Element-Binding*) estão principalmente envolvidos na regulação de estresses abióticos, no entanto poucos estudos estão disponíveis para os 54 genes desta subfamília no feijoeiro. Neste estudo foram investigados os polimorfismos de nucleotídeos ao longo de sequências de íntrons, éxons e UTR de genes *PvDREB* (*PvDREB1F*, *PvDREB2A, PvDREB5A* and *PvDREB6B*) em 17 genótipos de feijoeiro representando os dois principais centros de domesticação da espécie, Andino e Mesoamericanos. Também foi incluída uma linhagem de *P. acutifolius*. Concomitantemente, os marcadores SNP mais próximos de cada um dos 54 *PvDREB* foram analisados através da plataforma BARCBean6K_3 Bead Chip, previamente desenvolvida para o feijoeiro. Os resultados mostraram variações nucleotídicas (SNP, *frame shifts*, INDEL) em todos os quatro genes, especialmente *PvDREB6B*, que mostrou o número mais elevado de polimorfismos entre os genótipos de feijoeiro (18 SNP). Na análise do SNP chip foram encontrados 51 SNP não-redundantes próximos aos sítios de iniciação da transcrição de de todos os genes *DREB*. Diversos marcadores SNP estão a uma distância de menos de 10 kb do sítio de iniciação da transcrição de cada gene *DREB* do chip. Análise de coordenadas principais com 43 SNP de elevada qualidade possibilitaram nítida distinção entre genótipos Mesoamericanos e Andinos. Desse modo, o SNP chip mostrou ser uma plataforma de grande utilidade na identificação de variação nucleotídica próxima a genes *DREB* em feijoeiro. Estes resultados são importantes para próximos estudos direcionados ao mapeamento de caracteres associados à variação destes genes.

Palavras-chave: Variabilidade nucleotídica. INDEL. BARCBean6k_3 Bead Chip. Seleção assistida por marcadores.

4.1 Introduction

Common bean (*Phaseolus vulgaris* L.) is a crop of prime economic and nutritional interest for many developing countries. It is one of the basic components of the dietary routine in several areas of Latina America and Africa (BEEBE, 2012; BEEBE et al., 2014). It is also produced in many other regions throughout the world, such as in North America, Europe, and Asia. Beans are a rich source of iron, proteins and other essential components of a regular diet. Common bean is mostly grown by small farmers, which, in general, have no proper equipment or supplies for high throughput cultivation systems (BROUGHTON et al., 2003).

The low technology cultivation systems for common bean associated with several constraints from the environment account for several yield losses every year. Approximately 60% of the cultivated area is naturally prone to drought episodes (BEEBE et al., 2013). Nutritional deficiencies (BROUGHTON et al., 2003; GRAHAM; VANCE, 2003), high soil salinity (GHIEY, 2000; GAMA et al., 2009), excessive heating (RAINEY; GRIFFITHS, 2005) or cold (SRIVASTAVA, 2015), flooding (CAUDLE; MARICLE, 2012), among other abiotic stresses, represent significant challenges for common bean production. It urges the attention to understand and identify the molecular mechanisms implied in abiotic stress acclimation. A critical step is recognizing genes involved in the process. Therefore, the identification of molecular markers such as single nucleotide polymorphisms associated with stress-responsive genes might be helpful to understand the association between traits and genomic regions.

The evolution of molecular fingerprinting techniques over the last three decades (RFLP, RAPD, AFLP, SSR, ISSR) has allowed the study of genetic diversity and structure and the generation of various genetic maps for common bean (e.g. KOINANGE et al., 1996; FREYRE et al., 1998; HANAI et al., 2010). Furthermore, next generation sequencing methodologies have been supplying numberless genomic-scale data, giving rise to extensive single nucleotide polymorphisms datasets for genotyping and mapping (e.g. GALEANO et al., 2009; HYTEN et al., 2010; CORTÉS; CHAVARRO; BLAIR, 2011; GALEANO et al., 2012; BLAIR et al., 2013). SNP markers have also been analyzed within specific genes from common bean, enabling the detection of population structure and evolutionary features (NANNI et al., 2011; CORTÉS et al., 2012; GIOIA et al., 2013), establishing association with specific traits (e.g. KWAK and GEPTS, 2012; KWAK et al., 2012) and determining the origin and domestication centers of *P. vulgaris* (e.g. BITOCCHI et al., 2012; BITOCCHI et al., 2013; SCHMUTZ et al., 2014).

Despite the advances in mapping common bean populations with high-throughput marker systems and QTL detection for several traits of economic importance, only a few studies are currently available for genes associated with abiotic stresses, especially for drought-related traits. One of the first maps was developed for two RIL populations (Sierra x AC1028, Sierra x Lef-2RB), using RAPD markers. Lines differed in tolerance and adaptation to drought, and five markers exhibited association with yield under stress and non-stress conditions (SCHNEIDER et al., 1997). Moreover, a Mesoamerican intra-gene pool genetic map was developed with RAPD, AFLP and SSR markers screened on a F5 population derived from the cross between the drought-adapted BAT 477 and the drought-sensitive DOR 364. Significant QTL were encountered for yield components and phenology traits (BLAIR et al., 2012). The same population was used to map QTL associated with root patterning traits (ASFAW and BLAIR, 2012) and for photosynthates acquisition, remobilization and accumulation (ASFAW; BLAIR; STRUIK, 2012).

The development of the 6 K SNP BeadChip (BeanCAP Project, MICHIGAN STATE UNIVERSITY, 2012) and other SNP libraries (HYTEN et al., 2010; BLAIR et al., 2013) has provided markers for several ongoing studies with common bean populations for diverse goals. As regards abiotic stresses, the most recent studies have employed SNP markers to perform QTL and association mapping for drought tolerance traits. In the first case, a dense map spanning 1351 cM was developed with 2,122 SNP from the BeadChip for a RIL population segregating for drought performance (cross of SEA 5 and CAL 96). In total, 14 QTL were identified for phenology and yield components (MUKESHIMANA et al., 2014). Alternatively, an association study based on Single-Marker Analysis (SMA), a method derived from Genome-Wide Association Study (GWAS), was performed for a population derived from two drought resistant lines (Pinto Saltillo and Pinto Villa). From 169 SNP, 83 markers were significantly associated with phenotypes including flowering time, maturity, reproductive stage, biomass and yield components (VILLORDO-PINEDA et al., 2015).

Despite the advances, there is a lack of information about candidate genes involved in abiotic stress tolerance in common bean. The *DREB* (*Dehydration Responsive Element-Binding*) genes subfamily encompasses 54 putative members in common bean (Section 3) and has received little attention. Nayak et al. (2009) and Cortés et al. (2012) were the first researchers to study two putative *PvDREB* genes from *P. vulgaris*, deciphering nucleotide variation patterns in wild and domesticated beans. Two isolated genes referred as *PvDREB2A* and *PvDREB2B* were studied. *PvDREB2A*, however, was renamed as *PvDREB6B* (Section 3 of this thesis) based on phylogenetic analyzes of the recent sequences deposited on

Phytozome database (http://phytozome.jgi.doe.gov/pz/portal.html). Cortés et al. (2012) combined SNP markers with phenotypic variables to evaluate wild and domesticated beans for predictive analyses involving allele-based selection. It was the first attempt in using *DREB* as candidate genes for association with drought-related traits in common bean.

DREB genes were first described as loci involved in drought responses in *Arabidopsis thaliana* (YAMAGUCHI-SHINOZAKI; SHINOZAKI, 1994), but further studies elucidated their responsiveness to other abiotic stresses such as to low (LIU et al., 1998) and high temperatures (SAKUMA et al., 2006), and high salinity (LIU et al., 1998). Several new *DREB* were isolated from other species. In *M. truncatula*, the salinity and drought-inducible *MtDREB2A* was isolated (CHEN et al., 2010), as well as the cold-inducible *MtDREB1C* (CHEN et al., 2010). In soybean, several *DREB* genes have been isolated such as *GmDREB2* (drought-, salinity-, cold- and abscisic acid-inducible) (CHEN et al., 2007), *GmDREBa*, *GmDREBb* and *GmDREBc* (LI et al., 2005) and the cold-inducible *GmDREB3* (CHEN et al., 2009). Moreover, some studies revealed association between SNP markers and stress-tolerance traits. A SNP marker within the coding region of *SiDREB2*, from *Setaria italica*, was found correlated with drought tolerance (LATA et al., 2011). Nucleotide variants within the promoter region of *ZmDREB2.7* were significantly associated with drought tolerance (LIU et al., 2013). A SNP from a chickpea *DREB* gene was associated with seed dry weight from experiments performed with heat and drought stress (THUDI et al., 2014).

In the previous chapter (Section 3), we described four novel *PvDREB* loci and investigated their expression profiles under abiotic stresses. Genes *PvDREB1F*, *PvDREB2A*, *PvDREB5A* and *PvDREB6B* revealed different patterns of inducibility under dehydration, salinity and low temperature. Hereby, we investigated the nucleotide variability of the four *PvDREB* based on Sanger sequencing in a pre-defined set of common bean genotypes. Furthermore, we used the BARCBean6k_3 Bead Chip system developed for common bean to detect and genotype the nearest SNP to each one of the 54 located *PvDREB*. Our results provided potential molecular markers that might be used for linkage and association mapping strategies with traits influenced by abiotic stresses.

4.2 Material and Methods

4.2.1 Plant materials

The Open Reading Frame, intron and UTR of *PvDREB1F*, *PvDREB2A*, *PvDREB5A* and *PvDREB6B* were resequenced in a set of 17 common bean genotypes. BAT 93 (Mesoamerican) and Jalo EEP558 (Andean) are the parental lines of the core mapping

population of common bean (FREYRE et al., 1998; HANAI et al., 2010). Midas (domesticated Andean) and G12873 (wild Mesoamerican) were also included, representing the population used for mapping traits associated with the domestication syndrome (KOINANGE et al., 1996). Another accession used was PI311859, of Mesoamerican origin. The third set of parental lines was IAC-Una (Mesoamerican) and CAL 143 (Andean), used to develop a mapping population screened for growth habit in Brazil (CAMPOS et al., 2010). Another Mesoamerican line used is SEA-5, a drought-tolerant line developed at Centro Internacional de Agricultura Tropical (CIAT) and used in QTL mapping for drought-related traits (MUKESHIMANA et al., 2014). BAT 477 (Mesoamerican, developed at CIAT) was also used for resequencing, as it has been considered well-adapted to drought conditions (SPONCHIADO et al., 1989; WHITE et al., 1994). It has been used in QTL mapping for drought-related traits such as yield and photosynthesis components and rooting patterns (BLAIR et al., 2012; ASFAW; BLAIR, 2012; ASFAW; BLAIR; STRUIK, 2012). Two other lines were the Mesoamerican ICA-Bunsi (white pea bean developed at Instituto Colombiano Agropecuario) and SxB 405 (cream seeded, drought-tolerant and high-yielding breeding line from CIAT). Experiments with 78 inbred lines in Ethiopia showed differences in drought-tolerance levels based on pod harvest index (ASSEFA et al., 2013). Two lines developed at the University of California Davis, the Andean UCD 0801 and UCD Canario 707 were also sequenced. These cultivars seem to exhibit low drought tolerance in field trials (data not showed). Another Mesoamerican and drought-sensitive line included was the Brazilian Rosinha G2. Finally, the Andean G19833 was used as the reference genotype for comparing the sequences, since it is the line used for the common bean genome sequencing (SCHMUTZ et al., 2014). Moreover, the tepary bean (*P. acutifolius*) accession G40111 was used as an outlier for the phylogenetic analyses. Plants of each genotype were grown in pots filled with soil in a greenhouse. Leaf tissue was collected from the first trifoliate leaf (V3 stage) and lyophilized.

4.2.2 DNA extraction

DNA was isolated with a modified precipitation protocol using ammonium acetate (PALOTTA et al., 2003). Approximately 20 mg of lyophilized and ground leaf tissue were placed within Eppendorf tubes containing 700 µl of extraction buffer (0,01 M Tris-HCl pH 7.5, 0.05 M EDTA pH 8.0, 1.25% SDS). Samples were vigorously mixed and incubated at 65ºC for 30 min. Hereafter, tubes were left at room temperature for 15 min and aid by 300 µl

of ammonium acetate 6 M. Samples were placed in the fridge for 15 min. Afterwards, tubes were centrifuged at 16,000 g for 7 min, and the supernatant was transferred to another tube containing 360 µL of cold isopropanol. After another centrifugation, pellets were washed with ethanol 95% and air-dried on the bench. All samples were solubilized with TE + RNAse 10 mg ml^{-1}, mixed and incubated at 37°C for 30 min. DNA was precipitated with 500 µL ethanol 95%. All samples were centrifuged and all the supernatant was discarded. Pellets were air-dried and resuspended with sterilized water overnight. Quantification was performed with Nanodrop Lite spectrophotometer (Thermo Scientific), and quality evaluation was based on absorbance reads (ratio 260/280) and agarose 1% gel stained with ethidium bromide.

4.2.3 Primer design for *DREB* genes

Primers were designed for resequencing genes *PvDREB1F*, *PvDREB2A*, *PvDREB5A* and *PvDREB6B* (Figure 4.1), defined in a previous study according to sequences deposited on Phytozome (http://phytozome.jgi.doe.gov/pz/portal.html) (Section 3). *PvDREB1F* (Phvul.003G212800.1) is constituted by an ORF spanning 882 bp, intercalated with an intron of 288 bp. For direct sequencing of the entire region, we designed a primer flanking a 1,318 bp region between the promoter and downstream the stop codon of the gene (F: 5'-*TTGTCAAATGAAAGTGTTTTACTGTTT*-3', R: 5'-*GTACAAAGAACCAAAAACGAAAGG*-3'). *PvDREB2A* (Phvul.011G107800.1) first intron (650 bp) and the entire ORF (600 bp) were sequenced using two pairs of primers (F: 5'-*GCTGTACAGAGGTTTTGATGTTG*-3', R: 5'-*GAGTATGCAAGTTGCACTCCT*-3', and F: 5'-*ATGTTAGTGAAAAACCACAACA*-3', R: 5'-*TCAAGACAATGAAAGATGGG*-3', respectively). The ORF of *PvDREB5A* (Phvul.008G098900.1, 483 bp) was sequenced with F: 5'-: *ATGGAAGGAGAAGGTTTAGGAG*-3' and R: 5'-*CTAGTCTTCGGGTTTAGGA*-3'), which was easily covered by sequencing in both directions (forward and reverse). The entire ORF of *PvDREB6B* (Phvul.002G254500.1) was sequenced with a pair of primers (*F: 5'-CTCCTTCTCCTCTCTTTCCTCA-3*; *R: 5'-GCCAAAGACATCCAAACATAAC-3'*) flanking 1,148 bp from the 5'-UTR to the 3'-UTR portions of the gene.

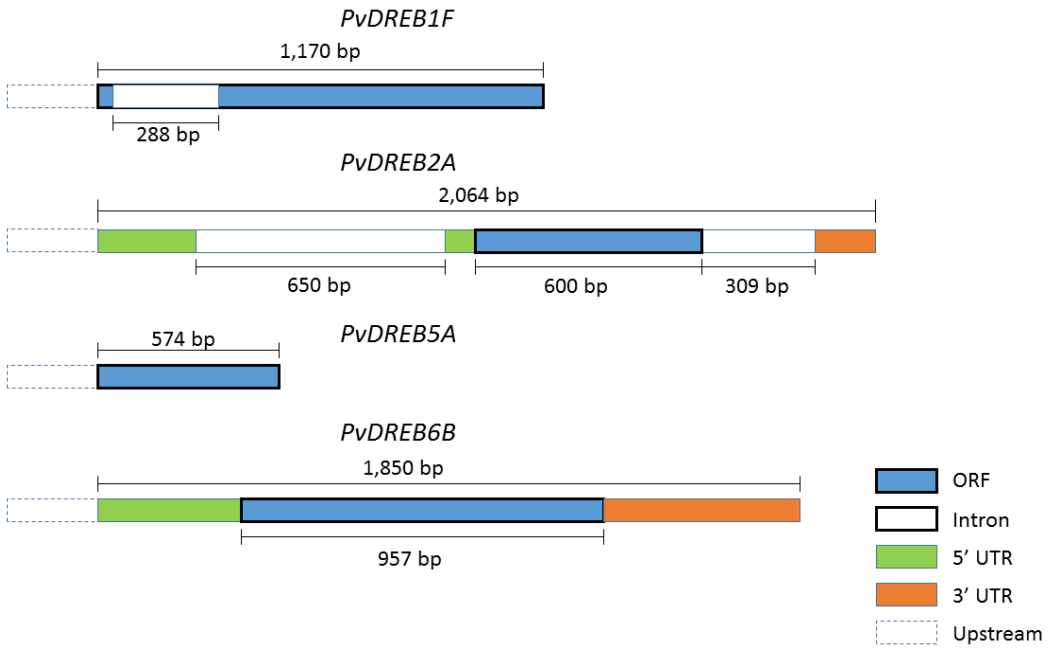


Figure 4.1 - Gene structure of four *PvDREB*, genes responsive to abiotic stresses in common bean.

4.2.4 Amplification and direct resequencing in common bean genotypes

DNA from all genotypes was diluted to 20 ng μl^{-1}. Each PCR reaction was prepared to a final volume of 50 µl for further purification of amplification products. Reactions contained 1X reaction buffer (2 mM Tris-HCl pH 8.4, 5 mM KCl), 3 mM $MgCl_2$, 0.2 mM dNTP, 0.2 µM of each primer and 1 U of High Fidelity Taq DNA Polymerase (Invitrogen). All primers designed for amplifying the *PvDREB* genes were set for annealing temperature at 59°C. Amplification conditions were set as follows: 94°C for 2 min for initial denaturation, 38 cycles of denaturation (94°C for 30 s), primer annealing (59°C for 30 s) and extension (68°C for 1 min), plus seven extra minutes for final extension at 68°C. All reactions were verified in agarose gel 1.2% for unique fragments at the expected size.

Amplification products were purified using the Wizard SV Gel and PCR Clean-Up System (Promega), following the instructions of the manufacturer. Purified samples were quantified and prepared for sequencing in both directions (forward and reverse) using the same pair of primers from the original amplifications. Sequencing was performed at the UC DNA sequencing facility, with the ABI 3730 Capillary Electrophoresis Genetic Analyzer using ABI BiDye Terminator v. 3.1. Cycle Sequencing kit.

4.2.5 Alignments, SNP detection and phylogenetic evaluation

All sequences were submitted to quality analysis with DNA Baser version 4.20.0.36 (Heracle BioSoft). Contig assembly was performed with forward and reverse sequences. Only sequences with a quality value higher than 20 were considered for the next steps. High-quality contigs were aligned with BioEdit Sequence Alignment Editor (HALL, 1999) with ClustalW multiple alignment algorithm. Alignment was also performed with CLC Sequence Viewer version 7.6 (QIAGEN Aarhus A/S), for double-checking. Alignments were used to analyze the presence of SNP among the genotypes. Polymorphic information content (PIC) of each SNP was calculated according to Nayak et al. (2009), with the equation

$$PIC = 1 - \sum_{i=1,\ n}^{n} fi^2 \qquad (6.1)$$

where *fi* stands for the frequency of the *i*th allele. The number of haplotypes was determined with DnaSP (LIBRADO; ROZAS, 2009).

Alignment files were submitted to phylogenetic analyses with Mega 6 (TAMURA et al., 2013). The evolutionary history was inferred using the Neighbor-Joining algorithm (SAITOU; NEI, 1987). Trees were designed to scale, with branch lengths in the same units as those of the evolutionary distances used for phylogenetic tree inference. The evolutionary distances were computed according to Kimura-2-parameter approach (KIMURA, 1980). All positions containing gaps and missing data were eliminated.

DNA sequences were translated into proteins in CLC Sequence viewer version 7.6 and proteins were aligned with ClustalW algorithm. Non-synonymous substitutions were checked with the alignments and their effect on protein function were predicted using PROVEAN (Protein Variation Effect Analyzer) (CHOI; CHAN, 2015).

4.2.6 BARCBean6K_3 Bead Chip analysis

We also traced SNP markers nearby all 54 putative *DREB* loci previously identified (Section 3), using the SNP position of the BARCBean6K_3 Bead Chip, a SNP array developed for common bean which comprises 5,398 markers distributed along the 11 chromosomes and some non-aligned scaffolds (BeanCAP). Since the chromosomal

positions of all SNP from the array are known, the nearest SNP to the transcription initiation site of each *DREB* was searched.

DNA samples from the 18 genotypes described were diluted to 100 ng μL^{-1}. Samples were genotyped with the SNP array at the Soybean Genomics and Improvement Laboratory (ARS/USDA) in Beltsville, Maryland. Intensity data were processed using Genome Studio software v.2011.1 (Illumina Inc., San Diego, CA, USA). Allele calls were performed with a no-call threshold of 0.15 with posterior clustering refining using heterozygotes of reference. Multivariate analysis using principal coordinate analysis (PCoA) was performed with genotypic data, based on a distance matrix, calculated with the Microsoft Excel macro GenAlEx 6.5 (PEAKALL; SMOUSE, 2012).

4.3 Results

4.3.1 Polymorphisms within *PvDREB1F*, *PvDREB2A*, *PvDREB5A* and *PvDREB6B*

The resequencing of *PvDREB1F* (ORF + intron), *PvDREB2A* (ORF + intron 1), *PvDREB5A* (ORF) and *PvDREB6B* (ORF) in 17 common bean genotypes and one *P. acutifolius* line evidenced different number of SNP markers and other nucleotide variants within each gene (Figure 5.2). The polymorphisms identified within *PvDREB1F* were located in the first exon from the start codon (positions +8, +9, +10, +23, +33, +38) and the intron between exon 1 and exon 2 (positions +87, +127, +154, +169, +214) (Figure 5.2 A). From the 11 polymorphic sites, one consisted of a frame shift (-), differing almost all Andean genotypes (presence of the nucleotide G), except Midas, from the Mesoamerican (gap in the alignment) (Figure 5.2 A). All the 10 SNP averaged PIC = 0.432, with five haplotypes (Table 5.1).

Nevertheless, we were not able to detect the first seven nucleotides from the ORF of *PvDREB1F* in all the Mesoamerican materials with an exception for the wild G12873, with the pairs of primers used for sequencing. Interestingly, the Andean Midas exhibited a similar pattern. In most polymorphic sites Midas had the same nucleotide than the Mesoamerican materials. Other primer pairs were designed for amplifying this region, and they revealed amplification product for the Andean genotypes except Midas. G12873 presented amplification (Supplementary Figure 4.1). The wild G12873 shared nucleotide polymorphisms between Andean (7) and Mesoamerican (4). No amplification and sequence were obtained for G40111 (*P. acutifolius*) in this study.

PvDREB2A exhibited a low number of polymorphic sites within the ORF of 600 bp (Figure 5.2 B). Only two SNP were detected within the common bean panel of genotypes. Additional five polymorphic sites were encountered among G40111 (*P. acutifolius*) and the

common bean panel (the sequence obtained for G40111, however, was not complete). Intron 1 from *PvDREB2A* showed seven SNP sites. In general, polymorphisms contrasted genotypes from Andean and Mesoamerican origin. Four SNP (+355, +356, +762 and +865) contrasted the wild G12873 from the other Mesoamerican materials (Figure 5.2 B). In average, high polymorphic information content was obtained for all the SNP (PIC = 0.412, six haplotypes) (Table 5.1).

The lowest number of SNP was detected within the ORF of *PvDREB5A*, a short fragment of 474 bp (reference genotype G19833 with sequence deposited on Phytozome). Only one SNP at position +33 completely distinguished the Mesoamerican from the Andean materials (PIC = 0.475, Table 5.1). However, an INDEL of 9 bp was encountered between the two gene pools. The short sequence (CGCAACAGCA) was absent in the Andean (ORF = 474 bp) and present in all Mesoamerican genotypes (ORF = 483 bp). The size of the INDEL was higher within G40111, with additional three nucleotides absent in comparison to the Mesoamerican sequences (Figure 5.2 C).

The highest number of polymorphic sites was detected within the ORF of *PvDREB6B* (Figure 5.2 D, Table 5.1). In total, 18 SNP were encountered among the common bean genotypes. More 10 SNP were detected among the *P. acutifolius* line and the common bean materials. Additionally, an INDEL of 9 bp (CACGTCAAT) was detected, being absent within the ORF of G40111 (Figure 5.2 D). The high variability of this gene has been previously explored by Nayak et al. (2009) and Cortés et al. (2012), which have named the gene as *PvDREB2A*. However, we renamed the gene as *PvDREB6B* (previous chapter) based on a detailed phylogenetic analysis of all sequences deposited on Phytozome and Genbank. Moreover, due to the lack of sequences available at the time, Nayak et al. (2009) were able to construct a contig of only 547 bp. The actual size of the open reading frame of *PvDREB6B* is 957 bp, aid by an 5'-UTR region of 386 bp and a 3'-UTR of 507, totalizing 1850 nucleotides.

A *PvDREB1F*

Gene pool	Genotype/Position	+8	+9	+10	+23	+33	+38	+87	+127	+154	+169	+214
	Reference (Phytozome)	C	A	A	C	G	T	T	C	C	G	C
ANDEAN	G19833	.	.	.	.	.	.	.	.	.	.	.
ANDEAN	Jalo EEP558	.	.	.	.	.	.	.	.	.	.	.
ANDEAN	CAL 143	.	.	.	.	.	.	.	.	.	.	.
ANDEAN	UCD 0801	.	.	.	.	.	.	.	.	.	.	.
ANDEAN	UCD Canario 707	.	.	.	.	.	.	.	.	.	.	.
ANDEAN	Midas	G	T	G	G	C	C	G	T	G	-	.
MESOAMERICAN	G12873	.	.	.	.	C	.	.	T	.	-	T
MESOAMERICAN	BAT 93	G	T	G	G	C	C	G	T	G	-	.
MESOAMERICAN	BAT 477	G	T	G	G	C	C	G	T	G	-	.
MESOAMERICAN	IAC-Carioca 80SH	G	T	G	G	C	C	G	T	G	-	.
MESOAMERICAN	RAB 96	G	T	G	G	C	C	G	T	G	-	.
MESOAMERICAN	Rosinha G2	G	T	G	G	C	C	G	T	G	-	.
MESOAMERICAN	IAC-Una	G	T	G	G	C	C	G	T	G	-	.
MESOAMERICAN	SEA-5	G	T	G	G	C	C	G	T	G	-	.
MESOAMERICAN	SxB 405	G	T	G	G	C	C	G	T	G	-	.
MESOAMERICAN	ICA-Bunsi	G	T	G	G	C	C	G	T	G	-	.
MESOAMERICAN	PI311859	G	T	G	G	C	C	G	T	G	-	.
G40111 (*Phaseolus acutifolius*)		No sequence										

C *PvDREB5A*

Gene pool	Genotype/Position	+33	+75	+76	+77	+78	+79	+80	+81	+82	+83	+84	+85	+86	+94	+308
	Reference (Phytozome)	G	-	-	-	-	-	-	-	-	-	A	C	A	G	C
ANDEAN	G19833	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ANDEAN	Jalo EEP558	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ANDEAN	CAL 143	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ANDEAN	UCD 0801	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ANDEAN	UCD Canario 707	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ANDEAN	Midas	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MESOAMERICAN	G12873	C	G	C	A	A	C	A	G	C	A	.	.	.	.	.
MESOAMERICAN	BAT 93	C	G	C	A	A	C	A	G	C	A	.	.	.	.	.
MESOAMERICAN	BAT 477	C	G	C	A	A	C	A	G	C	A	.	.	.	.	.
MESOAMERICAN	IAC-Carioca 80SH	C	G	C	A	A	C	A	G	C	A	.	.	.	.	.
MESOAMERICAN	RAB 96	C	G	C	A	A	C	A	G	C	A	.	.	.	.	.
MESOAMERICAN	Rosinha G2	C	G	C	A	A	C	A	G	C	A	.	.	.	.	.
MESOAMERICAN	IAC-Una	C	G	C	A	A	C	A	G	C	A	.	.	.	.	.
MESOAMERICAN	SEA-5	C	G	C	A	A	C	A	G	C	A	.	.	.	.	.
MESOAMERICAN	SxB 405	C	G	C	A	A	C	A	G	C	A	.	.	.	.	.
MESOAMERICAN	ICA-Bunsi	C	G	C	A	A	C	A	G	C	A	.	.	.	.	.
MESOAMERICAN	PI311859	C	G	C	A	A	C	A	G	C	A	.	.	.	.	.
G40111 (*Phaseolus acutifolius*)		C	.	.	.	.	.	.	.	.	.	-	-	-	C	G

B *PvDREB2A*

Gene pool	Genotype/Position	+284	+291	+298	+318	+339	+347	+355	+356	+357	+358	+364	+367	+399	+433	+449	+491	+517	+582	+598	+608	+609	+612	+648	+652	+665	+709	+762	+806	+849	+865	+870	+871	+1045	+1276	+1383	+1414	+1440	+1462	+1567
	Reference (Phytozome)	T	T	-	C	T	T	-	-	-	G	G	G	T	G	G	T	T	G	T	T	G	A	A	T	G	C	G	C	C	T	A	T	A	T	C	G	A	G	A
ANDEAN	G19833	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ANDEAN	Jalo EEP558	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ANDEAN	CAL 143	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ANDEAN	UCD 0801	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ANDEAN	UCD Canario 707	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T
ANDEAN	Midas	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T
MESOAMERICAN	G12873	.	.	T	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	
MESOAMERICAN	BAT 93	.	.	T	.	.	T	T	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	C	.	.	.	.	.	.	T	.	.	
MESOAMERICAN	BAT 477	.	.	T	.	.	T	T	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	C	.	.	.	.	.	.	T	.	.	
MESOAMERICAN	IAC-Carioca 80SH	.	.	T	.	.	T	T	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	C	.	.	.	.	.	.	T	.	T	
MESOAMERICAN	RAB 96	.	.	T	.	.	T	T	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	C	.	.	.	.	.	.	T	.	.	
MESOAMERICAN	Rosinha G2	.	.	T	.	.	T	T	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	C	.	.	.	.	.	.	T	.	.	
MESOAMERICAN	IAC-Una	.	.	T	.	.	T	T	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	C	.	.	.	.	.	.	T	.	T	
MESOAMERICAN	SEA-5	.	.	T	.	.	T	T	G	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	C	.	.	.	.	.	.	T	.	T	
MESOAMERICAN	SxB 405	.	.	T	.	.	T	T	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	C	.	.	.	.	.	.	T	.	T	
MESOAMERICAN	ICA-Bunsi	.	.	T	.	.	T	T	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	C	.	.	.	.	.	.	T	.	T	
MESOAMERICAN	PI311859	.	.	T	.	.	T	T	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	C	.	.	.	.	.	.	T	.	T	
G40111 (*Phaseolus acutifolius*)		G	C	T	T	A	G	T	T	T	-	.	A	A	T	A	G	C	T	A	C	C	T	G	.	A	-	.	G	T	.	G	C	-	C	T	A	T	A	.

Figure 5.2 – Nucleotide variants profile of four *DREB* genes from common bean (*PvDREB1F*, *PvDREB2A*, *PvDREB5A* and *PvDREB6B*) based on their direct resequencing on 17 genotypes with contrasting origin: six Andean and 11 Mesoamerican. An additional line, G40111 (*Phaseolus acutifolius*) was used as an outlier. Spots highlighted in green indicate the location and size of INDEL sites.

(Figure 5.2, end).

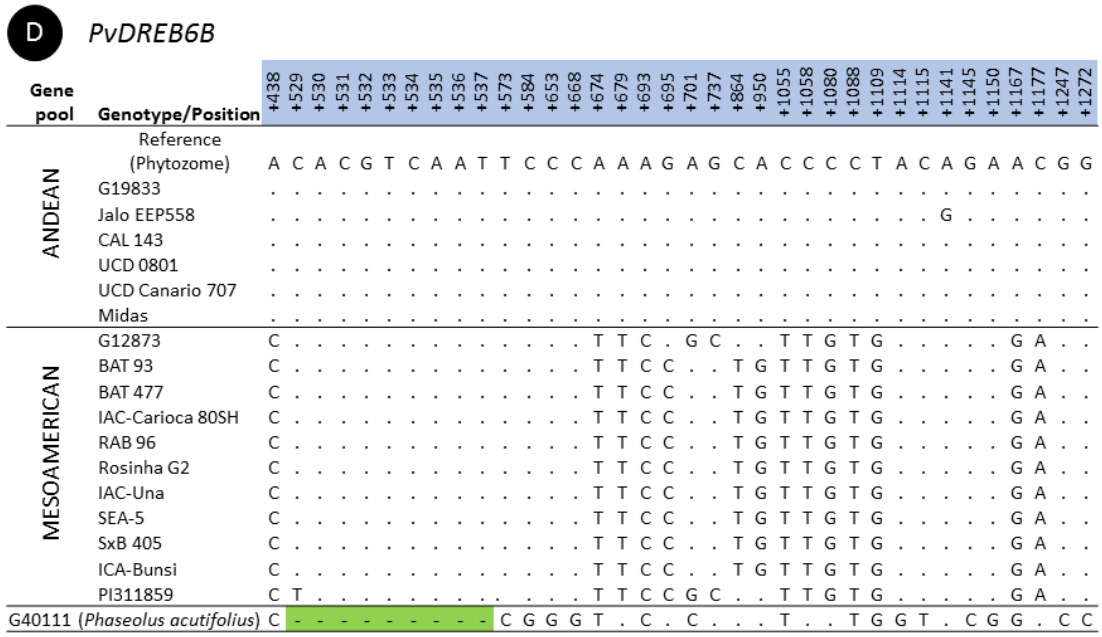


Table 5.1 – Single nucleotide polymorphisms, polymorphic information content (PIC) and number of haplotypes within the sequences of four *PvDREB* genes, based on 17 common bean genotypes.

Gene	Length of the amplified sequences (bp)	Number of SNP*	Average PIC	Number of haplotypes	Number of non-synonymous substitutions** caused by SNP and INDEL
PvDREB1F	ORF: 882, Intron: 288	10***	0.432	5	-*****
PvDREB2A	ORF: 600, Intron: 650	9***	0.412	6	1
PvDREB5A	ORF: 474, 483****	1	0.475	2	0 (SNP) and 3 (INDEL)
PvDREB6B	ORF: 957	18	0.408	5	9

* SNP from G40111 were not included for comparisons.
**Number of amino acids changed due point mutations or frame shifts and INDEL.
***Frame shifts were not accounted.
****The size of the ORF was different between Andean (474 bp) and Mesoamerican (483 bp) due to an INDEL not accounted here.
*****Not analyzed since no complete sequences were obtained.

Nucleotide variant sites resulted in non-synonymous substitutions when the ORF were translated to protein sequences (Table 5.1). Since no complete sequences were obtained for *PvDREB1F* in all genotypes, translations were performed only for *PvDREB2A*, *PvDREB5A* and *PvDREB6B* (Supplementary Figure 4.2). The point mutation +1440 within *PvDREB2A* resulted in a change from lysine (K) (Andean) to methionine (M) (Mesoamerican). The short INDEL sequence within *PvDREB5A* codes for three units of glutamine (Q). Therefore, while Andean genotypes presented four Q in a row, Mesoamerican had seven units of this amino acid. Finally, the high number of point mutations or frameshifts within *PvDREB6B* resulted in nine non-synonymous substitutions among the common bean genotypes. The INDEL from G40111 (*P. acutifolius*) represented three amino acids (S, R and Q) which appeared in common bean but not in G40111.

Phylogenetic analyses with the four *PvDREB* genes revealed their potential as phylogenetic indicators in common bean and for contrasting with *P. acutifolius* (Figure 5.3). The Andean and Mesoamerican gene pools were successfully separated from the nucleotide variants encountered within *PvDREB2A* (Figure 5.3 B), *PvDREB5A* (Figure 5.3 C) and *PvDREB6B* (Figure 5.3 D). Such result was partially accomplished with the analysis of *PvDREB1F* polymorphisms (Figure 5.3 A). Since Midas shared most haplotypes with the Mesoamerican and G12873 with the Andean at this gene, those genotypes were grouped in such gene pools, the contrary of what was observed for the other genes. We were able to obtain sequences for *PvDREB2A*, *PvDREB5A* and *PvDREB6B* for the *P. acutifolius* line. In those cases, the phylogenetic trees showed G40111 as an outlier with phylogenetic distances ranging from 0.0025 (*PvDREB5A*) to 0.012 (*PvDREB5A*). If only accounted the ORF sequences, the highest phylogenetic distance between G40111 and common bean was detected with *PvDREB6B*.

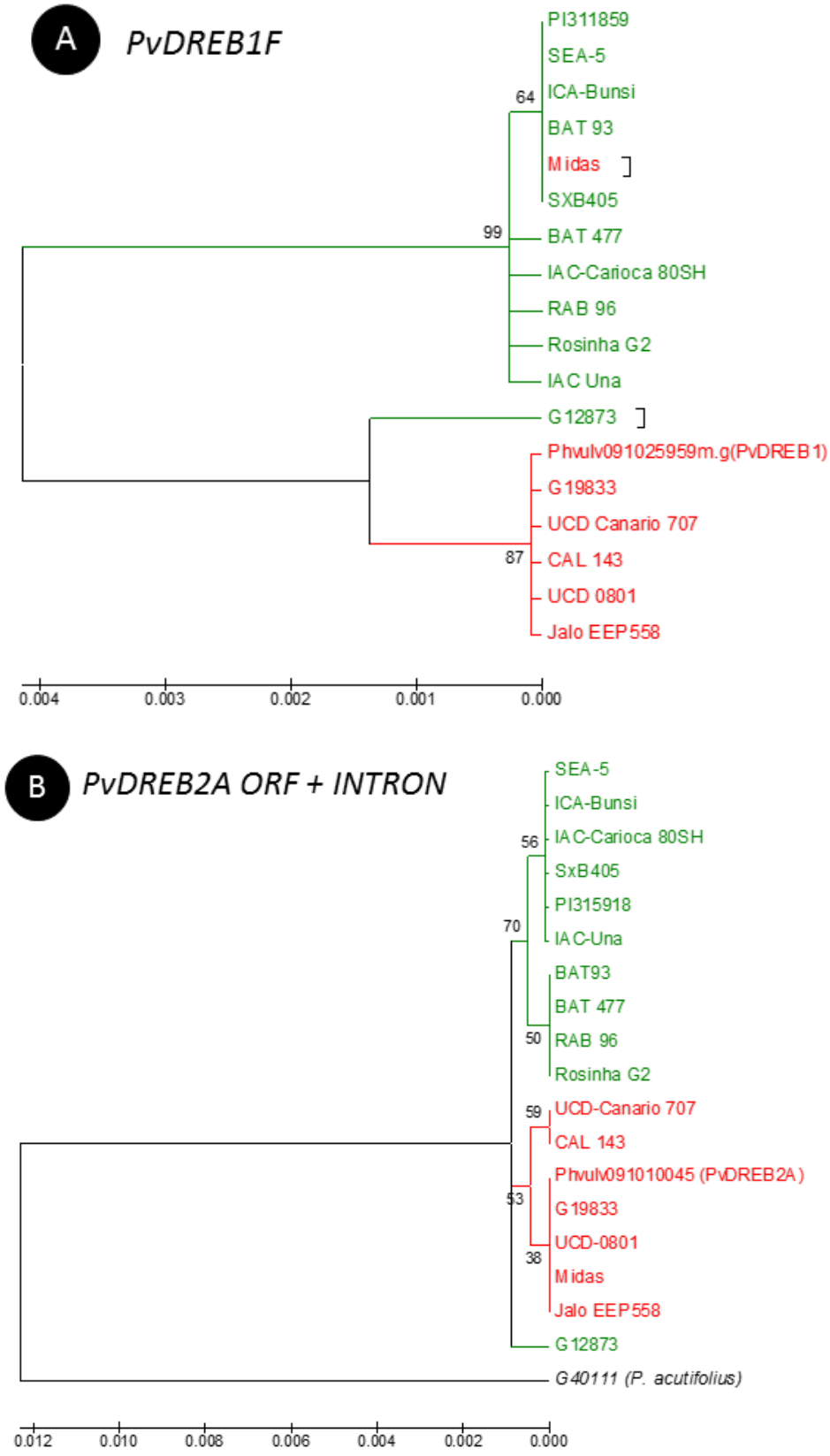
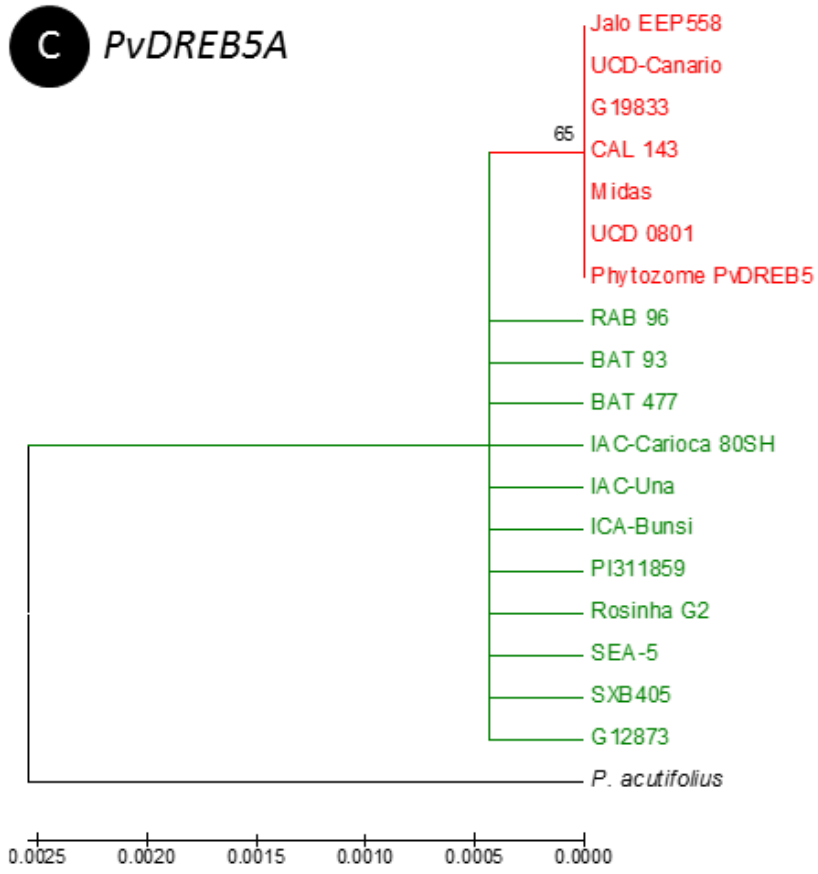
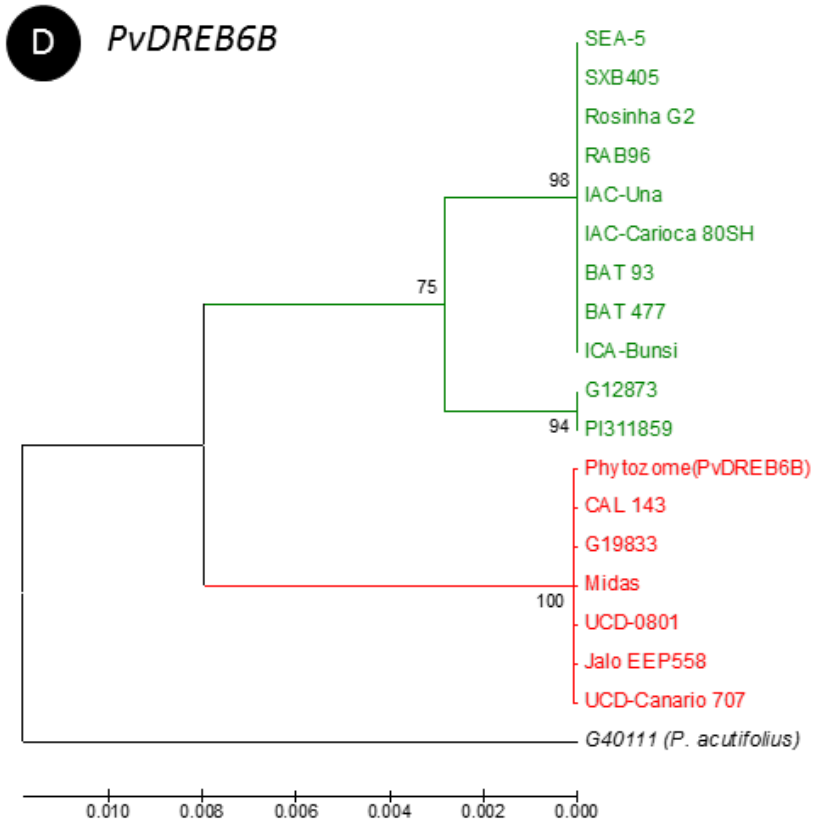


Figure 5.3 – Neighbor-joining trees based on sequence polymorphism of four *DREB* genes (ORF + intron from *PvDREB1F*, ORF + intron from *PvDREB2A*, ORF from *PvDREB5A* and ORF from *PvDREB6B*) in 17 *Phaseolus vulgaris* genotypes and one *P. acutifolius* line (G40111). No sequence of *PvDREB1F* was obtained for G40111. Green labels represent Mesoamerican genotypes while red labels indicate Andean lines (Continued).

Figure 5.3 – End.





4.3.2 Identification of SNP nearby the entire *DREB* gene subfamily with the BARCBean6K_3 Bead Chip

The closest SNP marker to each of the 54 *PvDREB* previously categorized (Section 3) were identified through the BARCBean6K_3 Bead Chip, developed for common bean (Table 5.2). Based on their chromosomal location, the distance between the transcription initiation site to the most proximal SNP from the chip ranged from 526 bp (Phvul.002G016700 to SNP ss715639434) to 362,854 bp (Phvul.010G146600 to SNP ss715645496). In fact, SNP ss715639434 was the only located within a distance less than 1,000 bp from a *PvDREB* gene (Phvul.002G016700, from A-5 subgroup according to Section 3). Several SNP were identified within a 10 kb distance from the initiation site of *PvDREB* genes: Phvul.001G010400 (4,378 bp), Phvul.001G073800 (9,353 bp), Phvul.001G187100 (3,074 bp), Phvul.002G036000 (8,525 bp), Phvul.002G056800 (2,638 bp), Phvul.003G212700 (5,786 bp), Phvul.003G212800 (1439 bp), Phvul.005G105200 (8,451 bp), Phvul.G126300 (7,091 bp), Phvul.005G170600 (5,533 bp), Phvul.007G255100 (5,819 bp), Phvul.008G098900 (8,262 bp), Phvul.008G165000 (8,380 bp) and Phvul.010G114900 (3,285 bp) (Table 5.2).

The nearest SNP to each of the four genes studied in this work (*PvDREB1F*, *PvDREB2A*, *PvDREB5A* and *PvDREB6B*) were also identified in the platform (Table 5.2). SNP ss715645943 was only 1,439 bp apart from the initiation site of *PvDREB1F* (Phvul.003G212800). SNP ss715639652 was located 20,886 bp apart from the initiation site of *PvDREB2A* (Phvul.011G107800). SNP ss715651042 was the closest marker to *PvDREB5A* (Phvul.008G098900), with a distance of 8,262 bp. SNP ss715649110 was the closest to *PvDREB6B* (Phvul.002G254500), with a distance of 10,194 bp.

Three SNP from the array were the closest markers to pairs of genes. SNP ss715649534 (chromosome 2) was the closest marker to both Phvul.002G153900 (11,816 bp) and Phvul.002G154000 (55,688 bp). SNP ss715647663 (chromosome 2) was nearby Phvul.002G035900 (24,621 bp) and Phvul.002G036000 (8,525 bp). SNP ss715646516 (chromosome 7) was located nearby Phvul.007G222500 (26,917 bp) and Phvul.007G222600 (34,056 bp). As a result, 51 distinct SNP markers were located as potential *DREB*-associated loci.

Table 5.2 – Relative position of SNP markers of the BarcBean 6k BeadChip (Illumina) from *DREB* genes in common bean genome.

***DREB* genes location**				***SNP location***		**Distance from SNP to the gene**	
Chromosome	**Initial position (bp)**	**Final position (bp)**	**Phytozome ID**	***Chromosome***	***Specific position***	**Distance from initiation site (kb)**	**Distance from end (kb)**
Chr01	931454	932855	Phvul.001G010400	Chr01	927076	-4.378	-5.779
Chr01	2101326	2102436	Phvul.001G023700	Chr01	2153411	52.085	50.975
Chr01	4680371	4681060	Phvul.001G044500	Chr01	4623673	-56.698	-57.387
Chr01	10050549	10051844	Phvul.001G073800	Chr01	10059902	9.353	8.058
Chr01	38195399	38196241	Phvul.001G136100	Chr01	38347891	152.492	151.650
Chr01	45289609	45290175	Phvul.001G187100	Chr01	45286535	-3.074	-3.640
Chr01	50886895	50888163	Phvul.001G251200	Chr01	50546984	-339.911	-341.179
Chr02	1837177	1838211	Phvul.002G016700	Chr02	1837703	0.526	-0.508
Chr02	3499002	3499535	Phvul.002G035100	Chr02	3448505	-50.497	-51.030
Chr02[1]	3545434	3545982	Phvul.002G035900	Chr02	3570055	24.621	24.073
Chr02[1]	3561530	3562521	Phvul.002G036000	Chr02	3570055	8.525	7.534
Chr02	5810982	5811876	Phvul.002G056800	Chr02	5808344	-2.638	-3.532
Chr02[2]	29528341	29528991	Phvul.002G153900	Chr02	29516525	-11.816	-12.466
Chr02[2]	29572213	29572719	Phvul.002G154000	Chr02	29516525	-55.688	-56.194
Chr02	30545248	30546790	Phvul.002G163700	Chr02	30457680	-87.568	-89.110
Chr02	42080727	42082576	Phvul.002G254500 (*PvDREB6B*)	Chr02	42070533	-10.194	-12.043
Chr02	47177131	47178333	Phvul.002G310200	Chr02	46918900	-258.231	-259.433
Chr03	42784293	42784829	Phvul.003G212700	Chr03	42790079	5.786	5.250
Chr03	42804542	42805711	Phvul.003G212800 (*PvDREB1F*)	Chr03	42803103	-1.439	-2.608
Chr03	44044998	44045690	Phvul.003G222600	Chr03	43979309	-65.689	-66.381
Chr03	44193652	44195231	Phvul.003G223600	Chr03	44265574	71.922	70.343
Chr03	46504117	46504556	Phvul.003G241700	Chr03	46585905	81.788	81.349
Chr04	39326716	39327951	Phvul.004G122000	Chr04	39341575	14.859	13.624
Chr04	45126736	45127899	Phvul.004G169800	Chr04	45170785	44.049	42.886
Chr05	31203406	31204629	Phvul.005G105200	Chr05	31211857	8.451	7.228
Chr05	32566424	32568115	Phvul.005G111200	Chr05	32370244	-196.180	-197.871
Chr05	35040940	35041680	Phvul.005G126300	Chr05	35048031	7.091	6.351
Chr05	35099239	35100039	Phvul.005G126600	Chr05	35065924	-33.315	-34.115
Chr05	39420680	39421777	Phvul.005G170600	Chr05	39426213	5.533	4.436
Chr06	22942518	22943272	Phvul.006G114100	Chr06	22953680	11.162	10.408

Chr07	5928454	5929059	Phvul.007G066500	Chr07	5986656	58.202	57.597
Chr07	33385083	33386444	Phvul.007G135300	Chr07	33621462	236.379	235.018
Chr07[3]	46214380	46215168	Phvul.007G222500	Chr07	46187463	-26.917	-27.705
Chr07[3]	46221519	46222906	Phvul.007G222600	Chr07	46187463	-34.056	-35.443
Chr07	49321858	49323149	Phvul.007G255100	Chr07	49327677	5.819	4.528
Chr08	9500856	9502893	Phvul.008G092800	Chr08	9519416	18.560	16.523
Chr08	10466381	10466854	Phvul.008G098900 (*PvDREB5A*)	Chr08	10458119	-8.262	-8.735
Chr08	23916631	23917738	Phvul.008G141000	Chr08	24044227	127.596	126.489
Chr08	42621369	42622218	Phvul.008G165000	Chr08	42612989	-8.380	-9.229
Chr08	45116054	45117855	Phvul.008G172200	Chr08	45096307	-19.747	-21.548
Chr08	53286141	53288298	Phvul.008G220400	Chr08	53313550	27.409	25.252
Chr08	53485082	53486071	Phvul.008G222400	Chr08	53651728	166.646	165.657
Chr09	2059353	2060102	Phvul.009G013200	Chr09	1896451	-162.902	-163.651
Chr09	6536069	6538397	Phvul.009G029600	Chr09	6610194	74.125	71.797
Chr09	13327339	13328429	Phvul.009G084400	Chr09	13305494	-21.845	-22.935
Chr09	16431520	16432907	Phvul.009G109600	Chr09	16444498	12.978	11.591
Chr09	18302891	18303953	Phvul.009G123300	Chr09	18434928	132.037	130.975
Chr09	33327983	33329052	Phvul.009G225000	Chr09	33353768	25.785	24.716
Chr10	8503366	8504597	Phvul.010G054000	Chr10	8541459	38.093	36.862
Chr10	38090570	38091511	Phvul.010G114900	Chr10	38093855	3.285	2.344
Chr10	41789538	41790170	Phvul.010G146600	Chr10	42153024	363.486	362.854
Chr11	9148323	9148991	Phvul.011G091400	Chr11	9133461	-14.862	-15.530
Chr11	12892620	12894683	Phvul.011G107800 (*PvDREB2A*)	Chr11	12913506	20.886	18.823
Chr11	19321983	19324035	Phvul.011G118600	Chr11	19373588	51.605	49.553

[1, 2, 3] Pairs of genes represented by the same SNP.

From the 51 SNP located in this study, only 43 presented high-quality calls for all the genotypes evaluated. It might be variable according to the genotype being genotyped and the quality of the sample that is hybridized to the chip. High-quality DNA samples and homogeneous quantities (dilution to 100 ng.μL^{-1}) are essential steps for accurate and efficient results.

From the analysis of the potential 43 *DREB*-associated SNP markers in the 18 genotypes, a clear separation between Andean and Mesoamerican genotypes was revealed, as shown by principal coordinate analysis (Figure 5.4 A). The wild G12873, however, was separated from the other Mesoamerican and close to G40111 (*P. acutifolius*). Similar results were obtained when 2,995 high-quality SNP calls (with no missing data among all genotypes) from the entire chip were used for the analysis (Figure 5.4 B). PCoA showed G12873 and PI311859 separated from the domesticated Mesoamerican lines. Thereby, the analysis of the 43 markers showed consistency in determining basic features of the genetic structure of common bean genotypes, as has been shown for whole-genome marker studies and sequence analysis of specific genes.

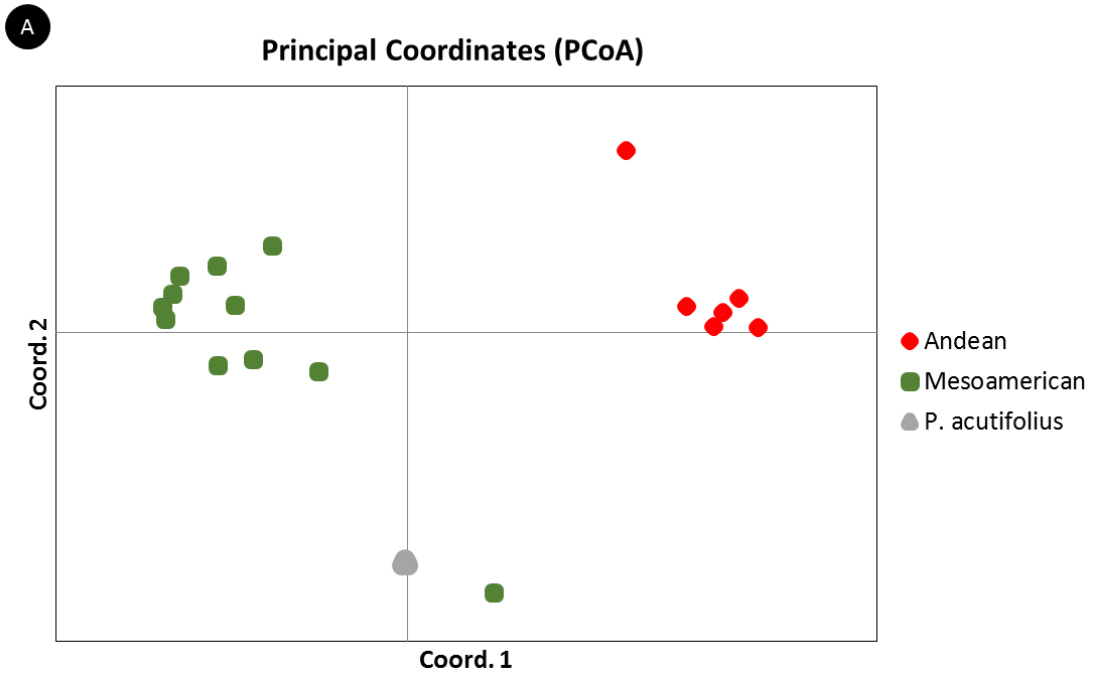
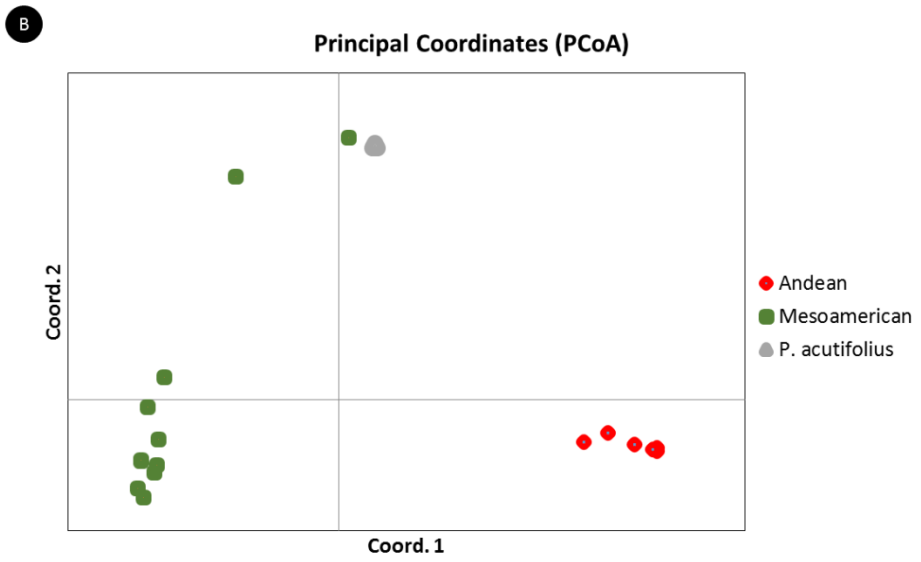


Figure 5.4 – Principal coordinate analysis (PCoA) plot of the genetic structure of 17 common bean genotypes from Andean (G19833, Jalo EEP558, Midas, UCD-0801, UCD-Canario 707, CAL 143) and Mesoamerican (G12873, PI311859, BAT 93, BAT 477, IAC-Carioca 80SH, RAB 96, Rosinha G2, IAC-Una, SEA 4, SxB 405, ICA Bunsi) background based on (A) 43 SNP markers nearby the initiation site of *PvDREB* genes and (B) 2,995 high quality SNP calls from the entire BARCBean6k_3 SNP array. An outier from *P. acutifolius* was added (line G40111) (Continued…)

Figure 5.4 – End.



4.4 Discussion

The SNP markers found for the four *PvDREB* genes, and nearby all the putative genes encoding DREB proteins in common bean (described in Section 3) distinguished Mesoamerican from Andean genotypes. Similar results have obtained with the sequence analysis of various other genes (McCLEAN et al., 2004; NANNI et al., 2011; KWAK et al., 2012; BITOCCHI et al., 2012; GIOIA et al., 2013), including *PvDREB* (CORTÉS et al., 2012). Similar genetic structure has been found when addressing the whole genome of common bean (SCHMUTZ et al., 2014).

The two intron sequences (within *PvDREB1F* and *PvDREB2A*) evaluated in this study exhibited higher number of polymorphisms than the open reading frames of the same genes (Figure 5.2). Although six polymorphic sites were found in the ORF of *PvDREB1F*, all were located at the beginning of the sequence, nearby the intron. Within *PvDREB2A* intron, seven SNP were detected, contrasting with only two within the ORF of common bean genotypes. In general, introns are generally under less selection pressure than protein-coding sequences and consequently have more point mutations. An intron might contain multiple INDELS and SNP (MCCLEAN; LEE; MIKLAS, 2004). Studies with other genes from common bean have shown higher number of polymorphic sites within intron as well, such as *INDEHISCENT* (GIOIA et al., 2013) and *PvTFL1* (KWAK et al., 2012). However, coding regions might have a significant number of polymorphisms, as shown for *PvDREB6B* in this study. In rye, a *DREB1/CBF* gene presented a high number of SNP (27 polymorphisms) spanning the coding

region of only 623 bp (LI et al., 2011). In our study, however, only 17 common bean genotypes were resequenced, most of them domesticated. The nucleotide diversity in a wild panel might be much higher.

The marked genetic structure observed through the polymorphisms detected within the coding region of *PvDREB* genes might have an impact on the conformation and function of the proteins encoded. The translated sequences of *PvDREB2A*, *PvDREB5A* and *PvDREB6B* exhibited non-synonymous substitutions (Table 5.1, Supplementary Figure 4.2), especially among gene pools. We tested for putative changes in protein function using PROVEAN tools. According to the prediction, all nine amino acid substitutions from *PvDREB6B* seem to be neutral (cutoff = -2.5). A neutral effect was also predicted for the only amino acid substitution (K/M, score 0.054, cutoff = -2.5) from PvDREB2A. On the contrary, the INDEL QQQ between amino acids 28 and 32 from the alignment of PvDREB5A proteins was predicted to be deleterious (score -2.819, cutoff = -2.5). All prediction are described in Supplementary Table 4.1. However, experimental evidence is a further step needed to confirm if such mutations effectively produce changes in the function of the encoded proteins.

Despite all considerations, our results were limited to a small set of genotypes. Further studies are required to have a deeper understanding of the evolutionary patterns shaping nucleotide diversity of the *DREB* genes and how they might be involved in adaptation to stress environments. Furthermore, these results point potential markers to be used for mapping populations for traits of interest. These markers might be incorporated with other SNP sets and used in genome-wide association studies (GWAS). The markers identified nearby all 54 *DREB* genes have potential to indicate their effect on QTL discovered for abiotic stress mapping studies in common bean.

4.5 Conclusion

Our study provided molecular markers associated with *PvDREB* genes, which are mostly associated with mechanisms to circumvent and adapt common bean to abiotic stresses. We identified SNP markers within the genes *PvDREB1F*, *PvDREB2A*, *PvDREB5A* and *PvDREB6B*. Moreover, 51 non-redundant SNP markers from the BARCBean6K_3 array were located nearby all the 54 members of the *PvDREB* gene subfamily. In general, the identified SNP were able to detect the basic genetic structure commonly observed for common bean genotypes, and even differentiate from other species, as *P. acutifolius*. The new molecular markers presented in this study might be used for genome-wide association studies for abiotic stress-related traits variation in common bean from a diverse range of backgrounds.

References

ASFAW, A.; BLAIR, M. Quantitative trait loci for rooting pattern traits of common beans grown under drought stress versus non-stress conditions. **Molecular Breeding**, Dordrecht, v. 30, n. 2, p. 681-695, 2012.

ASFAW, A.; BLAIR, M. W.; STRUIK, P. C. Multienvironment quantitative trait loci analysis for photosynthate acquisition, accumulation, and remobilization traits in common bean under drought stress. **Genes, Genomes & Genetics**, Bethesda, v. 2, n. 5, p. 579-595, 2012.

BEEBE, S. E.; RAO, I. M.; BLAIR, M. W.; ACOSTA-GALLEGOS, J. A. Phenotyping common beans for adaptation to drought. **Frontiers in Physiology**, Lausanne, v. 4, n. 3, p. 1–20, 2013.

BEEBE, S.; RAO, I.; DEVI, M.; POLANIA, J. Common beans, biodiversity, and multiple stress: challenges of drought resistance in tropical soils. **Crop and Pasture Science**, Clayton, v. 65, p. 667-675, 2014.

BITOCCHI, E.; NANNI, L. BELUCCI, E.; ROSSI, M.; GIARDINI, A.; ZEULI, P. S.; LOGOZZO, G.; STOUGAARD, J.; McCLEAN, P.; ATTENE, G.; PAPA, R. Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. **Proceedings of the Natural Academy of Sciences of the USA**, Washington, DC, v. 109, n. 14, p. 788–796, 2012.

BITOCCHI, E.; BELLUCCI, E.; GIARDINI, A.; RAU, D.; RODRIGUEZ, M.; BIAGETTI, E.; SANTILOCCHI, R.; SPAGNOLETTI ZEULI, P.; GIOIA, T.; LOGOZZO, G.; ATTENE, G.; NANNI, L.; PAPA, R. Molecular analysis of the parallel domestication of the common bean (*Phaseolus vulgaris*) in Mesoamerica and the Andes. **New Phytologist**, Oxford, v. 197, n. 1, p. 300–13, 2013.

BLAIR, M. W.; GALEANO, C. H.; TOVAR, E.; TORRES, M. C. M.; CASTRILLÓN, A. V.; BEEBE, S. E.; RAO, I. M. Development of a Mesoamerican intra-gene pool genetic map for quantitative trait loci detection in a drought tolerant x susceptible common bean (*Phaseolus vulgaris* L.) cross. **Molecular Breeding**, Dordrecht, v. 29, n. 1, p. 1-18, 2012.

BLAIR, M. W.; CORTÉS, A. J.; PENMETSA, R. V.; FARMER, A.; CARRASQUILLA-GARCIA, N.; COOK, D. R. A high-throughput SNP marker system for parental polymorphism screening, and diversity analysis in common bean (*Phaseolus vulgaris* L.). **Theoretical and Applied Genetics**, Heidelberg, v. 126, n. 2, p. 535–548, 2013.

BROUGHTON, W. J.; HERNÁNDEZ, G.; BLAIR, M.; BEEBE, S.; GEPTS, P.; VANDERLEYDEN, J. Beans (*Phaseolus* spp.) - Model food legumes. **Plant and Soil**, Dordrecht, v. 252, n. 1, p. 55–128, 2003.

CAMPOS, T.; OBLESSUC, P. R.; SFORÇA, D. A.; CARDOSO, J. M. K.; BARONI, R. M.; SOUSA, A. C. B.; CARBONELL, S. A. M.; CHIORATTO, A. F.; GARCIA, A. A. F.; RUBIANO, L. B.; SOUZA, A. P. Inheritance of growth habit detected by genetic linkage analysis using microsatellites in the common bean (*Phaseolus vulgaris* L.). **Molecular Breeding**, Dordrecht, v. 27, n. 4, p. 549–560, 2011.

CAUDLE, K. L.; MARICLE, B. R. Effects of flooding on photosynthesis, chlorophyll fluorescence, and oxygen stress in plants of varying flooding tolerance. **Transaction of the Kansas Academy of Science**, Lawrence, v. 115, n. 1/2, p. 5-18, 2012.

CHEN, J. R.; LÜ, J. J.; LIU, R.; XIONG, X. Y.; WANG, T. X.; CHEN, S. Y.; GUO, L. Bin; WANG, H. F. DREB1C from *Medicago truncatula* enhances freezing tolerance in transgenic *M. truncatula* and China Rose (*Rosa chinensis* Jacq.). **Plant Growth Regulation**, Dordrecht, v. 60, n. 3, p. 199–211, 2010.

CHEN, M.; WANG, Q. Y.; CHENG, X. G.; XU, Z. S.; LI, L. C.; YE, X. G.; XIA, L. Q.; MA, Y. Z. GmDREB2, a soybean DRE-binding transcription factor, conferred drought and high-salt tolerance in transgenic plants. **Biochemical and Biophysical Research Communications**, New York, v. 353, n. 2, p. 299–305, 2007.

CHEN, M.; XU, Z.; XIA, L.; LI, L.; CHENG, X.; DONG, J.; WANG, Q.; MA, Y. Cold-induced modulation and functional analyses of the DRE-binding transcription factor gene, *GmDREB3*, in soybean (*Glycine max* L.). **Journal of Experimental Botany**, Oxford, v. 60, n. 1, p. 121–135, 2009.

CHOI, Y.; CHAN, A. P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. **Bioinformatics**, Oxford, v. 31, n. 16, p. 2745-2747, 2015.

CORTÉS, A. J.; CHAVARRO, M. C.; BLAIR, M. W. SNP marker diversity in common bean (*Phaseolus vulgaris* L.). **Theoretical and Applied Genetics**, Heidelberg, v. 123, n. 5, p. 827–845, 2011.

CORTÉS, A. J.; THIS, D.; CHAVARRO, C.; MADRIÑÁN, S.; BLAIR, M. W. Nucleotide diversity patterns at the drought-related *DREB2* encoding genes in wild and cultivated common bean (*Phaseolus vulgaris* L.). **Theoretical and Applied Genetics**, Heidelberg, v. 125, n. 5, p. 1069–1085, 2012.

FREYRE, R.; SKROCH, P. W.; GEFFROY, V.; ADAM-BLONDON, A. F.; SHIRMOHAMADALI, A.; JOHNSON, W. C.; LLACA, V.; NODARI, R.O.; PEREIRA, P. A.; TSAI, S. M.; TOHME, J.; DRON, M.; NIENHUIS, J.; VALLEJOS, C. E.; GEPTS, P. Towards an integrated linkage map of common bean. 4. Development of a core linkage map and alignment of RFLP maps. **Theoretical and Applied Genetics**, Heidelberg, v. 97, p. 847–856, 1998.

GALEANO, C. H.; CORTÉS, A. J.; FERNÁNDEZ, A. C.; SOLER, Á.; FRANCO-HERRERA, N.; MAKUNDE, G.; VANDERLEYDEN, J.; BLAIR, M. W. Gene-Based Single Nucleotide Polymorphism Markers for Genetic and Association Mapping in Common Bean. **BMC Genetics**, London, v. 13, n. 1, p. 48, 2012.

GALEANO, C. H.; FERNÁNDEZ, A. C.; GÓMEZ, M.; BLAIR, M. W. Single strand conformation polymorphism based SNP and Indel markers for genetic mapping and synteny analysis of common bean (*Phaseolus vulgaris* L.). **BMC Genomics**, London, v. 10, p. 629, 2009.

GAMA, P. B. S.; TANAKA, K.; ENEJI, A. E.; ELTAYEB, A. E.; SIDDIG, K. El. Salt-induced stress effects on biomass, photosynthetic rate, and reactive oxygen species-scavenging enzyme accumulation in common bean. **Journal of Plant Nutrition**, New York, v. 32, n. 5, p. 837–854, 2009.

GHEYI, H. R. Problemas de salinidade na agricultura irrigada. In: OLIVEIRA, T .S.; ASSIS JUNIOR, R. N.; OMERO, R. E.; SILVA, J. R. C. (Ed.). **Agricultura, sustentabilidade e o semi-árido**. Fortaleza: Sociedade Brasileira de Ciências do Solo, 2000. p. 329- 346.

GIOIA, T.; LOGOZZO, G.; KAMI, J.; SPAGNOLETTI ZEULI, P.; GEPTS, P. Identification and characterization of a homologue to the *Arabidopsis INDEHISCENT* gene in common bean. **Journal of Heredity**, Washington, DC, v. 104, n. 2, p. 273–286, 2013.

GRAHAM, P. H.; VANCE, C. P. Update on legume utilization legumes: importance and constraints to greater use. **Plant Physiology**, Rockville, v. 131, n. 3, p. 872–877, 2003.

HALL, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. **Nucleic Acids Symposium Series**, Oxford, v. 41, p. 95-98, 1999.

HANAI, L. R.; SANTINI, L.; CAMARGO, L. E. A.; FUNGARO, M. H. P.; GEPTS, P.; TSAI, S. M.; VIEIRA, M. L. C. Extension of the core map of common bean with EST-SSR, RGA, AFLP, and putative functional markers. **Molecular Breeding**, Dordrecht, v. 25, p. 25-45, 2010.

HYTEN, D. L.; SONG, Q.; FICKUS, E. W.; QUIGLEY, C. V; LIM, J.-S.; CHOI, I.-Y.; HWANG, E.-Y.; PASTOR-CORRALES, M.; CREGAN, P. B. High-throughput SNP discovery and assay development in common bean. **BMC Genomics**, London, v. 11, p. 475, 2010.

KIMURA, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. **Journal of Molecular Evolution**, New York, v. 16, p. 111-120, 1980.

KOINANGE, E. M. K.; SINGH, S. P.; GEPTS, P. Genetic control of the domestication syndrome in common-bean. **Crop Science**, Madison, v. 36, p. 1037–1045, 1996.

KWAK, M.; TORO, O.; DEBOUCK, D. G.; GEPTS, P. Multiple origins of the determinate growth habit in domesticated common bean (*Phaseolus vulgaris*). **Annals of Botany**, Oxford, v. 110, n. 8, p. 1573–1580, 2012.

LATA, C.; BHUTTY, S.; BAHADUR, R. P.; MAJEE, M.; PRASAD, M. Association of an SNP in a novel DREB2-like gene SiDREB2 with stress tolerance in foxtail millet [*Setaria italica* (L.)]. **Journal of Experimental Botany**, Oxford, v. 62, n. 10, p. 3387–3401, 2011.

LI, X. P.; TIAN, A. G.; LUO, G. Z.; GONG, Z. Z.; ZHANG, J. S.; CHEN, S. Y. Soybean DRE-binding transcription factors that are responsive to abiotic stresses. **Theoretical and Applied Genetics**, Heidelberg, v. 110, n. 8, p. 1355–1362, 2005.

LI, Y.; HASENEYER, G.; SCHÖN, C.-C.; ANKERST, D.; KORZUN, V.; WILDE, P.; BAUER, E. High levels of nucleotide diversity and fast decline of linkage disequilibrium in rye (*Secale cereale* L.) genes involved in frost response. **BMC Plant Biology**, London, v. 11, n. 1, p. 6, 2011.

LIBRADO, P.; ROZAS, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphisms data. **Bioinformatics**, Oxford, v. 25, p. 1451-1452, 2009.

LIU, Q.; KASUGA, M.; SAKUMA, Y.; ABE, H.; MIURA, S.; YAMAGUCHI-SHINOZAKI, K.; SHINOZAKI, K. Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in *Arabidopsis*. **The Plant Cell**, Rockville, v. 10, n. 8, p. 1391–1406, 1998.

LIU, S.; WANG, X.; WANG, H.; XIN, H.; YANG, X.; YAN, J.; LI, J.; TRAN, L. S. P.; SHINOZAKI, K.; YAMAGUCHI-SHINOZAKI, K.; QIN, F. Genome-wide analysis of zmdreb genes and their association with natural variation in drought tolerance at seedling stage of *Zea mays* L. **Plos Genetics**, San Francisco, v. 9, n. 9, e1003790, 2013.

McCLEAN, P. E.; LEE, R. K.; MIKLAS, P. N. Sequence diversity analysis of dihydroflavonol 4-reductase intron 1 in common bean. **Genome**, Ottawa, v. 47, n. 2, p. 266–280, 2004.

MUKESHIMANA, G.; BUTARE, L.; CREGAN, P. B.; BLAIR, M. W.; KELLY, J. D. Quantitative trait loci associated with drought tolerance in common bean. **Crop Science**, Madison, v. 54, n. 3, p. 923–938, 2014.

NANNI, L.; BITOCCHI, E.; BELLUCCI, E.; ROSSI, M.; RAU, D.; ATTENE, G.; GEPTS, P.; PAPA, R. Nucleotide diversity of a genomic sequence similar to SHATTERPROOF (PvSHP1) in domesticated and wild common bean (*Phaseolus vulgaris* L.). **Theoretical and Applied Genetics**, Heidelberg, v. 123, n. 8, p. 1341–1357, 2011.

NAYAK, S. N.; BALAJI, J.; UPADHYAYA, H. D.; HASH, C. T.; KISHOR, P. B. K.; CHATTOPADHYAY, D.; RODRIQUEZ, L. M.; BLAIR, M. W.; BAUM, M.; MCNALLY, K.; THIS, D.; HOISINGTON, D. A.; VARSHNEY, R. K. Isolation and sequence analysis of *DREB2A* homologues in three cereal and two legume species. **Plant Science**, Amsterdam, v. 177, n. 5, p. 460–467, 2009.

PALLOTTA, M. A.; WARNER, P.; FOX, R. L.; KUCHEL, H.; JEFFERIES, S. J.; LANGRIDGE, P. Marker assisted wheat breeding in the southern region of Australia. In: INTERNATIONAL WHEAT GENETICS SYMPOSIUM, 10., 2003, Paestum, Italy. **Proceedings…** Sydney: Sydney University Press, 2003. p. 1-6.

PEAKALL, R.; SMOUSE, P. E. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. **Bioinformatics**, Oxford, v. 28, n. 9, p. 2537-2539, 2012.

SAITOU, N.; NEI, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Molecular Biology and Evolution**, Chicago, v. 4, p. 406-425, 1987.

SCHNEIDER, K. A.; BROTHERS, M. E.; KELLY, J. D. Marker assisted selection to improve drought resistance in common bean. **Crop Science**, Madison, v. 37, p. 51–60, 1997.

SPONCHIADO, B. N.; WHITE, J. W.; CASTILLO, J. A.; JONES, P. G. Root growth of four common bean cultivars in relation to drought tolerance in environments with contrasting soil types. **Experimental Agriculture**, Cambridge, v. 25, p. 249-257, 1989.

RAINEY, K. M.; GRIFFITHS, P. D. Differential response of common bean genotypes to high temperature. **Journal of the American Society for Horticultural Science**, Alexandria, v. 130, n. 1, p. 18–23, 2005.

REPINSKI, S. L.; KWAK, M.; GEPTS, P. The common bean growth habit gene PvTFL1y is a functional homolog of Arabidopsis TFL1. **Theoretical and Applied Genetics**, Heidelberg, v. 124, n. 8, p. 1539-1547, 2012.

SAKUMA, Y.; MARUYAMA, K.; QIN, F.; OSAKABE, Y.; SHINOZAKI, K.; YAMAGUCHI-SHINOZAKI, K. Dual function of an *Arabidopsis* transcription factor DREB2A in water-stress-responsive and heat-stress-responsive gene expression. **Proceedings of the National Academy of Sciences of the USA**, Washington, DC, v. 103, n. 49, p. 18822–18827, 2006.

SCHMUTZ, J.; MCCLEAN, P. E.; MAMIDI, S.; WU, G. A.; CANNON, S. B.; GRIMWOOD, J.; JENKINS, J.; SHU, S.; SONG, Q.; CHAVARRO, C.; TORRES-TORRES, M.; GEFFROY, V.; MOGHADDAM, S. M.; GAO, D.; ABERNATHY, B.; BARRY, K.; BLAIR, M.; BRICK, M. A.; CHOVATIA, M.; GEPTS, P.; GOODSTEIN, D. M.; GONZALES, M.; HELLSTEN, U.; HYTEN, D. L.; JIA, G.; KELLY, J. D.; KUDRNA, D.; LEE, R.; RICHARD, M. M. S.; MIKLAS, P. N.; OSORNO, J. M.; RODRIGUES, J.; THAREAU, V.; URREA, C. A.; WANG, M.; YU, Y.; ZHANG, M.; WING, R. A.; CREGAN, P. B.; ROKHSAR, D. S.; JACKSON, S. A. A reference genome for common bean and genome-wide analysis of dual domestications. **Nature**, London, v. 46, n. 7, p. 707–713, 2014.

SRIVASTAVA, V. Analysis on effect of cold stress in bean seeds (*Phaseolus vulgaris* L.). **American Journal of BioScience**, New York, v. 3, n. 4, p. 145-166, 2015.

TAMURA, K.; STECHER, G.; PETERSON, D.; FILIPSKI, A.; KUMAR, S. MEGA6: Molecular evolutionary genetics analysis version 6.0. **Molecular Biology and Evolution**, Chicago, v. 30, n. 12, p. 2725–2729, 2013.

THUDI, M.; UPADHYAYA, H. D.; RATHORE, A.; GAUR, P. M.; KRISHNAMURTHY, L.; ROORKIWAL, M.; NAYAK, S. N.; CHATURVEDI, S. K.; BASU, P. S.; GANGARAO, N. V. P. R.; FIKRE, A.; KIMURTO, P.; SHARMA, P. C.; SHESHASHAYEE, M. S.; TOBITA, S.; KASHIWAGI, J.; ITO, O.; KILLIAN, A.; VARSHNEY, R. K. Genetic dissection of drought and heat tolerance in chickpea through genome-wide and candidate gene-based association mapping approaches. **Plos One**, San Francisco, v. 9, n. 5, p. 40–50, 2014.

VILLORDO-PINEDA, E.; GONZÁLEZ-CHAVIRA, M. M.; GIRALDO-CARBAJO, P.; ACOSTA-GALLEGOS, J. A.; CABALLERO-PÉREZ, J. Identification of novel drought-tolerant-associated SNPs in common bean (*Phaseolus vulgaris*). **Frontiers in Plant Science**, v. 6, p. 1–9, 2015.

YAMAGUCHI-SHINOZAKI, K.; SHINOZAKI, K. A novel cis-acting element in an *Arabidopsis* gene is involved in responsiveness to drought, low-temperature, or high-salt stress. **The Plant Cell**, Rockville, v. 6, n. 2, p. 251–264, 1994.

WHITE, J. W.; OCHOA, R. M.; IBARRA, F. P.; SINGH, S. P. Inheritance of seed yield, maturity and seed weight of common bean (*Phaseolus vulgaris*) under semi-arid rainfed conditions. **Journal of Agricultural Science**, Cambridge, v. 122, p. 265-273, 1994.

5 SNP MARKERS INTERFERE WITH GENE EXPRESSION ANALYSIS: A CASE STUDY WITH A *DREB* GENE FROM COMMON BEAN

Abstract

Expression profiling of target genes in plants is a fundamental step for further understanding of their functional role. However, species such as common bean (*Phaseolus vulgaris* L.) hold high levels of genetic diversity and structure, implying in a high frequency of single nucleotide polymorphisms throughout their genomes. The extensive nucleotide variability within the Open Reading Frame of several genes in wild and domesticated accessions might lead to misinterpretation of expression profiling results. Primers flanking SNP sites among genotypes may produce mismatches that diminish the amplification potential of a target cDNA. Hereby, we show the case study analysis of an abiotic stress-inducible *DREB* (*Dehydration Responsive Element-Binding*) gene named *PvDREB6B*, which exhibits nucleotide polymorphisms among common bean genotypes. A systematic analysis demonstrated how polymorphic sites produced mismatches for primer annealing, impairing the amplification of cDNA samples with contrasting origin. These findings reveal the importance of sequencing the entire ORF of genes in all genotypes to be introduced in a gene expression profiling experiment. Ideally, primers should be designed in regions flanking no polymorphism, although variant sites generating weak mismatches might be acceptable if proved that no significant changes occur. This evaluation is essential for accurate analysis of expression profiles of genes from common bean and other plants.

Keywords: Single nucleotide polymorphisms. RT-qPCR. *PvDREB6B*. Mismatches.

Resumo

A determinação do perfil de expressão de genes-alvo em plantas é uma etapa fundamental para posterior estudo funcional dos genes. No entanto, espécies como o feijoeiro (*Phaseolus vulgaris* L.) apresentam elevada diversidade genética com alta frequência de polimorfismos de nucleotídeos simples ao longo dos genomas. A extensa variabilidade nucleotídica dentro da região codante de diversos genes em acessos selvagens e domesticados e pode levar a interpretação errônea dos resultados. *Primers* flanqueando SNP entre genótipos podem produzir *mismatches* que diminuem o potencial de amplificação do cDNA alvo. Este é um estudo de caso com um gene *DREB* (*Dehydration Responsive Element-Binding*) referido pelo nome *PvDREB6B*, induzido por estresses abióticos em feijoeiro. Este gene apresenta polimorfismos de nucleotídeos entre genótipos distintos. Uma análise sistemática demonstrou como SNP geraram *mismatches* e impediram a amplificação de cDNA de origem de genótipos contrastantes. Os resultados mostraram a importância de sequenciar a ORF completa de genes em todos os genótipos a ser incluídos em um experimento visando estudar perfis de expressão. Idealmente, *primers* devem ser desenhados em regiões não flanqueando polimorfismos, embora SNP gerando *mismatches* fracos possam ser aceitáveis se provado que não houve efeito significativo. Este estudo mostra ser essencial para avaliações acuradas no perfil de expressão de genes de feijoeiro e outras espécies de plantas.

Palavras-chave: Polimorfismos de nucleotídeos simples. RT-qPCR. *PvDREB6B*. Mismatches.

5.1 Introduction

The analysis of expression profiles of specific genes is an essential step toward unraveling their functional role in plants. Single gene expression analyzes have frequently been performed with reverse transcription quantitative-PCR (RT-qPCR). Procedures mainly involve automated systems, in which are used fluorescent dyes such as SYBR-Green I into PCR (PONCHEL et al., 2003). Moreover, large sets of genes have been analyzed through microarray technologies, such as in *Arabidopsis thaliana*, *Glycine max*, rice, etc. Probes designed for several genes are disposed in a microchip and hybridized with RNA from biological samples. After hybridization, color signals are produced, recorded and converted to expression values. Such numbers are used to determine whether transcripts accumulated or diminished in comparison to a control condition.

Nevertheless, several factors influence accurate determinations of absolute or relative amount of transcripts (REMANS et al., 2014). One researcher must optimize all conditions for PCR and hybridization, so proper comparisons are performed. Additionally, experimental conditions and sample manipulation from harvesting to RNA extraction have to be severely controlled. Quality control is essential for achieving trustworthy results (D'HAENE, 2010).

One critical factor for obtaining accurate results in gene expression analysis, though, is the design of oligonucleotides or probes with high specificity and efficiency for a gene. For primers design, several factors need to be considered, such as the primer length, the terminal nucleotide at the 3' end, the GC content and melting temperature, and the size of the fragment produced (DIEFFENBACH; LOWE; DVEKSLER, 1993). Furthermore, while designing primers researchers should account for the presence of single nucleotide polymorphisms within primer annealing sites (LEFEVER et al., 2013).

Experimental tests have proved SNP produce mismatches that potentially impair the generation of amplification products. Minimal or zero extension efficiency was reported when mismatches were introduced within the last three or four bases of the 3' ending. The low efficiency was probably due the inability of the DNA polymerase to bind to the target site (WU; HONG; LIU, 2009). Further examination showed the impact of mismatches was more pronounced according to the number of mismatched nucleotides and their distance from the 3' end. If mismatches are located > 5 bp from the 3'end, they have a moderate effect on qPCR amplification (LEFEVER et al., 2013).

The importance of evaluating SNP effects on qPCR analysis gains increased value as recent advances in next-generation sequencing technologies have revealed several SNP markers across plant and animal genomes. Comparing the expression pattern of one gene

across different genotypes of a species might be challenging if the transcript possesses nucleotide variation along its sequence. It may mask the interpretation of results from single genes to the level of microarrays (WALTER et al., 2007a). Walter et al. (2007b) identified at least one known SNP in 13,292 probes from a microarray platform for mice, and showed that 6,590 probes results were affected by the polymorphisms, accounting for 16 % of the total probes from the array.

Common bean (*Phaseolus vulgaris* L.) is a crop of fundamental economic importance since its grains are widely consumed throughout the world. *P. vulgaris* shows high genetic diversity from wild to domesticated accessions. Wild beans span from northern Mexico to northeastern Argentina (CHACÓN et al., 2007) and are structured in two main gene pools, the Mesoamerican and the Andean, from whose they were independently domesticated (BITOCCHI et al., 2013). Higher genetic diversity prevails within the Mesoamerican pool, which shows considerable intra-gene pool structure as well. A whole-genome comparison between pooled Mesoamerican and Andean samples revealed almost 9,000,000 SNP sites for the wild Mesoamerican accessions. Meanwhile, around 1,400,000 SNP were found for Andean landraces (SCHMUTZ et al., 2014). The immense number of SNP draws the attention to the need of having knowledge about the variation of genes in common bean before performing experiments aimed at expression analyzes of their transcripts.

This work is a case study, in which we showed how SNP polymorphisms interfere with the expression analysis a specific *DREB* (*DEHYDRATION RESPONSIVE ELEMENT-BINDING*) gene from common bean. Previous study has shown the expression profile of four *PvDREB* transcripts (*PvDREB1F*, *PvDREB2A*, *PvDREB5A*, and *PvDREB6B*) through RT-qPCR in common bean genotypes submitted to abiotic stresses (Section 3). Another study revealed the levels of polymorphism hold by the same genes (Section 4). Among all, *PvDREB6B* revealed a high number of nucleotide polymorphisms along its Open Reading Frame, considering a comparison between Mesoamerican and Andean genotypes. Hereby, we showed how polymorphic sites within the ORF of *PvDREB6B* have the potential to mislead the interpretation of RT-qPCR results through both semi-quantitative and automated approaches. A systematic analysis with two alternative pairs of primers revealed the primary aspects to be considered when performing gene expression profiling among distinct genotypes.

5.2 Material and Methods

5.2.1 Plant materials and experiment for gene expression analysis

In this study, we performed an experiment with seven common bean genotypes. BAT 93, BAT 477, RAB 96 and IAC-Carioca 80SH are Mesoamerican and domesticated cultivars while Jalo EEP558, Midas and G19833 have Andean background. G19833 was the genotype chosen for the whole genome sequencing of common bean (SCHMUTZ et al., 2014). BAT 93 and Jalo EEP558 are the parental lines from the core genetic map developed for common bean (FREYRE et al., 1998; HANAI et al., 2010). BAT 477 has been studied for its adaptation to drought (BLAIR et al., 2012). RAB 96 has been described as drought-sensitive (GUIMARÃES; STONE; BRUNINI, 2006) as well as IAC-Carioca 80SH (RECCHIA et al., 2013). Midas was used as one of the parental lines along with the wild bean G12873 to map traits associated with the domestication syndrome of the species (KOINANGE et al., 1996).

Our experiment consisted on the application of abiotic stresses to the seven common bean genotypes for further gene expression analyzes. Seed were germinated in pots containing a mixture of vermiculite and sand (1:1). Plants were grown in a controlled growth room (26^{o}C ± 3, around 60% air moisture) until they reached vegetative stage 3, defined by the complete expansion of the first trifoliate leaf. At this point, plants were submitted to the following stress-treatments, all in biological triplicates: polyethylene glycol 10% for dehydration; NaCl 250 mM for high salinity; cold chamber at 4^{o}C for low temperature; and abscisic acid 100 μM. Treatments lasted three hours (for reference, see Section 3). After that, leaf tissues were collected separately from each plant, and frozen in liquid nitrogen. Leaf tissues were ground and stored at -80^{o}C.

5.2.2 *PvDREB6B* resequencing and RT-qPCR analyses

In a prior study (Section 4) we showed *PvDREB6B* presented a high number of single nucleotide polymorphisms in its open reading frame sequence. The SNP markers were found by resequencing the ORF in 17 common bean genotypes from Andean (Jalo EEP558, G19833, CAL 143, UCD 0801, UCD Canario 707 and Midas) and Mesoamerican (BAT 93, BAT 477, IAC-Carioca 80SH, RAB 96, Rosinha G2, IAC-Una, SEA-5, SxB 405, ICA-Bunsi, PI311859 and G12873) background. The breeding line G40111 (*P. acutifolius*) was used as an outlier for the analysis (Section 4).

In this work, we used the sequences obtained in Section 4 to design primers flanking SNP sites and primers flanking no SNP to test whether the polymorphic sites interfered with the analysis of *PvDREB6B* expression. *PvDREB6B* was named in previous work (Section 3),

according to phylogenetic analyses. Its sequence was retrieved from Phytozome (www.phytozome.net) with the following identification code: Phvul.002G254500.1. The gene is constituted by an ORF frame of 957 bp coding for a peptide sequence of 318 amino acids. Primers for RT-qPCR analyses were designed with Primer 3.0. We defined two sets of primers: *RT-DREB6B-1* (flanking SNP markers: F: 5'–*AATTCTGCATCTCCCTCACG*-3', R: 5'-*GCTGGGCTTGATTTAGACGA*-3'), generating a fragment of 167 bp; and *RT-DREB6B-2* (flanking no SNP: F: 5'-*GCTACCTTCTTCCGTGGACT*-3', R: 5'-*CAGACACAGGCCATGACAAC*-3') amplifying a fragment with 227 bp. The sequence of the genotype BAT 477 was used for primer design.

5.2.3 RNA extraction and quantification

RNA extraction was conducted with 100 mg of leaf tissue using TRIzol® Reagent (Invitrogen) and following manufacturer's instructions. Quantification and quality evaluation was done with NanoDropTM 2000 (Thermo Scientific) spectrophotometer. Gel electrophoresis (agarose 1% in TAE buffer) was used for sample quality evaluation.

Analysis of expression profiles were performed with the two sets of primers (*RT-DREB6B-1* and *RT-DREB6B-2*) for amplifying *PvDREB6B* under all stress treatments. In addition, two reference genes, UBQ and IDE, were used for reactions, being chosen according to tissue analyzed and on stability previously studied (BORGES et al., 2011). A 100 ng RNA sample of each treatment was used to synthesize the first cDNA strand using Maxima First Strand cDNA Synthesis kit (Fermentas). Quantitative PCR reactions were prepared using 1 µL of newly synthesized cDNA, 0.25 µM of each primer and 1X SYBR® Green PCR Master Mix (Thermo Scientific). Amplifications were performed on StepOnePlus™ Real-Time PCR System (Applied Biosystems) equipment with the following steps: 10 min at 95ºC, 40 cycles of cDNA amplification at 95ºC for 15 s, 59ºC for 30 s and 72ºC for 20 s with fluorescence signal recording at this stage. A final step at 95ºC for 15 s and at 60ºC for 1 min, with fluorescence measurements at each 0.7ºC variation (from 60 to 95ºC), was included to obtain the melting curve. All reactions were performed in triplicates. The final reactions were analyzed on Step One Plus equipment (Applied Biosystems) as well as by agarose 1.2% for verification of the PCR products.

5.2.4 RT-qPCR procedures

Raw data with fluorescence levels were submitted to LinRegPCR software (RAMAKERS et al., 2003). Fluorescence was baseline corrected for each sample, and linear regression analysis of each amplification curve was performed. Samples efficiencies were calculated based on the slope of the line, ranging from 1.8 to 2.0 and with a correlation of at least 0.995, as ideal parameters. Relative expression data were obtained by REST software (PFAFFL; HORGAN; DEMPFLE, 2002) using average values of efficiency and Cq of target and reference genes.

5.2.5 DNA extractions and PCR reactions for gel electrophoresis

DNA samples were obtained from plants of the 18 genotypes (17 common beans and one line from *P. acutifolius*) previously evaluated (Section 4) for amplifications with the RT-qPCR primers. DNA extractions were performed according to Doyle and Doyle (1990), with modifications. Around 50 mg of leaf tissue from each genotype were ground in liquid nitrogen and placed in tubes with 700 µL of CTAB 2% buffer. After kept at 60°C for 30 min, samples were centrifuged at 14,000 rpm for 7 min. To each sample were added 650 µL of chloroform-isoamyl alcohol (CIA 24:1), followed by centrifugation. Supernatants were transferred to a new tube aid by the same volume of isopropanol for DNA precipitation. After centrifugation, samples were washed with absolute ethanol, air-dried and resuspended with Tris RNAse. DNA quantification and quality assay were performed with Nanodrop 2000c spectrophotometer.

DNA of all samples was diluted to 20 ng/µL. PCR reactions were prepared to 50 µL containing 10 X Rxn Buffer (20 mM Tris HCl pH 8.4, 50 mM KCl), 3 mM MgCl2, 0.2 mM dNTPs, 0.2 µM from each primers and 1 U of High Fidelity Platinum Taq DNA Polymerase. Amplification conditions were set as follows: 94°C for 2 min (initial denaturation), 38 cycles of 94°C for 30 s (denaturation), 59°C for 30 s (primer annealing) and 72°C for 1 min (extension), and final extension for 7 min. All amplification were verified for specific fragments with the expected size on agarose gel 1.2% stained with ethidium bromide (10 mg ml^{-1}).

5.3 Results

In prior work (Section 4), we searched for polymorphisms along the *PvDREB6B* ORF sequence of several Andean and Mesoamerican genotypes, using DNA samples. The pair of primers *RT-DREB6B-1* was designed in a region flanking polymorphic sites (Figure 5.1 A), creating potential mismatches for the amplification of the DNA and cDNA samples. The amplification of the 18 genotypes with these primers produced a pattern compatible with the presence of alternative nucleotides between Mesoamerican and Andean materials. A single fragment of the expected size was obtained for all Mesoamerican, while no band was detected in the Andean and the outlier G40111 (*P. acutifolius*) (Figure 5.1 B). The second pair of primers (*RT-DREB6B-2*) was able to amplify all DNA sequences, as it was designed in a region flanking no SNP (Figure 5.2B). Figure 5.1 C shows the former reverse *primer* (*RT-DREB6B-1*) was designed in a region with three mismatches (A/T, A/T and A/C), producing its inability to anneal and extend new copies of the fragments.

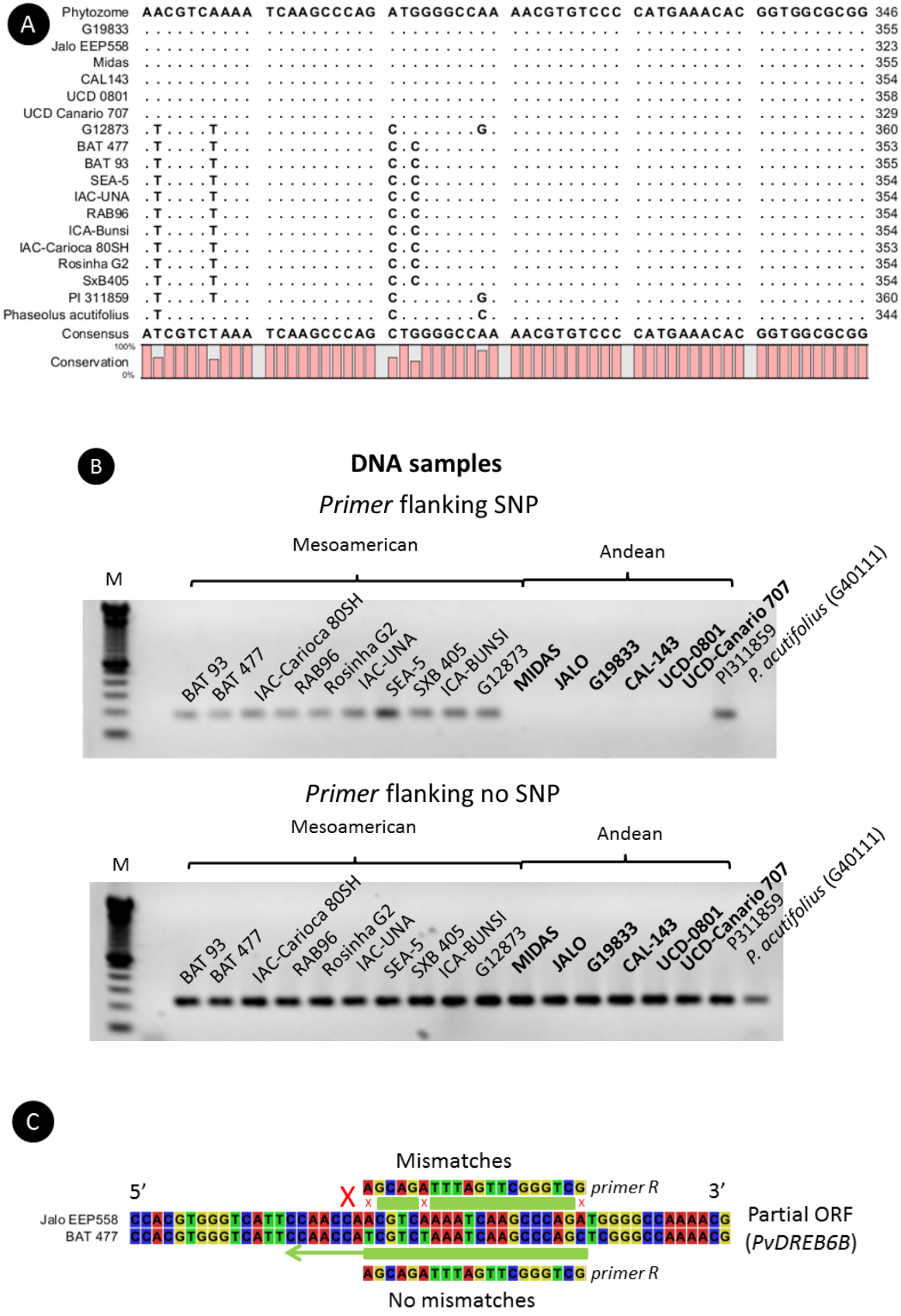


Figure 5.1 – Single nucleotide polymorphisms along the ORF of *PvDREB6B* and their influence on amplifications of one primer. A – Partial region of the ORF of *PvDREB6B* of 18 common bean genotypes showing SNP markers among materials. B – Amplifications products of two pairs of primers for *PvDREB6B*. C – The annealing site of the reverse primer flanking three SNP markers, creating mismatches and impairing the amplification. The red X indicates amplification does not proceed while the green arrow shows normal progress of the PCR.

We further compared the *RT-DREB6B-1* and *RT-DREB6B-2* primers with 40 cycles amplifications with cDNA samples from BAT 93 and Jalo EEP558 submitted to the four stress-treatments (Figure 5.2). The former set of primers (flanking three SNP among them) produced amplification products only with BAT 93. *RT-DREB6B-2* (no SNP among them) revealed fragments of strong intensity for both genotypes, with contrasting intensities among treatments. Furthermore, the samples were submitted to automated RT-qPCR with Step One Plus equipment (Applied Biosystems). The primers flanking SNP sites (*RT-DREB6B-1*) impaired curve ascendance with Jalo EEP558 samples while normal amplification curve was observed with BAT 93 samples. Some residues due primer self-annealing were detected after 32 cycles of amplification (Figure 5.3). With the primers flanking no SNP (*RT-DREB6B-2*), regular amplification curves were observed for both genotypes. Figure 5.3 shows control and cold-treated samples from the two genotypes, with the cold-treated having Ct values lower than the control.

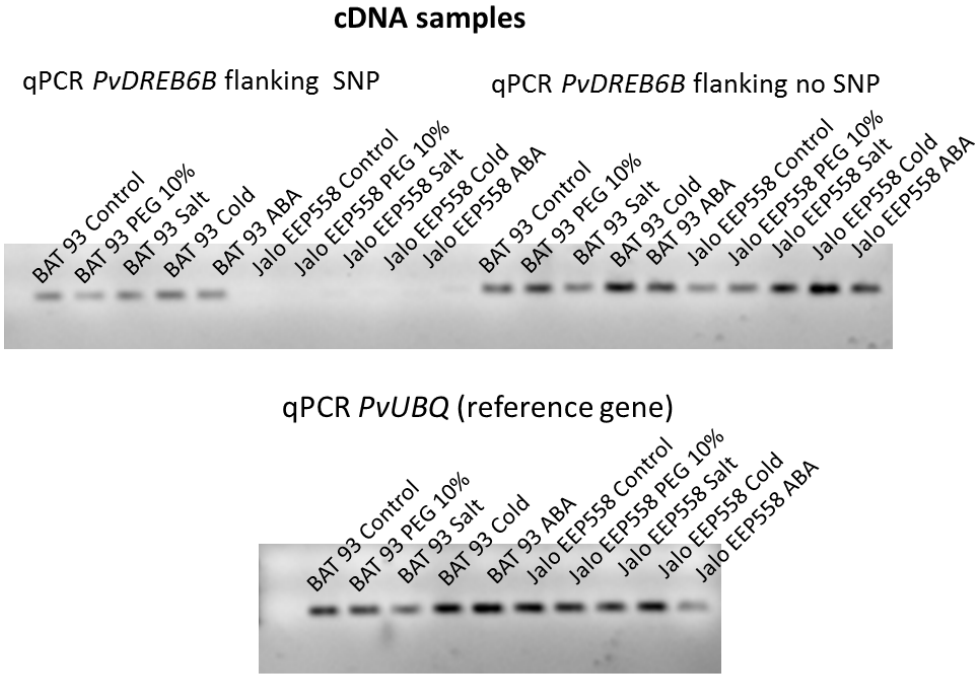


Figure 5.2 – Amplification products after 40 cycles of a RT-qPCR process with cDNA samples from two divergent common bean genotypes under distinctive abiotic stress treatments.

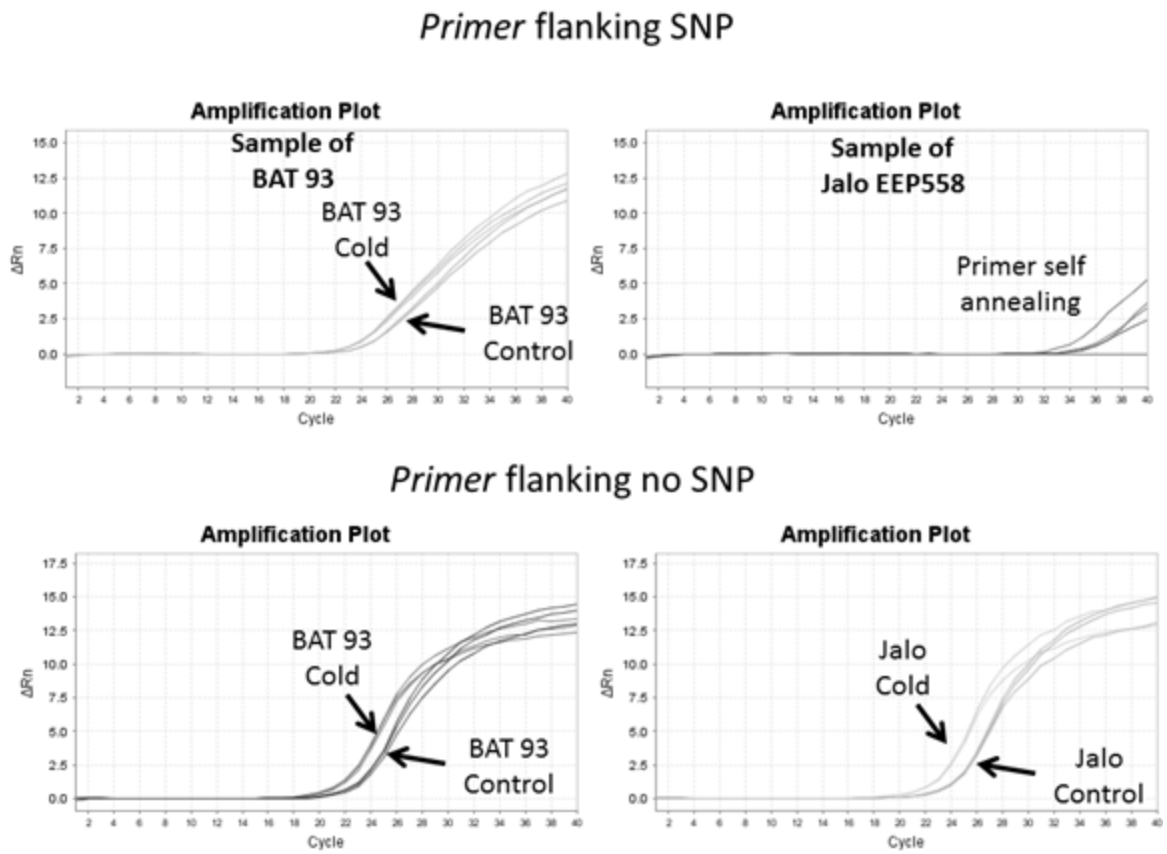


Figure 5.3 – Amplification plot after 40 cycles of qPCR reactions. The pair of primers of which the reverse flanked SNP positions impaired the amplification of cDNA samples of the genotype Jalo EEP558. Only primer annealing was detected. Conversely, normal amplification was observed for the primer flanking no SNP.

Finally, the set of primers flanking no SNP was used to perform the analysis of *PvDREB6B* expression with seven genotypes, including three Andean (Jalo EEP558, G19833 and Midas) and Mesoamerican (BAT 93, BAT 477, IAC-Carioca 80SH and RAB 96). A three-hour period of stress induction with PEG 10%, NaCl 250 mM, 4°C and ABA was applied, and leaf samples were analyzed. As in the previous study (Section 3), *PvDREB6B* transcripts were accumulated in leaves with cold treatment after three hours. RAB 96 and Jalo EEP558 presented a little change in expression, though. All the other treatments resulted in relative reduction in *PvDREB6B* transcript levels after three hours of stress (Figure 5.4).

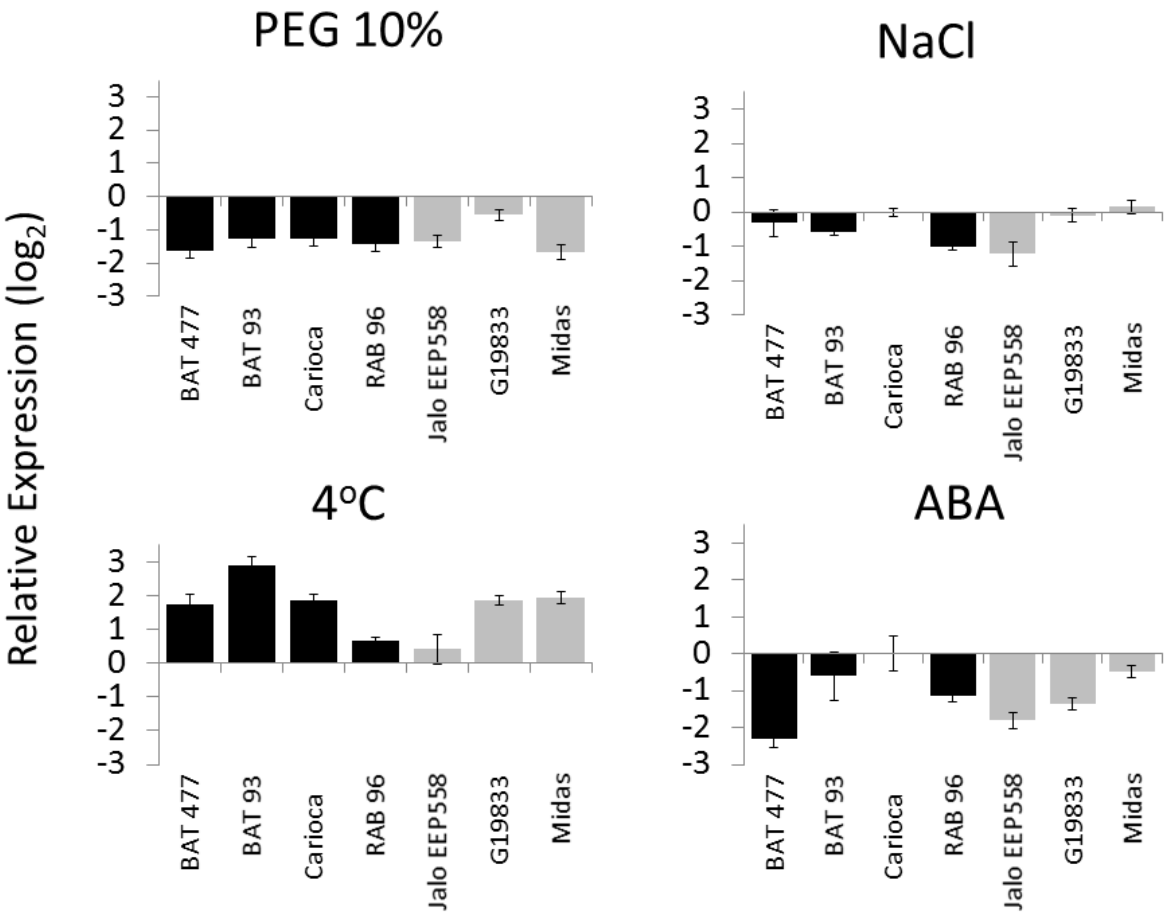


Figure 5.4 – Expression profile of *PvDREB6B* in common bean leaves of genotypes under abiotic stress treatments.

5.4 Discussion

As other studies have shown, RT-qPCR analyses are very sensitive and require several efforts to minimize variability at the technical and biological levels (REMANS et al., 2014). The Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines describe the major aspects to be considered for obtaining high-quality results with consistency between laboratories (BUSTIN, 2002). One point highlighted in this guide is the specificity of the oligonucleotides used for the RT-qPCR procedures. Any homology to pseudogenes or other unexpected targets might influence interpretation. Moreover, SNP markers flanking primers designed for such analyses have shown to interfere with the results (LEFEVER et al., 2013). Our study illustrated with a systematic analysis how SNP sites might lead to misinterpretation of relative expression results for target genes.

Several polymorphisms were encountered within the ORF of *PvDREB6B* and the reverse *primer* from the first set (*RT-DREB6B-1*) was flanking three polymorphic sites (G/A, A/A, A/A). According to Bui and Liu (2009) non-complementary nucleotide pairing produces destabilization forces, which ultimately impede extension of the DNA strand. For instance, G/A pairing has a maximum strength of destabilization while A/A produces medium strength. In our study, G/A mismatch was produced at the first base of the reverse primer (Figure 5.1 C). Therefore, the three mismatches reduced to a minimum or almost zero the amplification of *PvDREB6B* cDNA in the Andean genotypes, since the primers were designed based on the Mesoamerican BAT 477.

Amplification plots with primers with no prior screening for SNP on target genetic materials might result in many analytical errors, as we have shown in Figure 5.3, comparing *RT-DREB6B-1* and *RT-DREB6B-2*. If the primers from *RT-DREB6B-1* were used for expression analysis of BAT 93 and Jalo EEP558 we could have incurred in the wrong conclusion that no transcripts of *PvDREB6B* were detected in Andean genotypes. Instead, the use of *RT-DREB6B-2*, verified all conditions for RT-qPCR, revealed accurate results, as well as similar expression patterns of expression among genotypes (Figure 5.4). Therefore, we present a summary of the steps considered for the analyses of this work, and that could be applied to other genes and plant species (Table 5.1). Having a pre-defined list of genotypes to be analyzed is helpful in the process since all potential SNP sites interfering with the analyses will be known in advance.

Our work reinforced previous reports that were concerned with the molecular and technical effects of SNP markers on gene expression profiles. We highlighted the importance of a pre-screening of genotypes for nucleotide variant sites discovery before the expression studies. Furthermore, researches must be careful when designing primers and accounting for their specificity. Along with all quality checks that are usually performed, polymorphic sites should be avoided, or a pre-screening of the effects of mismatches should be considered. All these concerns will ultimately reflect in reliable results for the genes of interest.

Table 5.1 – Summary of the main steps for designing RT-qPCR primers not flanking SNP polymorphisms.

Steps	Procedures
1. Gene selection and sequencing	Check for gene exon-intron structure, UTR regions
	Check for the presence of protein conserved domain within the sequence
	Design primers to cover the entire Open Reading Frame or at least partial sequence of interest
2. Plant materials selection	Define a set of genotypes, varieties, cultivars, landraces, species, ecotypes, etc. to be compared
	Sequence the ORF or partial sequence of the genes in all genotypes through direct sequencing approach
	Align sequences and check for SNP sites
3. RT-qPCR primer design	Design primers flanking no SNP or the minimum number of polymorphic sites as possible for the region of interest. Conserved domains, though, should be avoided due to multiple targeting.
	If mismatches are eventually found within the primer flanking area, select the ones with the weakest effect (see BUI; LIU, 2009)
4. Analyses	Perform the qPCR reactions with high quality cDNA samples observing all standards required for accurate results

5.5 Conclusions

In this work, we showed how SNP sites interfere with the analysis of gene expression with the case study of *PvDREB6B*, a gene involved in responses to dehydration and cold stress in common bean. The results evidenced that one primer with three mismatches in the flanking region was not able to produce amplification products, while primer with no SNP in the flaking area produced normal amplification. Our work is a practical example of previous reports concerned with the effects of mismatches in expression analyses and provides suggestions for further experiments aimed at such directions.

References

BLAIR, M. W.; GALEANO, C. H.; TOVAR, E.; TORRES, M. C. M.; CASTRILLÓN, A. V.; BEEBE, S. E.; RAO, I. M. Development of a Mesoamerican intra-gene pool genetic map for quantitative trait loci detection in a drought tolerant × susceptible common bean (*Phaseolus vulgaris* L.) cross. **Molecular Breeding**, Dordrecht, v. 29, n. 1, p. 71–88, 2012.

BITOCCHI, E.; BELLUCCI, E.; GIARDINI, A.; RAU, D.; RODRIGUEZ, M.; BIAGETTI, E.; SANTILOCCHI, R.; SPAGNOLETTI ZEULI, P.; GIOIA, T.; LOGOZZO, G.; ATTENE, G.; NANNI, L.; PAPA, R. Molecular analysis of the parallel domestication of the common bean (*Phaseolus vulgaris*) in Mesoamerica and the Andes. **New Phytologist**, Oxford, v. 197, n. 1, p. 300–13, 2013.

BORGES, A.; TSAI, S. M.; CALDAS, D. G. G. Validation of reference genes for RT-qPCR normalization in common bean during biotic and abiotic stresses. **Plant Cell Reports**, Heidelberg, v. 31, n. 5, p. 827–838, 2012.

BUI, M.; LIU, Z. Simple allele-discriminating PCR for cost-effective and rapid genotyping and mapping. **Plant Methods**, London, v. 5, p. 1, 2009.

BUSTIN, S. A. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): Trends and problems. **Journal of Molecular Endocrinology**, Bristol, v. 29, n. 1, p. 23–39, 2002.

CHACÓN S., M. I.; PICKERSGILL, B.; DEBOUCK, D. G.; ARIAS, J. S. Phylogeographic analysis of the chloroplast DNA variation in wild common bean (*Phaseolus vulgaris* L.) in the Americas. **Plant Systematics and Evolution**, Wien, v. 266, n. 3-4, p. 175–195, 2007.

D'HAENE, B. The importance of quality control during qPCR data analysis. **International Drug Discovery**, Indianapolis, n. 9, p. 18–31, 2010.

DIEFFENBACH, C. W.; LOWE, T. M.; DVEKSLER, G. S. General concepts for PCR primer design. **PCR Methods and Applications**, Cold Spring Harbor, v. 3, n. 3, p. S30–S37, 1993.

FREYRE, R.; SKROCH, P. W.; GEFFROY, V.; ADAM-BLONDON, A. F.; SHIRMOHAMADALI, A.; JOHNSON, W. C.; LLACA, V.; NODARI, R. O.; PEREIRA, P. A.; TSAI, S. M.; TOHME, J.; DRON, M.; NIENHUIS, J.; VALLEJOS, C. E.; GEPTS, P. Towards an integrated linkage map of common bean. 4. Development of a core linkage map and alignment of RFLP maps. **Theoretical and Applied Genetics**, Heidelberg, v. 97, n. 5-6, p. 847–856, 1998.

GUIMARÃES, C. M.; STONE, L. F.; BRUNINI, O. Common bean (*Phaseolus vulgaris* L.) adaptation to drought. Revista Brasileira de Engenharia Agrícola e Ambiental, Campina Grande, v.10, n.1, p.70–75, 2006.

HANAI, L. R.; SANTINI, L.; CAMARGO, L. E. A.; FUNGARO, M. H. P.; GEPTS, P.; TSAI, S. M.; VIEIRA, M. L. C. Extension of the core map of common bean with EST-SSR, RGA, AFLP, and putative functional markers. **Molecular Breeding**, Dordrecht, v. 25, n. 1, p. 25–45, 2010.

KOINANGE, E. M. K.; KOINANGE, E. M. K.; SINGH, S. P.; SINGH, S. P.; GEPTS, P.; GEPTS, P. Genetic control of the domestication syndrome in common-bean. **Crop Science**, Madison, v. 36, p. 1037–1045, 1996.

LEFEVER, S.; PATTYN, F.; HELLEMANS, J.; VANDESOMPELE, J. Single-nucleotide polymorphisms and other mismatches reduce performance of quantitative PCR assays. **Clinical Chemistry**, Chicago, v. 59, n. 10, p. 1470–1480, 2013.

PFAFFL, M. W.; HORGAN, G. W.; DEMPFLE, L. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. **Nucleic Acids Research**, Oxford, v. 30, n. 9, p. e36, 2002.

PONCHEL, F.; TOOMES, C.; BRANSFIELD, K.; LEONG, F. T.; DOUGLAS, S. H.; FIELD, S. L.; BELL, S. M.; COMBARET, V.; PUISIEUX, A.; MIGHELL, A. J.; ROBINSON, P. a; INGLEHEARN, C. F.; ISAACS, J. D.; MARKHAM, A. F. Real-time PCR based on SYBR-Green I fluorescence: an alternative to the TaqMan assay for a relative quantification of gene rearrangements, gene amplifications and micro gene deletions. **BMC Biotechnology**, London, v. 3, p. 18, 2003.

RAMAKERS, C.; RUIJTER, J. M.; LEKANNE DEPREZ, R. H.; MOORMAN, A. F. M. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. **Neuroscience Letters**, Shannon, v. 339, n. 1, p. 62–66, 2003.

RECCHIA, G. H.; CALDAS, D. G. G.; BERALDO, A. L. A.; DA SILVA, M. J.; TSAI, S. M. Transcriptional analysis of drought-induced genes in the roots of a tolerant genotype of the common bean (*Phaseolus vulgaris* L.). **International Journal of Molecular Sciences**, Basel, v. 14, n. 4, p. 7155–7179, 2013.

REMANS, T.; KEUNEN, E.; BEX, G. J.; SMEETS, K.; VANGRONSVELD, J.; CUYPERS, a. Reliable Gene Expression Analysis by Reverse Transcription-Quantitative PCR: Reporting and Minimizing the Uncertainty in Data Accuracy. **The Plant Cell**, Rockville, v. 26, n. 10, p. 3829–3837, 2014.

SCHMUTZ, J.; MCCLEAN, P. E.; MAMIDI, S.; WU, G. A.; CANNON, S. B.; GRIMWOOD, J.; JENKINS, J.; SHU, S.; SONG, Q.; CHAVARRO, C.; TORRES-TORRES, M.; GEFFROY, V.; MOGHADDAM, S. M.; GAO, D.; ABERNATHY, B.; BARRY, K.; BLAIR, M.; BRICK, M. A.; CHOVATIA, M.; GEPTS, P.; GOODSTEIN, D. M.; GONZALES, M.; HELLSTEN, U.; HYTEN, D. L.; JIA, G.; KELLY, J. D.; KUDRNA, D.; LEE, R.; RICHARD, M. M. S.; MIKLAS, P. N.; OSORNO, J. M.; RODRIGUES, J.; THAREAU, V.; URREA, C. A.; WANG, M.; YU, Y.; ZHANG, M.; WING, R. A.; CREGAN, P. B.; ROKHSAR, D. S.; JACKSON, S. A. A reference genome for common bean and genome-wide analysis of dual domestications. **Nature**, London, v. 46, n. 7, p. 707–713, 2014.

TAMURA, K.; STECHER, G.; PETERSON, D.; FILIPSKI, A.; KUMAR, S. MEGA6: Molecular evolutionary genetics analysis version 6.0. **Molecular Biology and Evolution**, Chicago, v. 30, n. 12, p. 2725–2729, 2013.

WALTER, N. A. R.; MCWEENEY, S. K.; PETERS, S. T.; BELKNAP, J. K.; HITZEMANN, R.; BUCK, K. J. SNPs matter: impact on detection of differential expression. **Nature Methods**, London, v. 4, n. 9, p. 679–680, 2007a.

WALTER, N. A. R.; McWEENEY, S. K.; PETERS, S. T.; BELKNAP, J. K. HITZEMANN, R.; BUCK, K. J. Single-nucleotide polymorphism masking. **Nature Methods**, London, v. 4, n. 9, p. 270-217, 2007b.

WU, J. H.; HONG, P. Y.; LIU, W. T. Quantitative effects of position and type of single mismatch on single base primer extension. **Journal of Microbiological Methods**, Amsterdam, v. 77, n. 3, p. 267–275, 2009.

6 ANALYSIS OF NUCLEOTIDE POLYMORPHISMS OF *PvDREB6B* IN WILD MESOAMERICAN COMMON BEAN AND THEIR ASSOCIATION WITH GEOGRAPHIC AND BIOCLIMATIC VARIABLES

Abstract

We investigated the polymorphisms of *PvDREB6B* in a wild common bean collection. *PvDREB6B* has been found cold- and drought-inducible, and previous work has shown the diversity within partial sequence of the locus in wild and domesticated beans of Mesoamerican and Andean origin. In the present study, nucleotide diversity patterns were identified in 112 wild Mesoamerican beans distributed from northern Mexico to Colombia. The complete open reading frame of the gene was sequenced, revealing considerably high nucleotide diversity estimates (33 polymorphic sites, π = 0.00725, θ_W = 0.00652). Neutrality tests suggested no positive selection at the locus, considering the entire population, but specific codons might be under selection as revealed by prediction tools. Population structure analysis revealed seven major clusters based on *PvDREB6B* polymorphisms. In total, 37 haplotypes were detected, and haplotype networks were correspondent to the seven clusters. Furthermore, we showed association of the polymorphisms with geographic and bioclimatic variables of the locations the clusters were distributed. Polymorphisms partially followed a latitudinal gradient, and as a consequence, a gradient of humidity, from drier places in northern Mexico to humid locations in Central America and Colombia. Moreover, one cluster exhibited association with the altitude and the lowest average temperatures. Although no consistent evidence of selection was found at this locus, other genomic regions need to be searched so signatures of local adaptations could be effectively proved. Overall, our data provided possible signatures of cold and drought tolerance associated with the population structure determined by *PvDREB6B*. Such information might be useful for breeding programs aimed at selection for stress-tolerance in common bean.

Keywords: Adaptation. Abiotic stresses. Stress-tolerance. Low-temperature. *DREB* genes.

Resumo

Polimorfismos do gene *PvDREB6B* foram analisados em uma coleção de genótipos selvagens de feijoeiro. *PvDREB6B* é predominantemente induzido por baixa temperatura e déficit hídrico e trabalho prévio mostrou polimorfismos de nucleotídeos ao longo de parte da sequência do gene em genótipos selvagens e domesticados de origem Mesoamericana e Andina. No presente estudo, foram identificados padrões de diversidade nucleotídica em 112 acessos selvagens de origem Mesoamericana distribuídos desde o norte do México até a Colômbia. A região codante completa do gene foi sequenciada e elevados valores de diversidade nucleotídica foram encontrados (33 sítios polimórficos, π = 0.00725, θ_W = 0.00652). Testes de neutralidade não detectaram evidência de seleção sobre o gene, no entanto códons específicos podem estar sob seleção. Por sua vez, as análises de estrutura genética mostraram sete grupos principais a partir de polimorfismos de nucleotídeos do gene. No total, 37 haplótipos foram detectados e a análise de correlações entre haplótipos correspondeu à definição dos sete subgrupos. Além disso, associação parcial entre polimorfismos de nucleotídeos e variáveis geográficas e bioclimáticas foi detectada. Os polimorfismos seguiram parcialmente um gradiente latitudinal e, consequentemente, um gradiente de umidade, de climas mais secos do norte do México a climas úmidos na Colômbia. Adicionalmente, um dos subgrupos foi bem definido pela altitude e pelas mais baixas temperaturas médias do ano. Embora não tenham sido encontradas evidências consistentes de seleção sobre o gene, outras regiões genômicas devem ser exploradas para identificar possíveis evidências de adaptações locais. Estes resultados mostram prováveis fontes de tolerância ao frio e déficit hídrico associados com o gene *PvDREB6B*. Estas informações podem ser úteis para programas de melhoramento destinados a seleção para tolerância a estresses abióticos no feijoeiro.

Palavras-chave: Adaptação. Estresses abióticos. Tolerância a estresses. Baixa temperatura. Genes *DREB*.

6.1 Introduction

Wild common bean (*Phaseolus vulgaris* L.) populations are currently distributed from subtropical areas of northern Mexico to northeastern Argentina (CHACÓN et al., 2007). Plants are annual, viney and germinate in the midst of small trees and shrubs in forest clearings or disturbed environments. The growth cycle usually lasts between 8 and 10 months (CORTÉS et al., 2012). Wild beans exhibit high morphologic variation, which can be observed in flowers, seeds, leaves, rooting patterns, among several other characteristics (SINGH; DEBOUCK; GEPTS, 1991). High genetic variability has been detected in wild bean populations through the analysis of allozyme diversity (KOENIG; GEPTS, 1989), the protein phaseolin variability (GEPTS et al., 1986) and DNA-based markers (e.g. MACIEL et al., 2003; GONZÁLEZ et al., 2005; BLAIR et al., 2006; BENCHIMOL et al., 2007; BURLE et al., 2010; CORTÉS; CHAVARRO; BLAIR, 2011). It is a consensus that the diversity of wild common bean is structured in two major gene pools, the Mesoamerican and the Andean (KOENING; GEPTS, 1989; BENCHIMOL et al. 2007) with subdivisions within each group (BLAIR; SOLER; CORTÉS, 2012).

Mesoamerican beans are distributed from northern Mexico to Colombia in a diverse range of environments. In these areas, wild beans are subjected to diverse abiotic stresses. In tropical bimodal regions, wild beans are commonly affected by mid-term drought episodes. Conversely, on sub-tropical unimodal rainfall regions wilds may be growing for longer periods of water deficit (CORTÉS et al., 2013). Additionally, wild common beans span an altitude and temperature gradient along their distribution area. In the highlands of Mexico and Central America, the average minimum temperatures are near 0°C at some locations during the coldest period. Meanwhile, other areas have minimum temperatures averaging around 20°C and maximum over 30°C over the year (WorldClim database).

In general, changes in air and soil temperature, soil salinity and water supply challenge the development of plants such as common bean, limiting their growth and productivity. Such effects are caused by changes is morphology and physiology by inhibition of cell division, reduction of photosynthesis, among other consequences (SAIBO; LOURENÇO; OLIVEIRA, 2009). As wild common bean populations are distributed through a wide range conditions, evolutionary forces may have led beans to adapt to climate, soil and vegetation of these areas, developing mechanisms of tolerance. At the molecular level, it is well known that tolerance implicates in several regulatory processes that trigger the response to environmental factors

in plants. In involves signaling and gene regulation mechanisms trying to cope with the adversities of the environment (LATA; PRASAD, 2011; DUQUE; ALMEIDA; SILVA, 2013).

One group of genes that has been frequently studied in analyses of the effects of abiotic stresses is the *DREB* (*DEHYDRATION RESPONSIVE ELEMENT-BINDING*) subfamily. In common bean, it comprises 54 genes, according to the prior *in silico* categorization (Section 3). The entire set is divided into six subgroups (A-1 to A-6), based on phylogenetic relationships and comparisons with other species such as *Arabidopsis thaliana* (SAKUMA et al., 2002; NAKANO et al., 2006; MIZOI; SHINOZAKI; YAMAGUCHI-SHINOZAKI, 2012) and *Glycine max* (ZHANG et al., 2008). Although limited data are available for their expression profiles in common bean (Section 3), some general features have been reported across plant species. Usually *DREB1* (A-1) genes are up-regulated by low temperature, although osmotic stresses have been reported as well (LIU et al., 1998; SAKUMA et al., 2002; AKHTAR et al., 2012). *DREB2* (A-2) are predominantly induced by dehydration and osmotic stresses (LIU et al., 1998; NAKASHIMA; YAMAGUCHI-SHINOZAKI, 2006; LATA; PRASAD, 2011; MIZOI; SHINOZAKI; YAMAGUCHI-SHINOZAKI, 2012). The only A-3 member, *ABI4*, is involved in developmental processes and signaling for mechanisms other than abiotic stresses (SHKOLNIK-INBAR; BAR-ZVI, 2011). In general, A-4 genes show low stress-inducibility. A-5 and A-6 genes are commonly stress-inducible and have been found interacting with other pathways (MIZOI; SHINOZAKI; YAMAGUCHI-SHINOZAKI, 2012).

Expression profiles of common bean *DREB* member are limited to four genes (*PvDREB1*, *PvDREB2A*, *PvDREB5A* and *PvDREB6B*) (Section 3). Their expression profiles were determined under dehydration, salinity and low temperature. Furthermore, nucleotide polymorphisms were investigated within each of the transcripts and *PvDREB6B* showed the highest number of polymorphic sites along the open reading frame of domesticated genotypes evaluated (Section 4). *PvDREB6B* turns out to be equivalent to the *DREB2A* studied by (CORTÉS et al., 2012). However, our recent phylogenetic analyses supported the classification of this gene as an A-6 member (Section 3). Thereby, hereafter this gene is referred as *PvDREB6B* and it was found responsive to dehydration and low temperature, although depending on the duration of stress, it was also induced by salinity (Sections 3 and 5). Cortés et al. (2012) were able to sequence a fragment of only 547 bp from this gene, but with the release of the common bean genome (SCHMUTZ et al., 2014), its ORF was found spanning 957 bp in the reference Andean genotype G19833. In their work, Cortés et al. (2012)

show a considerable number of polymorphisms in wild and domesticated accessions with the 547 bp fragments obtained. They have also shown the first attempt in associating single nucleotide polymorphisms from this gene with drought tolerance levels in the populations examined, but no clear patterns were obtained.

In a different approach than Cortés et al. (2012), we addressed the nucleotide polymorphisms within the entire ORF of *PvDREB6B* in a larger collection of wild common bean accessions representing only the Mesoamerican gene pool. Our primary objective was to find single nucleotide polymorphisms in accessions distributed from northern Mexico to Colombia and how such nucleotide variants are associated with environmental variables. Temperature, precipitation, altitude, location parameters (latitude and longitude) and a series of derived bioclimatic indicators were used to verify possible association between environmental data and nucleotide variants structuring.

6.2 Material and Methods

6.2.1 Structure of *PvDREB6B* gene, homology with other genes and primer design

PvDREB6B was isolated according to descriptions in Section 3. We searched orthologs from other species such as *A. thaliana*, *G. max* and *M. truncatula* to reveal the degree of homology with common bean and the size and protein motifs shared among all orthologs, based on MEME suite prediction tool (http://meme-suite.org/tools/meme). A neighbor-joining tree was elaborated using MEGA 6.0 (TAMURA et al., 2013) with a set of *DREB* genes from common bean and other species, to verify relationships with *PvDREB6B*.

We investigated the nucleotide variability within the Open Reading Frame of *PvDREB6B,* which spans 957 bp from the start to the stop codon. It is located on chromosome 2, from positions 42080727 to 42082576 (forward direction) (Phytozome ID: Phvul.002G254500) (Section 3). The transcription initiation site starts with a 5' UTR region of 386 bp followed by the ORF and a 3'UTR region with 507 bp. In total, the gene comprises 1,850 bp (Section 4) and the ORF is translated into a protein with 318 amino acids.

A pair of primers was designed to amplify the entire coding region (ORF) of *PvDREB6B*. However, to obtaining high quality sequences covering the entire region through direct sequencing, we designed primers covering around 100 bp upstream and downstream from the ORF (F: *5'-CTCCTTCTCCTCTCTTTCCTCA-3'*; R: *5'-GCCAAAGACATCCAAACATAAC-3'*). This pair of primers was used to amplify all common bean accessions from this study.

6.2.2 Plant materials

DNA was extracted from 121 Mesoamerican wild beans accessions of a collection obtained from the Plant Introduction Station of the United States Department of Agriculture (USDA) at Pullman, WA, and from Centro Internacional de Agricultura Tropical (CIAT) at Cali, Colombia. These materials originated from natural populations of common bean spanning the Mesoamerican distribution area of the species, from northern Mexico to Colombia. Therefore, they are distributed from very dry to very humid climates, as well as different altitude gradients. Supplementary Table 6.1 presents passport information about each accession evaluated (country, latitude, longitude, altitude, precipitation). A breeding line (G40111) was used as a representative from *P. acutifolius* to be an outlier in the evolutionary analysis.

All plants were greenhouse-grown in pots filled with soil. Leaf tissues were collected when plants had their first trifoliate leaf completely expanded. Leaves were lyophilized in a VirTis Sentry 2.0 equipment for approximately 48 hours and stored at room temperature until their processing for DNA extractions.

6.2.3 DNA extraction and quantification

DNA was isolated with an adapted ammonium acetate precipitation strategy (PALOTTA et al., 2003). Approximately 20 mg of lyophilized and ground leaf tissue were placed within Eppendorf tubes containing 700 µl of extraction buffer (0,01 M Tris-HCl pH 7.5, 0.05 M EDTA pH 8.0, 1.25% SDS). Samples were vigorously shaken and incubated at 65ºC for 30 min. Afterwards, tubes were left at room temperature for 15 min and aid by 300 µl of ammonium acetate 6 M. Then, all samples were placed in the fridge for 15 min. Subsequently, tubes were centrifuged at 16,000 g for 7 min and the supernatant was transferred to other tube containing 360 µL of cold isopropanol. Tubes were centrifuged once more. Pellets were washed with ethanol 95% and air-dried on the bench. All samples were solubilized with TE + RNAse 10 mg ml^{-1}, mixed and incubated at 37ºC for 30 min. DNA was precipitated with 500 µL ethanol 95%, samples were centrifuged and all supernatant was discarded. Pellets were air-dried and dissolved with sterilized water overnight. Quantification was performed with Nanodrop spectrophotometer, and quality evaluation was based on absorbance reads (ratio 260/280) and agarose 1% gel stained with ethidium bromide.

6.2.3 Amplification of *PvDREB6B* ORF, purification and sequencing

DNA from all genotypes was diluted to 20 ng μl^{-1}. Each PCR reaction was prepared to final volume of 50 µl for further purification of amplification products. Reactions contained 1 X reaction buffer (20 mM Tris-HCl pH 8.4, 50 mM KCl), 3 mM $MgCl_2$, 0.2 mM dNTP, 0.2 µM of each primer and 1 U of High Fidelity Taq DNA Polymerase (Invitrogen). Amplification conditions were set as follows: 94°C for 2 min for initial denaturation, 38 cycles of denaturation (94°C for 30 s), primer annealing (59°C for 30 s) and extension (68°C for 1 min), and finally, 7 min for final extension at 68°C. All reactions were verified in agarose gel 1.2% for single bands matching the expected size.

Amplification products were purified using the Wizard SV Gel and PCR Clean-Up System (Promega), following the instructions of the manufacturer. Purified samples were quantified and submitted for sequencing. Sequencing was performed in both directions (forward and reverse) using the same pair of primers from the initial amplifications. It was conducted at the DNA core facility in the Division of Biological Sciences at the University of California in Davis, with the ABI PRISM 3730 Capillary Electrophoresis Genetic Analyzer using ABI BiDye Terminator v. 3.1. Cycle Sequencing kit.

6.2.4 Sequence alignments and phylogenetic analyses

All sequences were submitted to quality analysis with DNA Baser version 4.20.0.36 (Heracle BioSoft). Contig assembly was performed with forward and reverse sequences. Only sequences with quality value higher than 20 were trusted for the next steps. High-quality contigs were further aligned with BioEdit Sequence Alignment Editor (HALL, 1999) with ClustalW multiple alignment. To verify the accuracy of the alignment, we also performed the same analysis with CLC Sequence Viewer version 7.6 (QIAGEN Aarhus A/S). From the 121 wild accessions, we were able to obtain high-quality sequences for 112 accessions.

Alignment files were used to perform phylogenetic analyses with Mega 6 (TAMURA et al., 2013) with the 112 sequences. Evolutionary history was inferred using the Neighbor-Joining algorithm (SAITOU; NEI, 1987). Trees were designed to scale, with branch lengths in the same units as those of the evolutionary distances used for phylogenetic tree inference. The evolutionary distances were computed according to Kimura-2-parameter approach (KIMURA, 1980). All positions containing gaps and missing data were eliminated. NETWORK (version 4.6.1.3, BANDELT et al., 1999) was used for reconstructing a median-joining (MJ) network.

6.2.5 Nucleotide diversity estimates and neutrality tests

Nucleotide diversity estimates were calculated for the 112 wild accessions (Supplementary Table 6.1) using DnaSP version 5.10 (LIBRADO; ROZAS, 2009). The number of polymorphic sites (*S*), singleton variable sites, haplotypes and the haplotype diversity (*Hd*, NEI, 1987) were computed. The nucleotide diversity was determined according to two parameters. First, the expected heterozygosity per nucleotide site (π, NEI, 1987). Second, the θ, Watterson's theta estimator (WATTERSON, 1975). Another parameter was the number of recombination events, estimated using the four-gamete test (HUDSON; KAPLAN, 1985).

Selection on *PvDREB6B* was tested with multiple approaches, computed with DnaSP. Tajima's *D* (TAJIMA, 1989), Fu and Li's *D** and *F** (FU; LI, 1993) analyses were performed to test for excess and deficiency of mutations at segregating nucleotide sites, all polymorphic. Tajimas's *D* is calculated as the difference between two measures of genetic diversity (the mean number of pairwise differences and the number of segregating sites). The general purpose of this estimate is to identify sequences that do not fit in the neutral theory model at equilibrium between mutation and genetic drift. Therefore, *D* values equal to zero represent no evidence of selection. $D < 0$ indicates the presence of rare alleles at low frequencies, that might be related to recent selective sweep, population expansion after a recent bottleneck or linkage to a swept gene. $D > 0$ designates the presence of multiple alleles that might be explained by balancing selection or sudden population contraction. Fu and Li's *D** compares the number of mutations in internal and external branches with their expectations under selective neutrality (FU; LI, 1993). Another estimate obtained through DnaSP was Fu's F_S (FU, 1997). Negative values of F_S evidence excessive number of alleles, as would be expected from recent population expansion or genetic hitchhiking. Positive values are an evidence for deficiency of alleles, as would be expected from a recent bottleneck or overdominant selection. Finally, the Fay and Wu *H* test was performed to measure departures from neutrality that are reflected in the difference between high-frequency and intermediate-frequency alleles (FAY; WU, 2000). The significance of the *H* test was calculated 1000 coalescent simulations, with no recombination and threshold of $P = 0{,}05$ (HUDSON, 2000).

6.2.6 Predictive analysis of non-synonymous substitutions on protein function

The number of synonymous and non-synonymous substitutions were estimated on DnaSP. The effective amino acid changes were checked on translated sequences with

CLC Sequence Viewer 7.0. To test whether amino acid replacements among accessions have neutral or selective effect on *PvDREB6B*, the PROVEAN (Protein Variation Effect Analyzer) (http://provean.jcvi.org) algorithm was used. PROVEAN provides a generalized approach to predict functional effects of changes in protein sequences, enabling comparisons for single or multiple amino acid substitutions, and in-frame insertions and deletions (CHOI et al., 2015). The cutoff score used by PROVEAN is -2.5. Values lower than that indicate possible deleterious effects of amino acid changes, while values higher than that suggest no change (neutral effect).

6.2.7 Analysis of population structure

The population structure of the 112 wild beans was evaluated using Bayesian Analysis of Population Structure (BAPs) version 5.3. It uses the Bayesian clustering analysis method described by Corander et al. (2003) to identify hidden population structure. Individuals are clustered into K genetically distinguishable groups based on nucleotide frequencies (CORANDER et al., 2003; CORANDER; MARTTINEN, 2006; CORANDER; TANG, 2007). First, a genetic mixture analysis was carried out to determine the most probable number of populations (K) for all wild beans. Simulations of K values from 1 to 10 were performed. Further analysis of admixture determined the individual admixture proportions respectively to the most probable number of K clusters determined. In total, 100 iterations were performed to estimate the admixture coefficients. All runs were repeated five times to ensure consistent results. The clustering with BAPs is more reliable for single-gene analyses than other packages for structure analysis since it provides specific algorithms for clustering with linked loci.

To verify the dispersion of the haplotypes obtained for *PvDREB6B*, we performed a principal coordinate analysis (PCoA) with all the polymorphic sites data, using GeneAlEx 6.5 (PEAKALL; SMOUSE, 2012). PCoA graphics were obtained based on the genetic distances among all accessions. All haplotypes groups shown in PCoA and BAPs structure analysis were plotted with color-markers on a raster map indicating the precise location of each accession using Gengis 2.4.0.

6.2.8 Association with geographic and bioclimatic parameters

We attempted to relate polymorphism data obtained for *PvDREB6B* and a series of geographic and environmental variables regarding the specific location of each one of the wild accessions. Genesys (Gateway to Genetic Resources, https://www.genesys-pgr.org)

database was accessed to collect passport information for all the 112 wilds evaluated, including geographic location (latitude, longitude and altitude) parameters. Furthermore, a series of bioclimatic variables were retrieved from WorldClim (Global Climate Data) database (http://www.worldclim.org/) and filtered by the locations of each accession with Raster package on R (HIJMANS et al., 2015). Variables include temperature ranges over the year (average maximum and minimum for each month) and precipitation of each location (averages for all months of the year). Data represent averages for a 50-year-period of observations. Cortés et al. (2013) used the BioClim variables to define potential drought-tolerant accessions based on a modeling study. Their findings indicated wild accessions were distributed among different precipitation regimes following a latitude gradient, and ecological features of each collection site were associated with sub-populations. Thereby, we considered the seven groups identified through BAPs structure analysis as different subpopulations. Average values of each one of the geographic (latitude, longitude and altitude) and environmental variables (19 BioClim variables, Table 6.1) were calculated for each subpopulation and compared with ANOVA and F-test. Afterwards, a canonical discriminant analysis was performed with all variables, setting each group as the categories for discrimination. All analyses were performed using JMP® 12.0.1 (SAS Institute Inc.).

Table 6.1 – Bioclimatic variables used to perform association analysis with *PvDREB6B* haplotypes.

BioClim variables
BIO1 = Annual Mean Temperature
BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))
BIO3 = Isothermality (BIO2/BIO7) (* 100)
BIO4 = Temperature Seasonality (standard deviation *100)
BIO5 = Max Temperature of Warmest Month
BIO6 = Min Temperature of Coldest Month
BIO7 = Temperature Annual Range (BIO5-BIO6)
BIO8 = Mean Temperature of Wettest Quarter
BIO9 = Mean Temperature of Driest Quarter
BIO10 = Mean Temperature of Warmest Quarter
BIO11 = Mean Temperature of Coldest Quarter
BIO12 = Annual Precipitation
BIO13 = Precipitation of Wettest Month
BIO14 = Precipitation of Driest Month
BIO15 = Precipitation Seasonality (Coefficient of Variation)
BIO16 = Precipitation of Wettest Quarter
BIO17 = Precipitation of Driest Quarter
BIO18 = Precipitation of Warmest Quarter
BIO19 = Precipitation of Coldest Quarter

6.3 Results

6.3.1 *PvDREB6B* conservation across different species

PvDREB6B has been previously defined in a study dedicated to its expression profiling when it was found mainly involved in responses to dehydration and low temperature (Section 3). Nucleotide polymorphisms have been found in the open reading frame of a limited sample of common bean genotypes representing the Mesoamerican and Andean gene pools. *PvDREB6B* revealed considerable polymorphism within the ORF of these genotypes (Section 4). Hereafter, we examined phylogenetic relationships among related proteins from other species with the one encoded by *PvDREB6B* (Figure 6.1), with 318 aa. BLASTp search revealed high similarity with a hypothetical protein from *Vigna angularis* (gi920686711, KOM30694.1, 79% identity, E-value = 3 x 10^{-170}) as well as with an ethylene-responsive transcription factor (ERF060-like, gi|920686711, KOM30694.1, 76% identity, E-value = 6 x 10^{-140}) from *G. max* (Figure 6.2). It also had considerable similarity with a predicted ethylene responsive factor from *M. truncatula* (gi|922337685, |XP_013446832.1, 52% identity, E-value = 4 x 10^{-87}) and an AP2-containing protein from *Oryza sativa* (gi|115451269, NP_001049235.1, 40% identity, E-value = 2 x 10^{-52}) (Figure 6.2). It is also similar to *AtRAP2.4* from *A. thaliana* (LIN; PARK; WANG, 2008) and *GmDREBb* from soybean (LI et al., 2005), both from A-6 subgroup in the DREB subfamily. *PvDREB6B* presents 100% identity with *DREB2A* from Nayak et al. (2009), even with partial coverage. The gene has been referred as *PvDREB6B* since it clearly fits among representatives from A-6 subgroup (Sections 3, 4 and 5).

Although sharing common amino acids, protein sizes were variable among *PvDREB6B*-related sequences. ERF060-like from soybean presented 312 aa while *PvDREB6B* comprises 318 aa from the start to the stop codon. The sequence from *V. angularis* presented 334 aa, while *M. truncatula* sequence comprised only 286 aa and *O. sativa* 297 aa. All A-6 proteins share some common motifs along the amino acid chain, as presented in Figure 6.3 with MEME suite predictions. It becomes evident that close related species share 10 similar motifs (*P. vulgaris*, *G. max* and *V. angularis*). *M. truncatula* proteins shares seven motifs with common bean. The more distant species share less motifs with *PvDREB6B* and, as expected, the distant monocot *O. sativa* shares the least number of motifs. Motif 1 represents the AP2 domain, which is present in all sequences as a premise for being included in the ERF and DREB subfamilies.

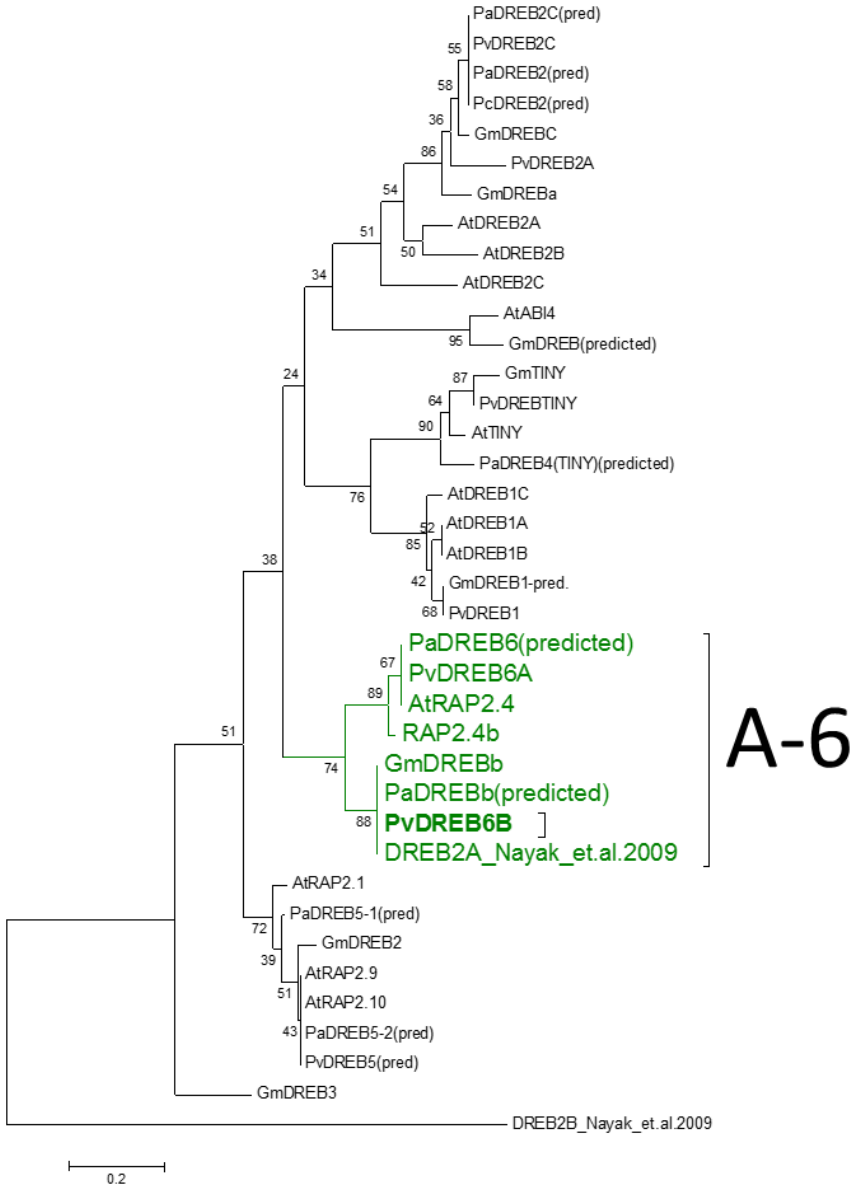


Figure 6.1 – Neighbor-joining tree with the main representative DREB sequences from each subgroup (A-1 to A-6). The common bean PvDREB6B is highlighted within group A-6 along with members from *Arabidopsis thaliana*, *Phaseolus acutifolius*,*P. coccineus* and *Glycine max*. Sequences were retrieved from NCBI/GenBank: AtDREB1B (AT4G25490); AtDREB1C (AT4G25470); AtDREB2A (AT5GO5410); AtDREB2B (AT3G11020); AtDREB2C (AT2G40340); AtABI4 (AT2G40220); AtTINY (AT5g11590); AtRAP2.1 (AT1G46788); AtRAP2.10 (AT2G36900); AtRAP2.4 (AT1G22190). *G. max* sequences: similar to DREB1 (EH260462.1); GmDREBa (AY542886.1); GmDREBc (AY244760.1); similar to ABI4 (DY577304.1); GmTINY (FJ895345.1); similar to RAP2.1 (Glyma01g39540); similar to RAP2.10 (Glyma06g04490); GmDREBb (AY296651.1); GmDREB1 (AF514908.1); GmDREB2 (DQ054363.1); GmDREB3 (NM_001250024.1); GmDREBb (AY296651.1). *P. acutifolius*: similar to DREB2A - PaDREB2 (CA902352.1); similar to DREB2C - PaDREB2C (HO784764.1); similar to TINY – PaTINY (HO808951); similar to RAP2.4 - PaDREB6(1) (HO779787.1); similar to RAP2.4 - PaDREB6(2) (HO794256.1). *P. coccineus*: similar to DREB2C (CA902352.1). The *P. vulgaris* sequences comprise proteins from previously defined genes: PvDREB1F, PvDREB2A, PvDREB5A and PvDREB6B, and also sequences similar to PvDREB2C, PvTINY (Phvulv091006839m.g) and PvDREB6A

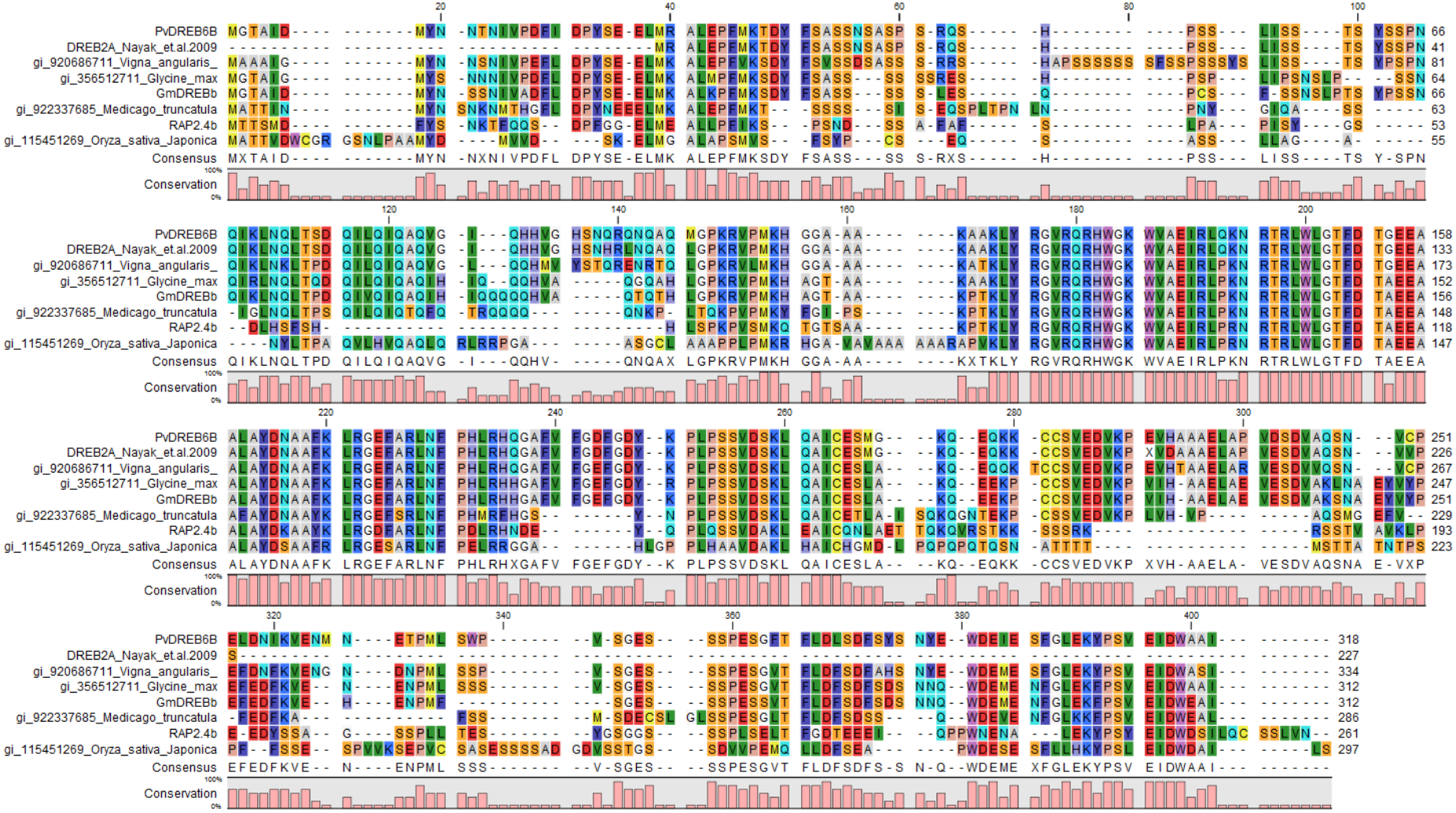


Figure 6.2 – Full protein alignments of PvDREB6B and related proteins from *Arabidopsis thaliana*, *Oryza sativa*, *Medicago truncatula*, *Vigna angularis* and *Glycine max*.

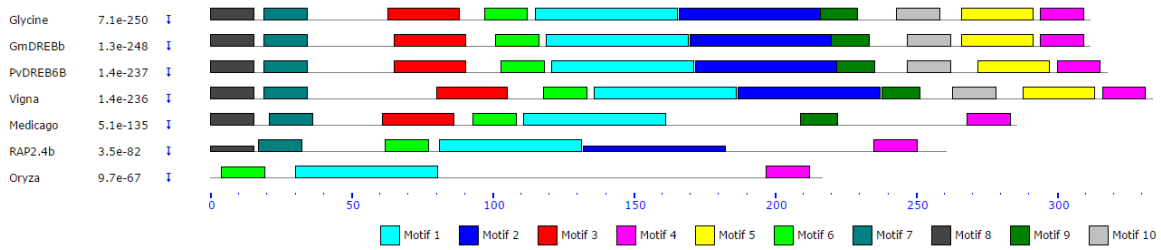


Figure 6.3 – Protein conserved motifs along the sequence of PvDREB6B and related proteins from *Arabidopsis thaliana*, *Oryza sativa*, *Medicago truncatula*, *Vigna angularis* and *Glycine max*.

6.3.2 *PvDREB6B* phylogeny in wild Mesoamerican accessions of common bean

In common bean, we addressed the nucleotide variability of *PvDREB6B* within 112 wild accessions of Mesoamerican origin distributed from northern Mexico to Colombia. The neighbor-joining analysis with 1,000 bootstrap samplings on MEGA 6.0 generated a tree in which most sequences were grouped with weak bootstrap values (≤ 70%) (Figure 6.4). We defined seven main clades based on the analyses. Clade 1 (58% bootstrap value) comprised 12 wild accessions and the reference G19833, which is Andean. G19833 was only included as the reference for *PvDREB6B* in Phytozome database. Although being included in this group, previous analysis of *PvDREB6B* polymorphisms supported its classification in the Andean gene pool (Section 4). Clade 2 (66% bootstrap value) presented 12 wilds as well. These wilds are distributed in a small area at the border between the states of Jalisco and the western side of Guanajuato in Mexico, with one exception, the accession G24576, located further south (Lat. 16.8000, Long. -96.5167). Some accessions of this area were referred as 'truly' wild common bean forms with an 'S' type phaseolin type (GEPTS et al., 1986; GEPTS, 1988). Clade 3 comprised 14 wilds with a high value of bootstrap (90%). Clade 4 was only composed by four wild accessions and was defined by a weak bootstrap value (52%). The highest bootstrap value was found for the 14 accessions from Clade 5 (97%). From this clade, seven accessions belonged to northern regions of Mexico in Durango and Chihuahua. The other seven were further south in Oaxaca. Finally, a bootstrap value of 62% defined other 57 genotypes (clusters C-6 and C-7). However, further analyses of structure with BAPs 5.3 software revealed two distinct subgroups within this cluster. Although represented by low bootstrap values, they are also distinguished by the phylogenetic tree (C-6 and C-7). C-6 included most accessions from Central America, distributed through Costa Rica, Honduras, Guatemala and Colombia. Most C-7 accessions were from Mexico, although a few representatives from Central America were grouped together (Figure 6.4).

The analysis of structure was carried out with BAPs 5.3 software. Structure calculations were performed without using information from the location of the wild accessions. According to the analyses performed for linked loci, seven distinct genetically subgroups were assigned as the most probable structure (K = 7) (Figure 6.5). Therefore, BAPs structure analysis reinforced the subdivision of the neighbor-joining tree in seven clades.

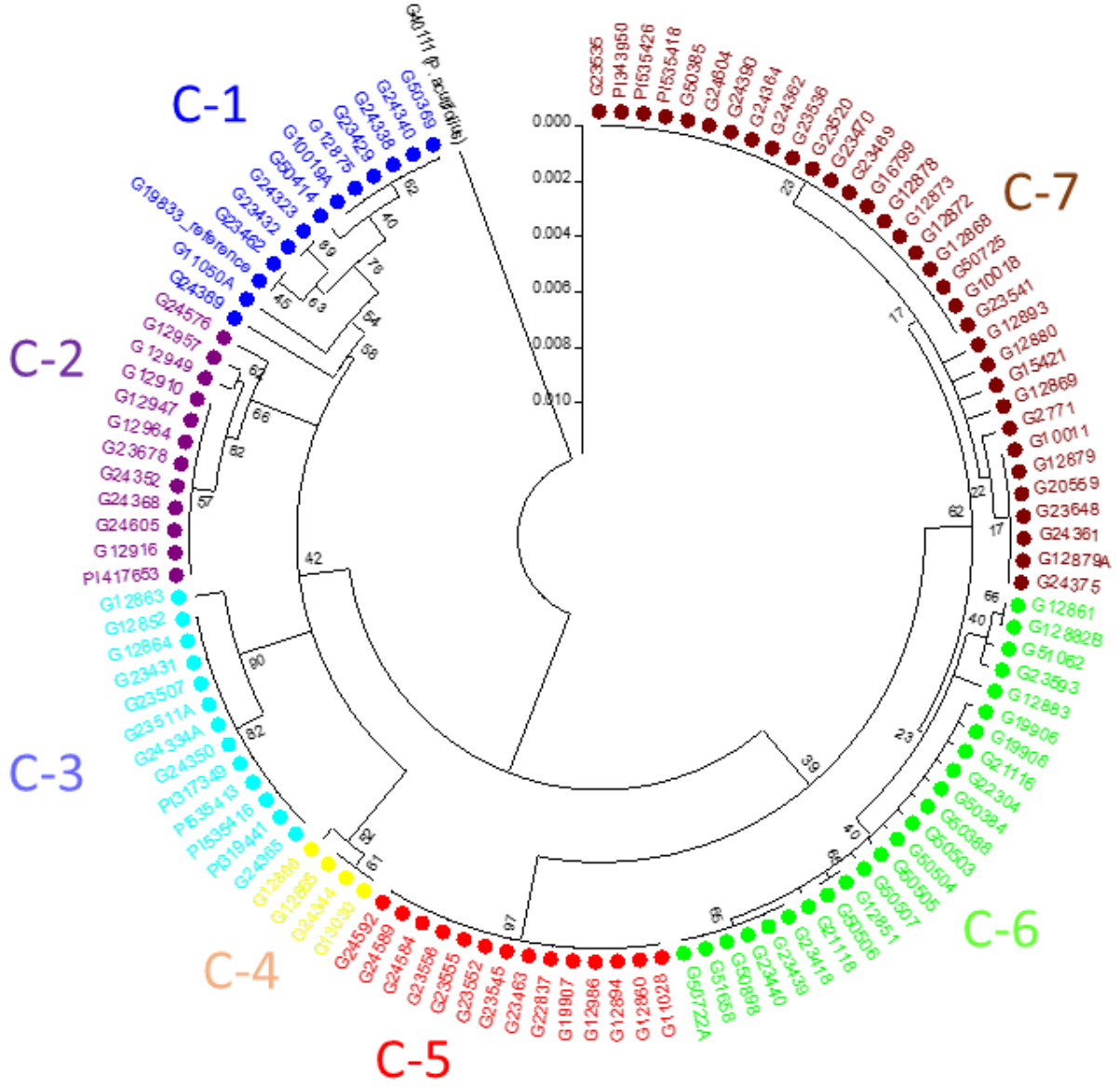


Figure 6.4 - Unrooted neighbor-joining tree with bootstrap analysis (1,000 replicates) of Kimura two-parameter distances calculated from *PvDREB6B* sequences of 112 wild *Phaseolus vulgaris* accessions from Mesoamerican genepool and one accession from *P. acutifolius* (G40111). Seven main clusters are indicated by different colors.

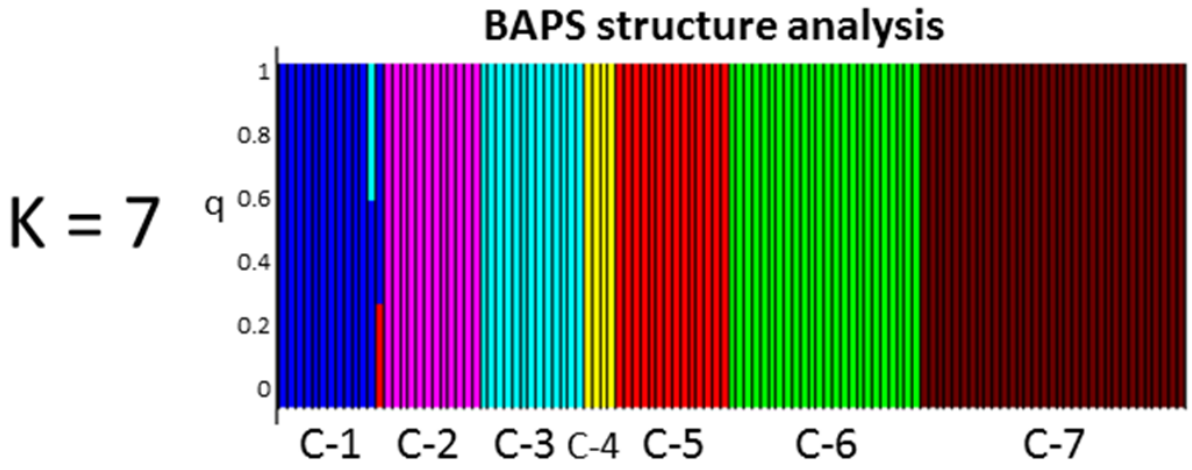


Figure 6.5 – Population structure of 112 wild common bean accessions as estimated with BAPs version 5.3. Each accession is represented by a vertical histogram divided into seven K groups, which corresponded to the seven clusters defined through the NJ tree. Colors represent the same clusters from the NJ tree.

Further evidence for the population structure revealed by NJ and BAPs was obtained with a principal coordinate analysis (PCoA) based on the genetic distances among all accessions. Sequence data were placed in GenAlEx software; genetic distances were calculated and principal coordinates were graphically represented. A similar pattern of structure was observed by PCoA (Figure 6.6 A). However, it suggested samples from C-6 and C-7 belonged to the same group. All other clusters (C-1 to C-5) were positioned apart from each other in the PCoA graph, indicating particular nucleotide patterns within each subgroup. Even though, we kept the seven clusters subdivision based on further analyses that suggested particular environmental differences between C-6 and C-7. The seven clusters were also represented in a raster map, showing their geographic location from northern Mexico to Colombia, using GenGIS software. Figure 6.6 B shows the positions of each accession marked by dots colored the same way that for NJ, BAPs and PCoA analyses. Although clusters are not entirely separated according their location, cluster C-2 samples (purple dots) were mostly concentrated in a narrow area between the states of Jalisco and Guanajuato in Mexico (highlighted in Figure 6.6 B). C-4 accessions were all located in Jalisco state as well. Moreover, C-6 samples were mostly distributed from Central America to Colombia. C-7 accessions were intermittently distributed between C-5 and C-1 samples.

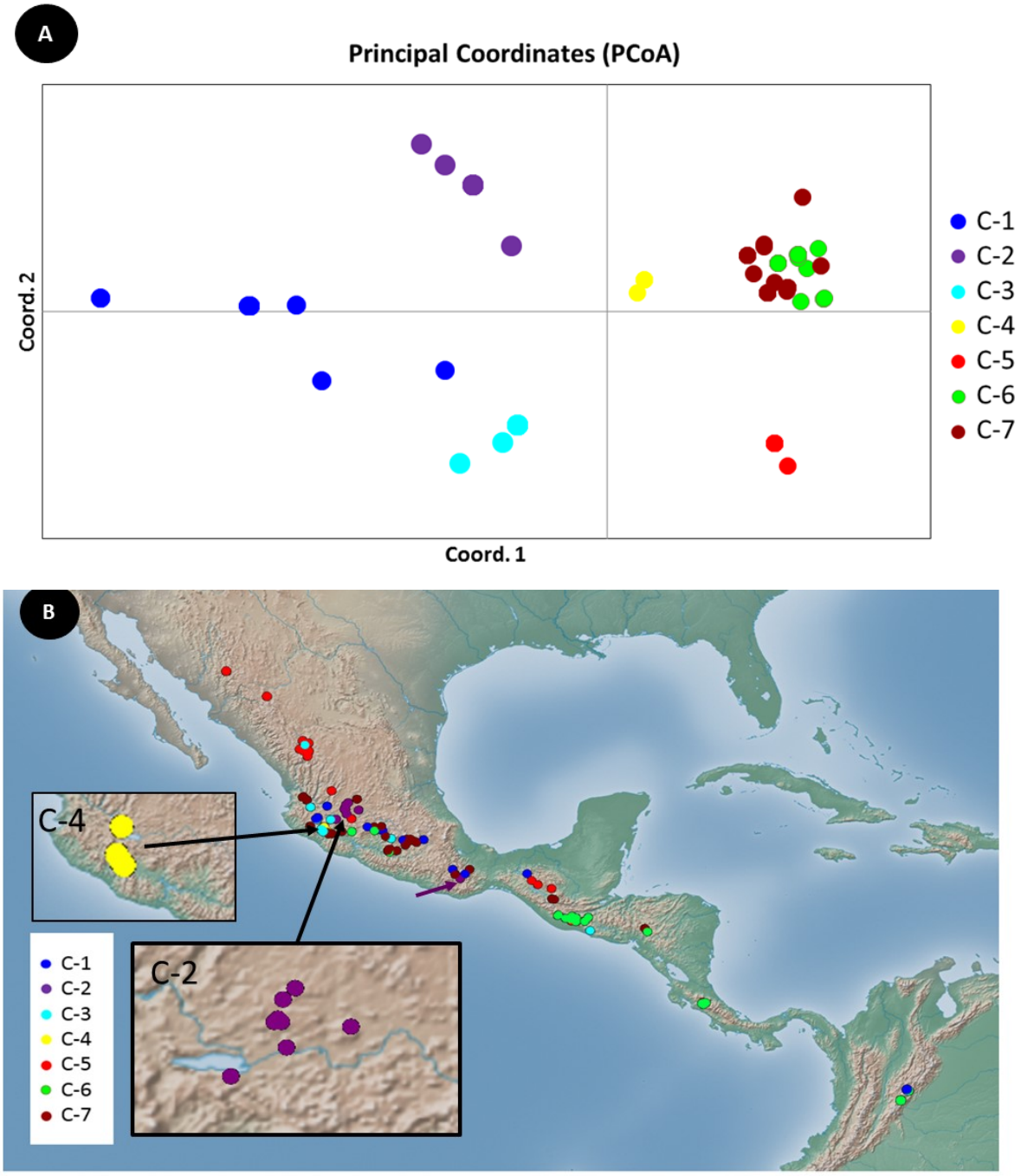


Figure 6.6 – (A) Principal coordinate analysis based on genetic distances among wild common bean accessions based on nucleotide polymorphisms within the open reading frame of *PvDREB6B*, as performed with GenAlEx. Clusters C-1 to C-7 are the same of previous NJ and BAPs analyses. (B) Geographic representation of all accessions colored by their respective clusters, as performed with GenGIS. The purple arrow indicates one line that shared haplotypes with C-2.

6.3.3 Nucleotide diversity estimates, haplotype networks and neutrality tests

The software DnaSP was used to compute nucleotide diversity estimates of the 112 sequences obtained from sequencing the ORF of *PvDREB6B*. Estimates were obtained for the whole set of genotypes. In total, 33 SNP sites were found in the 112 wild beans, resulting in an average frequency of one SNP in every 29 bp (Table 6.2). In total, 35 mutations were detected, implying two SNP positions with more than two alleles. SNP at position +52 was represented by three alleles (A/T/C) as was SNP 923 at position +923 (A/T/C) from the beginning of the ORF. The number of singleton sites accounted for 12% of polymorphic sites encountered in the wilds (Table 6.2). When the *P. acutifolius* line was included in the analysis, 41 nucleotide variants were detected for the whole set (112 wilds + G40111). An INDEL of 9 bp was detected between G40111 and the wild common beans, but it was not computed in the number of polymorphic sites. The short sequence (C/T)ACGTCAAT (+143 to +151) was not present within G40111.

The number of haplotypes detected by DnaSP analysis was of 37 (numbered from H-1 to H-37, Supplementary Table 6.2), considering all 112 wild accessions. Six genotypes from C-7 cluster shared haplotype H-1, and 15 genotypes from the same subgroup shared H-2. H-3 was found within six sequences from C-1 cluster. H-4 was only found within C-5 cluster. H-6 was exclusive to C-7 genotypes. Most members of A-3 groups presented H-11. H-12 and H-13 were encountered within C-4 members. C-2 presented H-19 within nine sequences. Haplotypes H-7, H-9, H-24 and H-27 were exclusively found within C-6 subgroup. H-28 was found within two members of C-1 cluster. The remaining haplotypes were found in distinct genotypes at the frequency of only one genotype per haplotype (Supplementary Table 6.2). Further analysis of haplotype networks was performed with Network 4.6.1.3 software. Figure 6.7 illustrates the networks found, which confirmed the analyses of NJ, BAPs and PCoA about the clusters definition for *PvDREB6B* polymorphisms. As shown in the NJ tree (Figure 6.4), networks were divided into two major subgroups, one encompassing clusters C-1, C-2, C-3, and the other with clusters C-5, C-6 and C-7. C-4 was shown in intermediate position. C-6 and C-7 samples were very close in the network. NJ and PCoA also indicated their proximity.

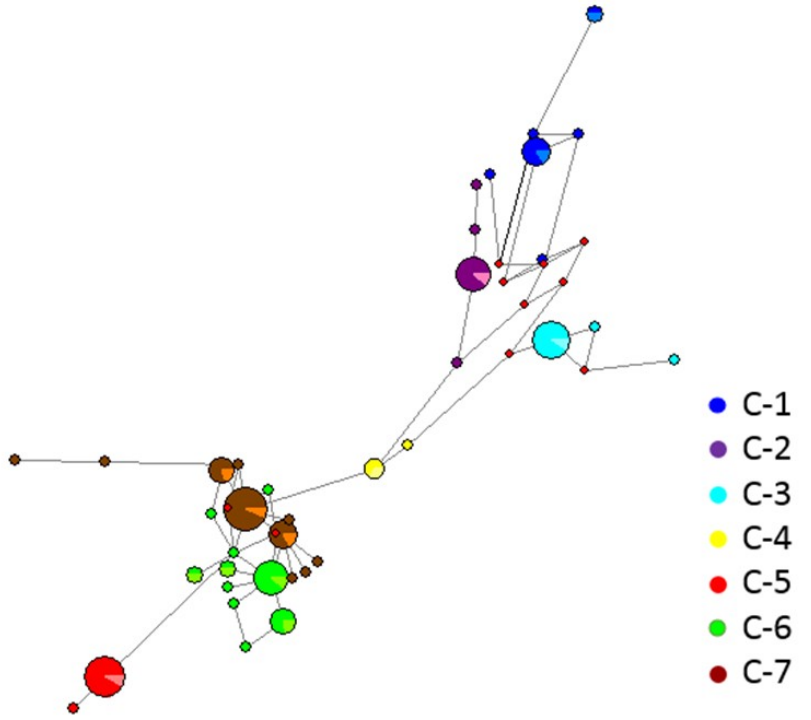


Figure 6.7 - Haplotype networks from *PvDREB6B* in wild common bean accessions of Mesoamerican origin and one accession of *Phaseolus acutifolius*. Colors represent once more the cluster structure proposed along this chapter. In total, 37 haplotypes networks are represented.

Table 6.2 – Nucleotide diversity estimates of *PvDREB6B* in 112 wild common accessions from Mesoamerican genepool

	All wilds
Number of sequences	112
Total number of sites	957
Polymorphic sites	33
Average frequency of SNP (bp/SNP)	29
Total number of mutations	35
Singletons sites	4
Parsimony informative sites	29
Haplotypes	37
Haplotype diversity (Hd)	0.941
Nucleotide variation	
Diversity (π/bp)	0.00725
Polymorphic (θ_W/bp)	0.00652
π (JC)	0.00730
Neutrality tests	
Tajima's *D*	0.15089
Fu and Li's *D**	0.51071
Fu and Li's *F**	0.43978
Fu's F_S	-11.015*
Fay and Wu's *H*	-1.09813
Ramos-Onsins and Roza's R^2	0.104016
Recombination	
Four-gamete types site pairs	80
Rm	6

Significance levels for neutrality tests: * $P < 0.05$.
Fay and Wu's *H* was calculated using an outlier from *P. acutifolius*.

The nucleotide diversity was described by two main measures, the expected heterozygosity per nucleotide site (π) and the number of polymorphic sites in a genotypic sample corrected for sample size (θ_W) (Table 6.2). The overall nucleotide variation was high within the whole set of accessions (π = 0.00725, θ_W = 0.00652). In comparison to the previous work of Cortés et al. (2012), which have examined only partial sequence of the gene in samples from contrasting gene pools (Andean and Mesoamerican), our results similar values, although the value of θ_W was considerably lower (π = 0.0088, θ_W = 0.0048, both in wild beans). However, as stated before, Cortés et al. (2012) analyzed a limited region of the gene (only 547 bp from the entire ORF, which encompasses 957 bp) and genotypes from both gene pools, revealing only 11 polymorphic sites. Our goal was the discovery of polymorphisms only within the Mesoamerican pool, but for the entire ORF, revealing number of polymorphisms three times higher (33 sites).

Neutrality tests were performed to verify if *PvDREB6B* was under positive selection (Table 6.2). Tajima's *D* (0.1589), Fu and Li's *D** (0.51071) and Fu and Li's *F** (0.43978) showed no significant difference from zero ($P > 0.10$), which suggested no evidence for changes in population size or any particular pattern of selection at *PvDREB6B*. However, a significant negative value of Fu's F_S was detected (F_S = -11.015, $P < 0.05$). Negative values of F_S suggest an excessive number of alleles, which could be an indication of recent population expansion, adaptive selection or genetic hitchhiking. To solve one of the questions, we tested for the possibility of hitchhiking with the Fay and Wu's *H* test. A negative but still non-significant value was obtained (*H* = -1.09813). As Tajimas's *D* was non-significantly different from zero and F_S was highly negative, *PvDREB6B* might have an excessive number of alleles, but no evident positive selection was detected. Population expansion was tested with Ramos-Onsins & Rozas R^2 test and it revealed no significant value (R^2 = 0.104016, P = 0.70). In general, these results are in agreement with the ones obtained by Cortés et al. (2012).

In further analysis, we also tested for the putative effect of non-synonymous substitutions on the function of PvDREB6B protein using PROVEAN tools (Table 6.3). The number of non-synonymous substitutions (16) was lower than the synonymous (19) and PROVEAN results revealed most non-synonymous changes had neutral effect (PROVEAN score > -2.5). Nevertheless, three substitutions indicated to be deleterious, which could have a negative impact in some wild beans. Although no signatures of positive selection were revealed by the neutrality tests for the whole *PvDREB6B* gene, some selective forces might be acting in specific codons.

Table 6.3 – Number of synonymous and non-synonymous substitutions within *PvDREB6B* open reading frame and its putative effect of the protein function, as predicted by PROVEAN algorithm

Synonymous and non-synonymous substitutions			
Number of synonymous SNP sites	19		
Number of non-synonymous SNP sites	16		
Positions of single amino acid substitutions	**Type of substitution**	**PROVEAN score**	**Prediction (cutoff= -2.5)**
10	Asn/Thr	-0.216	Neutral
18	Ile/Leu	-0.122	Neutral
42	Ser/Phe	-2.41	Neutral
48	Ser/Leu	-1.688	Neutral
96	His/Gln	0.877	Neutral
98	Leu/Gln	0.715	Neutral
103	Met/Leu	-1.924	Neutral
119	Ala/Ser	-0.206	Neutral
232	His/Asp	-0.156	Neutral
241	Glu/Asp	-1.243	Neutral
261	Val/Met	-0.119	Neutral
264	Asn/Thr	-0.157	Neutral
308	Tyr/Ser	-7.379	Deleterious
308	Tyr/Phe	-1.921	Neutral
314	Asp/Glu	-3.896	Deleterious
317	Ala/Asp	-2.796	Deleterious

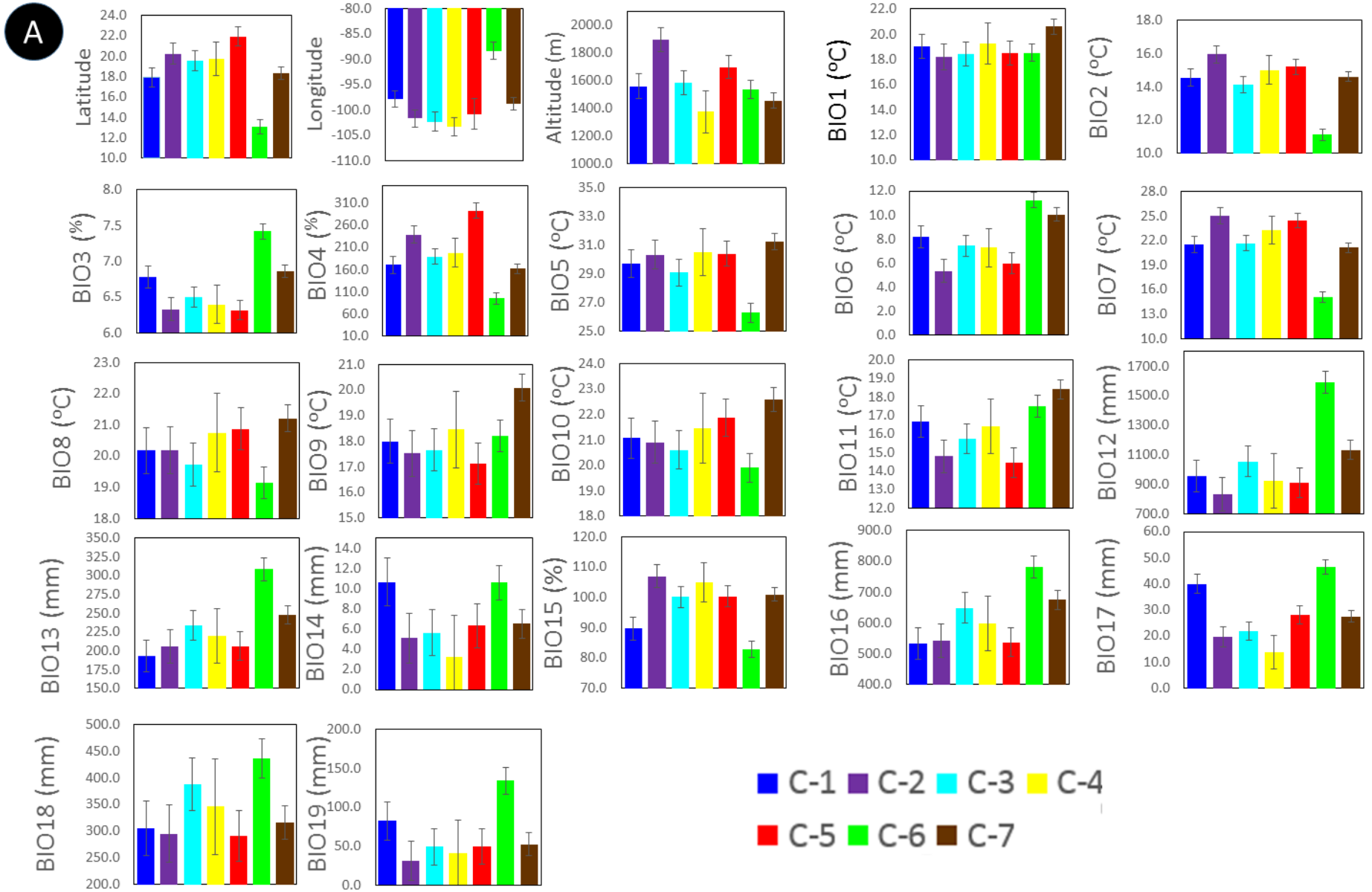
Although no consistent evidence of positive selection were encountered when analyzing the whole set of wild genotypes, some forces might be modeling local populations. Our previous phylogenetic and population structure analyses (NJ, PCoA, BAPs, Network and GenGIS) supported seven subgroups of wild accessions. Perhaps, increased representation within each subgroup is needed to obtain further evidence that actual selection forces are shaping individuals from each local.

In this line, we also looked at the environmental features that characterize the areas in which the accessions are distributed. A series of precipitation, temperature and other bioclimatic variables were analyzed within each of the seven clusters detected in our study. Analyses were performed by univariate and multivariate methods to detect possible variables associated with the polymorphism distribution.

6.3.4 Association with environmental variables and geographic distribution

Univariate analyses showed some interesting features about each cluster determined through *PvDREB6B* sequence analyses (Figure 6.8 A and Supplementary Table 6.3). The distribution of haplotypes partially followed a latitudinal gradient, as Cortés et al. (2012) also reported for Andean and Mesoamerican genotypes. The haplotypes from C-6 subgroup were located at the lowest latitudes and longitudes, in general from Central America to Colombia.

Hereby, we emphasized the altitude, temperature and precipitation variation as particular factors varying among genetic clusters (Figure 6.8). The average altitude of cluster C-2 (1,894 m) was the highest among all the other subgroups (ranging from 1,375 to 1,696 m). Almost all individuals from this cluster were located at the border between Jalisco and Guanajuato states in Mexico. In average, accessions of C-2 showed the lowest temperature of the coldest month of the year (BIO 6 = 5.34ºC), the lowest annual precipitation (BIO 12 = 831 mm) and the lowest precipitation of the coldest quarter (BIO 19 = 31 mm). The highest average mean diurnal range of temperature (BIO 2 = 15.96ºC) was also observed at the area the accession of C-2 are located. In a broader panel, as regards the minimum temperature of the coldest month (BIO 6), C-1, C-2, C-3, C-4 and C-5 presented the lowest averages, while subgroups C-6 and C-7 showed averages significantly higher. C-6 presented the most significant differences in several variables from the other subgroups. C-6 clearly exhibited the highest average isothermality over the year (BIO 3 = 7.4%), which reflected in the lowest average temperature of the wettest quarter (BIO 8 = 19.1) and the lowest average temperature of the warmest quarter (BIO 10 = 19.9) observed. Although, the most prominent variables to distinguish C-6 from the others were related to precipitation. Accession from C-6 are distributed in the areas with the highest average annual precipitation (BIO 12 = 1,591 mm) while C-7 members figure in second place (BIO 12 = 1,133 mm). Thereby, precipitation-derived bioclimatic variables were mostly highlighted for the locations of C-6 members.


A
Latitude
Longitude
Altitude (m)
BIO1 (ºC)
BIO2 (ºC)
BIO3 (%)
BIO4 (%)
BIO5 (ºC)
BIO6 (ºC)
BIO7 (ºC)
BIO8 (ºC)
BIO9 (ºC)
BIO10 (ºC)
BIO11 (ºC)
BIO12 (mm)
BIO13 (mm)
BIO14 (mm)
BIO15 (%)
BIO16 (mm)
BIO17 (mm)
BIO18 (mm)
BIO19 (mm)
C-1 C-2 C-3 C-4
C-5 C-6 C-7

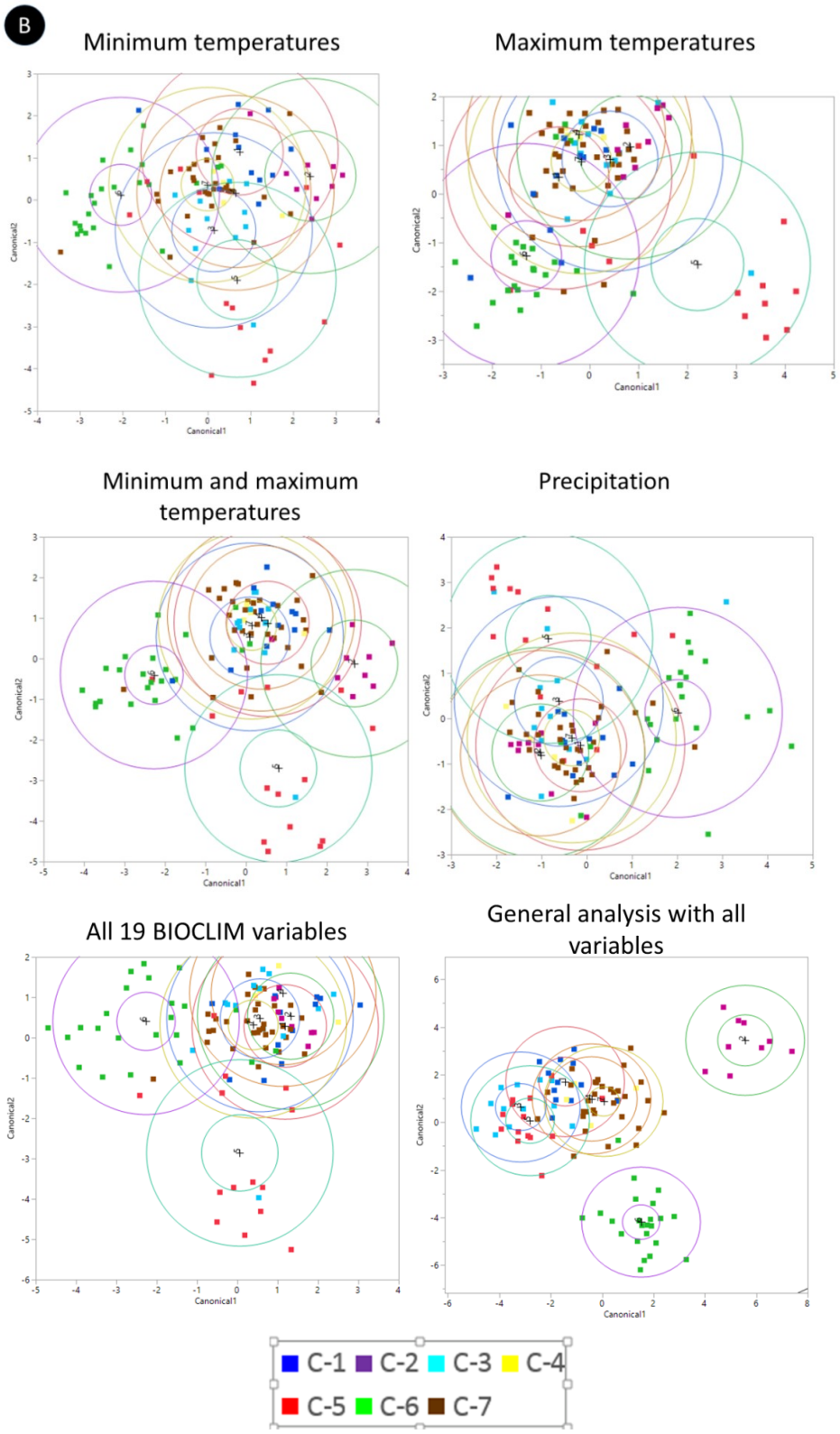


Figure 6.8 – (A) Average latitude, longitude, altitude and of 19 bioclimatic variables based on data retrieved from BioClim for the location of accessions distributed among seven genetic clusters (C-1 to C-7) defined by prior multivariate analysis (NJ tree, BAPs structure, haplotype networks and principal coordinate analysis). BIO1 (Annual mean temperature), BIO2 (Mean diurnal range), BIO3 (Isothermality), BIO4 (Temperature seasonality), BIO5 (Maximum temperature of warmest month), BIO6 (Minimum temperature of coldest month), BIO7 (Temperature annual range), BIO8 (Mean temperature of wettest quarter), BIO9 (Mean temperature of driest quarter), BIO10 (Mean temperature of warmest quarter), BIO11 (Mean temperature of coldest quarter), BIO12 (Annual precipitation), BIO13 (Precipitation of wettest month), BIO14 (Precipitation of driest month), BIO15 (Precipitation seasonality), BIO16 (Precipitation of wettest quarter), BIO17 (Precipitation of driest quarter), BIO18 (Precipitation of warmest quarter), BIO19 (Precipitation of coldest quarter). (B) Canonical discriminant analyses with determined variable combinations.

Canonic discriminant analyses combining different set of geographical and bioclimatic variables revealed interesting differences among clusters (Figure 6.8 B). The combination of all average minimum temperatures and their discrimination by each cluster showed clusters C-2, C-6 and C-5 more distinguished from the other subgroups. C-5 was more clearly separated when all maximum temperatures were combined. The clearest separation for C-5, however, were obtained when minimum and maximum temperature were combined. In fact, C-5 group was defined by accessions distributed mostly in drier areas from norther Mexico, however, some accessions from southern Mexico also fitted within this subgroup. Despite this, the accessions from the driest location upper north were clearly separated by the canonical plot with minimum and maximum temperatures. The univariate variables showed these accessions accounted for the highest temperature seasonality (BIO 4 = 293%), which refers to the annual range between the minimum and maximum temperatures. Thereby, some alleles might be more predominant in drier areas. C-6 was discriminated mainly by precipitation, as detected for most univariate analyses. The combination of minimum and maximum temperatures, precipitation over the year, altitude and the 19 bioclimatic variables resulted in a well-defined discrimination of C-6 and C-2 clusters. C-6 and C-2 were previously emphasized for most univariate analyses.

6.4 Discussion

Our study was dedicated to further investigate nucleotide variability patterns of *PvDREB6B* within a wild bean collection. This analysis consisted of an expansion from the prior work developed by Cortés et al. (2012). At the time, the authors classified two genes as a *DREB2A* and *DREB2B* and traced their sequence variation (NAYAK et al., 2009; CORTÉS et al., 2012). With the release of the common bean genome (SCHMUTZ et al., 2014) the complete sequence of both genes was obtained from Phytozome. Phylogenetic analysis with a series of *DREB* genes from *Arabidopsis* and *G. max* evidenced both *DREB2A* and *DREB2B* are actually representatives from A-6 subgroup. In previous work (Section 3), we proposed the replacement of the name of *DREB2A* to *PvDREB6B*, sustained by not only phylogenetic profiling, but also the expression patterns under specific abiotic stresses. Moreover, we analyzed the full open reading frame of *PvDREB6B*, spanning 957 bp, while Cortés et al. (2012) were able to cover 547 bp within the gene. We also centered our work on only wild beans from Mesoamerican origin, with higher number of samples for better representing the structure within the gene pool. Finally, the most significant effort presented in this work was

an attempt to associating *PvDREB6B* polymorphisms with a series of bioclimatic variables involved in cold and drought tolerance.

The nucleotide diversity estimates π and θ_W were considerably high for *PvDREB6B*. In comparison to Cortés et al. (2012), the number of polymorphic sites was three times higher due to the fact we were able to cover 410 bp more in length reaching the 957 that effectively constitute the open reading frame of *PvDREB6B*. When compared to other genes investigated for common bean, *PvDREB6B* diversity, on the basis of π, was intermediate. The nucleotide diversity within an intron within a gene coding for Dihydroflavonol 4-reductase averaged 0.05 (McCLEAN; LEE; MIKLAS, 2004). The gene *PvSHP1* (*SHATTERPROOF*) exhibited high diversity in a set of Andean and Mesoamerican genotypes (π = 0.0147) (NANNI et al., 2011). A large fragment of 2,043 bp was sequenced and identified as *PvIND*, a homolog to the *INDEHISCENT* gene. Mesoamerican and Andean wild and domesticated genotypes exhibited lower levels of nucleotide diversity (π = 0.0023) (GIOIA et al., 2013). Population genetic estimates of five loci including intron and partial coding regions also revealed variable levels of nucleotide diversity for *Leg044* (π = 0.0061), *Leg100* (π = 0.0226), *Leg 133* (π = 0.0051), *Leg223* (0.0032) and *PvSHP1* (0.0140) in wild genotypes of common bean from both gene pools (BITOCCHI et al., 2012). In general, these works have reported that intron regions have higher levels of nucleotide variation. Despite this, we showed several polymorphisms within the ORF of *PvDREB6B*. Investigations around the sequence of the gene could unravel even higher number of SNP, such as in the promoter of the gene.

Several polymorphisms identified at the DNA levels (16 sites) were translated into alternative amino acids in the protein sequences of the wild beans. We performed an analysis of the putative effects such polymorphisms could have on the protein function. PROVEAN tools provide a detailed search of sequences in databases and the possible effects of non-synonymous substitutions on gene functioning (CHOI; CHAN, 2015). Applying the PROVEAN tools, three amino acid replacements were predicted to be deleterious. Although neutrality tests have suggested that *PvDREB6B* is not being subjected to positive selection, specific codons such as the reported might be under some selection force. To effectively prove the functional effects of such mutations, further experimental corroboration is needed.

Neutrality tests suggested that the wild Mesoamerican bean from this study are not under positive selection at *PvDREB6B*. Cortés et al. (2012) encountered similar results, however, they adverted that local adaptation might reflect the haplotype networks detected for the gene. Furthermore, neither population structure or expansions after bottlenecks could fully explain the events shaping variation at the gene (CORTÉS et al., 2012). The sampling

size might also have influenced the results. Although we sample a higher number of wild beans only from Mesoamerican origin, the neutrality tests accounted for the whole set of genotypes. If we were to increase sampling within each cluster (C-1 to C-7), more consistency would be obtained about selection forces operating at *PvDREB6B*.

The population structure revealed by BAPs, haplotypes network and principal coordinate analysis reveal seven main clusters. As other works have detected, haplotypes distribution was partially associated with the latitude gradient in which accessions are distributed (CHACÓN et al, 2007; CORTÉS et al., 2012). As a consequence, this association can be extended to climatic variability, since higher latitudes have drier climates while Central America and Colombia have much higher average precipitation over the year. A previous work has shown the potential of bioclimatic variables in determining drought tolerance indexes for wild accessions, rather than performing field experiments with such goal (CORTÉS et al., 2013). Beans have a growth cycle varying from eight to ten months and, therefore, they are subjected to dramatic changes in meteorological conditions over the year. Tropical environmental with bimodal rainfall have a mid-season dry period. In this period, beans enter a cycle of slow growth and low physiological activity (CORTÉS et al., 2012).

Hereby, we dissected the potential association of several bioclimatic variables retrieved for specific location of each one of the wild accessions. *PvDREB6B* has been found induced by low temperature and dehydration treatments (Sections 3 and 5). Therefore, we extracted components of temperature and precipitation regimes to verify possible association between *PvDREB6B* nucleotide variants and low temperature and a gradient of humidity over the distribution area of the wilds. Univariate analyses suggested low temperatures and isothermality as variables associated with *PvDREB6B* variation. Moreover, precipitation was a determining factor for distinguishing cluster C-6 as constituted of the accessions located in the most humid areas in Central America and Colombia. Canonical discriminant analyses with a combination of all temperature, precipitation and altitude variables showed clusters C-2 and C-6 discriminated from all the others. C-6 discrimination was mostly explained by precipitation while C-2 was determined by both precipitation and temperature. Altitude of C-2 was a significant aspect for determining its structure as well.

These results provided insights about the possible association of the population structure revealed by *PvDREB6B* variation with low temperature and a gradient of humidity. However, to dissect the specific role of the gene on the adaptation to cold and drought further

experimental and modeling analyzes are needed. Although, considering all data, this locus seems to be a potential candidate for improving abiotic stress tolerance in common bean.

6.5 Conclusions

In this study, we identified nucleotide diversity patterns of *PvDREB6B* in wild accessions of common bean from the Mesoamerican gene pool. High levels of polymorphisms were detected within the open reading frame of this locus. Neutrality tests suggested no positive selection forces are affecting the gene, however some specific codons might be under selection. Further evidence is needed to attest for local adaptations and how they may have shaped the polymorphisms at single genes. Our analyses of population structure revealed seven clusters of wild Mesoamerican accessions, according to the distribution of *PvDREB6B* haplotypes. Haplotypes were partially associated with a series of environmental variables, especially as regards temperature oscillation and precipitation regimes. *PvDREB6B* might be related to cold and drought adaptation as we found significant association between haplotypes and accessions located at areas with the coldest temperatures and others located over a gradient of humidity from northern Mexico to Colombia. This gene might be useful to select sources of drought and cold tolerance for breeding programs of common bean.

References

AKHTAR, M.; JAISWAL, A.; TAJ, G.; JAISWAL, J. P.; QURESHI, M. I.; SINGH, N. K. DREB1 / CBF transcription factors : their structure , function and role in abiotic stress tolerance in plants. **Journal of Genetics**, London, v. 91, n. 3, p. 385–395, 2012.

BANDELT, H. J.; FORSTER, P.; RÖHL A. Median-joining networks for inferring intraspecific phylogenies. **Molecular Biology and Evolution**, Chicago, v. 16, p. 37–48, 1999.

BENCHIMOL, L. L.; CAMPOS, T. de; CARBONELL, S. A. M.; COLOMBO, C. A.; CHIORATTO, A. F.; FORMIGHIERI, E. F.; GOUVÊA, L. R. L.; SOUZA, A. P. de. Structure of genetic diversity among common bean (*Phaseolus vulgaris* L.) varieties of Mesoamerican and Andean origins using new developed microsatellite markers. **Genetic Resources and Crop Evolution**, Dordrecht, v. 54, n. 8, p. 1747–1762, 2007.

BITOCCHI, E.; NANNI, L.; BELLUCCI, E.; ROSSI, M.; GIARDINI, A., ZEULI, P. S.; PAPA, R. Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. **Proceedings of the National Academy of Sciences of the USA**, Washington, DC, v. 109, n. 14, p. E788–E796, 2012. doi:10.1073/pnas.1108973109.

BLAIR, M. W.; GIRALDO, M. C.; BUENDÍA, H. F.; TOVAR, E.; DUQUE, M. C.; BEEBE, S. E. Microsatellite marker diversity in common bean (*Phaseolus vulgaris* L.). **Theoretical and Applied Genetics**, Heidelberg, v. 113, n. 1, p. 100–109, 2006.

BLAIR, M. W.; SOLER, A.; CORTÉS, A. J. Diversification and population structure in common beans (*Phaseolus vulgaris* L.). **Plos One**, San Francisco, v. 7, n. 11, e49488, 2012.

BURLE, M. L.; FONSECA, J. R.; KAMI, J. A.; GEPTS, P. Microsatellite diversity and genetic structure among common bean (*Phaseolus vulgaris* L.) landraces in Brazil, a secondary center of diversity. **Theoretical and Applied Genetics**, Heidelberg, v. 121, n. 5, p. 801–813, 2010.

CHACÓN S., M. I.; PICKERSGILL, B.; DEBOUCK, D. G.; ARIAS, J. S. Phylogeographic analysis of the chloroplast DNA variation in wild common bean (*Phaseolus vulgaris* L.) in the Americas. **Plant Systematics and Evolution**, Wien, v. 266, n. 3-4, p. 175–195, 2007.

CHOI, Y.; CHAN, A. P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. **Bioinformatics**, Oxford, v. 31, n. 16, p. 2745-2747, 2015.

CORANDER, J.; MARTTINEN, P. Bayesian identification of admixture events using multilocus molecular markers. **Molecular Ecology**, Chichester, v. 15, p. 2833–2843, 2006.

CORANDER, J.; TANG, J. Bayesian analysis of population structure based on linked molecular information. **Mathematical Biosciences**, Philadelphia, v. 205, p. 19–31, 2007.

CORANDER, J.; WALDMANN, P.; SILLANPÄÄ, M. J. Bayesian analysis of genetic differentiation between populations. **Genetics**, Bethesda, v. 163, p. 367–374, 2003.

CORTÉS, A. J.; CHAVARRO, M. C.; BLAIR, M. W. SNP marker diversity in common bean (*Phaseolus vulgaris* L.). **Theoretical and Applied Genetics**, Heidelberg, v. 123, n. 5, p. 827–845, 2011.

CORTÉS, A. J.; MONSERRATE, F. A.; RAMÍREZ-VILLEGAS, J.; MADRIÑÁN, S.; BLAIR, M. W. Drought tolerance in wild plant populations: the case of common beans (*Phaseolus vulgaris* L.). **Plos One**, San Francisco, v. 8, n. 5, e62898, 2013.

CORTÉS, A. J.; THIS, D.; CHAVARRO, C.; MADRIÑÁN, S.; BLAIR, M. W. Nucleotide diversity patterns at the drought-related DREB2 encoding genes in wild and cultivated common bean (*Phaseolus vulgaris* L.). **Theoretical and Applied Genetics**, Heidelberg, v. 125, n. 5, p. 1069–1085, 2012.

DUQUE, A.; ALMEIDA, A. M. de; SILVA, A. B. da; SILVA, J. M. da; FARINHA, A. P.; SANTOS, D.; FEVEREIRO, P.; ARAÚJO, S. de S. Abiotic stress responses in plants: unraveling the complexity of genes and networks to survive. In: VAHDATI, K.; LESLIE, C. (Ed.). **Abiotic stresses** – Plant responses and applications in agriculture. Rijeka, Croatia: InTech, 2013. p. 49–101.

FAY, J. C.; WU, C. I. Hitchhiking under positive Darwinian selection. **Genetics**, Bethesda, v. 155, p. 1405–1413, 2000.

FU, Y. X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. **Genetics**, Bethesda, v. 147, p. 915–925, 1997.

GEPTS, P. A middle American and an Andean common bean gene pool. In: GEPTS, P. (Ed.). **Genetic resources of *Phaseolus* beans**. Dordrecht: Kluwer, 1988. p. 375–407.

GEPTS, P.; OSBORN, T. C.; RASHKA, K.; BLISS, F. A. Phaseolin-protein Variability in Wild Forms and Landraces of the Common Bean (*Phaseolus vulgaris*): Evidence for Multiple Centers of Domestication. **Economic Botany**, New York, v. 40, n. 4, p. 451–468, 1986.

GIOIA, T.; LOGOZZO, G.; KAMI, J.; SPAGNOLETTI ZEULI, P.; GEPTS, P. Identification and characterization of a homologue to the *Arabidopsis INDEHISCENT* gene in common bean. **Journal of Heredity**, Washington, DC, v. 104, n. 2, p. 273–286, 2013.

GONZÁLEZ, A.; WONG, A.; DELGADO-SALINAS, A.; PAPA, R.; GEPTS, P. Assessment of inter simple sequence repeat markers to differentiate sympatric wild and domesticated populations of common bean. **Crop Science**, Madison, v. 45, n. 2, p. 606–615, 2005.

HALL, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. **Nucleic Acids Symposium Series**, Oxford, v. 41, p. 95-98, 1999.

HUDSON, R. R. A new statistic for detecting genetic differentiation. **Genetics**, Bethesda, v. 155, p. 2011–2014, 2000.

HUDSON, R. R.; KAPLAN, N. L. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. **Genetics**, Bethesda, v. 111, p. 147–164, 1985.

KOENIG, R.; GEPTS, P. Allozyme diversity in wild *Phaseolus vulgaris*: further evidence for two major centers of genetic diversity. **Theoretical and Applied Genetics**, Heidelberg, v. 78, n. 6, p. 809–817, 1989.

LATA, C.; PRASAD, M. Role of DREBs in regulation of abiotic stress responses in plants. **Journal of Experimental Botany**, Oxford, v. 62, n. 14, p. 4731–4748, 2011.

LIBRADO, P.; ROZAS, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphisms data. **Bioinformatics**, Oxford, v. 25, p. 1451-1452, 2009.

LIU, Q.; KASUGA, M.; SAKUMA, Y.; ABE, H.; MIURA, S.; YAMAGUCHI-SHINOZAKI, K.; SHINOZAKI, K. Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in *Arabidopsis*. **The Plant Cell**, Rockville, v. 10, n. 8, p. 1391–1406, 1998.

MACIEL, F.; ECHEVERRIGARAY, S.; GERALD, L.; GRAZZIOTIN, F. Genetic relationship and diversity among Brazilian cultivar and of common beans (*Phaseolus vulgaris* L.) revealed by AFLP markers. **Genetic Resources and Crop Evolution**, Dordrecht, v. 50, p. 887–893, 2003.

McCLEAN, P. E.; LEE, R. K.; MIKLAS, P. N. Sequence diversity analysis of dihydroflavonol 4-reductase intron 1 in common bean. **Genome**, Ottawa, v. 47, n. 2, p. 266–280, 2004.

MIZOI, J.; SHINOZAKI, K.; YAMAGUCHI-SHINOZAKI, K. AP2/ERF family transcription factors in plant abiotic stress responses. **Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms**, Amsterdam, v. 1819, n. 2, p. 86–96, 2012.

NAKANO, T.; SUZUKI, K.; FUJIMURA, T.; SHINSHI, H. Genome-Wide Analysis of the ERF Gene Family. **Plant Physiology**, Rockville, v. 140, p. 411–432, 2006.

NAKASHIMA, K.; YAMAGUCHI-SHINOZAKI, K. Regulons involved in osmotic stress-responsive and cold stress-responsive gene expression in plants. **Physiologia Plantarum**, Copenhagen, v. 126, n. 1, p. 62–71, 2006.

NANNI, L.; BITOCCHI, E.; BELLUCCI, E.; ROSSI, M.; RAU, D.; ATTENE, G.; GEPTS, P.; PAPA, R. Nucleotide diversity of a genomic sequence similar to SHATTERPROOF (PvSHP1) in domesticated and wild common bean (*Phaseolus vulgaris* L.). **Theoretical and Applied Genetics**, Heidelberg, v. 123, n. 8, p. 1341–1357, 2011.

NAYAK, S. N.; BALAJI, J.; UPADHYAYA, H. D.; HASH, C. T.; KISHOR, P. B. K.; CHATTOPADHYAY, D.; RODRIQUEZ, L. M.; BLAIR, M. W.; BAUM, M.; MCNALLY, K.; THIS, D.; HOISINGTON, D. A.; VARSHNEY, R. K. Isolation and sequence analysis of *DREB2A* homologues in three cereal and two legume species. **Plant Science**, Amsterdam, v. 177, n. 5, p. 460–467, 2009.

NEI, M. **Molecular evolutionary genetics**. New York: Columbia University Press, 1987.

PALLOTTA, M. A.; WARNER, P.; FOX, R. L.; KUCHEL, H.; JEFFERIES, S. J.; LANGRIDGE, P. Marker assisted wheat breeding in the southern region of Australia. In: INTERNATIONAL WHEAT GENETICS SYMPOSIUM, 10., 2003, Paestum, Italy. **Proceedings…** Sydney: Sydney University Press, 2003. p. 1-6.

PEAKALL, R.; SMOUSE, P. E. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. **Bioinformatics**, Oxford, v. 28, n. 9, p. 2537-2539, 2012.

SAIBO, N. J. M.; LOURENÇO, T.; OLIVEIRA, M. M. Transcription factors and regulation of photosynthetic and related metabolism under environmental stresses. **Annals of Botany**, Oxford, v. 103, n. 4, p. 609–623, 2009.

SAITOU, N.; NEI, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Molecular Biology and Evolution**, Chicago, v. 4, p. 406-425, 1987.

SAKUMA, Y.; LIU, Q.; DUBOUZET, J. G.; ABE, H.; SHINOZAKI, K.; YAMAGUCHI-SHINOZAKI, K. DNA-binding specificity of the ERF/AP2 domain of *Arabidopsis* DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. **Biochemical and Biophysical Research Communications**, New York, v. 290, n. 3, p. 998–1009, 2002.

SCHMUTZ, J.; MCCLEAN, P. E.; MAMIDI, S.; WU, G. A.; CANNON, S. B.; GRIMWOOD, J.; JENKINS, J.; SHU, S.; SONG, Q.; CHAVARRO, C.; TORRES-TORRES, M.; GEFFROY, V.; MOGHADDAM, S. M.; GAO, D.; ABERNATHY, B.; BARRY, K.; BLAIR, M.; BRICK, M. A.; CHOVATIA, M.; GEPTS, P.; GOODSTEIN, D. M.; GONZALES, M.; HELLSTEN, U.; HYTEN, D. L.; JIA, G.; KELLY, J. D.; KUDRNA, D.; LEE, R.; RICHARD, M. M. S.; MIKLAS, P. N.; OSORNO, J. M.; RODRIGUES, J.; THAREAU, V.; URREA, C. A.; WANG, M.; YU, Y.; ZHANG, M.; WING, R. A.; CREGAN, P. B.; ROKHSAR, D. S.; JACKSON, S. A. A reference genome for common bean and genome-wide analysis of dual domestications. **Nature**, London, v. 46, n. 7, p. 707–713, 2014.

SHKOLNIK-INBAR, D.; BAR-ZVI, D. Expression of *ABSCISIC ACID INSENSITIVE 4* (*ABI4*) in developing *Arabidopsis* seedlings. **Plant signaling & Behavior**, Austin, v. 6, n. 5, p. 694–696, 2011.

SINGH, S. P.; DEBOUCK; GEPTS, P. Races of common bean (*Phaseolus vulgaris*, Fabaceae). **Economic Botany**, New York, v. 45, p. 379–396, 1991.

TAJIMA F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. **Genetics**, Bethesda, v. 123, p. 585–595, 1989.

TAMURA, K.; STECHER, G.; PETERSON, D.; FILIPSKI, A.; KUMAR, S. MEGA6: Molecular evolutionary genetics analysis version 6.0. **Molecular Biology and Evolution**, Chicago, v. 30, n. 12, p. 2725–2729, 2013.

WATTERSON, G. On the number of segregating sites. **Theoretical Population Biology**, New York, v. 7, p. 256-276, 1975.

ZHANG, G.; CHEN, M.; CHEN, X.; XU, Z.; GUAN, S.; LI, L. C.; LI, A.; GUO, J.; MAO, L.; MA, Y. Phylogeny, gene structures, and expression patterns of the ERF gene family in soybean (*Glycine max* L.). **Journal of Experimental Botany**, Oxford, v. 59, n. 15, p. 4095–4107, 2008.

7 PHENOTYPIC ANALYSIS OF DROUGHT-RELATED TRAITS IN WILD COMMON BEAN AND THEIR ASSOCIATION WITH GENETIC STRUCTURE AND *DREB*-LINKED LOCI

Abstract

This chapter is part of a collaborative work aimed at phenotyping drought-related traits in a wild common bean panel of Mesoamerican accessions from northern Mexico to Colombia and some domesticated genotypes; and find association between the traits and SNP markers from the BARCBean6K_3 BeadChip, developed for common bean. Thereby, we show the results of a greenhouse experiment consisting of a tube screening assay under two treatments, irrigation to field capacity and water deficit. The innovation of this experiment resided on sequential measurements of root and shoot growth for 34 days, along with biomass, chlorophyll and leaf area determinations for all the genotypes and treatments. We correlated the phenotypic data from the wild genotypes with environmental data of the location of each accession. The results indicated that root depth, plant height and biomass were positively correlated with latitude and, consequently, negatively correlated with precipitation, as the higher the latitude the lower the precipitation. Phenotypic variation was also analyzed for its association with the genetic structure determined through *PvDREB6B* haplotypes in the previous study. Overall, the genotypes from the driest areas, such as Durango and Jalisco, presented the highest root depths and plant heights, fitting in specific genetic clusters. Conversely, the genotypes with lower root depth and plant height belonged to genetic clusters distributed in areas with higher humidity. Ultimately, we attempted to find association between the drought-related traits and *DREB*-linked loci. Although some associations were found, a test of false rate discoveries attested their lack of significance. Our study presented important prospects for the use of genetic resources from a wild background into breeding programs aimed at drought tolerance in common bean, emphasizing the *DREB* subfamily as candidate genes for marker-assisted selection.

Keywords: Phenotyping. Drought-tolerance. Wild accessions. Association mapping.

Resumo

Este capítulo é parte de um trabalho de colaboração para avaliar caracteres fenotípicos em uma coleção selvagem de feijoeiro de origem Mesoamericana, com acessos distribuídos desde o norte do México até a Colômbia, e alguns genótipos domesticados. Entre os objetivos estão encontrar associação entre os caracteres e marcadores SNP da plataforma BARCBean6K_3 BeadChip, desenvolvida para feijoeiro. Desse modo, são apresentados os resultados de experimento de casa de vegetação consistindo de um sistema de fenotipagem com tubos plásticos com dois tratamentos, irrigado e sob déficit hídrico. O aspecto inovador deste experimento consistiu em medidas sequenciais do crescimento radicular e da parte aérea por 34 dias. Além disso, foram avaliados biomassa, clorofila e área foliar em todos os genótipos e tratamentos. Os dados fenotípicos de cada acesso foram correlacionados com dados ambientais de cada um dos acessos. Os resultados indicaram que o comprimento radicular, a altura das plantas e a biomassa foram positivamente correlacionados com latitude e, consequentemente, apresentaram correlação negativa com precipitação, pois quanto maior a latitude menor a precipitação. A variação fenotípica também foi associada com a estrutura populacional determinada através de haplótipos de *PvDREB6B* no capítulo anterior. No geral, os genótipos das áreas mais seca, como Durango e Jalisco, apresentaram os mais elevados comprimentos de raiz e altura de plantas. Contrariamente, os genótipos de menor comprimento radicular e altura de planta pertenceram aos subgrupos genéticos de regiões mais úmidas. Finalmente, foram buscadas associações entre parâmetros fenotípicos e marcadores associados a *DREB*. Embora algumas associações foram encontradas, teste de razão de falsas descobertas mostrou que as associações não foram significativas. Este estudo mostrou importantes resultados para o uso de recursos genéticos provenientes de genótipos selvagens para aplicações em programas de melhoramento para tolerância ao déficit hídrico em feijoeiro, enfatizando genes da família *DREB* como candidatos para seleção assistida por marcadores moleculares.

Palavras-chave: Fenotipagem. Tolerância à seca. Acessos selvagens. Mapeamento associativo.

7.1 Introduction

Drought episodes represent a significant challenge for common bean production throughout the world (BEEBE et al., 2013). Thereby, several projects have been focused on the understanding of the mechanisms of drought adaptation in the species, recurring to wild accessions as sources of drought tolerance (CORTÉS et al., 2013). Wild beans naturally undergo a longer life cycle (8 – 10 months) (CORTÉS et al., 2012) than domesticated genotypes, so it is a challenging process to determine drought-tolerance levels for wilds based on phenotyping approaches. Instead, a modelling approach has been proposed, considering several bioclimatic variables combined to define a drought index for wild common beans (CORTÉS et al., 2013).

On the other hand, experimental evaluations of drought tolerance have been restricted to cultivated common bean. Studies have been concerned with the identification of genomic regions underlying traits affected by drought stress, examining populations in field (BLAIR et al., 2012; ASFAW et al., 2012; MUKESHIMANA et al., 2014) and greenhouse (ASFAW; BLAIR, 2012) conditions. Greenhouse phenotyping has been done with a PVC tube screening assay (RAO et al., 2006), which enabled the discovery of QTL especially related with root deepening and biomass (ASFAW; BLAIR, 2012).

More recently, as several candidate genes have been reported for drought responses in common bean (RECCHIA et al., 2013), studies have attempted association analyses between drought-related traits and specific loci (CORTÉS et al., 2012; ARIANI; GEPTS, 2015; VILLORDO-PINEDA et al., 2015). In this context, *DREB* (*DEHYDRATION RESPONSIVE ELEMENT-BINDING*) genes have proved to be suitable candidates due to their involvement in the regulation of abiotic stress responses in plants (LATA; PRASAD, 2011; MIZOI et al., 2012). *DREB* genes have been extensively characterized in *Arabidopsis* (NAKANO et al., 2006), rice (NAKANO et al., 2006), soybean (ZHANG et al., 2008; MARCOLINO-GOMES et al., 2013) and several other species. Nucleotide polymorphisms have been found associated with specific traits such as with root dry weight and yield under drought and heat stress in chickpea (THUDI et al., 2014), plant biomass in wheat (EDAE et al., 2013), relative water content in foxtail millet (LATA et al., 2011), yield in rice (SELLAMMAL et al., 2013). In common bean, only one work attempted to associate nucleotide patterns of two *DREB* loci with drought-tolerance levels (CORTÉS et al., 2012). However, the analysis was based on the distribution of haplotypes detected for the genes across a panel of wild and domesticated beans with predicted levels of drought tolerance (CORTÉS et al., 2012). Further study provided more details of one of the genes, *PvDREB6B*, showing that the population structure

from areas of northern Mexico to Colombia might have been shaped by geographic and bioclimatic variation, with possible mechanisms of adaptation to the different environments wild beans have developed (Section 6). However, no phenotyping strategy for wild beans and no association mapping study have been performed so far.

Different approaches might be used for association mapping between molecular markers and phenotypic traits. One method employs the compressed mixed linear model (CMLM), developed by Zhang et al. (2010). A mixed linear model considers both fixed and random effects. When individuals are included as random effects, the MLM has the ability to include information about the relationships among individuals. This is revealed by the kinship matrix (K), used as the variance-covariance matrix for individuals. The kinship matrix may be combined with the population structure or with a principal component analysis. The combination improves the statistical power. CMLM decreases the effective population sizes and clusters individuals into groups. It reduces the computation that is usually required by usual MLM-based methods (ZHANG et al., 2010).

The development of this work is justified by the lack of information about phenotypic screening of drought-related traits in wild beans and association studies with candidate genes such as *DREB*. First, we adapted the greenhouse tube assay for screening drought-related traits in a wild bean collection with Mesoamerican accessions, at initial stages of development. Our experimental platform allowed the identification of variation for root deepening and biomass, plant height, leaf area, total biomass and several derived parameters under controlled irrigation setup. Second, we combined the phenotypic data, bioclimatic variables from the areas the wild beans naturally occur and the population structure previously determined for *PvDREB6B* (Section 6) to unravel possible association between the phenotypes and the genetic and environmental variation of the wild beans. Ultimately, we employed a software package implemented on R (GAPIT- Genomic Association and Prediction Integrated Tool) using compressed mixed linear model (CMLM) to find possible associations between *DREB*-related nucleotide polymorphisms and the traits measured through the greenhouse experiment. The association analysis was achieved by the combination of the genomic position of *PvDREB6B* polymorphisms identified in the previous section and the BARCBean6K_3 Beadchip (Illumina Infinium) SNP platform, which was used to genotype the entire wild population. The closest SNP to each of the 54 categorized *DREB* genes for common bean were identified in Section 4 and these positions were checked for association with specific traits of the experiment. However, in this work, we addressed associations to the levels of *DREB*-linked loci only. Association with other loci from the

entire SNP array were performed by J. C. Berny (University of California, Davis) and will be presented in details elsewhere.

7.2 Material and Methods

7.2.1 Plant materials

The experiment performed in this work comprised 132 genotypes. Our main goal was to evaluate the behavior of several traits in wild beans, with some domesticated included, subjected to two irrigation regimes, one keeping soil at field capacity and the other with progressive drought. From the total set of genotypes, 121 were wild accessions of Mesoamerican origin (Supplementary Table 6.1). Seeds were provided by CIAT, Colombia, and the Plant Introduction Station at Pullman, WA. The accessions were originally collected from their locations, distributed from northern Mexico to Colombia. Moreover, 11 domesticated materials were added for a brief comparison between wild and domesticated on the traits evaluated. The domesticated materials included eight parental lines from different market classes within the Mesoamerican gene pool, which were selected to increase genetic diversity, adaptation and productivity in environments affected by drought (SEA-5, Pinto San Rafael, Flor de Mayo Eugenia, SER 118, Matterhorn, UCD 9634, L88, Victor) (BERNY et al., 2014). To this purpose, an 8-way funnel cross has been performed with a modified conical design with all these parental lines (BERNY et al., 2014). The other lines included were BAT 477, Jalo EEP558 and IAC-Carioca 80SH. BAT 477 has been considered a drought-tolerant line while IAC-Carioca 80SH is drought-sensitive (RECCHIA et al., 2013). Jalo EEP558 was the only Andean genotype included as it was reported drought-sensitive (Section 3).

7.2.2 Experimental design and phenotyping

We designed an experiment adapted from the soil screening tube assay proposed by Rao et al. (2006), developed at CIAT for greenhouse evaluations. The experimental system employs transparent plastic tubes placed within PVC tubes with a small diameter and around 1 m of length for evaluation of root and shoot traits during the development of common bean plants. In our case, the experiment was designed in random blocks with three replicates in a greenhouse with controlled temperature (varying between 19°C and 30°C) and air moisture. The physical structure of the experiment consisted of three wire grid fences with dimensions 9.6 x 1.2 m each, sustained at a height of 1 m from the floor. Grid lines dimensions were 10 x 10 cm. Each one of the three blocks was covered in black and white plastic, with the

white surface turned upside for preventing surface heating, while the black surface was turned downside to prevent the roots from being exposed to light. Holes were made on the plastic to place plastic transparent tubes with 7.62 cm diameter and 1.1 m of length. In total, 48 rows of six tubes (288 tubes per block) were designed within each block. The two first rows from each side of the blocks were left as borders of the experiment (west and east side of the experiment). Tubes were filled in with a mixture of top soil and sand (2:1) and placed in all rows of each block. A thin layer of perlite covered all tubes to prevent evaporation.

All tubes were irrigated to field capacity and their initial weight was recorded before seeding. The distribution of the 132 genotypes was randomized within each block. Hence, we designed the experiment in a way to avoid errors in the irrigation events. Since we applied the two irrigation treatments (continued irrigation to field capacity and progressive drought stress), the 6-tube-lines were intercalated for each treatment, e.g. the first row was always irrigated, the second was subjected to drought stress, the third was irrigated and so on. At least two seeds from each accession were placed in opposite sides within individual tubes for both irrigated and drought treatments, totalizing 264 tubes per block. Two rows at each side of the blocks were set as borders for the experiment and seeded with a domesticated cultivar (Tio Canela 75). The experimental system is represented in Figure 7.1.

Before seeding, bamboo sticks were placed within each tube to prevent from causing damage to developing plants. At the time of seeding, all seeds were slightly injured with a scalpel blade on the hilum to accelerate water imbibition. This procedure was adopted to synchronize germination, an event with irregular timing in wild common bean. Due to the size of the experiment (792 tubes) and the time needed to perform measurements, we seeded each one of the three blocks on different days (Block 1: Day 1; Block 2: Day 2; Block 3: Day 2) to allow more precision. After seeding, all tubes were monitored on a daily-basis to account for time for emergence (the shoot was visible over the perlite layer). The number of days to reaching V3 stage was also registered (first trifoliate leaf completely expanded). The number of days for each developmental stage was accounted based on the day that the seeds were placed in each block.

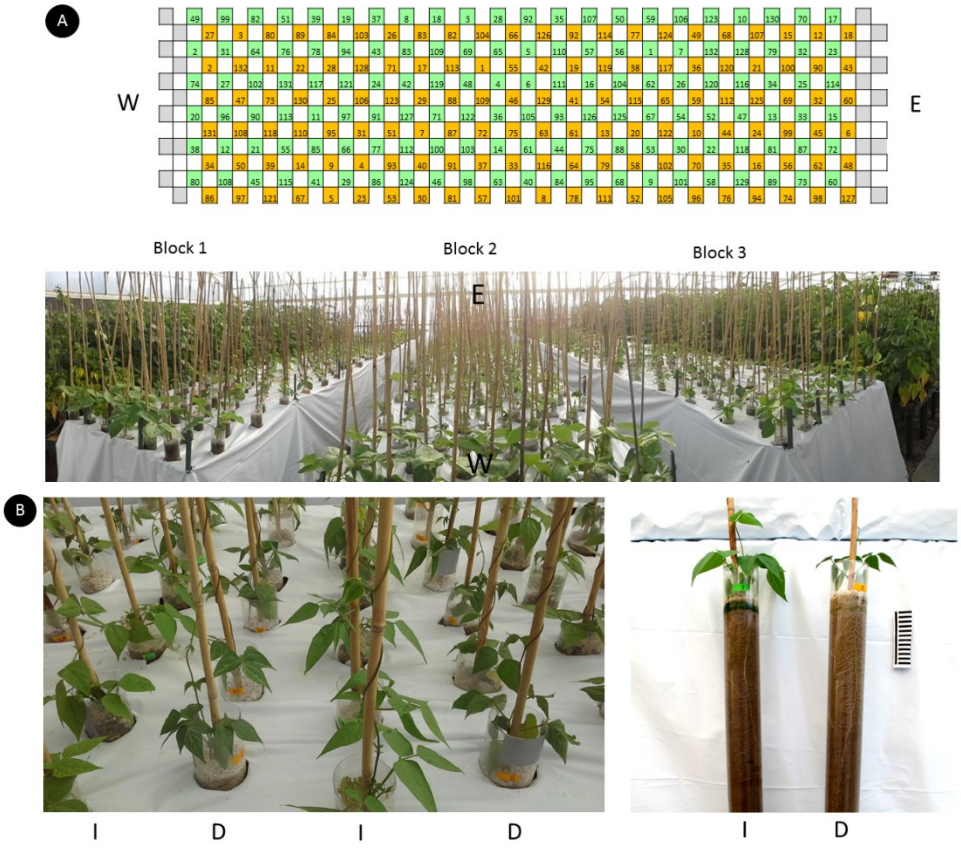


Figure 7.1 – Experimental design for screening drought-related traits in wild common bean accessions and some domesticated genotypes. A – Example of the distribution of experimental units within one of the blocks. Green labels indicate tubes that were constantly irrigated until the end of the experiment. The orange tubes had their irrigation suspended 16 days after planting. The picture below the scheme shows a partial view of the entire experiment. B – Partial view of the experiment showing plants under control (irrigated I) and drought (D) treatments, as well as the shoot and the rooting pattern of one genotype under irrigation and drought. As the majority of plants have indeterminate (climbing) growth habit, bamboo sticks were placed within each tube before the seeding, to avoid causing damage to the developing plants.

All plants were normally irrigated until the first trifoliate leaf started growing (between V2 and V3) in all plants. In general, this was observed for all plants after 16 days after planting, being the day the irrigation treatments started. During 16 days, plants from the control treatment were normally irrigated while the stress-treatment received no irrigation.

At the fourth day, we started to determine root lengths based on the depth visualized around the side of each transparent tube. Block 1 was measured on the fourth day, Block 2 on the fifth and Block 3 on the sixth. The next day was left for other measurements and data cataloguing. From the second to the eighth measurement the same pattern was maintained. During the first three measurements, we were not able to obtain data for all tubes, as some roots remained hidden in the middle of the tube. From the fourth to the eighth events all tubes exhibited visible roots and data were fully recorded. At the same time, plant height was also recorded when all root data were available. Therefore, six root length and plant height

measurements were considered for the whole period. Details of the measurements setup are presented in Supplementary Figure 7.1.

Hereby, we referred the root length as the apparent root length, since it was the one observed on the side of the tube. At harvest, both apparent and actual root length or final root length were determined. The final root length was measured after roots were carefully removed from the tube, occasion when it was possible to observe if roots were positioned in deeper layers in the middle of the tube. A correlation analysis has been performed between apparent root length and final root length to validate the use of the apparent root length while performing such type of experiment. We determined that the apparent measure was able to predict the actual length with a coefficient of determination of 92.3% (Supplementary Figure 7.2).

Between the seventh and the eighth measurements, SPAD measurements were taken with SPAD 502 Plus Chlorophyll Meter. Two measures were taken in alternate leaves (3^{rd} and 5^{th} insertion from the bottom). At harvest, several other parameters were determined. Individual total leaf area was determined following some standards. First, all leaves were taken from individual plants and placed separately over white paper surfaces with a red mark of 4 cm^2. Pictures with a Tab 4 Tablet (Samsung) were taken showing leaves from individual plants and including the red marks. All pictures were accordingly adjusted with several parameters of Easy Leaf Area software (EASLOM; BLOOM, 2014) which balances red and green colors to determine the area of green proportionally to the red surface. After taking pictures, leaves were placed in individual bags for each plant and dried in oven at 37°C for at least seven days. After removing leaves, stem samples were also individually placed in bags and dried. Roots were removed by carefully washing and removing all surrounding soil. All roots samples were dried separately as well. Dry weight was recorded for leaves, stem and roots and the total biomass was determined as the total dry weight. From these data were derived the proportion of leaf dry weight to the aboveground biomass and the proportion of root dry weight to the total biomass.

7.2.3 DNA extraction and genotyping

For DNA extractions, plants from each genotype were grown in pots until they reached V3 stage. Young leaves were harvested, lyophilized and grinded to a fine powder. DNA extractions were performed with an ammonium acetate precipitation method, as previously described (Sections 4 and 6; adapted from PALLOTTA et al., 2003). DNA was quantified with Nanodrop 1000 spectrophotometer. Only high quality samples based on

260/280 nm ratios over 1.8 and electrophoretic pattern on 1% agarose gel were used for genotyping. All samples were diluted to 100 $ng.\mu^{-1}$ and submitted to genotyping with the BARCBean6K_3 Beadchip (Illumina Infinium). The analyses were performed at the Soybean Genomics and Improvement Laboratory, ARS/USDA, in Beltsville, Maryland. This platform comprised 5,398 SNP selected based on previous sequences libraries (HYTEN et al., 2010; BeanCAP). Allele calls were performed with GenomeStudio™ with a non-call treshhold of 0.15 and further refinement considering known heterozygotes (BERNY et al., 2015).

SNP diversity, population and complete linkage disequilibrium analysis are going to be presented elsewhere (BERNY et al., in preparation) for all the accessions from this study. In this work, we used the chip data focusing on the SNP sites located nearby all the 54 *DREB* genes previously identified throughout the common bean genome (Section 4). Additionally, we determined the genomic position of each of the SNP sites encountered within the open reading frame of *PvDREB6B* in the previous study (Section 6) to create hapmap files for association mapping.

7.2.4 Statistics: evaluation of treatment effects and correlation between phenotypic and BIOCLIM data

Statistical significance of the differences between treatments and among genotypes were computed using JMP® version 12.0.1 (SAS Institute Inc.). To the analysis, blocks, treatments, genotypes and the interaction between treatments and genotypes were added to the model. Genotypes were set to analysis with random effect for a single environment. Analysis of variance was performed with a Residual Maximum Likelihood (REML) procedure. The distribution of data of each trait was performed separately for each trait. Correlations among traits were calculated based on Pearson's coefficient. Moreover, correlations were determined between all phenotypic parameters and the average values of several bioclimatic (minimum and maximum temperatures over the year, average precipitation over each month of the year and 19 derived BIOCLIM variables) and geographic (latitude, longitude and altitude) variables to predict the primary variables associated with the phenotypic variability of the wild beans from the experiment.

7.2.5 Univariate and multivariate analyses according to the population structure based on *PvDREB6B* variation

We used the previous information on population structure defined by 33 polymorphic sites within the open reading frame of *PvDREB6B* in a subsample of 112 wild beans to verify

putative differences in phenotypic data among each one of the seven subgroups defined (Section 6). Average values were calculated for each parameter with the 112 genotypes distributed within each one of the clusters (C-1 to C-7). Significance was verified with JMP adding the subgroups as a factor for the analysis. Complementary test of significant differences among average values of each cluster was performed with Tukey's HSD (honest significant difference) test.

Furthermore, a principal component analysis was performed on the correlations among traits, bioclimatic and geographic data. With this analysis, we wanted to verify if the subgroups defined by the *PvDREB6B* haplotypes could be differentiated by particular phenotypes correlated with the environment they occur.

Another step adopted to verify the differences among genetic clusters was the analysis of partial least squares (PLS) based on NIPALS (non-linear iterative partial least squares). NIPALS is an algorithm for computing the first two components in the PSL. This method is a linear predictive model for dependent and independent data sets. In our case, the dependent variables consisted of the drought-related traits and the independent variables comprised the environmental data of all accessions. The PSL produced a number of significant factors (set from 1 to 15) using a cross validation method, that divided the data into groups. The grouping method used was the "leave one out", which fitted the model on n-1 observations and used the one left out for validation (NASH; CHALOUD; LOPEZ, 2005).

7.2.6 Association analysis between phenotypic traits and *DREB* loci

Genotypic data from the SNP array and the polymorphic sites at *PvDREB6B* were converted to Hapmap format. In this format, SNP information (chromosome and location) is displayed in the rows and genotypes in columns. GWAS was performed with GAPIT (Genomic Association and Prediction Integrated Tool), implemented on R software (LIPKA et al., 2012). GAPIT was designed to perform accurate analyses of GWAS using compressed mixed linear model (ZHANG et al., 2010) and the complementary approach 'population parameters previously determined'(P3D), which eliminates the need to re-compute variance components (ZHANG et al., 2010). The MLM model used by the software was $Y = X\beta + Z\mu + e$, where Y is the vector of observed phenotypes, β is an unknown vector with fixed effects (includes the genetic marker, population structure and the intercept); μ is an unknown vector of random additive genetic effects from multiple QTL for individuals; X and Z are the known design matrices; and e is the unobserved vector of residuals (LIPKA et al., 2012). The variances of μ and e accounted for $G = \sigma^2_a K$ with σ^2_a as the additive genetic variance and K the kinship matrix. Instead, homogeneous variance was expected for the residual effect (e).

Principal component analyses (3 components) was performed to control for population structure. Mixed models were fitted to determine the optimum compressions levels as the one whose fitted mixed model yields the largest log likelihood function value (GAPIT documentation). Significant associations were first checked based on *p*-values lower than 0.01. Afterwards, *p-values* were adjusted following a false discovery rate (FDR)-controlling procedure (BEJAMIN; HOCHBERG, 1995).

7.3 Results

7.3.1 Phenotypic measurements

The general mean of each trait and the average values for wilds and domesticates are shown in Table 7.1. The irrigation treatments had distinct effects on each variable measured and the responses varied among genotypes. No significant differences were observed for days to emergence (DTE), since the treatments were started only when plants were reaching V3 stage. Contrarily, treatments affected days to reaching V3 stage (DTV3). An analysis of the distribution of data (Supplementary Figure 7.3) showed control plants took in general from 17 – 22 days to reaching V3 (mean = 19.9 days, Table 7.1), while the dataset for drought-treated plants showed a bimodal distribution with maximum frequencies at 19 and 24 days (mean = 21.3 days, Table 7.1). Among the domesticates, the only Andean line, Jalo EEP558, was the most affected by drought-treatment (DTV3 = 29 days, $p < 0.0001$).

Root length measurements were significantly affected by the treatments, but responses varied among genotypes. The data were distributed in a similar manner in both irrigated and non-irrigated treatments for all measurements (RL4, RL5, RL6, RL7 and RL8) (Supplementary Figure 7.3). This suggested that drought treatment had no unilateral effect on root depth. In fact, the wild accessions PI317349 (RL8 = 67.3 cm in control and 31.8 cm in drought), G12879A (RL8 = 68.6 cm in control and 32.2 cm in drought), G24389 (RL8 = 83.0 cm in control and 37.7 cm in drought) and G24589 (RL8 = 83.8 cm in control and 50.4 cm in drought) revealed the most significant reduction in the final root lengths. Conversely, accessions G12866 (RL8 = 46.8 cm in control and 76.1 cm in drought), G12873 (RL8 = 44.0 cm in control and 72.4 cm in drought) and G23556 (RL8 = 55.9 cm in control and 82.2 cm in drought) exhibited high increase in the final root depth in comparison to the controls. This variation was observed among the domesticated genotypes as well. Thereby, root length was reduced or augmented due to drought treatment and this response depended on the genotype.

Table 7.1 – General mean values of drought-related traits evaluated in a wild common bean collection and some domesticated genotypes in greenhouse conditions

Trait		Control			Drought			
	Abbreviation	Mean	Wilds	Domesticates	Mean	Wilds	Domesticates	*p* (treatments)
Days to emergence	DTE	6.6	6.6	6.1	6.6	6.7	6.0	0.9653 ns
Days to V3 stage	DTV3	19.9	19.9	20.6	21.3	21.0	25.7	< 0.0001 ***
Apparent root length (cm)	ARL	65.6	63.6	85.4	62.5	61.1	76.5	0.0189 *
Final root length (cm)	FRL	68.3	66.2	89.3	65.2	63.6	81.3	0.0216 *
Final plant height (cm)	FPH	113.5	120.4	33.6	45.7	48.2	17.2	< 0.0001 ***
Leaf area (cm2)	ELA	944.1	922.7	1137.2	158.7	157.8	168.4	< 0.0001 ***
Root dry weight (g)	RDW	1.260	1.111	2.841	0.811	0.748	1.488	< 0.0001 ***
Stem dry weight (g)	SDW	1.160	1.106	1.720	0.270	0.257	0.410	< 0.0001 ***
Leaf dry weight (g)	LDW	2.080	1.983	3.099	0.428	0.417	0.539	< 0.0001 ***
Total dry weight (g)	TDW	4.360	4.075	7.341	1.412	1.333	2.273	< 0.0001 ***
Aboveground/shoot biomass (g)	AGB	3.170	3.019	4.672	0.685	0.662	0.921	< 0.0001 ***
Proportion of leaf dry weight to total biomass	LDW/AGB	65.0	65.2	64.4	61.6	62.1	56.2	< 0.0001 ***
Proportion of root dry weight to total biomass	RDW/TDW	26.2	25.3	35.6	47.2	46.5	55.5	< 0.0001 ***
SPAD (Chlorophyll units)	SPAD	31.5	31.0	37.8	43.3	43.1	46.3	< 0.0001 ***
Root length 4th (cm)	RL4	21.7	20.6	32.1	20.5	19.6	30.4	0.2747 ns
Root length 5th (cm)	RL5	28.2	26.7	42.6	26.7	25.7	38.2	0.1266 ns
Root length 6th (cm)	RL6	40.9	39.1	58.2	40.2	39.1	53.3	0.7026 ns
Root length 7th (cm)	RL7	53.0	51.1	72.8	50.5	49.2	64.8	0.0192 *
Root length 8th (cm)	RL8	66.2	64.3	85.4	61.2	59.8	76.5	0.0027 **
Plant height 4th (cm)	PH4	7.0	6.7	9.7	7.5	7.3	9.5	0.0541 ns
Plant height 5th (cm)	PH5	12.7	12.7	10.9	12.0	12.2	10.7	0.3758 ns
Plant height 6th (cm)	PH6	42.2	44.1	17.7	21.0	21.7	13.2	< 0.0001 ***
Plant height 7th (cm)	PH7	77.7	82.5	21.3	32.3	33.9	14.5	< 0.0001 ***
Plant height 8th (cm)	PH8	112.5	118.7	30.4	45.6	48.2	15.5	< 0.0001 ***

Statistical significance: * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.0001$.

Contrarily, plant height was negatively affected by progressive water deficit. Distribution analysis (Supplementary Figure 7.3) clearly indicates frequencies concentrated at the left side of the graphs of the drought-treated plants, with low heights in general. The graphs illustrating the control treatment showed a progressive increase of frequency toward higher plant heights. The PH of the domesticated genotypes figured among the lowest observed in both treatments, as expected for genotypes with low tendency for climbing habit. On the contrary, wilds presented the highest PH values, reaching up to 225 cm in control treatment and 151 cm in drought.

Significant differences among treatments and genotypes were found for leaf area (ELA), root dry weight (RDW), stem dry weight (SDW), leaf dry weight (LDW), total biomass (TDW) and derivatives (LDW/TDW and RDW/TDW) and SPAD. Drought treatment implicated in dramatic reduction of leaf area and biomass components (RDW,

SDW, LDW and TDW) in both wild and domesticates (Table 7.1). As drought led to a decrease in shoot growth, the proportion of the shoot biomass (AGB) to the total biomass also decreased. Although, a significant reduction in root dry weight has been observed due to water deficit, the overall proportion of the root biomass to the total biomass was increased from control (26.2%) to drought (47.2%). Moreover, drought had higher impact on shoot biomass components (77% reduction in stem biomass and 79% reduction in leaf biomass in comparison to the control) than in root biomass (reduction in 36%). During water deficit, plants invested their root growth, being it lateral or to deeper areas within the tube, to find water. SPAD measurements were higher in stressed plants (Table 7.1, Supplementary Figure 7.3), which might be due to lower leaf area and consequently higher chlorophyll concentration.

Correlations among all variables are presented in Supplementary Table 7.1 for control and water deficit. Positive and significant correlations were verified among root depth (all measurements), plant height (all measurements) and biomass components (RDW, SDW, LDW, TDW and derivatives) in both treatments. In most cases, negative correlations were observed between DTE and DTV3 and the other traits. The apparent root length exhibited high positive correlation with the final root length in both control (0.978) and drought (0.975), with a high linear adjustment (r^2 = 0.925, Supplementary Figure 7.2). It validates our method of measuring root length over time based on the visible surface of the tube.

7.3.2 Correlation between traits and environmental variables of the wild beans

Correlation analysis was performed between traits and environmental variables corresponding to the original location of each wild accession and are represented by a heatmap with red colors indicating positive values and blue the negative correlations (Figure 7.1). Although not high, a positive and significant correlation was obtained between root length and latitude; and between plant height and latitude (RL8 = 0.168 and PH8: 0.447 in control; RL8 = 0.197 and PH8 = 0.517 in drought). The positive correlation with the latitude might produce a pattern of association with precipitation regimes as well. Higher latitudes (20 to 30°) encompass drier areas from Mexico. As latitude decreases, precipitation progressively increases, which is confirmed by a negative correlation (-0.621) between the two variables. Interestingly, our data on the phenotypic measurements follow a similar pattern. Precipitation-related variables exhibited negative and significant correlation with root length and plant height. For example, annual precipitation (BIO12) was negatively correlated with final root length (-0.182 in control and -0.212 in drought) and final plant height (-0.399

in control and -0.426 in drought) in both treatments. A negative correlation was also observed among root length, plant height, annual mean temperature (BIO1), mean temperature of the driest quarter (BIO9), mean temperature of the coldest quarter (BIO11), mean temperature of the coldest month (BIO6) and isothermality (BIO3). A positive and significant correlation, however, was observed between plant height and altitude (0.225 in control and 0.184 in drought) and root length and altitude in drought (0.143). Thereby, average root length and plant height tended to be higher for the accessions with original location to higher latitudes and altitudes, which represent areas with lower precipitation and higher seasonality (BIO4).

Biomass components behaved in a similar manner, as negative correlations were observed with precipitation and low temperatures and positive values were encountered with maximum temperatures, seasonality and latitude. Positive and significant correlations were also found among DTV3, annual mean temperature (BIO1), mean temperature of the driest quarter (BIO9), mean temperature of the coldest quarter (BIO11), mean temperature of the coldest month (BIO6), maximum temperature of warmest month (BIO5), mean temperature of the wettest quarter (BIO8), mean temperature of the warmest quarter (BIO10) and the precipitation components. The correlation between DTV3, latitude, altitude and temperature seasonality was negative.

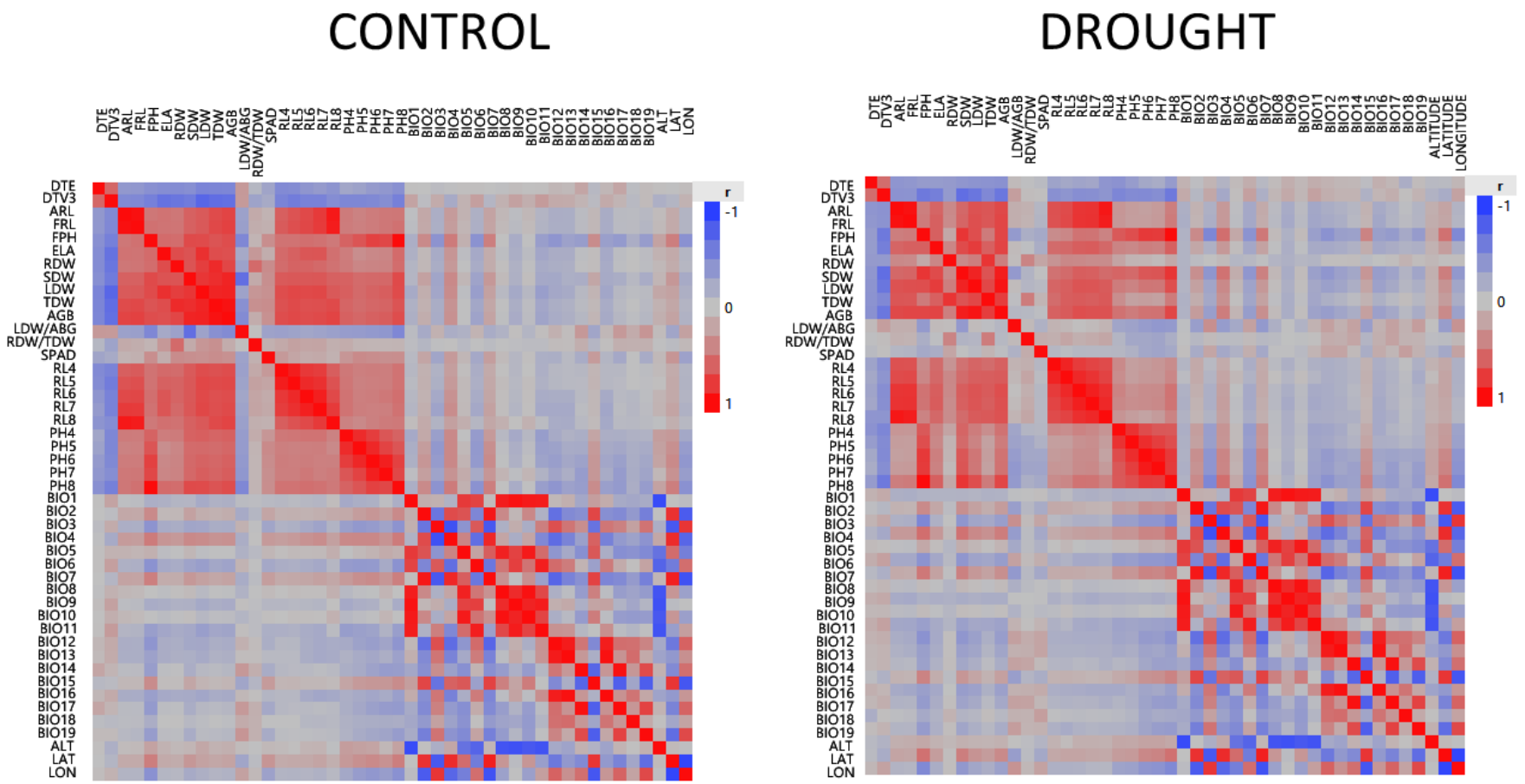


Figure 7.1 – Correlations heat map among drought-related traits of wild common bean (referred on Table 7.1) and environmental variables. BIO1 (Annual mean temperature), BIO2 (Mean diurnal range), BIO3 (Isothermality), BIO4 (Temperature seasonality), BIO5 (Maximum temperature of warmest month), BIO6 (Minimum temperature of coldest month), BIO7 (Temperature annual range), BIO8 (Mean temperature of wettest quarter), BIO9 (Mean temperature of driest quarter), BIO10 (Mean temperature of warmest quarter), BIO11 (Mean temperature of coldest quarter), BIO12 (Annual precipitation), BIO13 (Precipitation of wettest month), BIO14 (Precipitation of driest month), BIO15 (Precipitation seasonality), BIO16 (Precipitation of wettest quarter), BIO17 (Precipitation of driest quarter), BIO18 (Precipitation of warmest quarter), BIO19 (Precipitation of coldest quarter)

7.3.3 Association with population structure defined with *PvDREB6B*

In the previous chapter, we determined the population structure of the wild bean collection based on 33 polymorphic sites within the open reading frame of *PvDREB6B* (Section 6). Hereby, we attempted to find possible association with the variability observed for the traits and the subgroups proposed by the structure analysis of this gene. In Section 6, we defined seven main subgroups based on phylogenetic trees, haplotypes networks, principal coordinate analysis and BAPs population structure. The subgroups were partially associated with latitude, precipitation and temperature variation across the distribution areas of the wild beans. In this study, univariate statistics, principal component analysis (PCA) and partial least squares (PLS) based on NIPALS (non-linear iterative partial least squares) were used to address the association between traits and *PvDREB6B* subgroups.

Univariate analyses were performed with the original dataset converted to least squares values (Table 7.2). Significant differences were observed among the seven *PvDREB6B* clusters for most of the traits under both treatments. In general, C-5 cluster exhibited the highest mean values of root depth (FRL = 73.6 cm) and plant height (FPH = 137.7 cm) under the control treatment and among the highest values (FRL = 73.0 cm; FPH = 60.5 cm) under drought

(Table 7.2). Most accessions of C-5 subgroup are distributed throughout the Mexican state of Durango, which usually has drier climate than the other areas and has been considered the area where the most drought-resistant varieties have been developed. C-2 and C-4, with accessions distributed mainly in Jalisco also presented the highest mean values. The highest biomass under both control (TDW = 5.41 g) and water-stresses plants (TDW = 1.74 g) was verified for the accessions of cluster C-5. Conversely, the lowest final root depths (C-6 = 66.5 cm; C-7 = 63.6 cm, control; C-6 = 63.2 cm; C-7 = 62.8 cm, drought) and final plant heights (C-6 = 107.8; C-7 = 111.0 cm, control; C-6 = 38.7 cm; C-7 = 42.1 cm) were verified for the accessions of subgroups C-6 and C-7 (Table 7.2). Accessions from C-6 and C-7 are distributed in areas with higher annual precipitation and lower latitudes than C-2, C-4 and C-5.

Table 7.2 – Least square means of drought-related traits by genetic clusters defined according to phylogenetic and population structure analyses of *PvDREB6B* haplotypes. In total, data from 112 wild common bean accessions were compiled. DTE (days do emergence); DTV3 (days to V3 stage); ARL (apparent root length); FRL (final root length); ELA (leaf area based on Easy Leaf Area software); RDW (root dry weight); SDW (stem dry weight); LDW (leaf dry weight); TDW (total dry weight); AGB (aboveground biomass); LDW/ABG (proportion of leaf dry weight to the aboveground biomass); RDW/TDW (proportion of root dry weight to the total dry weight); SPAD (unit of measurement for chlorophyll); RL4, RL5, RL6, RL7 and RL8 (root length or depth at the fourth, fifth, sixth, seventh and eight measurements, respectively); PH4, PH5, PH6, PH7 and PH8 (plant height at the fourth, fifth, sixth, seventh and eight measurements, respectively)

	Control							Drought						
Trait	**C-1**	**C-2**	**C-3**	**C-4**	**C-5**	**C-6**	**C-7**	**C-1**	**C-2**	**C-3**	**C-4**	**C-5**	**C-6**	**C-7**
DTE (days)	6.7 a	6.8 a	6.4 a	6.2 a	6.6 a	6.6 a	6.9 a	6.8 a	6.2 a	6.5 a	5.8 a	6.6 a	6.6 a	7.0 a
DTV3 (days)	20.0 ab	19.6 ab	19.1 b	18.9 b	18.7 b	19.9 ab	20.6 a	20.9a	21.2a	20.2ab	18.8b	20.4a	20.7a	21.6a
ARL (cm)	64.0 a	64.5 a	60.6 a	66.6 a	71.5 a	63.7 a	61.4 a	57.4c	63.7bc	55.2c	74.3a	70.2ab	60.7c	60.1c
FRL (cm)	66.4 a	68.2 a	63.6 a	68.4 a	73.6 a	66.5 a	63.6 a	59.4cd	66.5bc	56.1d	78.3a	73.0ab	63.2bcd	62.8cd
FPH (cm)	124.6abc	134.0a	136.4a	129.5ab	137.7a	107.8c	111.0bc	48.6bcd	61.9b	56.1bc	76.5a	60.5b	38.7d	42.1cd
ELA (cm^2)	875.1b	872.7b	943.6b	985.3ab	1214.7a	947.1b	802.2b	185.8ab	165.6abc	140.9c	181.9abc	203.4a	149.7bc	138.0c
RDW (g)	1.06 bc	1.01 bc	1.34 ab	1.21abc	1.49 a	1.08 bc	0.89 c	0.75a	0.57a	0.64a	0.72a	0.90a	0.79a	0.75a
SDW (g)	1.09 bc	1.11 bc	1.11 bc	1.38 ab	1.64 a	1.04 c	0.90 c	0.27bcd	0.29bc	0.27bcd	0.36a	0.33ab	0.23cd	0.22d
LDW (g)	1.92 bc	2.05 bc	2.00 bc	2.19 ab	2.50 a	1.99 bc	1.72 c	0.42bc	0.45abc	0.40c	0.55a	0.51ab	0.41c	0.38c
TDW (g)	4.00bc	3.89bc	4.38abc	4.45 ab	5.41 a	4.10bc	3.37c	1.28bc	1.20c	1.23bc	1.58ab	1.74a	1.34bc	1.25bc
AGB (g)	3.02ab	2.95 b	3.05ab	3.33ab	4.02 a	3.03ab	2.53 b	0.64c	0.73bc	0.66c	0.91a	0.84ab	0.64c	0.59c
LDW/AGB	65 a	66 a	65. a	62 a	62 a	66 a	66 a	61bc	61bc	59c	61bc	61abc	64a	63ab
RDW/TDW	24 a	25 a	27 a	26 a	26 a	25 a	25 a	46a	37a	45a	41a	51a	48a	47a
SPAD	31.0 a	31.8 a	32.1 a	30.7 a	31.4 a	30.6 a	30.3 a	43.3a	41.7a	43.9a	41.1a	42.4a	44.7a	42.7a
RL4 (cm)	19.2b	20.8ab	18.2b	24.0a	24.4 a	20.7ab	18.4b	18.4a	17.8a	19.1a	26.6a	22.9a	19.2a	19.2a
RL5 (cm)	26.0b	25.9b	24.7b	29.3ab	31.8 a	26.8ab	24.5b	24.0bc	23.71c	24.6bc	32.4a	30.3ab	25.9bc	25.1bc
RL6 (cm)	37.6b	41.2ab	35.9b	41.7ab	47.7a	39.6b	36.4b	38.6bc	39.6bc	36.4c	50.1a	44.9ab	38.7bc	37.9bc
RL7 (cm)	49.9bcd	55.1abc	46.3d	57.2 ab	59.7a	51.7abcd	47.6cd	50.2bc	49.9bc	45.8c	62.5a	57.0ab	49.1c	46.8c
RL8 (cm)	66.4 a	68.2 a	63.6 a	68.4 a	73.6 a	66.5 a	63.6 a	59.4c	66.5bc	56.1c	78.3a	73.0ab	63.2bc	62.8c
PH4 (cm)	6.9bc	7.6abc	6.9bc	9.8a	8.9ab	6.3c	5.6c	8.6bc	9.8ab	7.2bc	11.3a	8.3bc	6.4c	5.9c
PH5 (cm)	13.7bc	19.0ab	14.4bc	20.1a	16.1ab	9.8c	9.9c	14.3abc	19.1a	15.1ab	19.8a	12.8bc	10.2bc	9.4c
PH6 (cm)	49.6abc	51.5ab	51.3ab	60.0a	55.2a	36.0c	38.5bc	24.7bcd	33.1ab	27.3abc	37.4a	24.1cd	16.4d	18.1cd
PH7 (cm)	86.2 a	85.3 a	86.7 a	103.4a	97.8 a	76.3 a	76.5 a	37.5bc	46.0ab	43.5ab	57.4a	39.3bcd	26.1d	28.4cd
PH8 (cm)	124.6abc	134.0a	136.4a	129.5ab	137.7a	107.8c	111.0c	48.6bcd	61.9b	56.1bc	76.5a	60.5b	38.7d	42.1cd

Statistical significance: Different letters indicate significant differences by Tukey test ($p < 0.05$).

PCA analysis combined average values of each one of the 24 traits and temperature, precipitation and bioclimatic variables derived (BIO1 to BIO19), latitude, longitude and altitude (Figure 7.2). PCA plots from the drought-treated plants showed more dispersion of data than the control. PCA showed that most of the members of subgroup C-5 belonging to Durango and Jalisco areas were located in one extreme of the graph, while most members from C-6 cluster, from Central America and Colombia, were in the opposite side of each graph (control and drought). Final plant height and final root length were combined along with other variables in one component that explains the distribution of most C-5, C-2, C-3 and C-4 members. As stated before, accessions from these clusters showed the deepest roots and the highest plants. On the contrary, C-6 and C-7 exhibited the lowest values of such parameters.

Some accessions of the C-5 cluster, however, were located at lower latitudes (G19907, G24584, G24589 and G24592) and presented different phenotypic features from the ones of Durango and Jalisco. PCA showed them amongst C-1, C-6 and C-7 clusters. Although they shared haplotypes with the ones from Durango, their phenotypes were compatible with the phenotypes of areas with more humidity. It must be stated that a very limited number of polymorphisms at a single gene were used to determine these clusters, so there are limitation in both genetic and phenotypic analyses.

Further investigation with partial least square analysis based on NIPALS was used to address the differences among subgroups as well. Traits measurements were treated as the Y response, while the X components consisted on the environmental variables. A similar pattern was obtained in comparison to the PCA analysis (Figure 7.2). C-5 accessions from Durango and Jalisco were placed apart from the ones at lower latitudes, which once more were mixed with C-1, C-6 and C-7 members. These analyses reinforce that a latitudinal and consequently a humidity gradient have shaped possible local adaptations of each genotype. It was reflected on their performance in a controlled environment, such as the greenhouse used for the experiment.

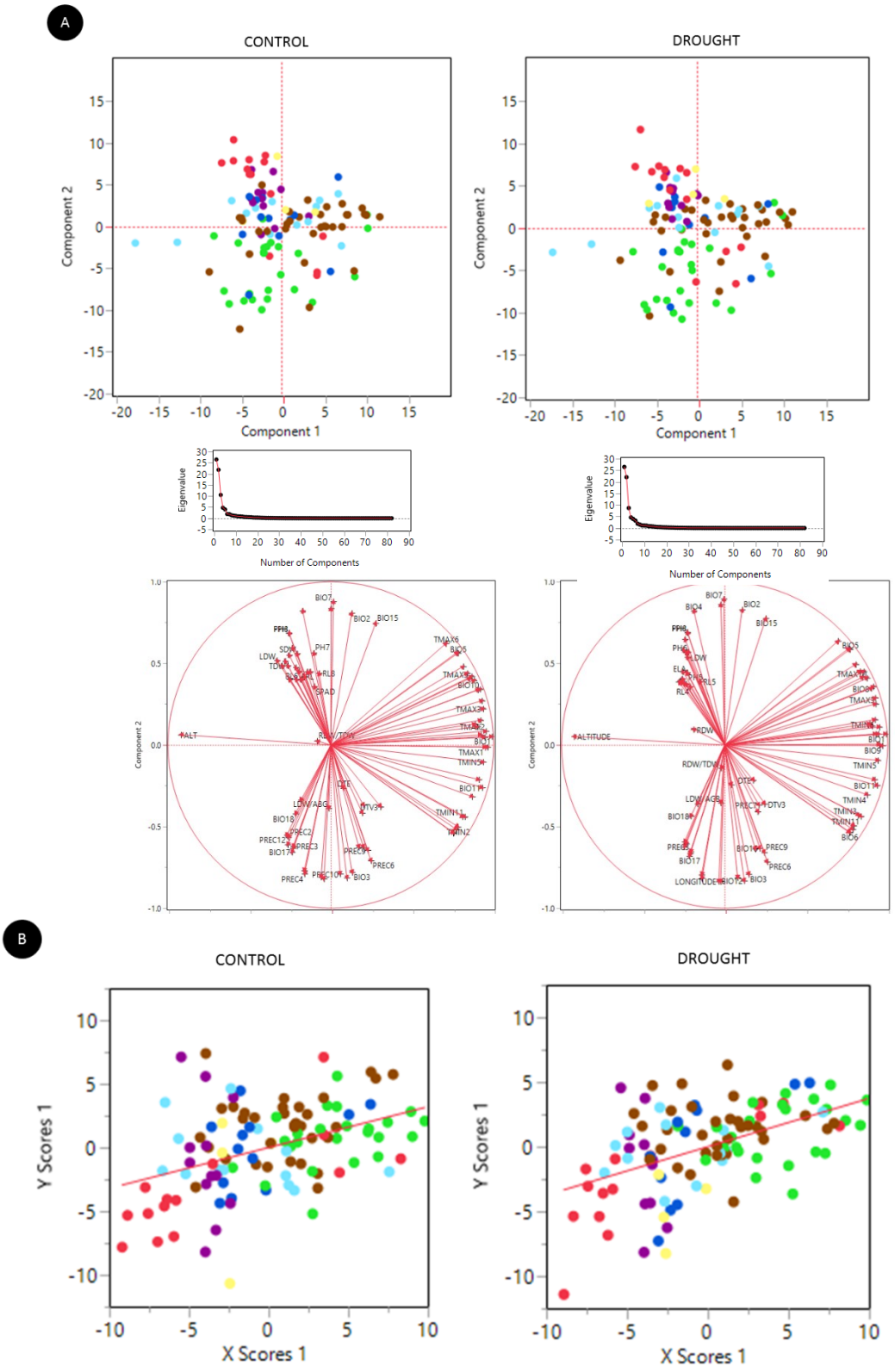


Figure 7.2 – (A) Principal components analysis combining drought-related traits and environmental variables. (B) Partial least squares based on NIPALS plot, with Y as the response component (drought-related traits) and X the environmental variables.

7.3.4 Association analysis

The association analysis using a CMLM provided significant associations between traits and SNP from the array. SNP sites were found associated with root length, plant height and leaf area (FDR < 0.05). Although, in this study we searched for association between *DREB*-linked SNP and the drought-related traits. The results indicated that the snp 923 of *PvDREB6B* was associated with the proportion of leaf biomass to the total biomass in drought (*p-value*: 9 x 10^{-4}) and root dry weight in control (*p-value*: 2.1 x 10^{-4}). Moreover, snp 143 of *PvDREB6B* was found associated with SPAD (*p-value*: 2.4 x 10^{-3}). The marker sc00004ln1947458, located nearby the A-4 gene Phvul.005G170600 was found associated with plant height at the fifth measurement (PH5, *p-value*: 5.1 x 10^{-4}). Although *p*-values indicated some degree of association, the application of the false discovery rate (FDR)-controlling test revealed no significant association. Supplementary Figure 7.4 shows Manhattan plots with significant associations between traits and SNP loci.

7.4 Discussion

The analysis of physiological traits in wild accessions of common bean to attest their drought tolerance levels might be impractical in field conditions. Wild beans have a longer cycle (8 – 10 months) than the cultivated, which would be difficult to follow experimentally due to the variation of temperature and precipitation over the year. Modelling approaches have been described based on bioclimatic variables of the areas such beans occur, defining drought indexes with the combination of such variables (CORTÉS et al., 2013). Nevertheless, in this work we showed the potential application of phenotyping wild beans during a limited time in a controlled greenhouse environment. One of the key aspects for obtaining trustworthy results was the strategy adopted for synchronizing as much as possible the germination time of the accessions. A small hole in the middle of the hilum area with a scalpel blade facilitated water imbibition, and the time to emergence exhibited low variability, averaging 6.7 days. Moreover, plants were submitted to stress-treatments and harvested before reaching flowering, so all measurements were compared within the vegetative period. Genotypes clearly exhibited differences in root and shoot traits under both irrigated and water deficit treatments.

Another innovative aspect of this work relied on the determination of root lengths over the entire period of the experiment. Previous study has shown several root traits, and amongst them was root depth, which showed variability among the lines derived from a population derived from the cross between the cultivated BAT 477 and DOR 364

(ASFAW; BLAIR, 2012). Hereby, we systematically evaluated root length from the beginning to the end, observing roots at the surface of the side of each tube. We were able to observe roots in all tubes from the fourth measurement to the final. At the same time, we followed plant height until the final measurement.

Most of the traits were significantly affected by the imposition of water deficit. Moreover, the analysis of the correlation between traits and environmental variables showed consistent patterns in both irrigated and drought treatments. There was a positive correlation between traits and latitude. Thereby, genotypes from higher latitudes exhibited higher root depth, plant height and biomass than the ones located in southern Mexico, Central America and Colombia. Conversely, there was a negative correlation between most traits and precipitation, especially plant height and root depth. It was intensified by the drought treatment. Similarly, Cortés et al. (2013) verified that wild beans were distributed among different precipitation regimes in a latitudinal gradient.

To reinforce the analysis, we combined the seven genetic clusters obtained from the haplotype diversity analysis of *PvDREB6B* (Section 6). The previous work has shown a partial association between the haplotypes and latitude and precipitation. The average values of each traits obtained for each one of the seven clusters revealed the differences observed with the environmental indicators. Principal component analysis and partial least squares combining both traits and bioclimatic and geographic parameters highlighted the differences among the clusters. Subgroups C-5 and C-6 were shown plotted in opposite sites on PCA and PLS graphs. Most genotypes from C-5 subgroup are originated from Durango state. C-2 and C-3 members were distributed around C-5, as represent genotypes distributed throughout Jalisco state. It is well known that cultivated races derived from the Jalisco-Durango complex possess high drought-tolerance, since they were bred in semi-arid areas of Mexico (DIAZ; BLAIR, 2006). Races Mesoamerica and Guatemala are distributed through low to mid altitudes and generally have higher water requirements for development (CORTÉS et al., 2012).

Our results provide important insights for the selection of desired materials for breeding programs aimed at improving drought tolerance in common bean. Introgression of wilds from Durango area, for example, could contribute for improving performance under drought episodes. The use of *DREB* genes could be useful in directing the selection, since the population structure revealed by the gene *PvDREB6B* helped to understand the distribution of phenotypes throughout variable environments, and enabled the detection of differences between drought-adapted and drought-sensitive materials.

7.5 Conclusions

Hereby, we employed a tube assay system for the evaluation of drought-related traits in a greenhouse experimental setup for both wild and domesticated genotypes. Our results revealed consistent correlation patterns between traits and environmental variables for the wild bean accessions. The genetic structure of *PvDREB6B* was partially associated with the phenotypic variability observed, enabling the detection of differences in plant growth potential among the genetic clusters. In general, genetic subgroups including accessions from drier areas exhibited deeper roots and higher plants and with higher drought-tolerance, while subgroups from areas with higher humidity exhibited less growth and more sensitivity to water deficit. It showed the potential of genetic structure determined through *DREB* genes for indicating differences in stress adaptation. Finally, polymorphisms linked to *DREB* genes were associated with drought-related traits, as revealed by compressed mixed linear model association analysis, but a FDR-test showed a lack of significance.

References

ARIANI, A.; GEPTS, P. Genome-wide identification and characterization of aquaporin gene family in common bean (*Phaseolus vulgaris* L.). **Molecular Genetics and Genomics**, Heidelberg, v. 290, p. 1771-1785, 2015.

ASFAW, A.; BLAIR, M. W.; STRUIK, P. C. Multienvironment Quantitative Trait Loci Analysis for Photosynthate Acquisition, Accumulation, and Remobilization Traits in Common Bean Under Drought Stress. **Genes, Genomes & Genetics**, Bethesda, v. 2, n. 5, p. 579–595, 2012.

ASFAW, A.; BLAIR, M. W. Quantitative trait loci for rooting pattern traits of common beans grown under drought stress versus non-stress conditions. **Molecular Breeding**, Dordrecht, v. 30, n. 2, p. 681–695, 2012.

BEEBE, S. E.; RAO, I. M.; BLAIR, M. W.; ACOSTA-GALLEGOS, J. A. Phenotyping common beans for adaptation to drought. **Frontiers in Physiology**, Lausanne, v. 4, n. 3, p. 1–20, 2013.

BENJAMIN, Y.; HOCKBERG, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. **Journal of the Royal Statistical Society**, London, v. 57, p. 289-300, 1995.

BERNY, J. C.; PALKOVIC, A.; GILVET, M. E.; GEPTS, P. Population development to investigate drought adaptation within the Mesoamerican gene pool of common bean. **Annual Report of the Bean Improvement Cooperative**, Prosser, WA, v. 57, p. 81-82, 2014.

BERNY, J.C.; KONZEN, E. R.; TSAI, S. M.; GEPTS, P. Potential use of the BarcBean6k_3 genotyping platform in the domesticated species of the *Phaseolus* genus. **Annual Report of the Bean Improvement Cooperative.** Prosser, WA, 2015. v. 58.

BLAIR, M. W.; GALEANO, C. H.; TOVAR, E.; TORRES, M. C. M.; CASTRILLÓN, A. V.; BEEBE, S. E.; RAO, I. M. Development of a Mesoamerican intra-genepool genetic map for quantitative trait loci detection in a drought tolerant × susceptible common bean (*Phaseolus vulgaris* L.) cross. **Molecular Breeding**, Dordrecht, v. 29, n. 1, p. 71–88, 2012.

CORTÉS, A. J.; MONSERRATE, F. a.; RAMÍREZ-VILLEGAS, J.; MADRIÑÁN, S.; BLAIR, M. W. Drought tolerance in wild plant populations: the case of common beans (Phaseolus vulgaris L.). Plos One, San Francisco, v. 8, n. 5, e62898, 2013. doi:10.1371/journal.pone.0062898.

CORTÉS, A. J.; THIS, D.; CHAVARRO, C.; MADRIÑÁN, S.; BLAIR, M. W. Nucleotide diversity patterns at the drought-related DREB2 encoding genes in wild and cultivated common bean (*Phaseolus vulgaris* L.). **Theoretical and Applied Genetics**, Heidelberg, v. 125, n. 5, p. 1069–1085, 2012.

DÍAZ, L. M.; BLAIR, M. W. Race structure within the Mesoamerican gene pool of common bean (*Phaseolus vulgaris* L.) as determined by microsatellite markers. **Theoretical and Applied Genetics**, Heidelberg, v. 114, n. 1, p. 143–154, 2006.

EASLON, H. M.; BLOOM, A. J. Easy leaf area: automated digital image analysis for rapid and accurate measurement of leaf area. **Applications in Plant Sciences**, St. Louis, v. 2, n. 7, p. 1400033, 2014. doi: 10.3732/apps.1400033.

EDAE, E. A.; BYRNE, P. F.; MANMATHAN, H.; HALEY, S. D.; MORAGUES, M.; LOPES, M. S.; REYNOLDS, M. P. Association Mapping and Nucleotide Sequence Variation in Five Drought Tolerance Candidate Genes in Spring Wheat. **Plant Genome**, Madison, v. 6, p. 1-13, 2013.

HYTEN, D. L.; SONG, Q.; FICKUS, E. W.; QUIGLEY, C. V.; LIM, J.-S.; CHOI, I.-Y.; HWANG, E.-Y.; PASTOR-CORRALES, M.; CREGAN, P. B. High-throughput SNP discovery and assay development in common bean. **BMC Genomics**, London, v. 11, p. 475, 2010.

LATA, C.; BHUTTY, S.; BAHADUR, R. P.; MAJEE, M.; PRASAD, M. Association of an SNP in a novel DREB2-like gene *SiDREB2* with stress tolerance in foxtail millet [*Setaria italica* (L.)]. **Journal of Experimental Botany**, Oxford, v. 62, n. 10, p. 3387–3401, 2011.

LATA, C.; PRASAD, M. Role of DREBs in regulation of abiotic stress responses in plants. **Journal of Experimental Botany**, Oxford, v. 62, n. 14, p. 4731–4748, 2011.

LIPKA, A. E.; TIAN, F.; WANG, Q.; PEIFFER, J.; LI, M.; BRADBURY, P. J.; GORE, M. a.; BUCKLER, E. S.; ZHANG, Z. GAPIT: Genome association and prediction integrated tool. **Bioinformatics**, Oxford, v. 28, n. 18, p. 2397–2399, 2012.

MARCOLINO-GOMES, J.; RODRIGUES, F. A.; OLIVEIRA, M. C. N.; FARIAS, J. R. B.; NEUMAIER, N.; ABDELNOOR, R. V.; MARCELINO-GUIMARÃES, F. C.;

NEPOMUCENO, A. L. Expression Patterns of GmAP2/EREB-Like Transcription Factors Involved in Soybean Responses to Water Deficit. **Plos One**, San Francisco, v. 8, n. 5, e62294, 2013.

MIZOI, J.; SHINOZAKI, K.; YAMAGUCHI-SHINOZAKI, K. AP2/ERF family transcription factors in plant abiotic stress responses. **Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms**, Amsterdam, v. 1819, n. 2, p. 86–96, 2012.

MUKESHIMANA, G.; BUTARE, L.; CREGAN, P. B.; BLAIR, M. W.; KELLY, J. D. Quantitative trait loci associated with drought tolerance in common bean. **Crop Science**, Madison, v. 54, n. 3, p. 923–938, 2014.

NAKANO, T.; SUZUKI, K.; FUJIMURA, T.; SHINSHI, H. Genome-Wide Analysis of the ERF Gene Family. **Plant Physiology**, Rockville, v. 140, n. 2, p. 411–432, 2006.

NASH, M. S.; CHALOUD, D. J.; LOPEZ, R. D. **Applications of canonical correlation and partial least square analyses in landscape ecology.** Research Triangle Parke: Environmental Sciences Division, USEPA, 2005. 82 p.

PALLOTTA, M. A.; WARNER, P.; FOX, R. L.; KUCHEL, H.; JEFFERIES, S. J.; LANGRIDGE, P. Marker assisted wheat breeding in the southern region of Australia. In: INTERNATIONAL WHEAT GENETICS SYMPOSIUM, 10., 2003, Paestum, Italy. **Proceedings…** Sydney: Sydney University Press, 2003. p. 1-6.

RAO, I. M.; POLANIA, J.; GARCIA, R.; BEEBE, S. Development of greenhouse soil tube method to quantify phenotypic differences among advanced lines in root development and distribution under drought stress. In: CENTRO INTERNACIONAL DE AGRICULTURA TROPICAL (Ed.). Annual Report 2006. ProjectIP-1: Bean Improvement for the Tropics. Cali, Colombia: CIAT, 2006. p. 19–25.

RECCHIA, G. H.; CALDAS, D. G. G.; BERALDO, A. L. A.; DA SILVA, M. J.; TSAI, S. M. Transcriptional analysis of drought-induced genes in the roots of a tolerant genotype of the common bean (*Phaseolus vulgaris* L.). **International Journal of Molecular Sciences**, Basel, v. 14, n. 4, p. 7155–7179, 2013.

SELLAMMAP, R.; VINOD, S. R. K. K.; RAVEENDRAN, M. Marker-trait association for physiological traits associated with drought resistance in rice under different water regimes. **Madras Agriculture Journal**, Coimbatore, v. 100, n. 9, p. 647–651, 2013.

THUDI, M.; UPADHYAYA, H. D.; RATHORE, A.; GAUR, P. M.; KRISHNAMURTHY, L.; ROORKIWAL, M.; NAYAK, S. N.; CHATURVEDI, S. K.; BASU, P. S.; GANGARAO, N. V. P. R.; FIKRE, A.; KIMURTO, P.; SHARMA, P. C.; SHESHASHAYEE, M. S.; TOBITA, S.; KASHIWAGI, J.; ITO, O.; KILLIAN, A. Genetic dissection of drought and heat tolerance in chickpea through genome-wide and candidate gene- based association mapping approaches. **Plos One**, San Francisco, v. 9, n. 5, p. 40–50, 2014.

VILLORDO-PINEDA, E.; GONZÁLEZ-CHAVIRA, M. M.; GIRALDO-CARBAJO, P.; ACOSTA-GALLEGOS, J. A.; CABALLERO-PÉREZ, J. Identification of novel drought-tolerant-associated SNPs in common bean (*Phaseolus vulgaris*). **Frontiers in Plant Science**, Lausanne, v. 6, n. 7, p. 1–9, 2015.

ZHANG, G.; CHEN, M.; CHEN, X.; XU, Z.; GUAN, S.; LI, L. C.; LI, A.; GUO, J.; MAO, L.; MA, Y. Phylogeny, gene structures, and expression patterns of the ERF gene family in soybean (*Glycine max* L.). **Journal of Experimental Botany**, Oxford, v. 59, n. 15, p. 4095–4107, 2008.

ZHANG, Z.; ERSOZ, E.; LAI, C.-Q.; TODHUNTER, R. J.; TIWARI, H. K.; GORE, M. A.; BRADBURY, P. J.; YU, J.; ARNETT, D. K.; ORDOVAS, J. M.; BUCKLER, E. S. Mixed linear model approach adapted for genome-wide association studies. **Nature Genetics**, New York, v. 42, n. 4, p. 355–360, 2010.

8 GENERAL CONCLUSIONS

This thesis presented several results concerning the identification of *DREB* genes in common bean, their expression profiling, nucleotide polymorphisms, correlation with environmental variables and association to phenotypic traits.

In Section 3, we provided a genome-wide catalogue of the *DREB* gene subfamily in common bean, based on *in silico* analyzes with sequences deposited on Phytozome. In total, 54 *PvDREB* were identified, being distributed along the 11 chromosomes of *P. vulgaris*. Four novel genes were cloned and their expression profile was addressed under four stress-treatments (dehydration, salinity, low-temperature and abscisic acid). *PvDREB1F* and *PvDREB5A* were induced by dehydration, salinity and low-temperature while *PvDREB2A* and *PvDREB6B* were mainly induced by dehydration and low-temperature. However, the expression was differentially modulated depending on the plant organ, period of stress and genotype under investigation. We verified that *PvDREB6B* was identical to previous gene published in literature as an A-2 gene, but our phylogenetic trees supported the categorization of this locus within the A-6 subgroup.

Section 4 was dedicated to present nucleotide polymorphisms within intron, UTR and coding regions of the four novel *PvDREB* under investigation. Among all loci, *PvDREB6B* showed the highest number of polymorphic sites between of set of Mesoamerican and Andean genotypes. Moreover, we identified SNP markers nearby all the 54 *PvDREB* previously identified through the BARCBean6K_3 Bead Chip platform. In total, 51 non-redundant SNP were identified, being able to distinguish between Mesoamerican and Andean gene pools.

In Section 5, we addressed a technical question about the interference of nucleotide polymorphisms in the analysis of gene expression profiles, using the case of *PvDREB6B*. The high number of polymorphisms within the gene led to interpretation errors of relative expression values between Andean and Mesoamerican genotypes, depending on the pairs of primers that were used. It is necessary to have prior knowledge of the polymorphisms among genotypes when designing a pair of primers aimed at expression studies.

In Section 6, we further evaluated *PvDREB6B* in a wild bean collection of Mesoamerican origin. We expanded the previous study of Cortés and collaborators by sequencing the entire open reading frame of the gene (957 bp), revealing higher number of polymorphisms (33 sites). The analysis of haplotype networks and population structure based on *PvDREB6B* variation revealed seven main clusters of wild accessions from northern Mexico to Colombia. An association analysis between nucleotide polymorphisms and

geographic and bioclimatic variables revealed that the DNA variation is partially correlated with latitude, and consequently with a humidity gradient from drier areas of northern Mexico to humid climates in Central America and Colombia. Moreover, some association degree was verified between DNA variants and the variation in temperature over the year.

Ultimately, in Section 7, we proposed and adaptation of a tube screening method for phenotyping wild beans and some domesticated in greenhouse conditions. The wild accessions analyzed in the previous section were evaluated under irrigated and water deficit treatments during vegetative developmental stages. Several traits were recorded, especially the gradual increase in root depth and plant height. We verified that the genotypes originated from the driest areas presented the highest root depth, plant height and biomass under both irrigated and drought treatments. Furthermore, we showed the phenotypes were correlated with environmental data (geographic and bioclimatic) of the original location of the wild accessions. As a result, the population structure determined based on *PvDREB6B* variants was able to distinguish between the most drought-tolerant and the drought-sensitive accessions. Association mapping analyses between traits and SNP from *PvDREB6B* and the SNP chip revealed some SNP possibly associated with phenotypic variation, but a false discovery rate test showed no significant association between *DREB*-linked SNP and the traits.

In general, this work provided insights for the use of *DREB* genes in improving abiotic stress tolerance in common bean and how to rescue sources of tolerance from wild accessions. This might be helpful in designing breeding strategies from improving commercial varieties, especially for drought tolerance, the most important issue in Brazil and several other countries. We must emphasize, however, the quantitative and complex inheritance of stress tolerance relies on several gene networks, and so further work with *DREB* genes should also address their interaction with other genes under the different approaches one researcher might follow: engineering genes for stress tolerance; fine-mapping of genes associated with stress responses and specific traits; marker-assisted selection combining *DREB*-linked markers and of other genes.

APPENDICES

Appendix A – Supplementary Files from Section 3

Supplementary Table 3.1 - BLASTp annotation file of all AP2-containing sequences from common bean. All protein sequences were retrieved from Phytozome.

	Phytozome number ID	AP2 domains	BLASTp	E-value	Putative DREB
1	Phvul.010G092300.1	1 AP2	ERF4-like protein soybean	1.00E-170	
2	Phvul.010G054000.1	**1 AP2**	**ERF034-like soybean**	**2.00E-124**	**X**
3	Phvul.010G159500.1	1 AP2	ERF003-like soybean	8.00E-87	
4	Phvul.010G050500.1	1 AP2	ERF13-like soybean	6.00E-89	
5	Phvul.010G146600.1	**1 AP2**	**DREB3-like soybean**	**1.00E-92**	**X**
6	Phvul.010G158700.1	1 AP2	AP2/ERF domain containing Populus trichocarpa	6.00E-40	
7	Phvul.010G050600.1	1 AP2	ERF2-like soybean	6.00E-79	
8	Phvul.010G130200.1	2 AP2	ERF AIL-like soybean	0.00E+00	
9	Phvul.003G223600.1	**1 AP2**	**ERF025-like soybean**	**3.00E-84**	**X**
10	Phvul.003G180000.1	1 AP2	ERF Medicago truncatula	1.00E-83	
11	Phvul.003G111800.1	AP2, B3	TEM1-like soybean	0.00E+00	
12	Phvul.003G165000.1	AP2	WIN1-like soybean	2.00E-130	
13	Phvul.009G084400.1	**AP2**	**DREB3 soybean**	**1.00E-38**	**X**
14	Phvul.009G206500.1	2 AP2	AP2-containing soybean	0.00E+00	
15	Phvul.009G240800.1	1 AP2	ERF098-like soybean	2.00E-64	
16	Phvul.009G262200.1	1 AP2	EREBP/AP2 soybean	0.00E+00	
17	Phvul.009G225000.1	**1 AP2**	**ERF017-like soybean**	**1.00E-111**	**X**
18	Phvul.009G029600.1	**1 AP2**	**Predicted RAP2.4 soybean**	**7.00E-160**	**X**
19	Phvul.009G161000.1	1 AP2	ERF114-like soybean	4.00E-104	
20	Phvul.009G109600.1	**1 AP2**	**DREB3-like soybean**	**2.00E-97**	**X**
21	Phvul.009G089300.1	1 AP2	PPLZ02-like soybean	4.00E-113	
22	Phvul.009G093600.1	1 AP2	SHINE 3-like soybean	7.00E-98	
23	Phvul.005G180600.1	1 AP2	ERF003-like soybean	1.00E-94	
24	Phvul.005G111200.1	**1 AP2**	**DREBa- soybean**	**5.00E-82**	**X**
25	Phvul.005G126300.1	**1 AP2**	**DREB1F - soybean**	**2.00E-98**	**X**
26	Phvul.005G170600.1	**1 AP2**	**DREB3-like soybean**	**5.00E-106**	**X**
27	Phvul.005G138300.1	2 AP2	RAP2.7	0.00E+00	
28	Phvul.005G126600.1	**1 AP2**	**DREB1D-like soybean**	**5.00E-93**	**X**
29	Phvul.011G091400.1	**1 AP2**	**DREB1F-like soybean**	**9.00E-117**	**X**
30	Phvul.011G187400.1	1 AP2	wrinkled-1 soybean	0.00E+00	
31	Phvul.011G125200.1	1 AP2	preditec ERF-9 soybean	2.00E-77	
32	Phvul.011G071100.1	2 AP2	RAP2.7-like soybean	0.00E+00	
33	Phvul.008G222400.1	**1 AP2**	**SGR soybean**	**1.00E-91**	**X**
34	Phvul.008G172200.1	**1 AP2**	**RAP.2.4 soybean**	**2.00E-135**	**X**
35	Phvul.008G246000.1	1 AP2	PTI6-like soybean	9.00E-130	
36	Phvul.008G131500.1	1 AP2	predited ERF- SOYBEAN	9.00E-123	
37	Phvul.008G141000.1	**1 AP2**	**ERF034 soybean**	**1.00E-102**	**X**
38	Phvul.008G046400.1	1 AP2	ERF2-like soybean	3.00E-100	
39	Phvul.008G214700.1	1 AP2	ERF3-like soybean	2.00E-77	
40	Phvul.008G220400.1	**1 AP2**	**DREB2A Vigna unguiculata**	**0.00E+00**	**X**
41	Phvul.008G159400.1	1 AP2	CRF3-like soybean	3.00E-136	
42	Phvul.008G039300.1	1 AP2	ERF Medicago truncatula	2.00E-148	
43	Phvul.008G043500.1	2 AP2	WRI1-like soybean	0.00E+00	
44	Phvul.008G092800.1	**1 AP2**	**DRE-binding protein soybean**	**0.00E+00**	**X**
45	Phvul.008G019600.1	1 AP2	ERF113-like soybean	3.00E-80	
46	Phvul.008G052000.1	2 AP2	AP2 BBM2-like soybean	0.00E+00	
47	Phvul.004G069900.1	1 AP2	ERF RAP2-3 soybean	4.00E-84	
48	Phvul.004G092100.1	1 AP2	AP2 EREBP Lotus japonicus	2.00E-70	
49	Phvul.004G169800.1	**1 AP2**	**AP2 DREB3 soybean**	**1.00E-104**	**X**
50	Phvul.004G031900.1	1 AP2	ERF Medicago truncatula	0.00E+00	
51	Phvul.004G081200.1	1 AP2	RAP2.11 Medicago truncatula	8.00E-82	
52	Phvul.004G122000.1	**1 AP2**	**DREB1-like soybean**	**1.00E-123**	**X**
53	Phvul.007G217800.1	1 AP2	ERF protein soybean	4.00E-180	
54	Phvul.007G102800.1	AP2, B3	TEM1-like soybean	0.00E+00	
55	Phvul.007G241600.1	1 AP2	ERF5 soybean	3.00E-120	
56	Phvul.007G193300.1	1 AP2	ERF1B-like soybean	2.00E-110	

57	Phvul.007G193700.1	1 AP2	ERF1B-like soybean	9.00E-112	
58	Phvul.007G222500.1	**1 AP2**	**DREB1F-like soybean**	**4.00E-93**	**X**
59	Phvul.007G272900.1	1 AP2	ERF098-like soybean	3.00E-58	
60	Phvul.007G128100.1	1 AP2	ERF098-like soybean	2.00E-61	
61	Phvul.007G135900.1	1 AP2	ERF Medicago truncatula	1.00E-111	
62	Phvul.007G255100.1	**1 AP2**	**DREB2F-like soybean**	**7.00E-161**	**X**
63	Phvul.007G272800.1	1 AP2	ERF098-like soybean	6.00E-49	
64	Phvul.007G240200.1	1 AP2	RAP2-7	0.00E+00	
65	Phvul.007G121200.1	1 AP2	AP2/ERF domain containing Populus trichocarpa	2.00E-43	
66	Phvul.007G193900.1	1 AP2	ERF098-like soybean	0.00E+00	
67	Phvul.007G135300.1	**1 AP2**	**ERF062-like soybean**	**0.00E+00**	**X**
68	Phvul.007G127800.1	1 AP2	ERF1B-like soybean	3.00E-102	
69	Phvul.001G111800.1	1 AP2	RAP2.11-like soybean	3.00E-156	
70	Phvul.001G010400.1	**1 AP2**	**DREB2B-like soybean**	**2.00E-68**	**X**
71	Phvul.001G099700.1	1 AP2	LEP-like soybean	7.00E-83	
72	Phvul.001G160400.1	1 AP2	ERF098-like soybean	7.00E-56	
73	Phvul.001G131300.1	superfamília AP2	AP2-like ERF soybean	0.00E+00	
74	Phvul.001G137500.1	2 AP2	PLT2-like soybean	5.00E-37	
75	Phvul.001G084000.1	1 AP2	ERF113-like soybean	2.00E-84	
76	Phvul.001G174400.1	2 AP2	Floral homeotic protein AP2 soybean	0.00E+00	
77	Phvul.006G173500.1	1 AP2	AP2-containing Populus tricocarpa	5.00E-157	
78	Phvul.006G183100.1	1 AP2	ERF4-like soybean	3.00E-102	
79	Phvul.006G163100.1	1 AP2	ERF086-like soybean	2.00E-110	
80	Phvul.006G179800.1	1 AP2	PTI5-like soybean	3.00E-99	
81	Phvul.006G047800.1	1 AP2	ERF3-like soybean	9.00E-98	
82	Phvul.006G110200.1	1 AP2	CRF4-like	3.00E-94	
83	Phvul.006G179700.1	1 AP2	ERF091-like soybean	6.00E-159	
84	Phvul.002G016900.1	2 AP2	Floral homeotic protein AP2 soybean	0.00E+00	
85	Phvul.002G162500.1	1 AP2	CRF2-like	6.00E-131	
86	Phvul.002G310200.1	**1 AP2**	**DREB3-like soybean**	**9.00E-100**	**X**
87	Phvul.002G009100.1	2 AP2	ANT-like soybean	0.00E+00	
88	Phvul.002G237300.1	1 AP2	ERF015-like soybean	1.00E-113	
89	Phvul.002G036000.1	**1 AP2**	**ERF021-like soybean**	**7.00E-83**	**X**
90	Phvul.002G281300.1	1 AP2	ERF114-like soybean	1.00E-75	
91	Phvul.002G035100.1	**1 AP2**	**ERF020-like soybean**	**7.00E-68**	**X**
92	Phvul.002G055800.1	1 AP2	ERF5-like soybean	1.00E-120	
93	Phvul.002G016700.1	**1 AP2**	**DREB1-like soybean**	**3.00E-77**	**X**
94	Phvul.002G056800.1	**1 AP2**	**DREB3-like soybean**	**1.00E-62**	**X**
95	Phvul.002G149500.1	1 AP2	ESR2-like	4.00E-174	
96	Phvul.002G154000.1	**1 AP2**	**ERF025-like soybean**	**9.00E-70**	**X**
97	Phvul.002G267800.1	1 AP2	ERF Medicago truncatula	2.00E-55	
98	Phvul.002G315900.1	1 AP2	ERF Medicago truncatula	6.00E-21	
99	Phvul.002G153900.1	**1 AP2**	**DREB1F-like soybean**	**4.00E-93**	**X**
100	Phvul.002G254500.1	**1 AP2**	**ERF060-like soybean, DREB em Populus**	**2.00E-140**	**X**
101	Phvul.010G124700.1	1 AP2	ERF soybean	0.00E+00	
102	Phvul.010G035200.1	1 AP2	CRF4-like soybean	2.00E-105	
103	Phvul.010G050700.1	1 AP2	ERF2-like soybean	2.00E-74	
104	Phvul.010G050800.1	1 AP2	ERF13-like soybean	1.00E-62	
105	Phvul.010G154900.1	1 AP2	WIN1-like soybean	2.00E-110	
106	Phvul.010G114900.1	**1 AP2**	**ERF023-like soybean**	**2.00E-87**	**X**
107	Phvul.003G241900.1	2 AP2	Floral homeotic protein AP2 soybean	0.00E+00	
108	Phvul.003G241700.1	**1 AP2**	**ERF008-like**	**7.00E-63**	**X**
109	Phvul.003G102500.1	1 AP2	ERF086-like soybean	0.00E+00	
110	Phvul.003G288500.1	2 AP2	PLT2-like soybean	0.00E+00	
111	Phvul.003G035500.1	1 AP2	ERF118-like soybean	4.00E-93	
112	Phvul.003G064800.1	1 AP2	ESR2-like	4.00E-88	
113	Phvul.003G212800.1	**1 AP2**	**DREB1F-like soybean**	**3.00E-130**	**X**
114	Phvul.003G153100.1	2 AP2	AP2-ERF factor soybean	0.00E+00	
115	Phvul.003G069000.1	2 AP2	ANT-like soybean	0.00E+00	
116	Phvul.003G292400.1	1 AP2	ERF109-like soybean	7.00E-111	
117	Phvul.003G222600.1	**1 AP2**	**DREB6-like soybean**	**1.00E-82**	**X**
118	Phvul.003G107900.1	4 AP2	RAP2-3 like soybean	4.00E-35	
119	Phvul.003G232600.1	2 AP2	ANT-like soybean	0.00E+00	
120	Phvul.003G212700.1	**1 AP2**	**ERF026-like soybean**	**5.00E-80**	**X**

121	Phvul.003G144500.1	2 AP2	AIL5-like soybean	0.00E+00	
122	Phvul.009G240900.1	1 AP2	ERF098-like soybean	1.00E-61	
123	Phvul.009G013200.1	**1 AP2**	**DREB2D soybean**	**4.00E-115**	**X**
124	Phvul.009G196900.1	1 AP2	WIN1-like soybean	1.00E-108	
125	Phvul.009G074300.1	2 AP2	ANT-like soybean	0.00E+00	
126	Phvul.009G251600.1	1 AP2	RAP2-3 like soybean	6.00E-95	
127	Phvul.009G137000.1	1 AP2	CRF4-like soybean	0.00E+00	
128	Phvul.009G137900.1	1 AP2	ERF109-like soybean	4.00E-96	
129	Phvul.009G123300.1	**1 AP2**	**ERF012-like soybean**	**3.00E-101**	**X**
130	Phvul.005G128200.1	Adaptin-N	AP-2 complex subunit alpha-2-like soybean	0.00E+00	
131	Phvul.005G105200.1	**1 AP2**	**ERF054-like soybean, DREB5 soybean**	**0.00E+00**	**X**
132	Phvul.005G183700.1	1 AP2	AP2 domain containing Populus trichocarpa	3.00E-124	
133	Phvul.005G074700.1	1 AP2	ERF4-like soybean	2.00E-85	
134	Phvul.L006500.1	1 AP2	SHINE 3-like soybean	5.00E-94	
135	Phvul.011G058000.1	2 AP2	PLT2-like soybean	0.00E+00	
136	Phvul.011G118600.1	**1 AP2**	**ERF054-like soybean, DREB (Malus domestica)**	**0.00E+00**	**X**
137	Phvul.011G162900.1	1 AP2	ERF4-like soybean	9.00E-75	
138	Phvul.011G107800.1	**1 AP2**	**DREB2A-like soybean**	**5.00E-98**	**X**
139	Phvul.008G271100.1	1 AP2	ERF3-like soybean	7.00E-74	
140	Phvul.008G253600.1	1 AP2	AP2 ERF soybean	5.00E-175	
141	Phvul.008G185400.1	2 AP2	Floral homeotic protein AP2 soybean	9.00E-119	
142	Phvul.008G098900.1	**1 AP2**	**DREB2 soybean**	**4.00E-87**	**X**
143	Phvul.008G165000.1	**1 AP2**	**ERF012-like soybean**	**1.00E-98**	**X**
144	Phvul.008G046500.1	1 AP2	ERF2-like soybean	3.00E-66	
145	Phvul.004G157600.1	1 AP2	CRF4-like soybean	1.00E-39	
146	Phvul.004G068900.1	1 AP2	LEP-like soybean	5.00E-147	
147	Phvul.007G193800.1	1 AP2	ERF1B-like soybean	5.00E-104	
148	Phvul.007G273000.1	1 AP2	ERF1B-like soybean	1.00E-80	
149	Phvul.007G222600.1	**1 AP2**	**ERF024-like soybean**	**4.00E-96**	**X**
150	Phvul.007G128000.1	1 AP2	ERF098-like soybean	3.00E-78	
151	Phvul.007G241800.1	1 AP2	ERF8 Medicago sativa	2.00E-118	
152	Phvul.007G002900.1	AP2, B3	AP2/ERF domain containing Medicago truncatula	3.00E-116	
153	Phvul.007G082000.1	1 AP2	RAP2.6 - like soybean	3.00E-64	
154	Phvul.007G066500.1	**1 AP2**	**DREB7-like soybean**	**2.00E-98**	**X**
155	Phvul.007G193400.1	1 AP2	ERF098-like soybean	4.00E-45	
156	Phvul.007G086600.1	1 AP2	ERF098-like soybean	8.00E-71	
157	Phvul.007G027000.1	1 AP2	ERF118-like soybean	5.00E-165	
158	Phvul.001G160300.1	1 AP2	ERF1B-like soybean	8.00E-71	
159	Phvul.001G136100.1	**1 AP2**	**DREB2F-like soybean**	**1.00E-175**	**X**
160	Phvul.001G160500.1	1 AP2	ERF15-like soybean	2.00E-101	
161	Phvul.001G073800.1	**1 AP2**	**TINY soybean**	**4.00E-100**	**X**
162	Phvul.001G046900.1	1 AP2	ERF003-like soybean	2.00E-114	
163	Phvul.001G160200.1	1 AP2	ERF1B-like soybean	3.00E-100	
164	Phvul.001G251200.1	**1 AP2**	**ERF060-like soybean, DREB em** *Trifolium repens*	**3.00E-137**	**X**
165	Phvul.001G023700.1	**1 AP2**	**DREB protein soybean**	**9.00E-88**	**X**
166	Phvul.001G160100.1	1 AP2	ERF098-like soybean	6.00E-54	
167	Phvul.001G044500.1	**1 AP2**	**RAP2.9-like soybean**	**4.00E-49**	**X**
168	Phvul.001G157600.1	1 AP2	ERF8-like soybean	1.00E-63	
169	Phvul.001G137600.1	1 AP2	PLT2 *Arabidopsis thaliana*	5.00E-21	
170	Phvul.001G187100.1	**1 AP2**	**ERF024-like soybean**	**2.00E-112**	**X**
171	Phvul.001G031200.1	2 AP2	ANT-like soybean	0.00E+00	
172	Phvul.006G080000.1	1 AP2	PTI6-like soybean	7.00E-93	
173	Phvul.006G114100.1	**1 AP2**	**ERF023-like soybean**	**2.00E-96**	**X**
174	Phvul.006G183200.1	1 AP2	ERF12-like soybean	2.00E-90	
175	Phvul.006G106100.1	1 AP2	ERF1B-like soybean	5.00E-71	
176	Phvul.002G035900.1	**1 AP2**	**ERF021-like soybean**	**3.00E-111**	**X**
177	Phvul.002G295700.1	1 AP2	CRF2-like	2.00E-121	
178	Phvul.002G055700.1	1 AP2	ERF1A-LIKE	3.00E-141	
179	Phvul.002G168900.1	1 AP2	AP2/ERF Populus trichocarpa	2.00E-59	
180	Phvul.002G146400.1	2 AP2	AIL6-like soybean	0.00E+00	
181	Phvul.002G163700.1	**1 AP2**	**ERF061-like soybean**	**3.00E-145**	**X**

Supplementary Table 3.2 – Primers designed for *PvDREB* cloning and RT-qPCR analyses.

Gene	Phytozome accession	Cloning primers (5' - 3')	Expected size (bp)	qRT-PCR primers (5'- 3')	Expected size (bp)
PvDREB1F	Phvulv091025959m.g	F: ATGAGCACAAAATCAGTTGTAG	882	F: TGCGTCGAGCAATTAGAGAA	153
		R: TCAGATGGAGAAGCTCCA		R: TCCTGATGCGTCTGGTATTG	
PvDREB2A	Phvulv091010045m.g	F: ATGTTAGTGAAAAACCACAACA	600	F: TTGGGTACTTTTCCCACTGC	229
		R: TCAAGACAATGAAAGATGGG		R: GCCTTCCATGTCATCATCCT	
PvDREB5A	Phvulv091010162m.g	F: ATGGAAGGAGAAGGTTTAGGAG	462	F: GATGAGGAAGTGGGGGAAGT	177
		R: CTAGTCTTCGGGTTTAGGA		R: TCGTCTTGGGAGAGCAGTTC	
PvDREB6B	Phvulv091016691m.g	F: ATGGGAACCGCTATAGATATGT	957	F: GCTACCTTCTTCCGTGGACT	227
		R: TCATATAGCCGCCCAATC		R: CAGACACAGGCCATGACAAC	

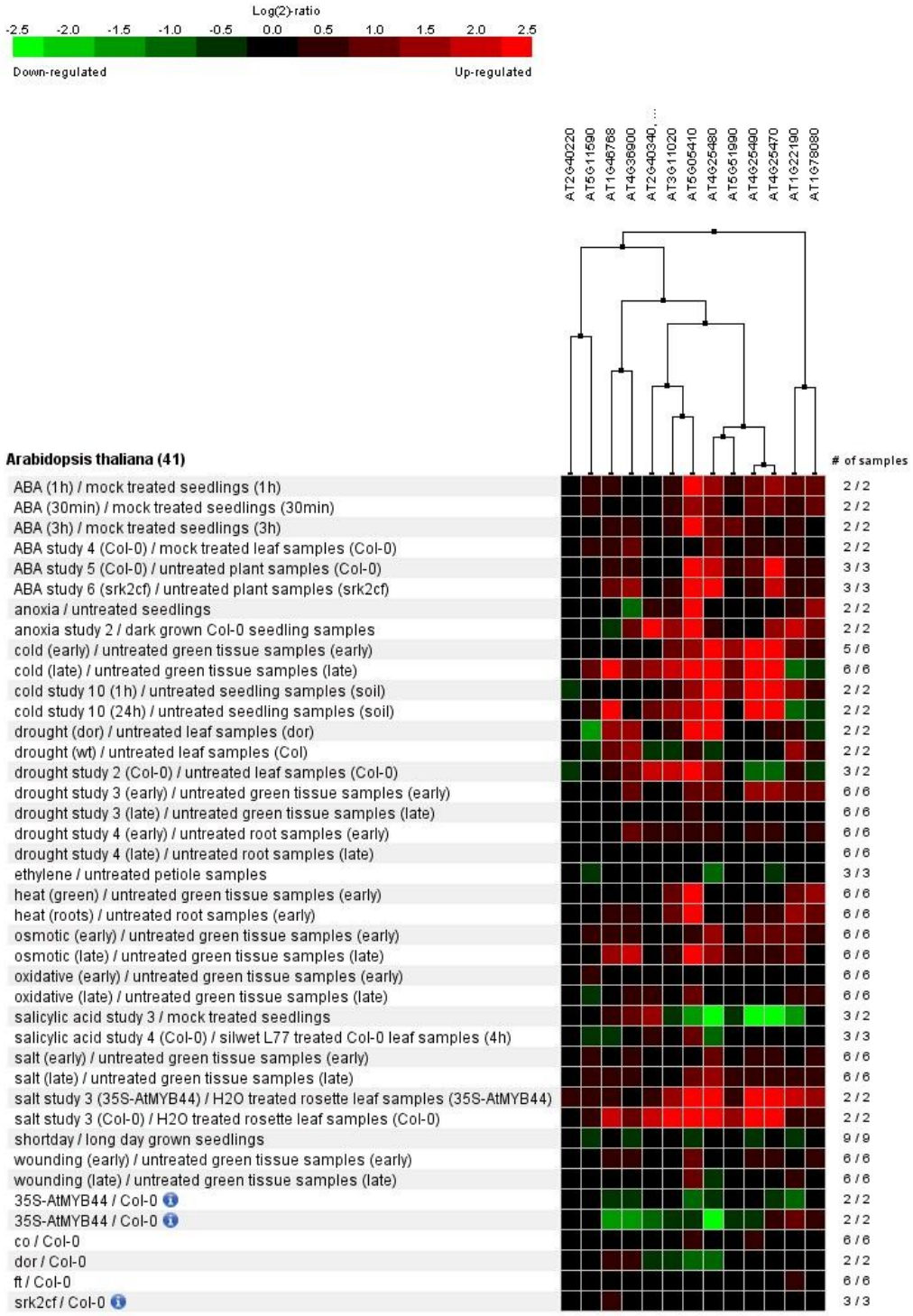
Supplementary Table 3.3 – Protein conserved motifs found within 54 putative DREB proteins from common bean.

Motif	Width	Best possible match	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	26	RIWLGTFPTPEMAARAYDVAAYCLKG	-	0.3	0.2	0.3	0.2	0.2	0.2	0.2	0.3	0.1	0.2	0.2	0.2	0.2
2	9	WGKWVCEIR	0.3	-	0.2	0.2	0.1	0.2	0.3	0.3	0.2	0.1	0.2	0.3	0.2	0.4
3	16	GGPENRHCVYRGVRQR	0.2	0.2	-	0.2	0.2	0.3	0.4	0.2	0.2	0.2	0.1	0.2	0.2	0.3
4	33	RSACLNFPDLAWRLPRPATCSPRDIQAAAAKAA	0.3	0.2	0.2	-	0.5	0.3	0.3	0.2	0.3	0.2	0.2	0.2	0.1	0.2
5	14	EDAILNFPNLLHNM	0.2	0.1	0.2	0.5	-	0.1	0.2	0.2	0.2	0.2	0.1	0.3	0.3	0.2
6	16	KKVPAKGWKKGCMRGK	0.2	0.2	0.3	0.3	0.1	-	0.4	0.4	0.2	0.2	0.1	0.1	0.1	0.2
7	5	EPRKK	0.2	0.3	0.4	0.3	0.2	0.4	-	0.3	0.2	0.3	0.3	0.4	0.2	0.3
8	28	DMSADSIRKKATQVGARVDALQTALHHH	0.2	0.3	0.2	0.2	0.2	0.4	0.3	-	0.4	0.2	0.2	0.2	0.1	0.3
9	20	FGEYKPLPSSVDAKLQAICQ	0.3	0.2	0.2	0.3	0.2	0.2	0.2	0.4	-	0.2	0.2	0.2	0.1	0.2
10	21	LNHLTPPQVHQIQAQIQIQKQ	0.1	0.1	0.2	0.2	0.2	0.2	0.3	0.2	0.2	-	0.2	0.2	0.1	0.2
11	22	EFVFVDPVDTWMYPPPWCDCIY	0.2	0.2	0.1	0.2	0.1	0.1	0.3	0.2	0.2	0.2	-	0.3	0.2	0.1
12	49	MMEPWYGLDDLQSPKYVDQMLSCAFFDIDSTQLLDDLYEESDIRLWSFC	0.2	0.3	0.2	0.2	0.3	0.1	0.4	0.2	0.2	0.2	0.3	-	0.2	0.2
13	21	YWEDDSDHFNLQKYPSYEIDW	0.2	0.2	0.2	0.1	0.3	0.1	0.2	0.1	0.1	0.1	0.2	0.2	-	0.2
14	24	HSKGDGSKSVADTLAKWKEYNAQL	0.2	0.4	0.3	0.2	0.2	0.2	0.3	0.3	0.2	0.2	0.1	0.2	0.2	-

Supplementary Table 3.4 – Blast2Go annotation file.

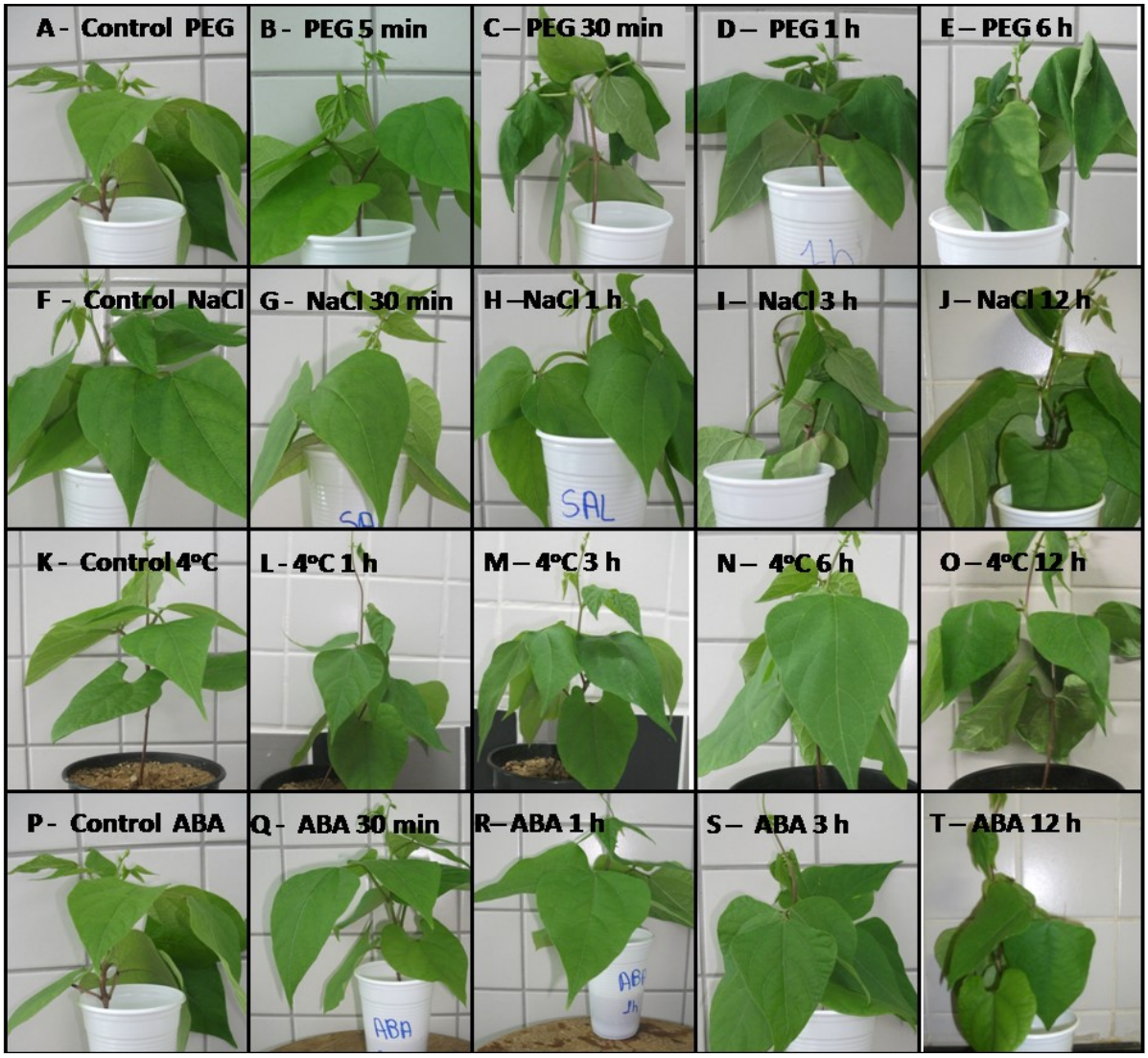
SeqName	SeqDesc	GOTerms	Enzymes
Phvul.010G054000	ethylene-responsive transcription factor erf034	nucleus	DNA binding
Phvul.010G146600	dehydration-responsive element-binding protein 3-like	nucleus	DNA binding
Phvul.003G223600	ethylene-responsive transcription factor erf027-like	nucleus	DNA binding
Phvul.009G084400	ethylene-responsive transcription factor erf018	nucleus	DNA binding
Phvul.009G225000	ethylene-responsive transcription factor erf017-like	nucleus	DNA binding
Phvul.009G029600	ethylene-responsive transcription factor rap2-4	nucleus	DNA binding
Phvul.009G109600	dehydration-responsive element-binding protein 3-like	nucleus	DNA binding
Phvul.005G111200	dehydration-responsive element-binding protein 2a	nucleus	DNA binding
Phvul.005G12630	c-repeat binding factor 3	nucleus	DNA binding
Phvul.005G170600	dehydration-responsive element-binding protein 3-like	nucleus	DNA binding
Phvul.005G126600	dehydration-responsive element-binding protein 1f-like	nucleus	DNA binding
Phvul.011G091400	dehydration-responsive element-binding protein 1f-like	nucleus	DNA binding
Phvul.008G172200	ethylene-responsive transcription factor rap2-4-like	nucleus	DNA binding
Phvul.008G141000	ethylene-responsive transcription factor erf034	nucleus	DNA binding
Phvul.008G220400	dehydration-responsive element-binding protein 2c	nucleus	DNA binding
Phvul.008G092800	dehydration responsive element binding protein	nucleus	DNA binding
Phvul.004G169800	dehydration-responsive element-binding protein 3-like	nucleus	DNA binding
Phvul.007G222500	dehydration-responsive element-binding protein 1f-like	nucleus	DNA binding
Phvul.007G255100	dehydration-responsive element-binding protein 2f-like	nucleus	DNA binding
Phvul.007G135300	ethylene-responsive transcription factor erf062	nucleus	DNA binding
Phvul.001G010400	dehydration-responsive element-binding protein 2d-like	nucleus	magnesium ion binding
Phvul.002G310200	dehydration-responsive element-binding protein 3-like	nucleus	DNA binding
Phvul.002G036000	ethylene-responsive transcription factor erf022-like	nucleus	DNA binding
Phvul.002G035100	ethylene-responsive transcription factor erf020-like	nucleus	DNA binding
Phvul.002G016700	dehydration responsive element-binding protein 1	nucleus	DNA binding
Phvul.002G056800	dehydration-responsive element-binding protein 3	nucleus	DNA binding
Phvul.002G154000	ethylene-responsive transcription factor erf026-like	nucleus	DNA binding
Phvul.002G153900	c-repeat dehydration-responsive element-binding factor 4	nucleus	DNA binding
Phvul.002G254500	ethylene-responsive transcription factor erf060-like	nucleus	DNA binding
Phvul.010G114900	ethylene-responsive transcription factor erf023-like	nucleus	DNA binding
Phvul.003G241700	ethylene-responsive transcription factor erf008-like	nucleus	DNA binding
Phvul.003G212800	dehydration-responsive element-binding protein	nucleus	DNA binding
Phvul.003G222600	ethylene-responsive transcription factor erf034-like	nucleus	DNA binding
Phvul.003G212700	ethylene-responsive transcription factor erf026-like	nucleus	DNA binding
Phvul.009G013200	dehydration-responsive element-binding protein 2d	nucleus	DNA binding
Phvul.009G123300	ethylene-responsive transcription factor erf012-like	nucleus	DNA binding
Phvul.005G105200	ethylene-responsive transcription factor erf053-like	nucleus	DNA binding
Phvul.011G118600	ethylene-responsive transcription factor erf053-like	nucleus	DNA binding
Phvul.011G107800	dehydration-responsive element-binding protein 2a	nucleus	DNA binding
Phvul.008G098900	dehydration-responsive element binding protein 2	nucleus	DNA binding
Phvul.008G165000	ethylene-responsive transcription factor erf012-like	nucleus	DNA binding
Phvul.007G222600	ethylene-responsive transcription factor erf024-like	nucleus	DNA binding
Phvul.007G066500	drought responsive element binding protein 1	nucleus	DNA binding
Phvul.001G136100	dehydration-responsive element-binding protein 2f-like	nucleus	DNA binding
Phvul.001G073800	dehydration-responsive element-binding protein 3-like	nucleus	DNA binding
Phvul.001G251200	ethylene-responsive transcription factor erf060-like	nucleus	DNA binding
Phvul.001G023700	transcription factor dreb2 protein	nucleus	DNA binding
Phvul.001G044500	ethylene-responsive transcription factor erf018	nucleus	DNA binding
Phvul.001G187100	ethylene-responsive transcription factor erf024-like	nucleus	DNA binding
Phvul.006G114100	ethylene-responsive transcription factor erf023-like	nucleus	DNA binding
Phvul.002G035900	ethylene-responsive transcription factor erf021-like	nucleus	DNA binding
Phvul.002G163700	ethylene-responsive transcription factor erf061-like	nucleus	DNA binding
Phvul.008G222400	ethylene-responsive transcription factor abi4-like	nucleus	DNA binding
Phvul.004G122000	drought responsive element binding protein 1	nucleus	DNA binding
Phvul.008G131500	integrase-type dna-binding superfamily protein isoform 1	nucleus	DNA binding
Phvul.006G179800	pathogenesis-related genes transcriptional activator pti5-like	nucleus	DNA binding
Phvul.003G292400	ethylene-responsive transcription factor erf109-like	nucleus	DNA binding
Phvul.003G111800	ap2 erf and b3 domain-containing transcription repressor	nucleus	DNA binding

Supplementary Figure 3.1



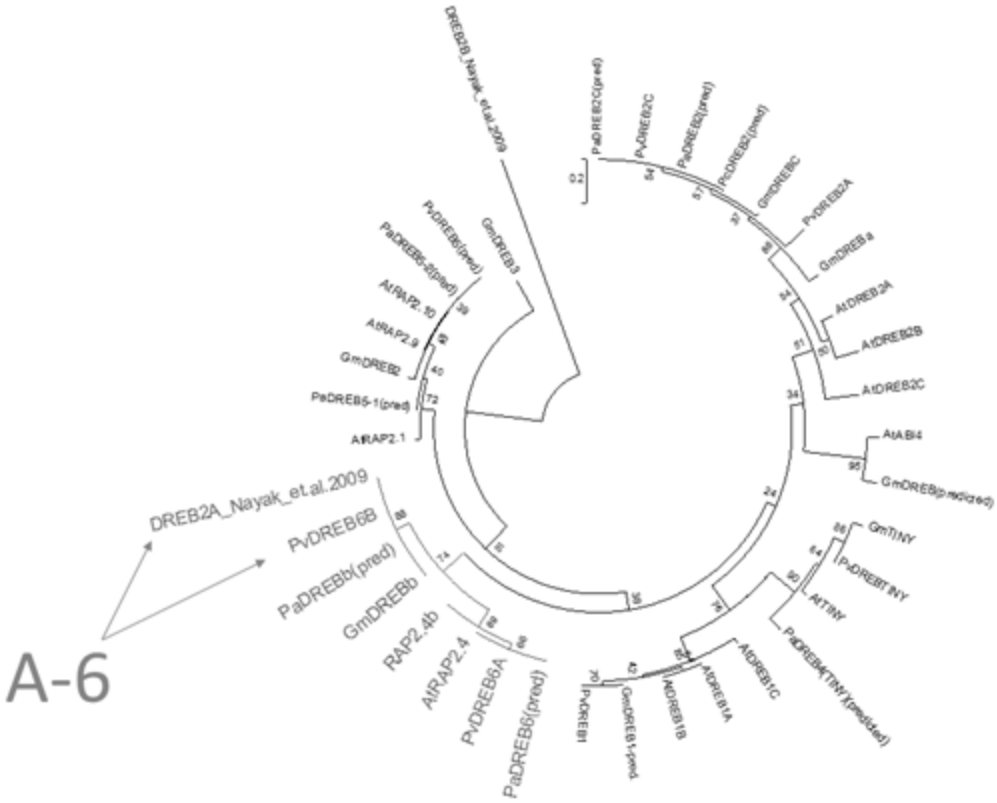
SF3.1 - Microarray expression profile (up-regulated is red and down-regulated is green) of *DREB* genes of *Arabidopsis thaliana*, under several abiotic treatment, generated by Genevestigator analysis tool. Codes above heat maps represent the following genes: AT2G40220 - gene *ABI4*; AT5g11590 - gene *TINY2*; AT1G46788 - gene *RAP2.1*; AT2G36900 - gene *RAP2.10*; AT2G40340 - gene *DREB2C*; AT3G11020 - gene *DREB2B*; AT5GO5410 - gene *DREB2A*; AT4G25480 - gene *DREB1A/CBF3*; AT5G51990 - gene *DREB1D/CBF4*; AT4G25490 - gene *DREB1B/CBF1*; AT4G25470 - gene *DREB1C/CBF2*; AT1G22190 - gene *RAP2.4*; AT1G78080 - gene *RAP2.4/WIND1*.

Supplementary Figure 3.2



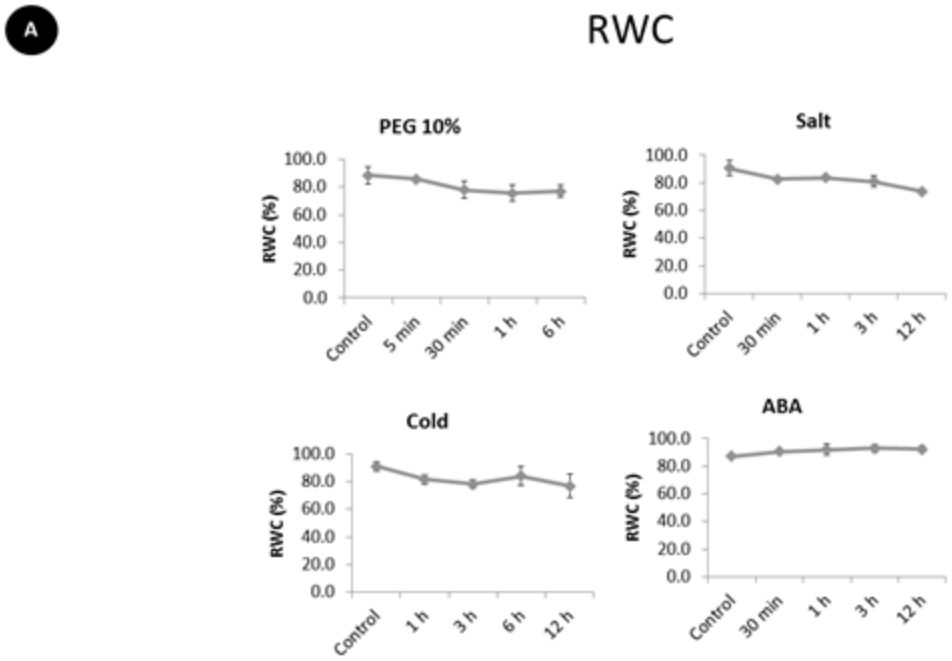
SF3.2 – Experiment pictures showing BAT 477 plants submitted to four abiotic stresses in controlled conditions: dehydration (PEG 10%), salinity (NaCl 250 mM), low temperature (4°C) and abscisic acid (ABA 100 µM). Picture show plants at different time periods of stress induction.

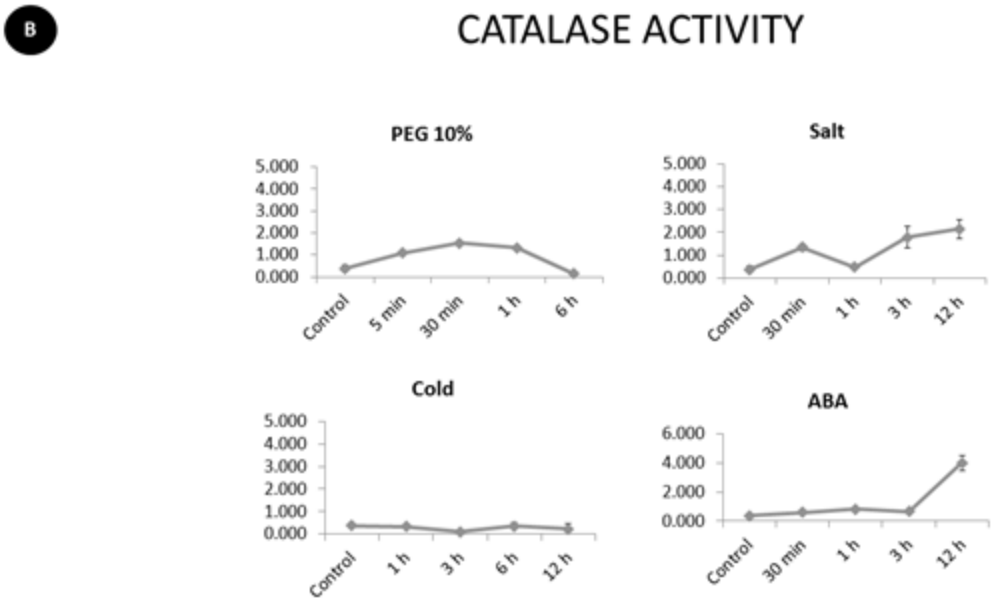
Supplementary Figure 3.3



SF3.3 – Neighbor-joining phylogenetic tree showing the main genes from each of the *DREB* subgroups (A-1 to A-6). Sequences *DREB2A* and *DREB2B* are included in the analysis. *DREB2A* is actually within subgroup A-6 and its name has been replaced by *PvDREB6B*, which shows to the equivalent sequence (although with no full coverage when aligned).

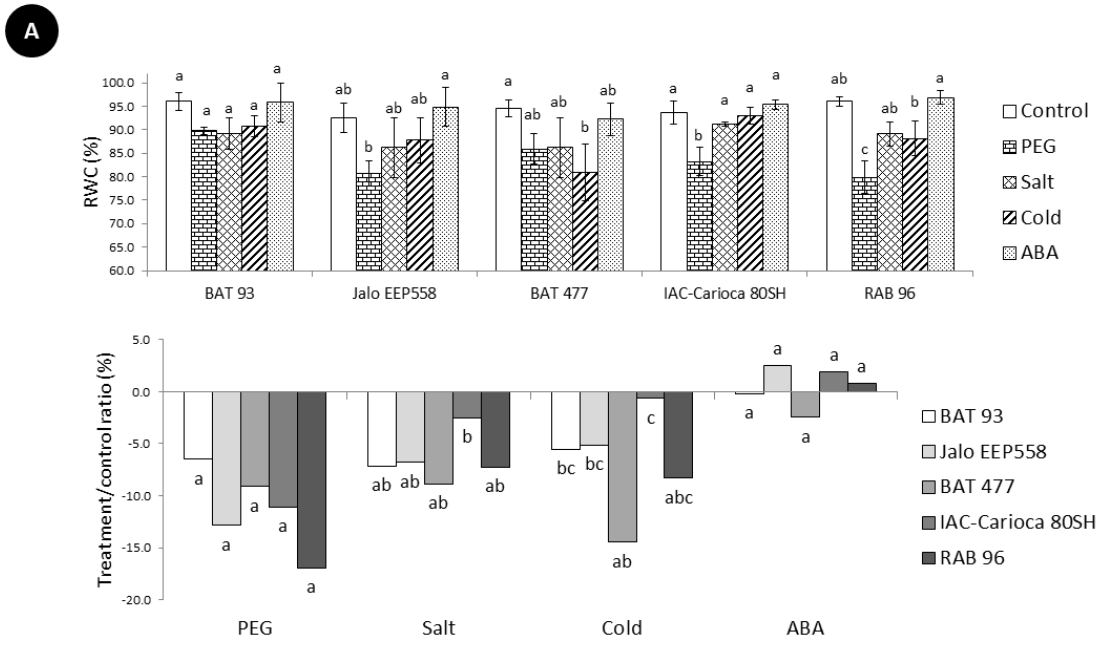
Supplementary Figure 3.4



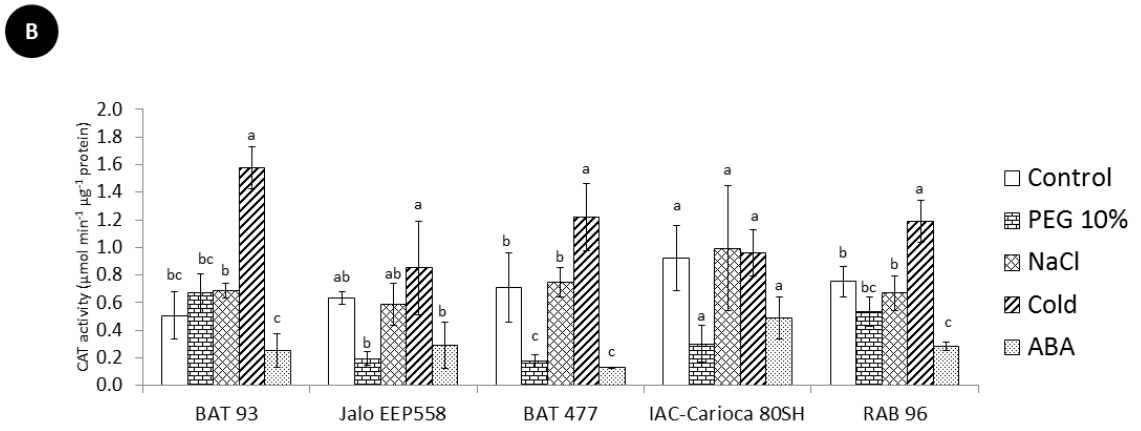


SF3.4 – Temporal analysis of the effects of four abiotic stresses on the relative water content (RWC) and enzyme activity (catalase) of BAT 477 leave samples. A – RWC (previous page) and B – Catalase activity.

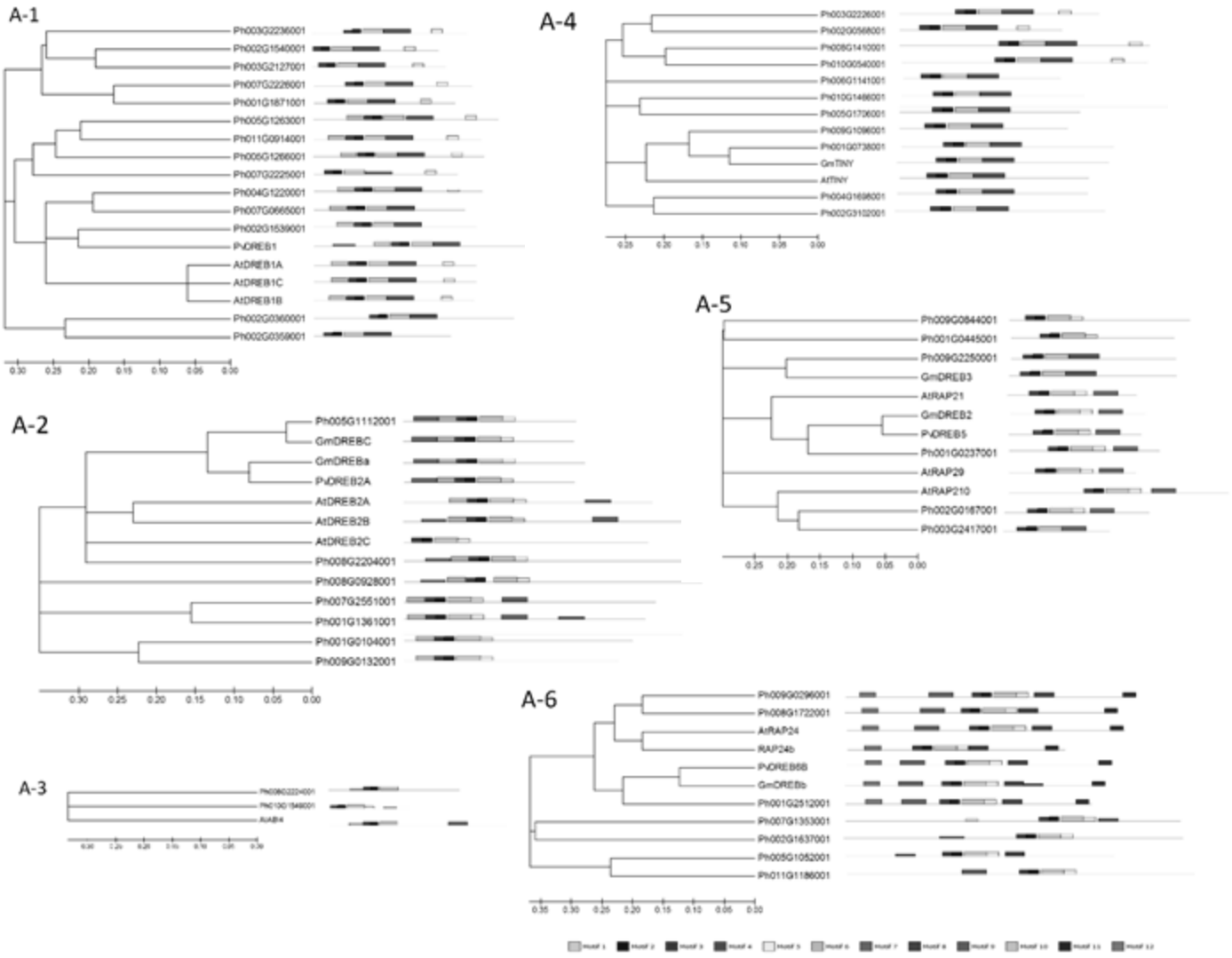
Supplementary Figure 3.5



SF3.5 – Effects of four stress treatments on the relative water content and the percentage of reduction in water content (A) and catalase activity (B) of five common bean genotypes. Catalase activity is shown on next page.

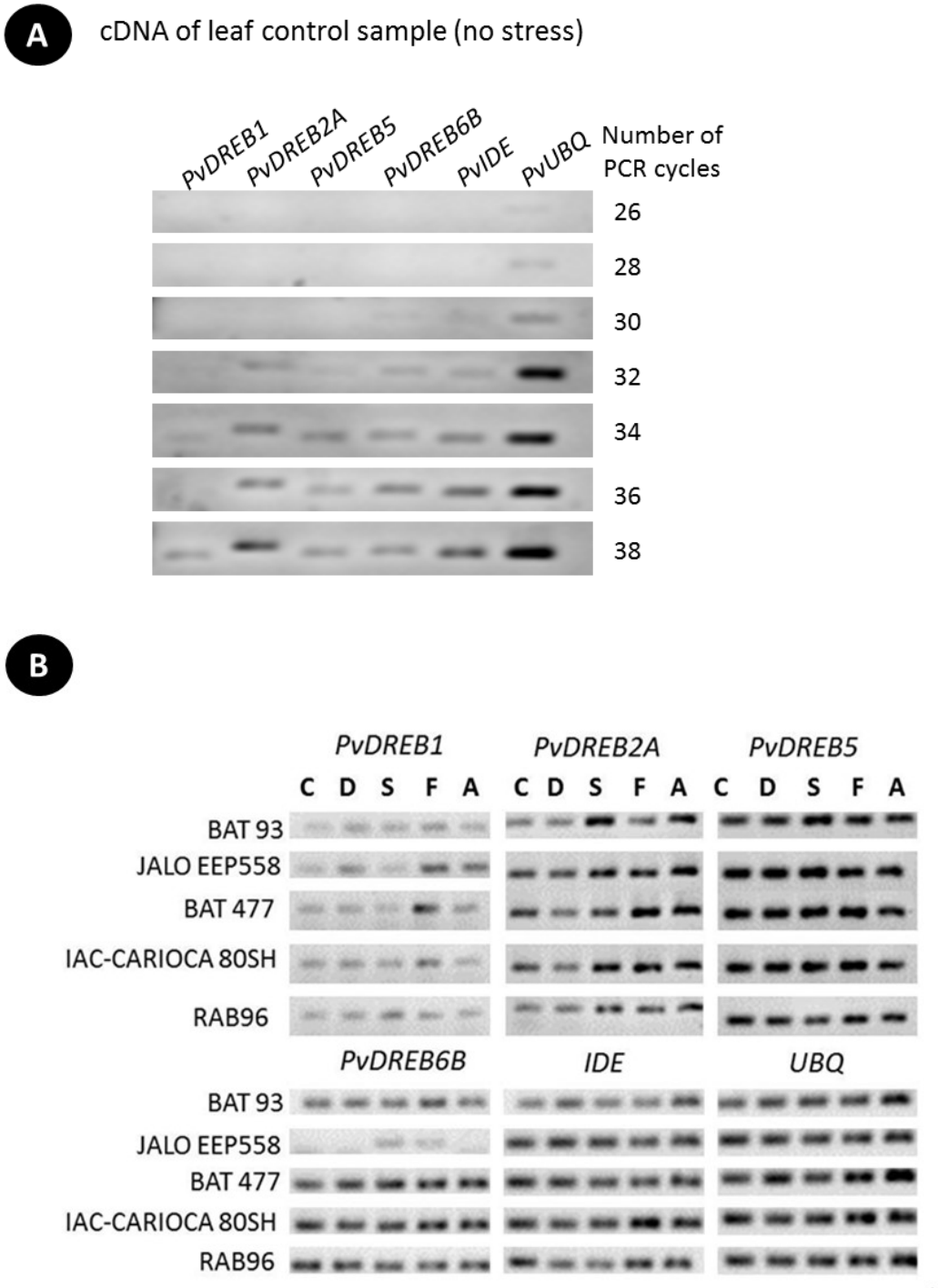


Supplementary Figure 3.6



SF3.6 – Pre-analysis of protein motif predictions with MEME tool. Sequences include putative DREB from common bean, *Arabidopsis thaliana* and *Glycine max*.

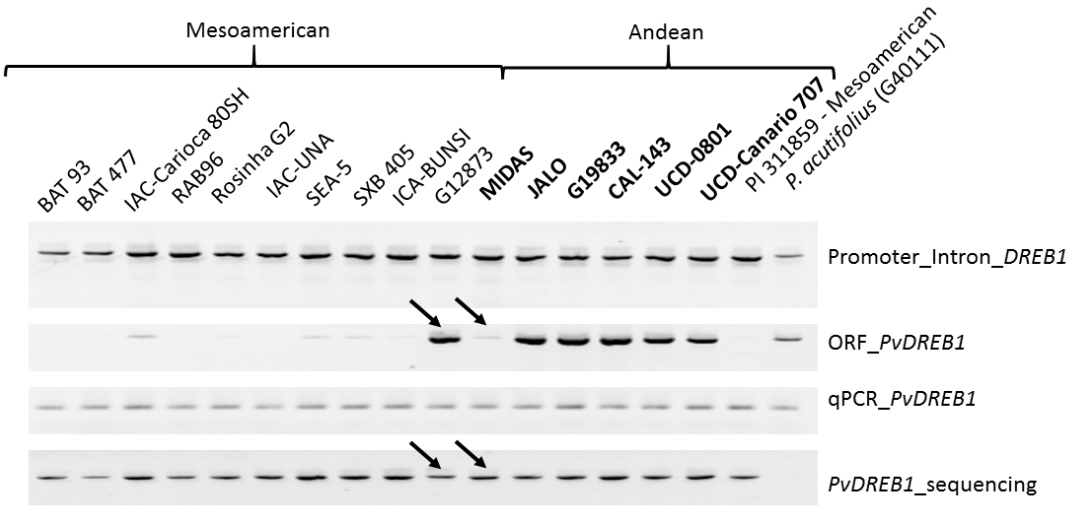
Supplementary Figure 3.7



SF3.7 – (A) Amplification of cDNA samples submitted to no treatment with *PvDREB* genes reveal differences in the initial amounts of transcripts of each gene. (B) A comparison between amplification of *PvDREB1F* and the other genes under stress-treatments with leaf samples from five genotypes under control (C), dehydration (D), salinity (S), freezing (F) and ABA (A).

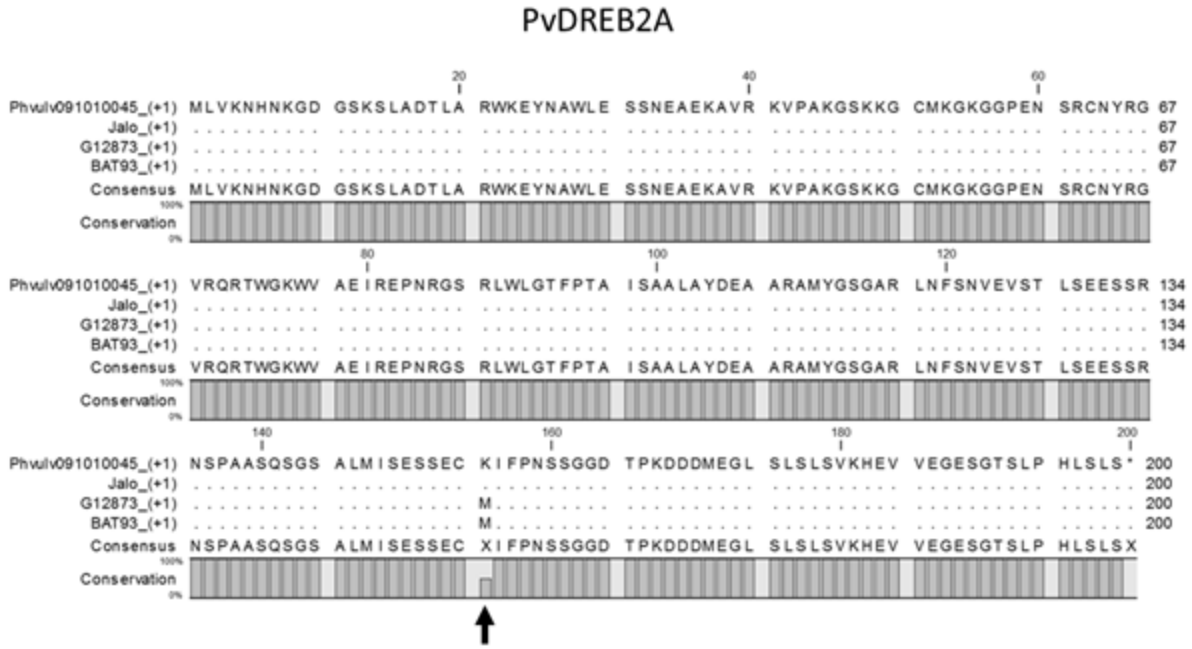
Appendix B – Supplementary Files from Section 4

Supplementary Figure 4.1

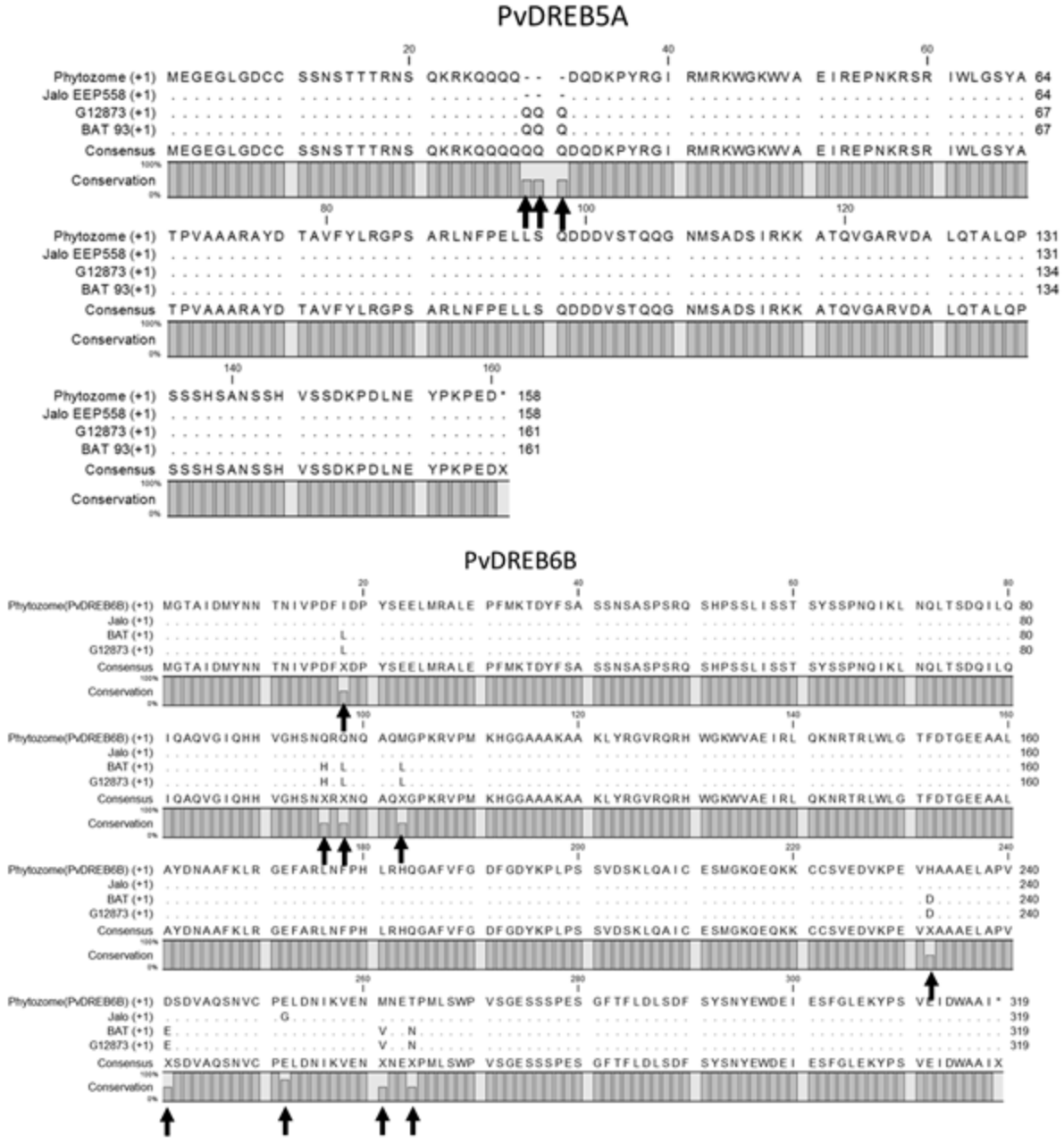


SF4.1 – Gel profile of primers designed for amplifying *PvDREB1F*. The pair of primers referred as ORF_*PvDREB1F* was designed to amplify from the start to the stop codon of the Open Reading Frame of *PvDREB1F*. No efficient amplification occurred with DNA from Midas, while G12873 sample was successfully amplified along with the Andeans Jalo EEP558, G19833, Cal 143, UCD 0801 and UCD Canario 707. Amplification product was also detected for *P. acutifolius* line.

Supplementary Figure 4.2



SF4.2 – Translated *PvDREB2A*, *PvDREB5A* and *PvDREB6B* sequences (this and next page). Black arrows indicate sites with amino acid changes. Alignments were performed with CLC Sequence Viewer.



Supplementary Table 4.1 – Predictions of the effects of amino acid changes within PvDREB proteins using PROVEAN tools.

Variant	PROVEAN score	Prediction (cutoff = -2.5)
	PvDREB2A	
K155M	0.054	Neutral
	PvDREB5A	
Q27_Q28insQQQ	-2.819	**Deleterious**
	PvDREB6B	
I18L	0.097	Neutral
Q98L	-0.547	Neutral
M103L	1.84	Neutral
H232D	-0.037	Neutral
D241E	0.991	Neutral
E252G	-1.571	Neutral
M261V	0.094	Neutral
T264N	0.126	Neutral
S48L	-1.73	Neutral

Appendix C – Supplementary Files from Section 6

Supplementary Table 6.1 - General information about wild bean accessions used for *PvDREB6B* nucleotide diversity analysis.

Number	Accession	Latitude	Longitude	Origin	Altitude (m)	Average annual precipitation (mm)
1	G10011	18.8833	-99.1500	Mexico	1430	973
2	G10018	19.6833	-100.9167	Mexico	888	1324
3	G10019A	19.4667	-100.4833	Mexico	1250	780
4	G11028	24.4667	-104.5833	Mexico	2020	530
5	G11050A	19.6833	-101.2667	Mexico	2040	808
6	G11052	20.8000	-103.4000	Mexico	1390	945
7	G50725	14.0333	-87.0500	Honduras	1085	1038
8	G11115	19.4000	-103.4000	Mexico	1140	1030
9	G12851	14.2833	-90.3000	Guatemala	2000	1932
10	G12852	13.9167	-89.8500	El Salvador	800	1747
11	G12860	23.8833	-104.2667	Mexico	1800	464
12	G12861	19.4167	-102.0667	Mexico	1341	1614
13	G12863	20.7833	-104.1833	Mexico	1420	747
14	G12864	19.4667	-103.5833	Mexico	1350	837
15	G12865	19.3333	-103.25	Mexico	1219	1126
16	G12866	19.6833	-103.4833	Mexico	1524	635
17	G12867	21.0833	-104.5000	Mexico	823	876
18	G12868	21.3333	-104.5833	Mexico	762	995
19	G12869	19.4167	-102.5833	Mexico	1310	971
20	G12872	18.9667	-99.1000	Mexico	1828	961
21	G12873	19.0000	-99.2500	Mexico	1981	964
22	G12875	17.3167	-96.9000	Mexico	1676	758
23	G12877	18.9500	-99.2167	Mexico	1920	1048
24	G12878	18.3500	-99.7667	Mexico	1402	1143
25	G12879	18.3500	-99.9833	Mexico	1585	1077
26	G12879A	18.3500	-99.9833	Mexico	1585	1077
27	G12880	18.4333	-100.1000	Mexico	1620	1201
28	G12882B	18.2833	-100.1500	Mexico	853	1130
29	G12883	19.4667	-100.9000	Mexico	1350	1073
30	G12893	21.2167	-101.8000	Mexico	ND	673
31	G12894	20.1333	-102.0833	Mexico	ND	912
32	G12910	20.6167	-101.7167	Mexico	1829	737
33	G12914	20.5333	-103.1833	Mexico	1524	889
34	G12916	20.6667	-102.4500	Mexico	ND	932
35	G12947	20.6667	-102.3833	Mexico	1829	948
36	G12949	20.6667	-102.3833	Mexico	1829	948
37	G12957	20.9000	-102.3667	Mexico	2134	838
38	G12964	21.0167	-102.2500	Mexico	ND	739
39	G12986	21.6833	-103.1000	Mexico	ND	679
40	G13029	19.7333	-103.1167	Mexico	1280	1046
41	G13030	19.4667	-103.3500	Mexico	1219	1022
42	G15421	18.3500	-100.2167	Mexico	1290	1155
43	G16799	18.4333	-100.1667	Mexico	1560	1170
44	G19906	14.4500	-90.7000	Guatemala	1550	1377
45	G19907	14.4500	-90.8167	Guatemala	1280	2178
46	G19908	14.5333	-90.8333	Guatemala	1550	1313
47	G20559	9.9333	-84.0833	Costa Rica	1171	1901
48	G21116	5.0167	-73.5500	Colombia	1600	1203
49	G21118	5.0167	-73.5500	Colombia	1900	1203
50	G22304	4.4833	-73.9333	Colombia	1750	1496
51	G22837	26.9333	-106.4167	Mexico	1750	536
52	G23418	9.8667	-84.1167	Costa Rica	1560	2103
53	G23429	18.9667	-98.3833	Mexico	1430	851
54	G23431	19.0333	-100.0500	Mexico	1630	1218
55	G23432	18.9500	-99.4333	Mexico	1890	891
56	G23439	14.4333	-90.1333	Guatemala	1430	1695
57	G23440	14.4167	-90.5833	Guatemala	1740	1497

58	G23462	5.0833	-73.6167	Colombia	1900	1060
59	G23463	28.3333	-108.5000	Mexico	1530	763
60	G23464	18.9000	-99.0333	Mexico	1250	922
61	G23469	18.9333	-99.1333	Mexico	1400	982
62	G23470	17.3167	-96.0333	Mexico	950	2548
63	G23507	19.5333	-103.5833	Mexico	1920	1156
64	G23511A	19.7167	-104.2000	Mexico	1300	901
65	G23512	20.4667	-103.8333	Mexico	1420	864
66	G23520	19.0000	-99.0667	Mexico	1900	939
67	G23535	18.9500	-99.0833	Mexico	1450	936
68	G23536	18.8167	-98.7500	Mexico	1800	1074
69	G23541	18.6667	-99.3167	Mexico	1950	993
70	G23545	24.3333	-104.2833	Mexico	1990	525
71	G23552	23.6000	-104.3333	Mexico	1460	470
72	G23555	23.9167	-104.5667	Mexico	1950	579
73	G23556	24.0000	-104.7333	Mexico	1860	504
74	G23593	14.6500	-89.9500	Guatemala	ND	1650
75	G23648	19.700	-100.5500	Mexico	ND	864
76	G23678	20.9000	-102.3667	Mexico	2100	838
77	G24323	17.0833	-93.0833	Mexico	1280	1519
78	G24334A	20.0833	-103.8500	Mexico	1650	986
79	G24338	20.1333	-103.8500	Mexico	1680	1034
80	G24340	20.1333	-103.9000	Mexico	1420	1024
81	G24344	20.5833	-103.3333	Mexico	1540	909
82	G24345	21.0333	-104.4833	Mexico	1330	855
83	G24350	19.6167	-104.2000	Mexico	1410	1218
84	G24352	20.1000	-102.8667	Mexico	1650	825
85	G24361	19.2833	-103.2500	Mexico	1600	1145
86	G24362	19.3000	-103.0833	Mexico	1460	1244
87	G24364	19.4000	-103.3333	Mexico	1520	987
88	G24365	19.4000	-103.5333	Mexico	1400	1042
89	G24368	20.4000	-102.3333	Mexico	1550	831
90	G24375	19.1500	-100.3333	Mexico	1400	1105
91	G24389	20.8333	-103.3333	Mexico	1050	915
92	G24390	21.3500	-104.5333	Mexico	640	1161
93	G24576	16.8000	-96.5167	Mexico	1800	581
94	G24584	16.4667	-92.5333	Mexico	1150	1414
95	G24589	16.2500	-91.8333	Mexico	1525	1666
96	G24592	16.7000	-92.8500	Mexico	1660	1524
97	G24604	17.0333	-96.7667	Mexico	1800	691
98	G24605	20.7000	-102.4167	Mexico	1960	936
99	G2771	21.1667	-104.3667	Mexico	1050	920
100	G50369	17.0667	-96.2667	Mexico	1480	1009
101	G50384	15.6667	-91.7833	Guatemala	1180	2495
102	G50385	15.6500	-91.7000	Guatemala	1540	1989
103	G50388	14.7500	-91.5167	Guatemala	1760	1915
104	G50414	20.2000	-103.8167	Mexico	1630	975
105	G50503	14.6167	-90.5667	Guatemala	1550	1168
106	G50504	14.7000	-90.7833	Guatemala	1330	1147
107	G50505	14.7167	-90.7833	Guatemala	1350	1062
108	G50506	14.6000	-91.1167	Guatemala	1330	2636
109	G50507	14.7833	-91.4833	Guatemala	2050	1299
110	G50722A	13.8500	-86.9167	Honduras	1650	1624
111	G50898	9.8167	-84.0500	Costa Rica	1540	1991
112	G51062	9.9000	-83.9333	Costa Rica	1500	1880
113	G51658	9.9000	-83.9833	Costa Rica	1500	1696
114	PI 317349	19.4333	-103.5167	Mexico	1730	1124
115	PI 319441	24.2300	-104.4700	Mexico	2000	588
116	PI 343950	15.6831	-91.8169	Guatemala	850	1600
117	PI 535413	20.0833	-103.1500	Mexico	ND	898
118	PI 535416	19.5500	-103.6333	Mexico	1920	1272
119	PI 535418	19.7167	-104.2000	Mexico	1040	901
120	PI 535426	18.8667	-98.8167	Mexico	1750	1035
121	PI 417653	20.6167	-101.7167	Mexico	2000	737

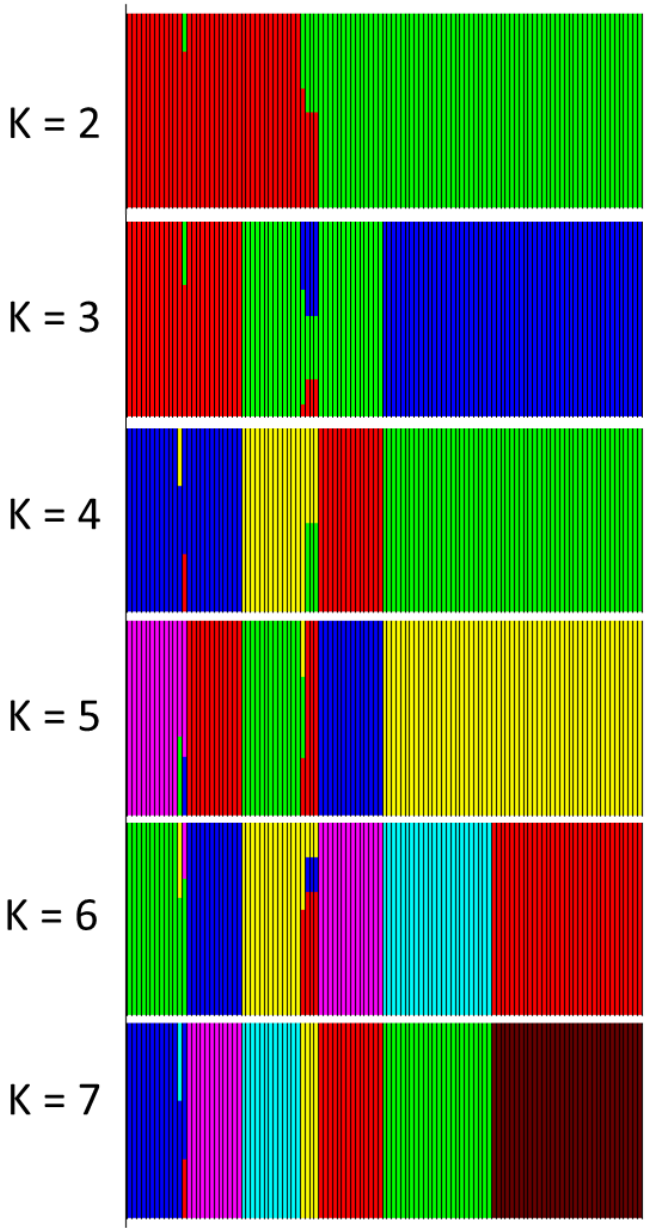
Supplementary Table 6.2 – *PvDREB6B* haplotypes analyzed through DnaSP and their distribution within the wild bean collection.

Hap_1: 6 [G10011 G12879 G23648 G24361 G12879A G24375]
Hap_2: 15 [G10018 G12872 G12873 G16799 G23469 G23470 G23520 G23536 G24362 G24364 G24390 PI535418 PI535426 G23535 G23541]
Hap_3: 6 [G10019A G12875 G23429 G24338 G24340 G50369]
Hap_4: 13 [G11028 G12860 G12894 G12986 G19907 G22837 G23545 G23552 G23555 G23556 G24584 G24589 G24592]
Hap_5: 1 [G11050A]
Hap_6: 5 [G50725 G12878 G24604 G50385 PI343950]
Hap_7: 2 [G12851 G50506]
Hap_8: 1 [G12852]
Hap_9: 2 [G12861 G12882B]
Hap_10: 1 [G12863]
Hap_11: 11 [G12864 G23431 G23507 G23511A G24334A G24350 PI317349 PI535413 PI535416 G24365 PI319441]
Hap_12: 3 [G12865 G13030 G24344]
Hap_13: 1 [G12866]
Hap_14: 1 [G12868]
Hap_15: 1 [G12869]
Hap_16: 1 [G12880]
Hap_17: 1 [G12883]
Hap_18: 1 [G12893]
Hap_19: 9 [G12910 G12947 G12964 G23678 G24352 G24368 G24605 PI417653 G12916]
Hap_20: 1 [G12949]
Hap_21: 1 [G12957]
Hap_22: 1 [G15421]
Hap_23: 1 [G19906]
Hap_24: 9 [G19908 G21116 G22304 G50384 G50388 G50503 G50504 G50505 G50507]
Hap_25: 1 [G20559]
Hap_26: 1 [G21118]
Hap_27: 5 [G23418 G23439 G50898 G51658 G50722A]
Hap_28: 2 [G23432 G24323]
Hap_29: 1 [G23440]
Hap_30: 1 [G23462]
Hap_31: 1 [G23463]
Hap_32: 1 [G23593]
Hap_33: 1 [G24389]
Hap_34: 1 [G24576]
Hap_35: 1 [G2771]
Hap_36: 1 [G50414]
Hap_37: 1 [G51062]

Supplementary Table 6.3 – Summary of bioclimatic variables averages for each one of the clades (C-1 to C-7) defined for wild bean accessions based on *PvDREB6B* haplotypes.

Number	Variable	C-1	C-2	C-3	C-4	C-5	C-6	C-7	*p*
1	Latitude	17.91	20.24	19.56	19.76	21.94	13.08	18.33	0.0001***
2	Longitude	-97.86	-101.70	-102.39	-103.35	-100.80	-88.35	-98.87	0.0001***
3	Altitude	1560.50	1894.83	1582.92	1375.50	1696.85	1536.63	1455.34	0.0024*
BIO1	Annual mean temperature	19.05	18.18	18.44	19.27	18.52	18.52	20.59	0.0515 ns
BIO2	Mean diurnal range	14.56	15.96	14.12	15.02	15.22	11.12	14.61	0.0001***
BIO3	Isothermality	6.78	6.33	6.50	6.40	6.32	7.42	6.86	0.0001***
BIO4	Temperature seasonality	170.03	238.72	189.13	198.05	293.17	94.77	162.00	0.0001***
BIO5	Maximum temperature of warmest month	29.73	30.33	29.10	30.52	30.42	26.31	31.26	0.0001***
BIO6	Minimum temperature of coldest month	8.20	5.34	7.42	7.27	5.97	11.23	10.05	0.0001***
BIO7	Temperature annual range	21.52	24.98	21.67	23.25	24.44	15.07	21.12	0.0001***
BIO8	Mean temperature of wettest quarter	20.18	20.19	19.72	20.75	20.87	19.15	21.21	0.098 ns
BIO9	Mean temperature of driest quarter	17.99	17.52	17.66	18.45	17.12	18.20	20.08	0.0263*
BIO10	Mean temperature of warmest quarter	21.07	20.90	20.61	21.47	21.87	19.91	22.58	0.0005**
BIO11	Mean temperature of coldest quarter	16.65	14.79	15.74	16.40	14.45	17.49	18.40	0.0001***
BIO12	Annual precipitation	953.16	831.18	1055.30	923.00	910.00	1591.20	1133.15	0.0001***
BIO13	Precipitation of wettest month	193.17	205.91	234.00	219.75	206.43	308.54	247.90	0.0001***
BIO14	Precipitation of driest month	10.66	5.09	5.61	3.25	6.35	10.58	6.51	0.0001***
BIO15	Precipitation seasonality	89.67	107.00	100.07	105.00	100.28	82.91	100.93	0.0001***
BIO16	Precipitation of wettest quarter	532.16	541.81	649.00	597.75	536.78	781.87	675.12	0.0002**
BIO17	Precipitation of driest quarter	40.00	19.72	21.92	13.75	28.07	46.50	27.45	0.0469*
BIO18	Precipitation of warmest quarter	305.08	294.63	387.07	345.75	290.71	436.62	315.87	0.0033**
BIO19	Precipitation of coldest quarter	82.33	31.09	48.92	40.75	49.71	134.17	52.49	0.0047**

Supplementary Figure 6.1



SF1 – Detailed BAPs analysis of population structure with clustering for K = 1 to K = 7 subpopulations.

Appendix D – Supplementary Files from Section 7

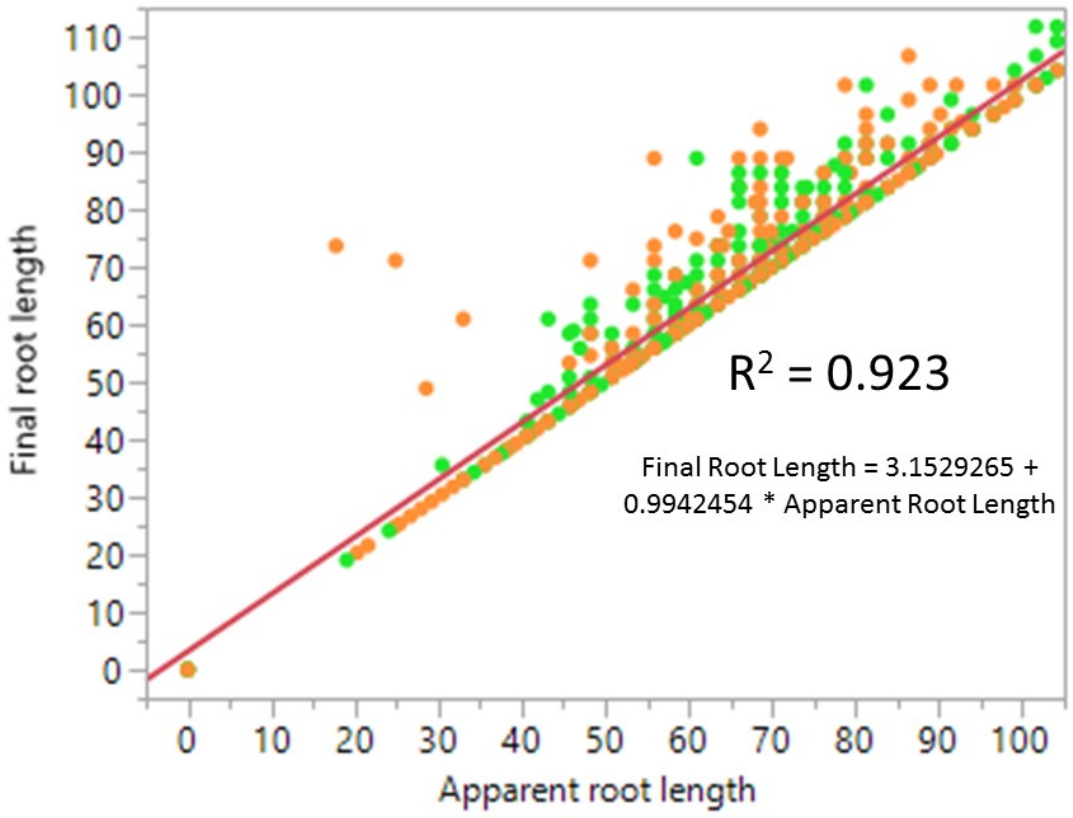
Supplementary Figure 7.1

	Day																
Block	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
1	Seeding			RL1				RL2				RL3				RL4/PH4		
2		Seeding			RL1				RL2				RL3				RL4/PH4	
3			Seeding			RL1				RL2				RL3				RL4/PH4

Block	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**
1		RL5/PH5				RL6/PH6				RL7/PH7				FRL/FPH		
2			RL5/PH5				RL6/PH6				RL7/PH7				FRL/FPH	
3				RL5/PH5				RL6/PH6				RL7/PH7				FRL/FPH

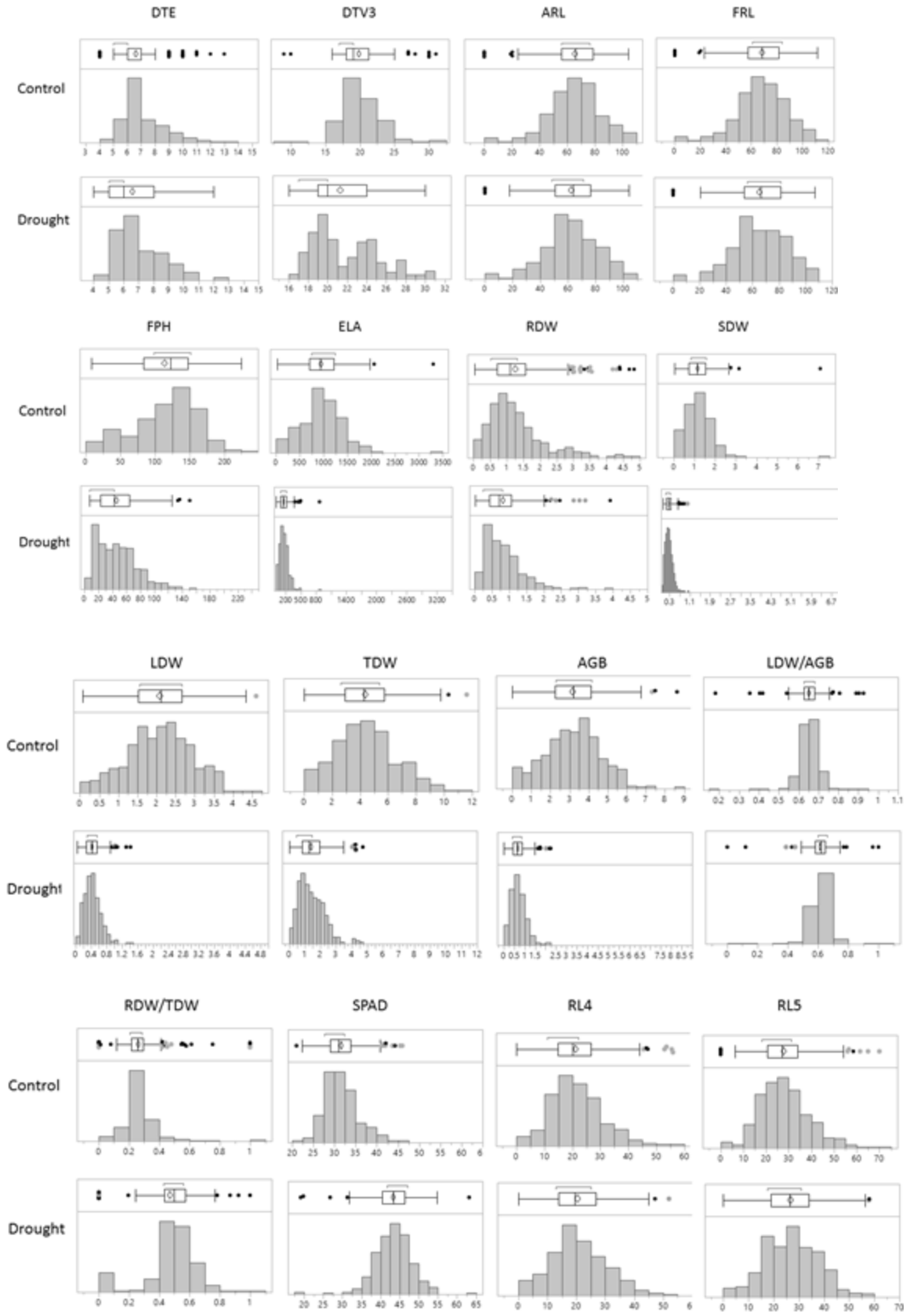
SF1 – Schematic representation of the sequential measurements of root and plant height. RL1 to RL7 stand for root length, while PH1 to PH7 stand for plant height. FRL is the final root length or RL8 and FPH is the final plant height measurement or PH8.

Supplementary Figure 7.2

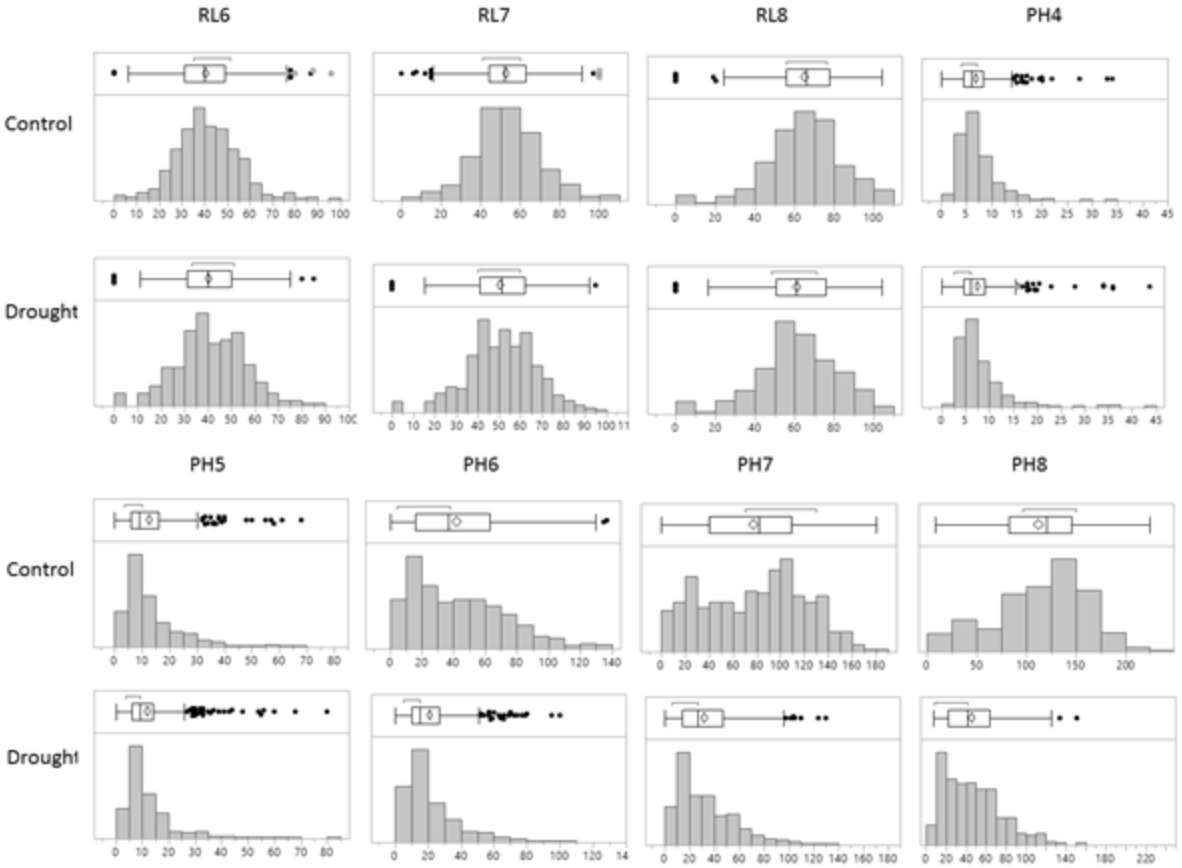


SF7.2 – Adjustment between apparent root length and final root length. Green dots indicate measurements of control plants while orange dots represents stressed plants.

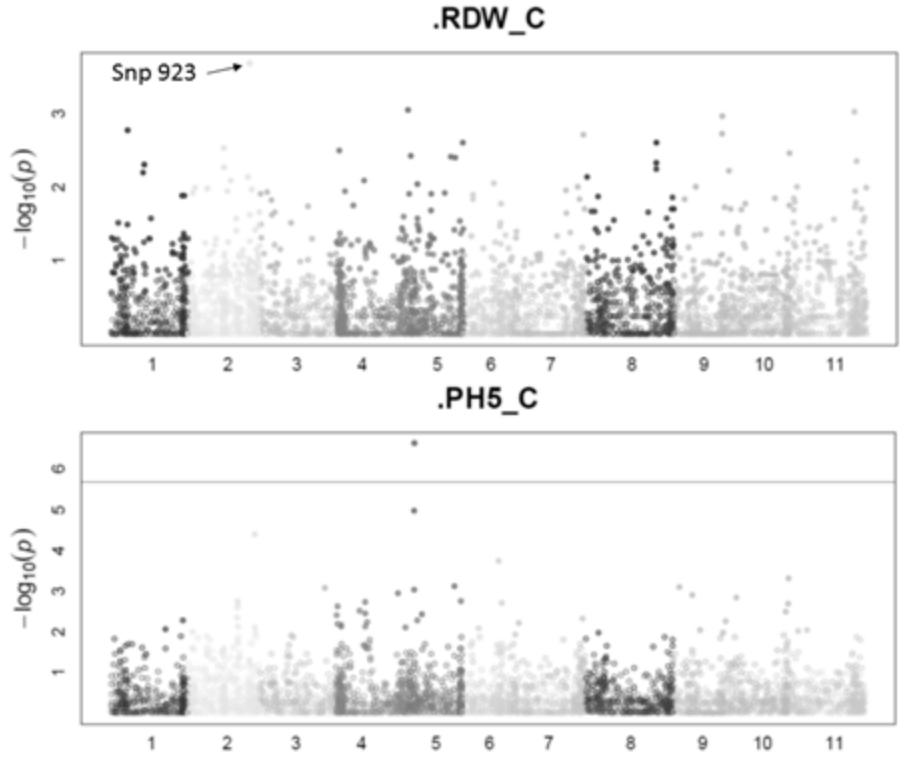
Supplementary Figure 7.3



SF7.3 – Distribution of the data obtained for 24 drought-related traits (Continued…).



Supplementary Figure 7.4



SF7.4 – Manhattan plot showing associations between root dry weight and plant height and SNP markers. Snp 923 is s marker within the coding region of *PvDREB6B*.

Supplementary Table 7.1 – Correlations among phenotypic measurements of wild common bean accessions of Mesoamerican origin submitted to two irrigation regimes in greenhouse conditions. In total, 24 variables were evaluated: DTE (days do emergence); DTV3 (days to V3 stage); ARL (apparent root length); FRL (final root length); ELA (leaf area based on Easy Leaf Area software); RDW (root dry weight); SDW (stem dry weight); LDW (leaf dry weight); TDW (total dry weight); AGB (aboveground biomass); LDW/ABG (proportion of leaf dry weight to the aboveground biomass); RDW/TDW (proportion of root dry weight to the total dry weight); SPAD (unit of measurement for chlorophyll); RL4, RL5, RL6, RL7 and RL8 (root length or depth at the fourth, fifth, sixth, seventh and eight measurements, respectively); PH4, PH5, PH6, PH7 and PH8 (plant height at the fourth, fifth, sixth, seventh and eight measurements, respectively). Only from the fourth measurement ahead, it was possible to measure visible roots in all tubes, so data are presented for both from this point. All correlation values highlighted in gray were non-significant, while all the others were significant ($p < 0.05$).

CONTROL

DROUGHT

	DTE	DTV3	ARL	FRL	FPH	ELA	RDW	SDW	LDW	TDW	AGB	LDW/ABG	RDW/TDW	SPAD	RL4	RL5	RL6	RL7	RL8	PH4	PH5	PH6	PH7	PH8
DTE		0.570	-0.372	-0.381	-0.438	-0.466	-0.469	-0.449	-0.534	-0.524	-0.519	0.279	-0.096	-0.268	-0.484	-0.471	-0.395	-0.355	-0.407	-0.257	-0.292	-0.355	-0.381	-0.438
DTV3	0.539		-0.489	-0.502	-0.565	-0.665	-0.639	-0.603	-0.749	-0.697	-0.682	0.275	-0.283	-0.215	-0.536	-0.593	-0.547	-0.546	-0.466	-0.513	-0.512	-0.497	-0.457	-0.565
ARL	-0.397	-0.419		0.978	0.487	0.509	0.484	0.603	0.640	0.715	0.747	-0.299	0.169	0.102	0.613	0.676	0.723	0.794	0.946	0.460	0.424	0.454	0.481	0.487
FRL	-0.383	-0.414	0.975		0.518	0.501	0.496	0.605	0.644	0.727	0.759	-0.315	0.148	0.127	0.633	0.691	0.731	0.791	0.927	0.454	0.436	0.459	0.483	0.518
FPH	-0.346	-0.613	0.445	0.460		0.552	0.539	0.718	0.715	0.670	0.703	-0.509	0.091	0.351	0.508	0.508	0.529	0.512	0.475	0.594	0.664	0.760	0.730	1.000
ELA	-0.343	-0.568	0.583	0.607	0.609		0.710	0.674	0.840	0.706	0.690	-0.227	0.286	0.148	0.557	0.602	0.661	0.666	0.521	0.453	0.353	0.386	0.402	0.552
RDW	-0.353	-0.360	0.383	0.372	0.184	0.401		0.668	0.777	0.868	0.739	-0.293	0.554	0.360	0.544	0.593	0.585	0.562	0.511	0.465	0.400	0.408	0.416	0.539
SDW	-0.391	-0.696	0.566	0.581	0.815	0.714	0.383		0.812	0.850	0.894	-0.691	0.168	0.346	0.641	0.647	0.647	0.655	0.602	0.596	0.552	0.562	0.567	0.718
LDW	-0.410	-0.639	0.677	0.696	0.701	0.786	0.444	0.895		0.894	0.900	-0.289	0.299	0.334	0.704	0.735	0.724	0.733	0.641	0.614	0.553	0.593	0.620	0.715
TDW	-0.417	-0.501	0.658	0.655	0.462	0.610	0.821	0.676	0.735		0.969	-0.436	0.338	0.354	0.683	0.718	0.690	0.694	0.720	0.584	0.544	0.569	0.585	0.670
AGB	-0.429	-0.632	0.722	0.737	0.737	0.748	0.440	0.923	0.960	0.780		-0.482	0.155	0.333	0.705	0.729	0.704	0.717	0.747	0.602	0.573	0.602	0.623	0.703
LDW/ABG	0.048	0.246	0.136	0.137	-0.360	0.047	0.035	-0.362	0.044	-0.028	-0.099		0.019	-0.284	-0.302	-0.282	-0.256	-0.276	-0.306	-0.337	-0.349	-0.364	-0.385	-0.509
RDW/TDW	-0.020	0.033	0.113	0.100	-0.182	-0.024	0.411	-0.130	-0.111	0.455	-0.087	0.057		0.125	0.174	0.230	0.202	0.205	0.132	0.153	0.091	0.086	0.088	0.091
SPAD	-0.066	0.020	-0.105	-0.123	-0.230	-0.193	0.204	-0.012	-0.077	0.092	-0.036	-0.154	0.153		0.284	0.202	0.158	0.175	0.114	0.302	0.376	0.355	0.342	0.351
RL4	-0.284	-0.435	0.681	0.681	0.364	0.549	0.312	0.511	0.583	0.545	0.604	0.102	0.130	-0.102		0.923	0.802	0.780	0.620	0.480	0.439	0.449	0.446	0.508
RL5	-0.239	-0.401	0.715	0.723	0.395	0.566	0.386	0.550	0.611	0.619	0.635	0.069	0.202	-0.052	0.898		0.869	0.838	0.677	0.488	0.426	0.440	0.444	0.508
RL6	-0.273	-0.454	0.818	0.823	0.509	0.617	0.380	0.619	0.671	0.633	0.700	0.019	0.132	-0.096	0.790	0.875		0.917	0.724	0.510	0.441	0.450	0.477	0.529
RL7	-0.250	-0.444	0.827	0.829	0.500	0.627	0.364	0.610	0.662	0.618	0.682	0.051	0.160	-0.089	0.746	0.825	0.916		0.795	0.508	0.446	0.434	0.442	0.512
RL8	-0.396	-0.402	0.948	0.924	0.415	0.566	0.366	0.549	0.636	0.645	0.681	0.094	0.193	-0.093	0.713	0.753	0.845	0.876		0.425	0.380	0.432	0.463	0.475
PH4	-0.421	-0.520	0.299	0.296	0.646	0.408	0.207	0.584	0.459	0.351	0.510	-0.261	-0.115	-0.103	0.375	0.349	0.416	0.387	0.313		0.861	0.779	0.660	0.594
PH5	-0.340	-0.480	0.253	0.252	0.700	0.384	0.098	0.553	0.417	0.229	0.455	-0.287	-0.246	-0.148	0.242	0.247	0.332	0.352	0.255	0.830		0.879	0.749	0.664
PH6	-0.237	-0.487	0.234	0.250	0.797	0.441	0.095	0.653	0.489	0.248	0.507	-0.339	-0.246	-0.248	0.266	0.289	0.366	0.361	0.217	0.710	0.855		0.914	0.760
PH7	-0.199	-0.458	0.269	0.276	0.788	0.419	0.055	0.662	0.485	0.232	0.515	-0.377	-0.269	-0.273	0.303	0.296	0.370	0.379	0.267	0.646	0.776	0.938		0.730
PH8	-0.346	-0.613	0.445	0.460	1.000	0.609	0.184	0.815	0.701	0.462	0.737	-0.360	-0.182	-0.230	0.364	0.395	0.509	0.500	0.415	0.646	0.700	0.797	0.788	